Returns to Scale, Productivity and Competition: Empirical Evidence from U.S. Manufacturing and Construction Establishments*,**

Wei Gao† and Matthias Kehrig‡

15 February 2015

Abstract

We build a model of firm entry and exit and show how returns to scale shape firm survival, the equilibrium productivity and size distributions and firm concentration. High productivity dispersion and high concentration ratios need not reflect inefficiencies when returns to scale are strongly decreasing. We apply a broad set of structural and reduced-form estimation techniques to establishment-level data from the U.S. Census of Construction and Manufacturing to assess returns to scale and productivity dispersion across establishments. Indeed, industries with lower returns to scale are characterised by higher productivity dispersion and lower concentration ratios as predicted in the model. Returns to scale are 0.96 on average, but range from 0.86 in Non-metallic Minerals to 1.3 in Semiconductors. Returns to scale tend to be highest in durable manufacturing, medium in non-durable manufacturing and lowest in the production of housing. An economy characterised by such differences in its sectoral production structure will exhibit long-run structural change away from construction and non-durable manufacturing, as in the data, and endogenously replicate the cyclical behavior of relative prices in the U.S.: relative durables prices are countercyclical while relative housing prices are procyclical.

JEL CODES: E2, L1, L2, L6, L7

KEYWORDS: Returns to scale, production functions, productivity dispersion, sectoral change.

*We are very grateful to Morris Davis and John Haltiwanger for helpful comments and discussions. Morris Davis gracefully shared several of his datasets on regional housing price indexes. Michael Blake and Susan Bucci provided help with the Census of Construction Industries. All errors are our own. The latest version of this paper can be downloaded at http://ssrn.com/abstract=SSRNID

**Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

†University of Texas at Austin, wgao@utexas.edu.

‡Corresponding Author, University of Texas at Austin, Department of Economics, 2225 Speedway Stop C3100, Austin, TX 78712-1690, matthias.kehrig@austin.utexas.edu.

1 Introduction

Differences across firm productivity are large and pervasive. Syverson (2004) documents that within narrowly defined U.S. industries the establishment at the top decile of the total factor productivity (TFP) distribution produces almost twice the output produced by the firm at the bottom decile of the TFP distribution that uses the same amount of inputs. That this remarkable heterogeneity in productivity is not only large but also persistent has been documented in various industries, countries and time periods (see Syverson (2011) for an overview). This heterogeneity has attracted a lot of attention recently when several studies quantified this heterogeneity result in significant efficiency losses, see for example Restuccia and Rogerson (2008); Hsieh and Klenow (2009). Consistent with this view, a large literature has argued that reallocation of resources away from the unproductive tail of the distribution to the top would increase efficiency, output and welfare (see for example Bloom et al. (2012); Akcigit et al. (2013)).

But to what extent dispersion in TFP persist depends on the long-run productivity characteristic of an industry: returns to scale. Firms with a low total factor productivity may survive competition with their more productive peers when returns to scale are strongly decreasing; then, small size compensates for low TFP. Although there has been a lot of empirical work documenting the existence of TFP dispersion, its cyclical fluctuations and how it affects macroeconomic outcomes, there is much less using the same micro-level data to document returns to scale. This paper fills that gap and empirically studies returns to scale across a large set of U.S. construction and manufacturing industries. We utilize several methods to estimate returns to scale at the industry level and document how returns to scale vary with TFP dispersion and other characteristics of competition in the industry such as the equilibrium size distribution and firm concentration.

We also study the impact of returns to scale on firm survival, productivity dispersion and firm concentration theoretically. To do that, we build a model of firm entry and exit where returns to scale differ across industries and so do equilibrium productivity dispersion and concentration ratios etc. We use the rich industry data to empirically support the predictions of main of our model: Indeed, industries with lower returns to scale are characterised by higher productivity dispersion and lower concentration ratios as predicted in the model. These results mean that one needs to be careful when assessing the efficiency and welfare consequences of measured productivity dispersion and firm concentration. When returns to scale are strongly decreasing, reallocation of resources to productive larger firms may do more harm than good.

Knowing the empirics of returns to scale is important for a number of additional reasons:

First, a recent strand of macro models builds "micro-founded" models, that is, frictions and economic behaviour are modelled at the level of the firm and then aggregated into macroeconomic outcomes. There is a heated debated whether or not heterogeneity across firms matters for macro aggregates or not: Khan and Thomas (2008); Thomas (2002); Bloom et al. (2012); Bloom (2009); Bachmann et al. (2013). Whether firm-level heterogeneity matters for macroeconomic aggregates crucially depends on returns to scale at the firm level because it shapes the curvature of the profit function. Typically, the literature calibrates a firm-level production function. Yet, we know little of the empirical counterpart, the production function estimated using firm-level data. In this paper, we attempt to use micro-level data and estimate returns to scale for a large part of the economy. While we find that most industries are characterised by moderately decreasing returns to scale of 0.94, there is considerable heterogeneity across industries: non-metallic minerals show returns to scale on the lower end while semiconductors have significantly increasing returns to scale of 1.3.

Second, returns to scale matter for the shape of the cost curve and thus supply curve. Differences in returns to scale will then translate into differences in relative prices and how they change over the business cycle. Changing relative prices have traditionally been attributed to exogenously asymmetric technology shocks. There is a large literature for example on biased technology shocks, most notably investment-specific technology shocks (see Justiniano et al. (2010, 2011); Basu et al. (2010)). This literature explains a considerable share of aggregate fluctuations with exogenous shocks to the technology of investment goods (durable manufacturing) which are identified as fluctuations in the relative price of durable goods. While this explanation is consistent with the data, we complement that view with supply-side differences in returns to scale which equally lead to the observed movement in relative prices when the economy is hit by short-run symmetric demand fluctuations. Indeed, we find that durable goods industries tend to exhibit the highest returns to scale while construction, the production of new housing, tends to exhibits the lowest returns to scale. If markups in all these industries behave similarly over the cycle, then the relative price of industries with higher returns to scale (such as durable manufacturing goods) would fall while that of industries with lower returns to scale (such as construction) would rise. Thus, short-run fluctuations in the relative price of durables could emerge from demand fluctuations that move along supply curves with different slopes.

Productivity studies using micro data have traditionally focused on the manufacturing sector. We are among the first to assess returns to scale and productivity among construction establishments. Understanding the productivity characteristics of construction will help understand the cost structure of firms that sector of the economy, an important aspect in the pricing of these firms. Considering the supply-side factors in the housing sector of the economy has been very limited (Saiz (2010) is a notable exception) and we contribute to that literature. It seems important to understand supply-side factors because pervisou research (Glaeser et al. (2012, 2013) among others) find that demand-side factors such as low interest rates, easy access to credit cannot explain the bulk of the housing price fluctuations.

Third, a long-standing literature since Marschak and Andrews (1944) has tried to estimate production functions. Assessing returns to scale has been a long and contentious debate surrounding the proper identification of production functions (see among others Burnside et al. (1995); Bils and Klenow (1998); Hall (1990); Basu and Fernald (1997); Basu (1996); Burnside (1996)). As a consequence, researchers in industrial organization have developed structural estimation techniques that overcome endogeneity and selection problems in the estimation of production functions: Olley and Pakes (1996); Levinsohn and Petrin (2003); Ackerberg et al. (2006); Wooldridge (2009); Gandhi et al. (2013); Grieco et al. (forthcoming). We apply most of these estimators to a large set of establishment-level data and compare the estimation techniques in terms of their results and their computational stability. Previous research has estimated returns to scale using industry-level data (Basu and Fernald (1995, 1997); Burnside et al. (1995); Burnside (1996)) or focused on a particular industry (Olley and Pakes (1996)). To date, a comprehensive estimation of establishments-level returns to scale in the U.S. economy is missing and we fill this gap. In addition to that, we add to the empirical literature to broaden the scope beyond manufacturing which often is the main sector of analysis (Burnside (1996)).

2 Theoretical motivation

We are interested in how returns to scale and productivity matter for the the competitive environment where firms struggle for survival. To assess the impact of returns to scale, we construct a heterogeneous firm model with firm entry and exit. This model is a simplified version from Kehrig (2015), but it suffices to study the relationship between returns to scale and the productivity and distribution of firms in equilibrium. Our model will feature a fixed overhead factor of production as in Melitz (2003); Ghironi and Melitz (2005) which will give rise to a survival productivity level, below which firms do not survive. The model features endogenous exit and firm-level heterogeneity in productivity. Importantly, firm survival, the measure of active firms in the economy and the size and productivity distribution in the economy will depend on the returns to scale. In the model, we adopt the assumption from Olley and Pakes (1996) that all cross-sectional profitability differences are driven by underlying technological differences rather than mark-up differences. For the purposes of studying industry characteristics of productivity and size it is enough to study this economy in partial industry equilibrium.

2.1 Firms

2.1.1 Production

A continuum of perfectly competitive firms produces a homogeneous output good which is sold to households. In period t, a measure N_t of firms is active. Each firm hires labour l_{it}, purchases materials m_{it} and rents capital k_{it} in order to produce output y_{it} using the following technology:

$$y_{it} = A_t z_{it} k_{it}^{\alpha} l_{it}^{\gamma} m_{it}^{\nu}$$

where A_t is an aggregate productivity level, z_{it} is the firm's idiosyncratic productivity and α, γ and ν are the production elasticities of all inputs. We refer to $\xi \equiv \alpha + \gamma + \nu$ as the returns to scale. For the time being, we assume $\xi < 1$ so that the industry equilibrium does not collapse to a degernate productivity and size distribution with only one firm. Since this is a purely real model, we assume that the output good is the numéraire and normalize its price to unity.

Aggregate growth and or business cycles would be driven by changes in the common productivity component, A_t. In line with the empirical results, we assume that each of the two productivity components follows an AR(1) stochastic process:

$$\log A_t = (1-\rho^A)\log \overline{A} + \rho^A \log A_{t-1} + \eta_t \quad (1)$$

$$\log z_{it} = (1-\rho^z)\log \overline{z}_i + \rho^z \log z_{it-1} + \varepsilon_{it} \quad (2)$$

where η_t and ε_{it} are the aggregate and idiosyncratic components of a technology shock, respectively. Both η and ε are drawn from a time-invariant distribution, F_η and F_ε respectively. The distribution of idiosyncratic technology shocks ε_{it} is time invariant and, importantly, independent of the level of aggregate technology A_t. $\overline{A}$ and $\overline{z}_i$ are the long-run averages of each productivity component.

Key to the production of intermediates is a fixed overhead input, denoted c_f, which must be hired every period by all firms regardless of their size, productivity or other characteristics. Since overheads are required to produce at all but are not related to the production amount, it is hard to identify their impact in the data.[1] Part of why that is the case is that overheads might be required in terms of any production input – capital, labour, materials, blueprints, reputation, legal licenses – or any combination of those. Examples for capital overheads could be the financing and maintenance cost of capital structures; examples of overhead labour are managerial workers, human resources, finance, organization, advertising or management etc. Regardless of their origin, these costs are most appropriately modeled as independent of the firm's production volume decision. Though overheads are ubiquitous, we will focus on overhead labour, e.g. managers, in the model. Kehrig (2015) found empirical support to attribute some of the overheads on non-production labour.

Conditional on being active and hiring overhead inputs, every firm i optimally chooses production labour, l_{it}, capital, k_{it}, and materials, m_{it}, in perfectly competitive markets in order to maximize static profits

$$\pi_{it} = A_t z_{it} k_{it}^{\alpha} l_{it}^{\gamma} - r_t k_{it} - w_t l_{it} - \omega_t c_f - q_t m_{it}$$

where r_t, w_t, q_t, ω_t are the prices of capital, labour, materials and the overhead input in period t. Standard profit maximization leads to the firm's factor demand:

$$r_t = \alpha A_t z_{it} k_{it}^{\alpha-1} l_{it}^{\gamma} m_{it}^{\nu} \quad \Leftrightarrow \quad k_{it} = \left[A_t z_{it} \left(\frac{\alpha}{r_t}\right)^{1-\gamma-\nu} \left(\frac{\gamma}{w_t}\right)^{\gamma} \left(\frac{\nu}{q_t}\right)^{\nu} \right]^{\frac{1}{1-\xi}} \quad (3)$$

and likewise of materials and labour. Note that firms will not optimize over overhead inputs as

[1] Overhead inputs are similar in nature to "intangible inputs" in that the econometrician observes more their indirect than their direct effects. Yet, researchers attribute important firm decisions to them: Atalay et al. (2014), for example, propose intangibles as a main driver of decisions of vertical integration, and Eisfeldt and Papanikolaou (2013) propose "organizational capital" – a concept similar to overhead inputs – as an important driver of a firm's investment and financing decisions.

they are required as long as the firm is active at all.

2.1.2 Profitability, exit and entry

While the previous section described the optimal behavior of active firms, we also model the life-cycle pattern of firms. This seems crucial because entry and exit are at the heart of the Schumpeterian models of business cycles. They rest on the idea that recessions are times when unproductive firms become unprofitable and are “cleansed out” of the economy. Obviously, productivity-driven exit has direct consequences for the cyclical nature of the productivity distribution. A key objective of my model is hence to feature endogenous exit along profitability. Although they are related in the present model, profitability, rather than productivity, determines whether or not a firm survives. This is in line with related empirical findings by Foster et al. (2008) who document that even unproductive firms may survive as long as they are able to extract a sufficient mark-up. To understand exit along the profitability margin, it is instructive to look at maximized firm profits:

$$\pi_{it} = (1-\xi)\left[A_t z_{it}\left(\frac{\alpha}{r_t}\right)^{\alpha}\left(\frac{\gamma}{w_t}\right)^{\gamma}\left(\frac{\nu}{q_t}\right)^{\nu}\right]^{\frac{1}{1-\xi}} - \omega_t c_f. \tag{4}$$

Equation (4) shows that more productive firms (higher z) make larger profits and that the profit function is convex in productivity z as long as returns to scale are decreasing ($0 < \xi < 1$). The most productive firms then post the highest profits and the least productive firms (z close to zero) post gross profits close to zero.

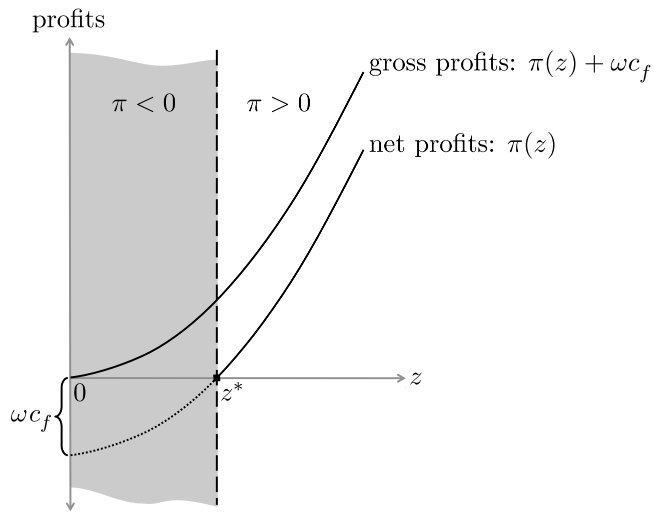
The overhead costs are responsible for why some unproductive firms become unprofitable at all. If there were no overhead costs, even the least productive firm would generate profits amounting to a $(1-\xi)$-share of their sales. When there are overheads, however, only the more productive firms break even after paying the overheads and post positive net profits. The least productive firms, which are also the smallest producers, do not generate enough gross profits to also cover the fixed overhead costs. They become unprofitable and exit. This immediate exit can be motivated by credit constraints: firms that temporarily make negative profits will be unable to pay their production factors without credit and have to shut down operations. Alternatively, one may think that firms may become idle and “wait out” unprofitable times until the macro-economic environment becomes profitable again. Such a strategy makes sense if there are no sunk costs of re-entry. Whatsoever scenario one picks, firms that are idle or temporarily shut down operations due to lack of finance do not appear in the data and hence they would be counted as exiting firms exactly like in the present model where firms permanently exit once they become unprofitable.

The convex gross profits and constant fixed overhead costs mean that there is a well-defined productivity threshold, denoted z_t^*, which yield zero profits and below which firms would make negative profits:

$$\pi(z_t^*) = 0 = (1-\xi)\left[A_t z_t^* \left(\frac{\alpha}{r_t}\right)^{\alpha}\left(\frac{\gamma}{w_t}\right)^{\gamma}\left(\frac{\nu}{q_t}\right)^{\nu}\right]^{\frac{1}{1-\xi}} - \omega_t c_f$$
$$\Leftrightarrow z_t^* = \frac{1}{A_t}\left(\frac{\omega_t c_f}{1-\xi}\right)^{1-\xi}\left(\frac{r_t}{\alpha}\right)^{\alpha}\left(\frac{w_t}{\gamma}\right)^{\gamma}\left(\frac{q_t}{\nu}\right)^{\nu} \tag{5}$$

Equation (5) is a key relationship as it regulates the measure of active firms in the economy, $1 - F(z_t^*)$, and the productivity distribution of active firms. Unproductive firms (firms with a low z_{it}) survive more easily when aggregate technology rises, i.e. when A_t is large. Then, the economy is in a boom and the cutoff level z_t^t is low. Conversely, a downturn poses harsher conditions for unproductive firms that may have to exit. This aspect reflects the view that recessions are weeding out unproductive firms – the "cleansing effect of recessions" (Caballero and Hammour (1994)). Note that endogenous factor prices r_t, w_t and ω_t will dampen the effect of A_t, but they will not overturn it. Figure 1 illustrates gross and net profits as a function of z and the productivity cutoff z^* for positive profits.

Figure 1: Firm productivity and firm profitability



Note: Firm profits gross and net of the overhead costs, ωc_f, as a function of the firm-specific productivity z. Any firm with productivity level below z^* will make losses (shaded area) because they are too small to generate enough revenue to cover both production and overhead costs. The productivity cutoff z^* is computed as the productivity level generating zero net profits.

Above, we have described how firms endogenously exit according to their productivity level. Key to the exit decision are negative profits which is consistent with the notion that re-entry is costless. Such an interpretation is consistent with temporarily unprofitable firms that merely become idle and do not have to pay entry costs. If there are no entry costs, however, we have to assume an exogenously fixed measure of entrants every period; this keeps the equilibrium measure of active firms stationary.[2] Upon entry, each firms receives its long-run productivity draw $\overline{z}_i$. The idiosyncratic productivity will be reverting to that long-run firm-specific mean.

The cutoff also determines the measure of active firms in the economy. The total measure of firms in the economy, N_t, is composed of surviving incumbents active last period , N^*_{t-1}, and new entrants, N^E_t: $N_t = N^*_{t-1} + N^E_t$. Since all incumbents receive productivity draws from the same distribution as the incumbents, their productivity distribution looks the same. Whether a firm is an incumbent or an entrant, it will only survive if its productivity is above this period's survival cutoff z^*_t, so the measure of firms that are active this period is:

$$N^*_t = [1 - F_z(z^*_t)]N_t = [1 - F_z(z^*_t)](N^*_{t-1} + N^E_t) \tag{6}$$

To keep the measure of active firms, N^*_t, stationary, we assume that the measure of entrants, N^E, equals the measure of exiting firms in steady state, i.e. that measure of firms below the steady-state cutoff level.

2.1.3 The firm size distribution and aggregation

Profit-maximising firm behavior determines some features of the cross section of active firms. We will start by showing that more productive firms will be larger in general, hiring more inputs and producing more output. Since all firms operate the same technology (summarized by α, γ and ν) and since all firms face the same prices r_t, w_t and q_t, their capital intensities, $\frac{k_{it}}{l_{it}}$ and $\frac{k_{it}}{m_{it}}$, must be the same. This can be simplified to writing the relative input choices of two distinct firms, i and j, as a function of their relative productivity only:

$$\alpha A_t z_{it} k_{it}^{\alpha-1} l_{it}^{\gamma} m_{it}^{\nu} = r_t = \alpha A_t z_{jt} k_{jt}^{\alpha-1} l_{jt}^{\gamma} m_{jt}^{\nu}$$
$$\Leftrightarrow \frac{k_{it}}{k_{jt}} = \left(\frac{z_{it}}{z_{jt}}\right)^{\frac{1}{1-\xi}}$$

[2] As an alternative, one may assume that firms do have to pay an entry cost c_e, denominated in units of labour for example. While the conclusions about the cyclicality of the productivity distribution would remain unchanged, the exit behavior of firms would be more complicated: Firms with a long-run productivity draw $\overline{z}_i$ below the steady state-level of z^* would immediately exit after entry, those with a long-run productivity draw above the steady state-level of z^* would only exit if their discounted future profits cover the temporary negative profits. The movements of the survival cutoff z^* over the business cycle would be weaker quantitatively, but unchanged qualitatively. A weaker movement in the cutoff, however, would strengthen the counter cyclicality of productivity dispersion.

and similarly for labour. Thus, all relative factor inputs and hence the relative size of firm output are a function of the relative productivities:

$$\frac{k_{it}}{k_{jt}} = \frac{l_{it}}{l_{jt}} = \frac{m_{it}}{m_{jt}} = \frac{y_{it}}{y_{jt}} = \left(\frac{z_{it}}{z_{jt}}\right)^{\frac{1}{1-\xi}} \qquad \forall i, j \in [0, N_t],\ i \neq j. \tag{7}$$

It will be convenient to compute firm size at the cutoff, $y(A_t, z_t^*)$ using the firm first-order conditions and the expression for z_t^*:

$$\begin{aligned} y(A_t, z_t^*) &= A_t z_t^* k(A_t, z_{it})^{\alpha} l(A_t, z_{it})^{\gamma} m(A_t, z_{it})^{\nu} \\ &= \frac{\omega_t c_f}{1-\xi} \end{aligned} \tag{8}$$

Aggregation In order to close the model, we need to aggregate firm-level production and employment to the economy-wide level. We can use the above-listed expressions for optimal firm production and employment and use the production function to obtain aggregates inputs and aggregate output. Despite the firm heterogeneity and the non-convex production structure, production and inputs aggregate nicely. In general, aggregates will be determined by the underlying productivity distribution of firms, the survival cutoff z_t^* and the measure of firms in the economy. Once, we know aggregate production and factor demand, will posit aggregate factor supply curves to compute factor prices and close the model. Aggregate output is

$$Y_t = \int_0^{N_t} y_{it} di.$$

We now rewrite aggregate output in various ways: First, We commute the integration to z rather than i; second, we take into account that only firms above the productivity threshold z_t^* are active in the economy; third, we make use of the expression for relative fir size in equation (7); fourth, we use the expression for output at the cutoff (8).

$$
\begin{aligned}
Y_t &= N_t \int_0^\infty y(A_t, z_{it}) dF_z(z) \\
&= N_t \int_{z_t^*}^\infty y(A_t, z_{it}) dF_z(z) \\
&= N_t^* \frac{1}{1 - F_z(z_t^*)} \int_{z_t^*}^\infty y(A_t, z_{it}) dF_z(z) \\
&= N_t^* y(A_t, z_t^*) \frac{1}{1 - F_z(z_t^*)} \int_{z_t^*}^\infty \frac{y(A_t, z_{it})}{y(A_t, z_t^*)} dF_z(z) \\
&= N_t^* y(A_t, z_t^*) \frac{1}{1 - F_z(z_t^*)} \int_{z_t^*}^\infty \left(\frac{z_{it}}{z_t^*}\right)^{\frac{1}{1-\xi}} dF_z(z) \\
Y_t &= N_t^* \frac{\omega_t c_f}{1 - \xi} \left[\frac{\zeta(z_t^*)}{z_t^*}\right]^{\frac{1}{1-\xi}} \qquad (9)
\end{aligned}
$$

where N_t is the total number of firms (active and exiting) and N_t^* is the measure of firms above the cutoff and who are producing output; N_t^* is what we see in the data. $\zeta(z_t^*) \equiv \left(\frac{1}{1-F_z(z_t^*)} \int_{z_t^*}^\infty z^{\frac{1}{1-\xi}} dF_z(z)\right)^{1-\xi} = \left(E\left[z^{\frac{1}{1-\xi}} \middle| z > z_t^*\right]\right)^{1-\xi}$ is an aggregate productivity index that is useful in expressing aggregate quantities. Aggregate output increases with the number of active firms, N_t^*, the firm size at the survival cutoff, $y(A, z^*)$, and declines with the cutoff z^* (Leibniz rule).

In a similar vein, one can compute aggregate demand for capital, labour and materials as

$$K_t = \int_{i=0}^{N_t} k_{it} di = \frac{\alpha}{r_t} Y_t \qquad (10)$$

$$L_t = \int_{i=0}^{N_t} l_{it} di = \frac{\gamma}{w_t} Y_t. \qquad (11)$$

Aggregate demand for overheads is given by the measure of active firms, N_t^*, because every active firm needs overhead inputs c_f in order to produce:

$$M_t = N_t^* c_f. \qquad (12)$$

Aggregate profits are

$$
\begin{aligned}
\Pi_t &= \int_{i=0}^{N_t} \pi_{it} di \\
&= \int_{i=0}^{N_t} (1 - \xi) y_{it} - \omega_t c_f di \\
&= (1 - \xi) Y_t - N_t^* \omega_t c_f
\end{aligned}
$$

I assume that net profits Π are distributed lump-sum to consumers who also provide all production factors to firms. That way, final consumers are paid income $r_t K_t + w_t L_t + N_t^* \omega_t c_f + \Pi_t$ which equals the value of aggregate output in the economy Y_t.

2.2 Returns to scale and industry characteristics

The industry equilibrium allows us to study the effect on returns to scale on moments in the observable data that are related to competition among firms: the number and average size of firms, the dispersion of total factor productivity across firms and the concentration ratio among firms. Most of these moments depend on the survival cutoff z^* which in turn depends on returns to scale ξ. The higher the returns to scale, the higher the survival cutoff and the lower the resulting TFP dispersion.

In the following we need to make some assumptions to relate returns to scale to the above listed data moments. In particular, we assume that the fixed overhead costs, c_f, the distribution from which the establishments draw their productivity shocks, F_z, and all factor prices are the same across industries. Further, we assume that the fixed overhead input is small enough so that higher production elasticities imply a higher productivity cutoff.[3] This means that higher returns to scale imply a higher survival cutoff. Intuitively, the closer returns to scale are to constant, the less unproductive firms with a low productivity z can play out their size advantage. In the extreme case of constant returns to scale, only the top firm with the highest productivity would survive.[4]

As long as returns to scale are decreasing, industries with returns to scale closer to unity should have a higher productivity cutoff. This means they would be characterized by fewer firms, larger firms, less TFP dispersion and higher concentration.

3 Empirical evidence on returns to scale and productivity

3.1 Data

We use the Census of Construction industries (CCN) and the Census of Manufactures (CMF) for our estimation. We describe details on the coverage, sampling and the measurement of inputs, outputs, industry, firm affiliation and establishment age in Appendix A. We focus our estimation at the plant level and limit our study to the 1982 ad 1987 Census. Only in these two Censuses do we observe hours worked for construction establishments; to maintain a consistent sample, we also focus on these two years in the manufacturing data. Some estimators require knowledge when a firm will exit or not (Olley and Pakes (1996)). We get this information from the Longitudinal Business Database (LBD): we define an establishment as exiting next period if that establishment

[3] This assumption corresponds to $\frac{d \log z^*}{d\alpha} = \log\left(\frac{r}{\omega c_f}\right) + \log\left(\frac{1-\xi}{\alpha}\right) > 0$ and similarly for γ and ν. Numerical simulations show that these are plausible assumptions.

[4] Strictly speaking, it would be hard to reconcile constant returns to scale, perfect competition and only one firm at the top.

exits in one of the subsequent four years until the next Census is taken. Summary statistics are displayed in Table 1.

Table 1: Summary Statistics

	Full Sample	Construction	Non-durable Manufacturing	Durable Manufacturing
Mean				
Output	6,693.89	2,504.92	12,145.89	7,044.72
Employment	39.187	17.41	52.84	50.88
Capital	4,047.65	676.26	8,287.17	4,429.63
Materials	3,163.75	743.74	6,721.93	3,092.20
Standard Deviation				
Output	67,358.18	14,093.51	98,557.08	72,223.04
Employment	224.83	94.96	207.86	308.61
Capital	6,349.74	87,133.13	38,952.73	
Materials	50,305.80	4,676.69	77,421.71	38,387.56
N	995,800	362,700	254,300	378,800

Note: This table reports summary statistics for the sample that is used for main specification. Output, Capital, and Materials are measured in thousand year-2005 dollars. Employment equals the total number of employees. Census disclosure rules require rounding the number of observations to the nearest hundred.

3.2 Results: Returns to scale

We assume, each firm i in an industry operates the same Cobb-Douglas production function

$$y_{it} = z_{it} k_{it}^{\beta_k} l_{it}^{\beta_l} m_{it}^{\beta_m} e_{it}^{\beta_e} \tag{13}$$

where k_{it}, l_{it}, m_{it}, e_{it} denote a firm's capital stock, hours worked, materials and energy input, respectively and the β's are production elasticities of each firm. Consistent with the theoretical model, we assume that each firm draws an idiosyncratic productivity level z_{it} around industry means A_t.

We choose the estimator suggested by Olley and Pakes (1996) as our benchmark model. It has the advantage to utilize information contained in an establishment's investment and to account for entry and exit. The former feature requires non-zero investment data which is not a very pressing problem in our data. The latter feature is important in the construction sector which is characterized by higher establishment turnover than manufacturing. Table 2 displays the estimated production functions for every sectoral subgroup – construction, nondurable manufacturing and durable manufacturing.

Overall returns to scale are slightly decreasing on average. Construction exhibits the lowest returns to scale (0.91), durable manufacturing the highest (0.97) and non-durable manufacturing

Table 2: Returns to Scale

	Whole Sample	Construction	Non-Durable Manufacturing	Durable Manufacturing
β_k	0.1291***	0.0592***	0.1342***	0.1374***
	(0.0273)	(0.0208)	(0.0286)	(0.0374)
β_l	0.2474***	0.3850***	0.2214***	0.2443***
	(0.0140)	(0.0067)	(0.0149)	(0.0146)
β_m	0.4980***	0.3903***	0.5082***	0.5087***
	(0.0227)	(0.0063)	(0.0257)	(0.0231)
β_e	0.0840***	0.0849***	0.0851***	0.0831***
	(0.0171)	(0.0050)	(0.0184)	(0.0182)
Returns to scale	0.9586	0.9194	0.9489	0.9735
N	995,800	362,700	285,900	347,200

Note: Results reported in this table are estimated using the strategy suggested by Olley and Pakes (1996). Output, capital, materials, and energy are measured in thousand year-2005 dollar value, employment in total hours worked. *, **, *** significantly different from 0 at the 10%, 5%, 1% level, respectively.

somewhere in between.

If we abstract from other aspects of marginal costs for a moment, this means that the slope of marginal costs is slightly increasing, especially for construction establishments. This means that demand fluctuations would lead to larger swings in marginal costs in construction compared to manufacturing. If the the cyclicality of markups does not differ too much in these sectors, then the price of construction output (the price of structure investment) fluctuate more procyclically than that of manufacturing. Indeed, this is true in the data where the price of structure investment relative to consumption goods is procycliacl, but the price of equipment investment relative to consumption goods fluctuates countercyclically. The traditional approach to explain these phenomena was to rely on equipment-specific technology shocks (see for example Greenwood et al. (2000); Fisher (2006)); Justiniano and Primiceri (2008) also found that shock to be an important driver of the Great Moderation. But given these differences in the slopes of marginal costs, sector-specific shocks – though they may play an additional role – are not necessary to explain the behavior of relative prices.

3.3 Comparing structural estimators

The empirical industrial organization literature has developed several approaches to structurally estimate production functions. Starting with the work by Olley and Pakes (1996), researchers have used assumptions on the timing of input choices and other variables outside the production function that depend on a firm's productivity (such as investment) to make inference on the production elasticities in (13) and TFP. Levinsohn and Petrin (2003) and Ackerberg et al. (2006) have expanded

Table 3: A Comparison of Different Estimation Strategy

	A.			B.		
	(I) OP	(II) LP		(III) GNR	(IV) ACF w/ investment	(V) ACF w/o investment
β_k	0.1291***	0.0762***	β_k	0.2382	0.1082	0.1074
	(0.0273)	(0.0331)				
β_l	0.2474***	0.2350***	β_l	0.3968	0.0932	0.1226
	(0.0140)	(0.0129)				
β_m	0.4980***	0.6311***	β_{ii}	0.3701	0.7853	0.7710
	(0.0227)	(0.0481)				
β_e	0.0840***	0.0703***				
	(0.0171)	(0.0107)				
Return To Scale	0.9586	1.0126	Return To Scale	1.0050	0.9867	0.5255
TFP Std.	0.3690	0.6109	TFP Std.	0.6229	0.5559	1.0010
N	995800	1008800	N	996,600	912,300	897,600

Note: Results in (I)-(III) are estimated using strategies suggested by OP1996, LP2005, GNR2013,respectively. Estimates in (IV) and (V) comes from ACF(2009). Specification (V) uses investment information to back out productivity, while (IV) doesn't. Capital is measured in 2005 dollar value. Employment is measured by total hours worked (th). For the first two group of estimates, material and energy are treated as two different inputs. For the last three group of estimates, material and energy are aggregated into one common intermediate input, because estimation result is very unstable if the two inputs are included seperately. Sample size differs across estimations because certain strategy doesn't run for some industries. $^{*, **, ***}$ significantly different from 0 at the 10%, 5%, 1% level, respectively.

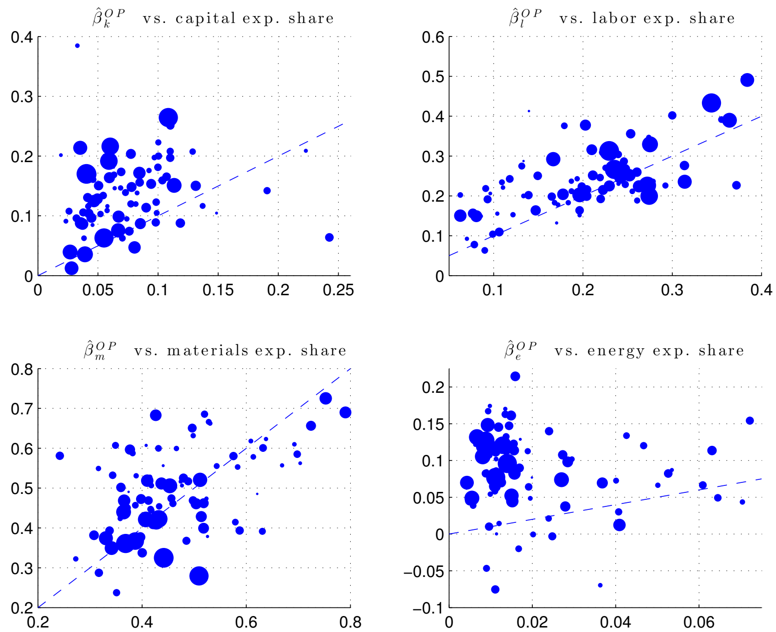
on that idea while Gandhi et al. (2013) utilize a firm's explicit first-order condition. Identification problems in some of these procedures are addressed in Wooldridge (2009). Grieco et al. (forthcoming) account for the biases caused by heterogeneous input prices. We employ and contrast some of the above-mentioned techniques on our fairly large dataset and compare their results and how stable they run.

Olley and Pakes (1996) This procedure does not run very smoothly in all industries: we only estimates for 90 out of the 95 industries; in another eight industries the procedure estimates negative coefficients for capital and/or energy. It is hard to pinpoint why the estimation does not run smoothly. We attribute this to the additional information contained in investment used in this procedure. If investment is not a strongly monotone function in TFP, then the inversion of the investment policy function is not well-defined and using investment information may cause more problems than it helps. A variant of the estimation procedure proposed by Ackerberg et al. (2006) which also uses information in investment suffers from similar problems. Since the Olley and Pakes (1996)-procedure also accounts for the selection step along productivity, it has problems estimating the hazard function in short panels when industries have either few observations or when the selection along productivity is weak – a feature that gains empirical support by Foster et al. (2013) for the recent recessions. Lastly, we find this estimation has convergence problems when we use employees rather than hours worked which may be due to the fact that many establishments may have a small-N problem in employees. To put the Olley-Pakes estimates into perspective, we plot them against the expenditure share on each input in Figure 2.

Levinsohn and Petrin (2003) This procedure was developed to circumvent the problem that investment data may be missing or zero. In our sample, however, this is not such a pressing probe and we do not loose a lot of information. So the estimation runs smoothly for all industries and it has no trouble with short or unbalanced panels. This is in some contrast to Olley and Pakes (1996), so we have the impression that not estimating the selection step and omitting investment is responsible for this. We do not find significant results when we separate materials and energy inputs in the production function or combine them into a single intermediate input. We find that the gross output specification overall runs more stable than the value added specification. Returns to scale are overall increasing (1.013) which is in contrast to almost all other procedures which estimate returns to scale to be decreasing. We plot the Levinsohn-Petrin estimates against the expenditure share on each input in Figure 3.

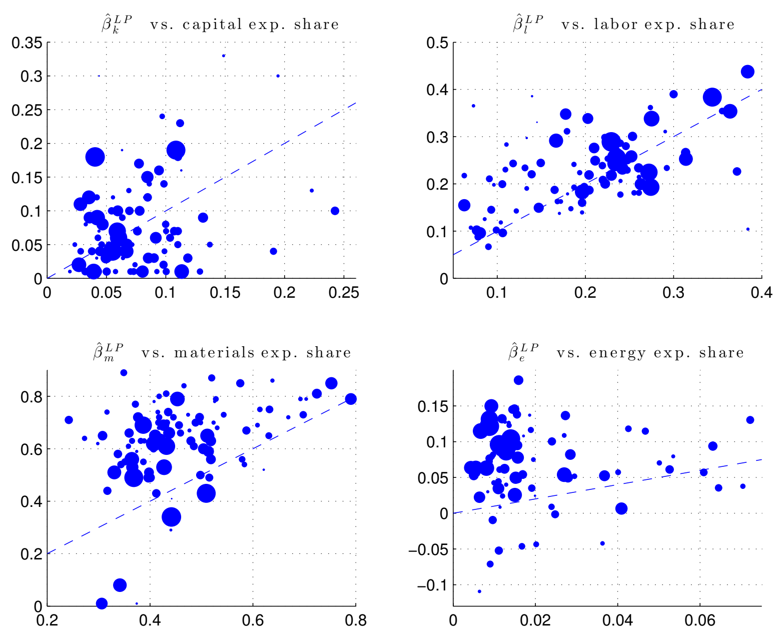
Ackerberg et al. (2006) Ackerberg et al. (2006) try to overcome the identification problems on Olley and Pakes (1996); Levinsohn and Petrin (2003) and add a timing assumption. This estimation comes in a version that utilizes investment (like Olley and Pakes (1996)) and one without. We find that both variants, especially the variant with investment (like Olley-Pakes), encounter severe convergence problems in our data. We therefor combine materials and energy into one intermediate

Figure 2: Olley-Pakes estimates vs. expenditure shares



Note: Production function elasticities are estimated using the method proposed by Olley and Pakes (1996) for 90 NAICS-4 industries in the manufacturing and construction sector. The size of the balls corresponds to an industry's size in terms of that industry's employment share. Choosing industry output or industry capital does not change the strongly positive relationship.

Figure 3: Levinsohn-Petrin estimates vs. expenditure shares

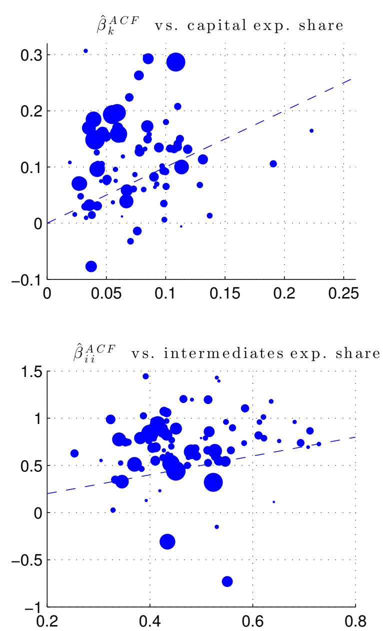


input:

$$y_{it} = z_{it} k_{it}^{\beta_k} l_{it}^{\beta_l} ii_{it}^{\beta_{ii}} \tag{13'}$$

This reduces dimensionality in the GMM step of the estimation and improves convergence, so that we get estimates in 87 industries, but a whopping 33 (!) of those estimate negative production elasticities. This is particularly true about the production elasticity of labor and negative estimates appear whether we use hours worked or employees in a long panel of all eight Censuses. In the variant with investment, we only obtain estimates in 80 industries; again, negative estimates emerge in 33 industries. Using the longer panel involving all eight Censuses makes the estimation run a bit more smoothly but does not help with the negative estimates, so estimating the stochastic process for TFP does not seem to be the primary cause for convergence problems or negative estimates. In the future, we plan to use all ASM years to see if the procedure runs more smoothly and produces better results with high-frequency and more data. We plot the Ackerberg-Caves-Frazer estimates against the expenditure share on each input in Figure 4.

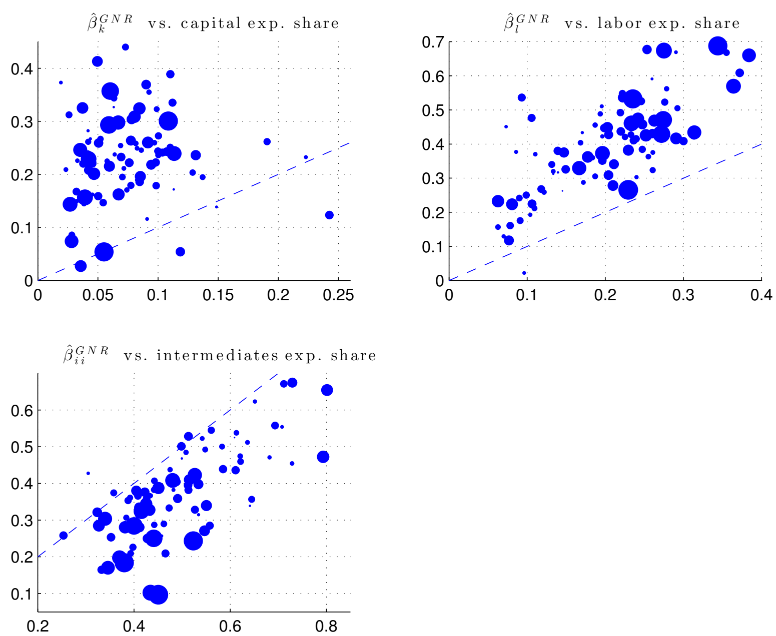
Figure 4: Ackerberg-Caves-Frazer estimates vs. expenditure shares



Gandhi et al. (2013) In contrast to the previous estimation techniques, Gandhi et al. (2013) use a firm's first-order condition for materials in addition to similar timing assumptions made above. Overall, this estimation procedure runs smoothly (convergence problems only in one industry), never produces negative production elasticities and it is quite fast too. But like with Ackerberg et al. (2006) above, we have to combine materials and energy into a combine intermediate input, otherwise, this processed is extremely unstable even in our large sample. It appears that the additional timing assumption and using the explicit first order condition helps reliably estimate the stochastic process for TFP; we ran the estimation on both a short and a long panel and didn't find significant differences in returns to scale or convergence problems.

One downside of this estimation is that the coefficient on materials will always be biased downward if the expenditure share on materials within an industry are very spread out. This means the residuals in the first step of the estimation (using the FOC on materials) are very dispersed and/or skewed to the right (this seems to be the case in the Census data), so β_i will be biased downward. As a consequence, β_k and β_l will be biased upward. Using our Census data, we find these general patterns to be the case if we compare these estimates to expenditure shares. The production elasticity on capital, for example, ranges from 0.15 to 0.3 in a gross output specification. When materials and energy are included as a separate inputs, the estimation not only has convergence problems, but these biases seem to become more attenuated. Lastly, the authors' main claim is that their estimation procedure correctly uses a firm's profit maximization conditions which leads to more narrow estimates of TFP dispersion. We cannot confirm this: the resulting standard deviation of TFP within narrow industries is 0.39 compared to 0.36 (Olley-Pakes) or 0.39 (Levinsohn-Petrin). Returns to scale are estimated to be more or less constant: 1.005. We plot the Gandhi-Navarro-Rivers estimates against the expenditure share on each input in Figure 5.

Figure 5: Gandhi-Navarro-Rivers estimates vs. expenditure shares

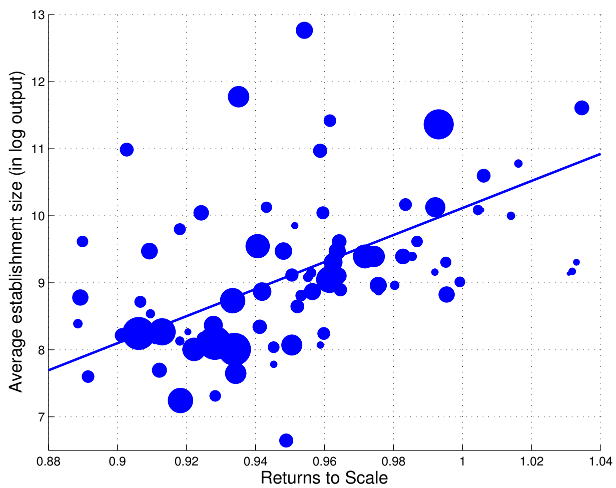


3.4 Industry evidence: Returns to scale and the firm distribution

In the model section, we established a relationship between returns to scale on the one hand and several data moments on the other hand. We showed that higher returns to scale should be associated with fewer, larger establishments, a higher establishment concentration and lower TFP dispersion. We will now utilize the industry heterogeneity in returns to scale to find out about the predictions of our model. We estimate returns to scale using our benchmark specification, the method suggested by Olley and Pakes (1996), for each of the 90 NAICS-4 industries in the construction and manufacturing sector.[5]

3.4.1 Higher returns to scale imply larger units

Figure 6: Returns to scale vs. establishment size



Note: Returns to scale are estimated using the method proposed by Olley and Pakes (1996) for 90 NAICS-4 industries in the manufacturing and construction sector. Size is measured in terms of gross output (in logs of 1,000 year-2005 dollars). The size of the balls corresponds to an industry's size in terms of that industry's employment share. Choosing industry output or industry capital does not change the strongly positive relationship. For robustness purposes we drop the top/bottom four industries (roughly the top/bottom 5%) of industries with outlier returns to scale: Grain and Oilseed Milling (3112), Apparel Accessories and Other Apparel Manufacturing (3159), Printing and Related Support Activities (3231), Lime and Gypsum Product Manufacturing (3274), Nonferrous Metal (except Aluminum) Production and Processing (3314), Computer and Peripheral Equipment Manufacturing (3341), Semiconductor and Other Electronic Component Manufacturing (3344) as well as Manufacturing and Reproducing Magnetic and Optical Media (3346). Not surprisingly, some of these industries are not considered manufacturing anymore (3231) or have extremely spotty data due to extensive outsourcing (3159).

First, we will study the prediction that higher returns to scale imply larger establishment

[5] We drop five industries – Land Subdivision (2372), Other Heavy and Civil Engineering Construction (2379), Beverage Manufacturing (3121), Tobacco Manufacturing (3122) and Iron and Steel Mills and Ferroalloy Manufacturing (3311) – because too few or too spotty information prevent the GMM estimation step from converging. Coincidentally, these are also industries with so few firms (as opposed to establishments) that would pose disclosure problems according to Census guidelines.

sizes. This is because it is more efficient to concentrate production in larger production units. Establishments with a low total factor productivity z cannot compensate their inefficiency with a lower production scale which they would do when returns to scale are strongly decreasing.

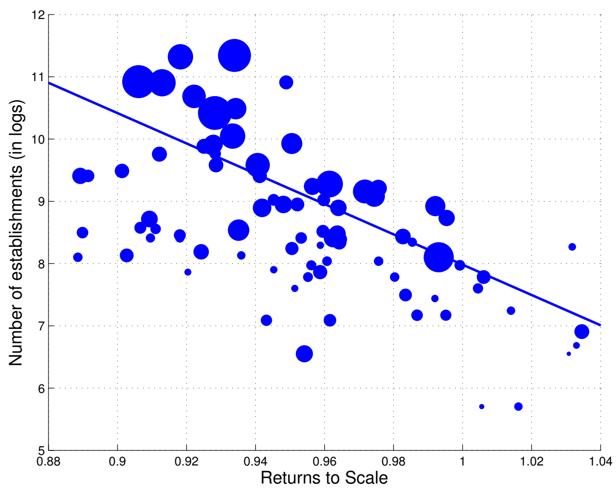
Figure 6 shows the relationship between returns to scale and the average establishment size. In fact, industries that are characterized by higher returns to scale tend to be populated by larger establishments. This size difference is significant: The inter-decile range across industries in terms of returns to scale is 0.113 which is associated with a log-size difference of 2.28. This means that the average establishment in the top decile industry (returns to scale equal to 1.02) is $e^{2.28} = 9.8$ times as large as the average establishment in the bottom quartile industry (returns to scale equal to 0.9). Although these numbers were obtained with measuring size in terms of gross output, the same results hold when choosing employment or the capital stock as a measure of establishment size.

3.4.2 Higher returns to scale imply fewer units and more concentration

Next, we turn to how many establishments populate an industry and how concentrated they are. This aspect is closely related to the average establishment size considered before. When industries are characterized by higher returns to scale, one would expect that several small establishments merge and thus leverage the size advantages of a technology that is closer to increasing returns. Likewise, large establishments in industries with strongly decreasing returns to scale may find it profitable to split up into many smaller production units. We plot the relationship between returns to scale and the number of establishments in Figure 7 and that between returns to scale and concentration ratio measures in Figure 8. As predicted, there are considerably fewer production units in industries with higher returns to scale and these industries are also more concentrated. The difference in terms of numbers is very significant: the inter-decile range across industries is 2.75 in logs; this means there are about $e^{2.75} = 15.7$ times as many establishments in the bottom decile industry (returns to scale 0.9) compared to the top decile industry (returns to scale 1.02).

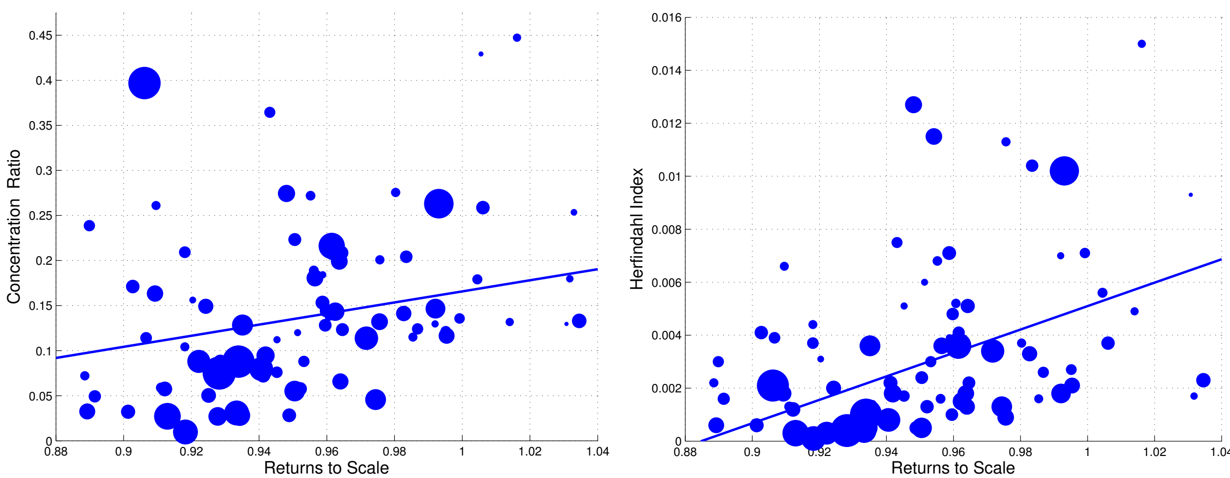
The relationship between returns to scale and the concentration of establishments is similar though not as strong as that between returns to scale and the number of establishments. We consider two concentration measures plotted in Figure 8: the market share of the two larges units in the industry (left panel) and the industry's Herfindahl index (right panel). The inter-decile range is 0.0696 which measn that the market share of the two larges units is about 7% larger than in the bottom decile industry. The Herfindahl index is also larger in higher returns to scale industries, but its level is quite low in general not exceeding 0.015. We want to emphasize that we consider the concentration across establishments, not firms; so this number hardly compares to the antitrust literature which considers firm concentration.

Figure 7: Returns to scale vs. the number of establishments



Note: For details see notes to Figure 6.

Figure 8: Returns to scale vs. concentration

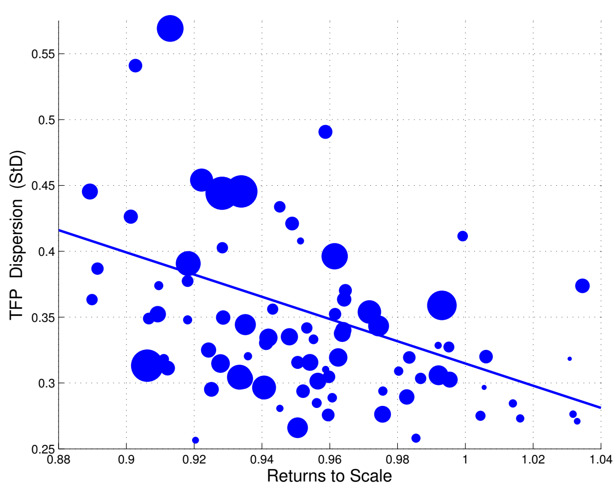


Note: Left panel displays the concentration ratio of the top two production units. We drop Vehicle Manufacturing (3361) which has an outlier concentration ratio of 0.8; this does not change the slope or significance of the regression line. The right panel displays the Herfindahl index. In the right panel we drop outlier industries Audio and Video Equipment Manufacturing (3343) and Railroad Rolling Stock Manufacturing (3365) which have an Herfindahl index of 0.050 and 0.032, respectively; including them would strengthen the positive relationship. For other details see notes to Figure 6.

3.4.3 Higher returns to scale imply less TFP dispersion

Next we turn to the relationship between returns to scale and TFP dispersion. This data moment is important because TFP dispersion is commonly interpreted as a sign of inefficiency. But the model in Section 2 showed that high TFP dispersion may persist in perfectly competitive industries when returns to scale are sufficiently decreasing. We will now treat this implication of the model and plot the relationship between returns to scale and TFP dispersion. Again, the model predictions are borne out in the data: higher returns to scale mean low-TFP establishments can hardly survive thus truncating the TFP distribution on the left and lowering TFP dispersion. Take the industry at the bottom decile (returns to scale 0.9) which typically has a standard deviation of TFP of about 0.4. This means an establishment that is one standard deviation less productive than the average establishment produces $e^{0.4} \approx 50\%$ less with the same inputs than the average establishment in that industry. That such relatively unproductive establishments manage to survive here is due to the strongly decreasing returns to scale which allows them to compensate their low TFP with being small. How does that compensation look like for the industry with just above constant returns to scale (the top decile industry has returns to scale of 1.02)? The standard deviation of TFP within that industry is 0.3 which means that an establishment with TFP of one standard deviation below the average TFP establishment produces only 35% less with the same inputs. That is, the least surviving establishments in constant-returns-to-scale industries are not as unproductive as their least surviving peers in decreasing-returns-to-scale industries.

Figure 9: Returns to scale vs. TFP dispersion



Note: For details see notes to Figure 6.

4 Conclusion

We have estimated returns to scale using various modern structural estimation techniques and found them to range between 0.86 and 1.3. These differences are remarkable and shape the competitive environment that determines the distribution and differences across firms in an industry. In particular, we have shown, that higher returns to scale are associated with a more compressed size distribution of firms, larger and fewer firms as well as a lower dispersion of total factor productivity. This information is relevant in assessing the efficiency and welfare consequences of TFP dispersion.

We hope that our extensive empirical work will serve as a useful calibration source for the literature on firm heterogeneity. The dynamics at the firm and aggregate levels as well as industry outcomes such as dispersion and concentration in these models greatly depend on returns to scale.

References

Daniel A. Ackerberg, Kevin Caves, and Garth Frazer. Structural identification of production functions. *Working Paper*, 2006.

Ufuk Akcigit, Daron Acemoglu, Nicholas Bloom, and William Kerr. Innovation, reallocation and growth. *NBER Working Paper No. 18993*, April 2013.

Enghin Atalay, Ali Hortaçsu, and Chad Syverson. Vertical integration and input flows. *American Economic Review*, 104(4):1120–1148, April 2014.

Rüdiger Bachmann, Ricardo J. Caballero, and Eduardo M. R. A. Engel. Aggregate implications of lumpy investment: New evidence and a DSGE model. *American Economic Journal: Macroeconomics*, 5(4):29–67, October 2013.

Susanto Basu. Procyclical productivity: Increasing returns of cyclical utilization? *Quarterly Journal of Economics*, 111(3):719–751, August 1996.

Susanto Basu and John G. Fernald. Are apparent productive spillovers a figment of specification error? *Journal of Monetary Economics*, 36(1):165–188, August 1995.

Susanto Basu and John G. Fernald. Returns to scale in U.S. production: Estimates and implications. *Journal of Political Economy*, 105(2):249–283, April 1997.

Susanto Basu, John G. Fernald, Jonas D. M. Fisher, and Miles S. Kimball. Sector-specific technical change. *Working Paper*, April 2010.

Mark Bils and Peter J. Klenow. Using consumer theory to test competing business cycles models. *Journal of Political Economy*, 106(2):233–261, April 1998.

Nicholas Bloom. The impact of uncertainty shocks. *Econometrica*, 77(3):623–685, May 2009.

Nicholas Bloom, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen Terry. Really uncertain business cycles. *NBER Working Paper No. 18245*, July 2012.

Craig Burnside. Production function regressions, returns to scale and externalities. *Journal of Monetary Economics*, 37(2):177–201, April 1996.

Craig Burnside, Martin Eichenbaum, and Sergio Rebelo. Capital utilization and returns to scale. In Ben S. Bernanke and Julio J. Rotemberg, editors, *NBER Macroeconomics Annual*, volume 10, chapter 2, pages 67–110. MIT Press, January 1995.

Ricardo J. Caballero and Mohamad L. Hammour. The cleansing effect of recessions. *American Economic Review*, 84(5):1350–68, December 1994.

Andrea L. Eisfeldt and Dimitris Papanikolaou. Organization capital and the cross-section of expected returns. *Journal of Finance*, 68(4):1365–1406, August 2013.

Jonas D. M. Fisher. The dynamic effects of neutral and investment-specific technology shocks. *Journal of Political Economy*, 114(3):413–451, June 2006.

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

Lucia Foster, Cheryl Grim, and John Haltiwanger. Reallocation in the Great Recession: Cleansing or not? *Center for Economic Studies Working Papers 13-42*, 2013.

Jeremy T. Fox and Valérie Smeets. Does input quality drive measured differences in firm productivity? *International Economic Review*, 52(4):961–989, November 2011.

Amit Gandhi, Salvador Navarro, and David Rivers. Identifying production functions using restrictions from economic theory. *Working Paper*, June 2013.

Fabio Ghironi and Marc J. Melitz. International trade and macroeconomic dynamics with heterogeneous firms. *Quarterly Journal of Economics*, 120(3):865–915, August 2005.

Simon Gilchrist and John C. Williams. Putty-clay and investment: A business cycle analysis. *Journal of Political Economy*, 108(5):928–960, October 2000.

Edward Glaeser, Joshua D. Gottlieb, and Kristina Tobio. Housing booms and city centers. *American Economic Review Papers and Proceedings*, 102(3):127–133, May 2012.

Edward Glaeser, Joshua D. Gottlieb, and Joseph Gyourko. Can cheap credit explain the housing boom? In Edward Glaeser and Todd Sinai, editors, *Housing and the Financial Crisis*, chapter 7, pages 301–360. University of Chicago Press, August 2013.

Jeremy Greenwood, Zvi Hercowitz, and Per Krusell. The role of investment-specific technological change in the business cycle. *European Economic Review*, 44(1):91–115, January 2000.

Paul L. E. Grieco, Shengyu Li, and Hongsong Zhang. Production function estimation with unobserved input price dispersion. *International Economic Review*, forthcoming.

Robert E. Hall. Invariance properties of Solow's residual. In Peter A. Diamond, editor, *Growth, Productivity, Unemployment: Essays to Celebrate Bob Solow's Birthday*, chapter 5, pages 71–111. MIT Press, 1990.

Chang-Tai Hsieh and Peter J. Klenow. Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics*, 124(4), November 2009.

Alejandro Justiniano and Giorgio E. Primiceri. The time varying volatility of macroeconomic fluctuations. *American Economic Review*, 98(3):604–641, June 2008.

Alejandro Justiniano, Giorgio E. Primiceri, and Andrea Tambalotti. Investment shocks and business cycles. *Journal of Monetary Economics*, 57(2):132–145, March 2010.

Alejandro Justiniano, Giorgio E. Primiceri, and Andrea Tambalotti. Investment shocks and the relative price of investment. *Review of Economic Dynamics*, 14(1):102–121, January 2011.

Matthias Kehrig. The cyclicality of productivity dispersion. *Working Paper*, 2015.

Aubhik Khan and Julia K. Thomas. Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics. *Econometrica*, 76(2):395–436, March 2008.

James Levinsohn and Amil Petrin. Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70(2):317–341, April 2003.

Jacob Marschak and William H. Andrews. Random simultaneous equations and the theory of production. *Econometrica*, 12(3&4):143–205, July-October 1944.

Marc J. Melitz. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725, November 2003.

G. Steven Olley and Ariel Pakes. The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297, November 1996.

Jeffrey S. Passel. The size and characteristics of the unauthorized migrant population in the U.S.: Estimates based on the march 2005 current population survey. Technical report, Pew Hispanic Center, March 2006.

Diego Restuccia and Richard Rogerson. Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11(4):707–720, October 2008.

Albert Saiz. The geographic determinants of housing supply. *Quarterly Journal of Economics*, 125 (3):1253–1296, August 2010.

Valérie Smeets and Frederic Warzynski. Estimating productivity with multi-product firms, pricing heterogeneity and the role of international trade. *Journal of International Economics*, 90(2): 237–244, July 2013.

Chad Syverson. Product substitutability and productivity dispersion. *Review of Economics and Statistics*, 86(2):534–550, May 2004.

Chad Syverson. What determines productivity? *Journal of Economic Literature*, 49(2):326–365, June 2011.

Julia K. Thomas. Is lumpy investment relevant for the business cycle? *Journal of Political Economy*, 110(3):508–534, June 2002.

Jeffrey M. Wooldridge. On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters*, 104(3):112–114, September 2009.

A Data and measurement

A.1 Data description

In this appendix we describe the four datasets collected by the U.S. Census Bureau: the Aunnal Survey of Manufactures (ASM), the Census of Manufactures (CMF), the Census of Construction Industries (CCN), and the Longitudinal Business Database (LBD). We will explain below how we construct the estimation sample from them. Additional data come from the National Bureau of Economic Research Manufacturing Database (NBER-CES), the price deflator series of the Bureau of Labor Statistics (BLS) and asset data from the Bureau of Economic Analysis (BEA).

The main data sources are the CMF and CCN. Both are national economic surveys conducted at quinquennial frequency (in years ending in 2 and 7). All existing and legally registered establishments in the construction and manufacturing sectors (NAICS 23 and 31-33) are surveyed via mail where establishment is defined as any distinct whose the predominant activity is production/construction. Pure adiministrative establishments are hence excluded. Each establishment is obliged to complete the survey for itself, and if a parent company owns more than one establishments, each of its subsidiaries will have to submit a seperate report on its own operational details. In each Census, CMF covers 300,000-350,000 establishments and CCN includes 150,000-200,000 plants. We choose not to incorporate ASM data due to disparity in sampling method and survey frequency between ASM and the census surveys. The ASM is conducted in non-census years merely for a sample about 50,000-60,000 establishments in the manufacturing sector.

In CMF and CCN, establishments are identified by a unique code that doesn't change in case of ownership change or temporary plant shutdown. This identifier, named Permanent Plant Number (`PPN`) before 2002, is replaced by the Survey Unit ID (`SURVU_ID`) afterwards. If an establishment exits permanently, its identifier will not be reassigned to a new-born establishment. We carefully map the more recent identifier to `PPN` using `LBDNUM` which uniquely identify an establishment in the LBD, the comprehensive panel of the entire universe of establishments. The fact that both `PPN` and `SURVU_ID` can be mapped to `LBDNUM` not only helps match `PPN` with `SURVU_ID` but enables us to observe whether an establishment is an entrant, a continuer, or one to exit in the next period. Besides, establishments that belong to the same legal firm carry the same firm identifier `FIRMID`.

Construction Establishments We have eight Censuses of Construction Industries available from 1972 to 2007 each of which covers the full cross section of existing establishments in that sector in that year. The Census Bureau has changed questionnaires multiple times throughout the years. Generally speaking, all questionnaires collect information on revenues and input costs as well as the number of employees and hours worked. However, the way Census records sales and costs differs across years, and some information is only collected in some years (for example, hours worked only in 1982 and 1987). These disparities add to the difficulty of getting variables necessary for estimation such as output and all kinds of input. We made every effort to create a consistent sample over years. We will next introduce how important variables are defined and created.

Manufacturing Establishments We have access to the analogous Census of Manufactures available from 1972 until 2007; additionally, there is an annual subset of manufacturing establishments – the Annual Survey of Manufactures (ASM) – which helps us impute some relevant information. Previous work has greatly facilitated the longitudinal consistency of observed establishments and the proper measurement of inputs and outputs (see for example the data appendix to Kehrig (2015)). The raw construction data, in contrast, we had to clean and make consistent from scratch.

A.2 Measurement of entry and exit

The last piece of information necessary to complete our sample is entry and exit behaviors. Exit information is needed when we use the estimator proposed by Olley and Pakes (1996) and Ackerberg et al. (2006). The relationship between entry and exit rates and return to scales is also of interest. Thus we create two dummies for establishments being an entrant and being ready to exit in the next year. To do this, we match the Census data to the Longitudinal Business Database (LBD), which records registration information of establishments in all industries at an annual frequency. The LBD identifies establishments with a unique LBD number which does not change over time, and the Census records `LBDNUM` in all eight Censuses. This makes it possible to match establishments in the two datasets by LBD number. We define an establishment in a certain year as an entrant when its `LBDNUM` is not observed in either dataset in the year before this year. An establishment is ready to exit in a certain year when it is not observed in the LBD the next year. When a firm enters (exits) more than once, we don't distinguish between the first entry (exit) and subsequent entry (exit).

A.3 Measurement of production

Construction Establishments Construction establishments generate revenues from various sources that are displayed in Table 4. Receipts for construction work carried out by the establishment itself (Item II.a.1) is the main source of cash flow. But establishments also receive cash flow for structures they subcontracted out to other construction establishments (II.a.2) whether or not they belong to same firm. Of course the compensation to these other establishments are not recorded under sales but in the materials section of the Census. Combined, these two items make up the "Receipts from Construction Work" (Item II.a). The Census of Construction also records a "Speculative Value of Construction" (Item II.b). This includes the estimated value of work done on projects that have not been completed during the Census year or projects that are not entirely sold yet. Receipts from construction work and the value of speculative construction combine to the "Total Value of Construction Work" put in place by an establishment. Some construction establishments also carry out non-construction activities such as building service and maintenance; sales from these side businesses are reported as "Receipts from Non-Construction Work." Over four

Table 4: Breakdown of a construction establishment's revenues

<table>
<tr><td colspan="4">I: Total Receipts</td></tr>
<tr><td colspan="3">II: Total Value of Construction Work</td><td rowspan="3">III: Receipts from Non-Construction Work</td></tr>
<tr><td colspan="2">II.a: Receipts from Construction Work</td><td rowspan="2">II.b: Value of speculative construction</td></tr>
<tr><td>II.a.1: Receipts for Construction Done In-house</td><td>II.a.2: Receipts for Construction Subcontracted Out</td></tr>
</table>

decades of data collection in the economic census, some of the above listed variables have been added and or have changed their label over time. Table 5 displays which variable appears under what name in what year.[6]

[6] We thank Susan Bucci and Michael Blake for extensive discussions and help with understanding how each Census of Construction industries was collected and recorded.

Table 5: Labels of revenue sources

Census	I	II	II.a	II.a.1	II.a.2	II.b	III
1972-1987	TR	-/-	CR	NCR	SO	-/-	BR
1992-1997	TR	CV	CR	NCR	SO	SV	BR
2002	RCPCTOT	-/-	RCPCWRK	RCPNCW	CSTSCNT	-/-	RCPOTH
2007	TVS	-/-	RCPCWRK	RCPNCW	CSTSCNT	-/-	RCPOTH

We are interested in a measure of an establishment's construction output which is the sum of "Receipts for Construction Done In-house" and the "Value of Speculate Construction." But the latter is only reported in two Census years and quantitatively pales in comparison to the former. For the sake of longitudinal consistency, we ignore the value of speculative work which also does not play a significant role quantitatively.

Net receipts from construction work done (NCW) is surveyed in all eight Censuses, as well as other sales measurements including total revenue from construction work (RCPNCW) and receipt from construction work subcontracted to others (CSTSCNT). In cases where the reported net receipts are missing, zero or negative, we calculate it by subtracting "Receipts from Construction Work Subcontracted to Others" (CSTSCNT) from "Total construction receipts" (RCPNCW). When calculating output, it is important to net out contract work which is much more prevalent in construction than it is in manufacturing.

We then deflate net construction receipts to get the real value of output. Our primary price index is an industry-specific price deflator for construction from the National Income and Product Accounts (NIPA) database. NIPA provides indices for residential and non-residential sectors as well as total outputs in the two sectors. Ideally, sales should be deflated by sector-specific price index. However, most establishments carry out construction projects in multiple sectors, and although the Census surveys revenue by type of activity after 1997, data in this category is missing for a large proportion of establishments. Thus we produce an industry average price index from the residential and non-residential price indices weighted by total outputs. Specifically, we calculate

$$P^{cstr} = \frac{P^{res} \times Q^{res} + P^{nres} \times Q^{nres}}{Q^{res} + Q^{nres}}.$$

in which P^{res} and P^{nres} are price indices for residential and non-residential sectors, respectively, and Q^{res} and Q^{nres} are total outputs for the two sectors; both are taken from Section 1 tables in the NIPA's information on industry input and outputs.[7] We use this averaged index as the NIPA price deflator for the whole industry.

For the industry "Residential and Non-Residential Buildings" (NAICS 236), we alternatively deflate the nominal receipts with a series of more localized price indices from the RS Means Database. RS Means provide the regional price indices per quality-adjusted square foot of construction for 199 cities and CSBAs (Core Based Statistical Area) across the country. We are able to get county-level price deflators by mapping the cities or CBSAs to counties, and then use geographical information included in CCN to match establishments to this set of deflators. Real outputs deflated by NIPA indices and the RS Means indices are not significantly different from each other. To keep consistency across all construction data, we use real output deflated by NIPA indices to construct our

[7]NIPA Section 1 tables can be found here: http://www.bea.gov/national/nipaweb/DownSS2.asp.

estimation sample. Output in the construction Censuses is thus defined as

$$\texttt{Q} = \begin{cases} \frac{\texttt{NCW}}{P^{cstr}} & \text{if } \texttt{NCW} \text{ exists} \\ \frac{\texttt{RCPNCW}-\texttt{CSTSCNT}}{P^{cstr}} & \text{if } \texttt{NCW} \text{ missing;} \end{cases}$$

where P^{cstr} is the NIPA price index of construction.

Manufacturing Establishments In the manufacturing data, we correct sales (TVS) by subcontracted work (CR) as well as changes in final and intermediate goods inventories (FIE − FIB and WIE − WIB, respectively).

$$\texttt{Q} = \frac{\texttt{TVS} - \texttt{CR} + \texttt{FIE} - \texttt{FIB} + \texttt{WIE} - \texttt{WIB}}{P^{ship}} \tag{14}$$

where TVS means total value of shipments (corresponding to RCPNCW in the 2002/07 CCN), CR cost of resales (corresponding to CSTSCNT in CCN) and FIE − FIB and WIE − WIB are changes in final and work-in-progress inventories which do not play a significant role in construction. P^{ship} is an industry-specific price deflator from the NBER-CES Manufacturing database.

A.4 Measurement of labor input

Construction Establishments Our main estimation utilizes hours worked as a measurement for labor input as explained in the body of this paper. All eight Censuses record total employment (TE), the number of construction workers (CE), total wages (TW), and wages paid to construction workers (CW). However, information on hours worked is merely collected in two census years, 1982 and 1987, and Census surveyed hours worked by construction workers (CH) instead of total hours worked by all employees. All hours measures appear to include leased workers as well.[8] We compute total hours worked under the assumption that wage rates are the same for construction and non-construction workers. More specifically, total hours worked (TH) is calculated as

$$\texttt{TH} = \texttt{CH}\frac{\texttt{TW}}{\texttt{CW}}$$

Here we assume that hourly wages paid to construction workers are the same as average wages for all workers in the construction sector. We believe this to be a reasonable assumption for two reasons. First, we compute and compare earnings per worker among construction employees, $\frac{\texttt{CW}}{\texttt{CE}}$, and average earnings of all employees, $\frac{\texttt{TW}}{\texttt{TE}}$. The two average earnings are roughly the same. Second, according to employment statistics published by the BLS, hourly wages for all occupations in the construction sector are similar to wages for construction workers only. In 2013, for example, the average wage in the construction sector is \$20.11 and the hourly wage for construction workers is \$19.13. The first data on occupational employment for the construction sector is collected in 1997, when the industry-level average wage rate is \$15.47 and the hourly wage for construction workers is \$14.33.

We take the following path to construct all the variables needed to infer total hours worked from the raw data. Take construction employment (CW) as an example; Census records construction employment in each quarter (EMPQCW) and the average annual employment (AVGEMPCW) which

[8] For example, Census distinguishes between own and leased employed ("Total Workers" versus "Total workers, non-leased"), but there is no similar distinction for hours worked. We assume that leased construction worker hours worked are contained in total construction hours worked.

by definition should equal the average of employments in the four quarters. In practice, average employment doesn't always equal the mean of quarterly employments. This disparity may be attributed to the fact that the Census has edited quarterly employment and corrected some implausible values. To that end, we construct number of construction employees by computing the mean of quarterly employment, and in cases where the computed mean is non-positive, we use reported average employment to supplement:

$$\texttt{AVGEMPCW} = \frac{1}{4}\sum_{t=1}^{4}\texttt{EMPQCW(t)}.$$

The same procedure is also applied to TW, CW and CH.

Manufacturing Establishments For manufacturing establishments, we construct analogously:

$$\texttt{TH} = \texttt{PH}\,\frac{\texttt{SW} + \texttt{0.5} \times \texttt{WW}}{\texttt{1.5} \times \texttt{WW}}$$

where TH are total hours worked, PH production hours, SW all salaries and wages paid to the workforce, WW the wage bill for production workers. Again, we assume that non-production workers earn a premium of 1.5 times the hourly wage rate of that of production workers.

A.5 Measurement of capital input

Construction Establishments Capital input is measured by the real value of capital stock available for production. Theoretically, the capital stock should equal the replacement value of fixed assets in constant dollars. In all years but 1972, the Census surveys the book value of total assets at the beginning of the year. In 1972, the raw data contains only the book value of total assets at the end of the year (TAE), total depreciation changes (TD), and total capital expenditure (TCE) which is equivalent to investments. We back out the book value of total assets at the beginning of the year (TAB) with the variables above. In principle, the accumulation of assets follows: $\texttt{TAE} = \texttt{TAB} + \texttt{TCE} - \texttt{TD}$, so $\texttt{TAB} = \texttt{TAE} + \texttt{TD} - \texttt{TCE}$. We then follow similar transformation as in Kehrig (2015) to construct deflators for asset values and investment price. Briefly speaking, the replacement value of an establishment's capital stock in constant dollars will be calculated as:

$$K_t = \texttt{TAB}_t \frac{CC_t}{HC_t}\frac{1}{P_t^{Inv}}$$

where HC_t and CC_t are historical-cost and current-cost estimates of capital stocks of 3-digit NAICS, respectively, and P_t^{Inv} is the investment price deflator. All indices are published by the Bureau of Economic Analysis (BEA).[9]

CCN records the total book value of both structure (buildings) and equipments (machinery) combined. However, the BEA publish price indices for structures and equipments seperately rather than one series of composite price deflators. We transform TAB into a market value of the capital stock in constant dollars assuming that establishments in construction sector share a similar

[9]The indices can be found in Section 5 of the NIPA tables: http://www.bea.gov/national/nipaweb/DownSS2.asp.

structure-to-equipment ratio (in real values). This means we assume

$$\frac{K_i^{st}}{K_i^{eq}} = \frac{CC^{st}/P^{st}}{CC^{eq}/P^{eq}} \qquad \forall i \text{ construction establishments.}$$

CC^{st} and CC^{eq} are the industry-wide stock of nominal structures and equipment, respectively, and P^{st} and P^{eq} are the price deflators for structure and equipment investment, respectively. This condition will help us to pin down the capital stock. Let BAB and MAB denote the book value of structure and equipment asset stocks, respectively. We observe their sum, $\texttt{TAB} = \texttt{BAB} + \texttt{MAB}$, in the data. Then we can write,

$$\begin{aligned} K &= K^{st} + K^{eq} \\ &= \texttt{BAB}\frac{CC^{st}}{HC^{st}}\frac{1}{P^{st}} + \texttt{MAB}\frac{CC^{eq}}{HC^{eq}}\frac{1}{P^{eq}} \\ &= \texttt{BAB}\left(\frac{CC^{st}}{HC^{st}}\frac{1}{P^{st}} - \frac{CC^{eq}}{HC^{eq}}\frac{1}{P^{eq}}\right) + \texttt{TAB}\frac{CC^{eq}}{HC^{eq}}\frac{1}{P^{eq}} \\ &= \texttt{TAB}\left(\frac{1}{P^{st}}\frac{CC^{st}}{HC^{st}+HC^{eq}} + \frac{1}{P^{eq}}\frac{CC^{eq}}{HC^{st}+HC^{eq}}\right). \end{aligned}$$

where the last line follows from the assumption that all firms in an industry share the same ratio of real structures to real equipment capital.

Manufacturing Establishments In principle the same procedure can be applied that we used for the construction establishment. But our data task is facilitated by the fact that the Census of Manufactures collected the book values of structures (BAB) and equipment (MAB) separately until 1987:

$$\begin{aligned} K &= K^{st} + K^{eq} \\ &= \texttt{BAB}\frac{CC^{st}}{HC^{st}}\frac{1}{P^{st}} + \texttt{MAB}\frac{CC^{eq}}{HC^{eq}}\frac{1}{P^{eq}} \end{aligned}$$

Since then, we apply the same imputation as for the construction establishments described above.

A.6 Measurement of materials input

Construction Establishments Material input equals the real value of materials consumed in the surveyed period. Materials consumed are either purchased from the market, which is recorded by the Census as cost of materials and parts (CP), or taken out from a firm's materials inventories. The difference between inventories of materials at the beginning and the end of the year is denoted by $\texttt{CHANGEM} = \texttt{MIB} - \texttt{MIE}$. In the case where MIB and/or MIE are missing, we use the difference between beginning-of-year and end-of-year inventories from administrative reports (i.e., $\texttt{CHANGEM} = \texttt{INVBOY\_A} - \texttt{INVEOY\_A}$) as an alternative, assuming that all changes in inventory are due to changes in materials inventories. This is likely a reasonable assumption because construction establishments can hardly hold inventories of output goods (as manufacturers do) and typically purchase energy when necessary (most energy inputs such as gasoline, crude or liquid gas are not cheaply stored). Meanwhile, fuel storage has never been mentioned or reported. We calculate nominal value of materials consumed by adding up CP and CHANGEM. When this value turns out negative, we assume it to be a measurement error and replace the negative value with zero.

The price deflator for material input comes from the Federal Reserve Bank of St. Louis which offers a market price index for materials specific to the construction industry. This index reveals how the market price for materials changes over time. Rigorously speaking, materials taken out from the inventory should ideally be deflated by an inventory price index, an object that we failed to find for construction sector. This is probably because inventories account merely for a small proportion of total inputs or even of material inputs in this industry. As can be seen from the data, only 13% (185,400 out of 1,423,600) of the establishments reported a non-zero change in inventories at all, and among those observations who did report a change in inventories, the change in inventory only accounts for 2% of total materials consumed. To that end, using index of market price for materials as the deflator may not lead to significant bias in the measurement of inputs.

Therefore, the material input, m, is computed as:

$$\mathtt{M} = \max\left\{0, \frac{\mathtt{CP} + \mathtt{CHANGEM}}{P^{mat}}\right\},$$

where P^{mat} is the industry-specific market price deflator of materials from the Federal Reserve Bank of St. Louis.

Manufacturing Establishments For manufacturing we follow a similar strategy the always define the change in inventories as $\mathtt{CHANGEM} = \mathtt{MIB} - \mathtt{MIE}$ because a manufacturing establishment may well have goods inventories. Another variable that matters for manufacturing establishments is contract work ($\mathtt{CW}$) which has to be added in materials usage:

$$\mathtt{M} = \frac{\mathtt{MIB} - \mathtt{MIE} + \mathtt{CP} + \mathtt{CW}}{P^{mat}}.$$

A.7 Measurement of energy input

Construction Establishments Energy input is measured by real value of energy consumption. In all but the 1972 Census, the CCN reports total cost of power ($\mathtt{CSTEFT}$) and a group of costs from specific sources, including the cost of natural gas ($\mathtt{CSTFUNG}$), of other fuels ($\mathtt{CSTFUOT}$), of electricity ($\mathtt{EE}$), and costs of gas on and off highway ($\mathtt{CSTFUOFF}$ and $\mathtt{CSTFUON}$). By definition, costs from all sources sum up to total cost of power, $\mathtt{CSTEFT} = \mathtt{CSTFUNG} + \mathtt{CSTFUOT} + \mathtt{EE} + \mathtt{CSTFUOFF} + \mathtt{CSTFUON}$. Since the cost of gas on highway ($\mathtt{CSTFUON}$) is only linked to transport equipment to construction sites, we decide not to treat it as energy input. Hence we calculate nominal value of energy inputs as reported total cost of power minus cost of gas on highway, that is, $\mathtt{CSTEFT} - \mathtt{CSTFUON}$. In the cases where the difference is non-positive, we substitute it with the sum of reported costs of natural gas, of other fuels, of electricity, and of gas off highway, i.e., $\mathtt{CSTFUNG} + \mathtt{CSTFUOT} + \mathtt{EE} + \mathtt{CSTFUOFF}$. This nominal energy input is then deflated by price index for energy consumption specific to the construction sector from the Federal Reserve Bank of St. Louis denoted by $\mathtt{PIEN}$:

$$\mathtt{E} = \frac{\mathtt{CSTFUNG} + \mathtt{CSTFUOT} + \mathtt{EE} + \mathtt{CSTFUOFF}}{P^{en}}.$$

Manufacturing Establishments We assume that fuels inventory (recorded as part of materials inventory in the CMF) is unchanged. This means that all fuel purchases are immediately consumed in production or in electricity generation. Total energy expenditures ($\mathtt{VE}$) comprise those for fuels

(`CF`) and electricity (`EE`); the nominal value is:

$$\begin{aligned} \mathtt{VE} &= \mathtt{CF} + \mathtt{EE} \\ \mathtt{E} &\equiv \frac{\mathtt{VEN}}{P^{en}} = \frac{\mathtt{CF} + \mathtt{EE}}{P^{en}} \end{aligned} \tag{15}$$

where we use P^{en}, the industry-specific energy price deflator from the NBER-CES productivity database, to obtain real energy input, `E`.

A.8 Measurement of investment

In addition to output and input variables discussed above, we also need investment to process estimation strategies suggested by Olley and Pakes (1996) and Ackerberg et al. (2006). Similarly, we need a measure of the cost per unit of capital to apply the method suggested by Gandhi et al. (2013).

Census surveys the nominal value of investment and record it under the name "total capital expenditures" (`TCE`), but investments made in equipment and structures are not recorded separately. To account for different capital embodied technical change in structure and equipment investment, we perform a similar transformation we did for the capital stocks. We make the analogous assumption that the ratio of real structure to equipment investment is the same within across establishments within an industry. Let `BAB` and `MAB` be nominal structure and equipment investments, respectively. Then, we can write real investment as

$$\begin{aligned} I &= I^{st} + I^{eq} \\ &= \frac{\mathtt{BAB}}{P^{st}} + \frac{\mathtt{MAB}}{P^{eq}} \\ &= \left(\frac{1}{P^{st}} \frac{I^{st}}{I^{st} + I^{eq}} + \frac{1}{P^{eq}} \frac{I^{eq}}{I^{st} + I^{eq}} \right) \mathtt{TCE} \end{aligned}$$

where the last line follows from the assumption that $\frac{I_i^{st}}{I_i^{eq}} = \frac{I^{st}}{I^{eq}}$. Here I^{st} and I^{eq} stand for the real value of investment in structure and equipment, respectively, both of which can be backed out by the nominal investment grand totals published by the BEA along with price indices P^{st} and P^{st} for every NAICS-3 industry.

The user cost of capital is defined as $\frac{r_t K_t}{P^{inv}}$ where K is the real capital stock (in year-2005 dollars) constructed as described above and r_t denotes the *nominal* rental rate (year-t dollars rent paid per one year-2005 dollar worth of capital). Multiplying this rental rate, r_t, by the real capital stock, K_t, gives the nominal period-t capital cost of financing the stock in period t. This makes it accord with the other nominal values. The rental rate is constructed from the BLS Capital Tables[10] by dividing corporate capital income (Table 3a) by the real capital stock (Table 4a). The latter variable is expressed in constant (year-2005 dollar), while the former is expressed in current-period dollars, so rK are the capital cost expressed in period-t dollars. Everything is then turned into a real value by dividing by the price index of investment. Note that capital cost merely includes rent and depreciation, not physical utilisation cost which is captured in the energy cost share.[11]

[10]"Capital by Asset Type for NIPA-level Manufacturing Industries" downloaded from http://www.bls.gov/mfp/mprdload.htm.

[11]This obviously assumes that depreciation is not influenced by utilisation.

B Industry estimates

We conduct analysis using the same approach of our main specification at the 4-digit NAICS industry level to recover return to scale and total factor productivity used in the body of this paper. Return to scale equals the sum of the production elasticities $\beta_k + \beta_l + \beta_m + \beta_e$. Total factor productivity corresponds to $e^{\omega_{it}+\varepsilon_{it}}$ in the production function, and it can be estimated along with other parameters in the production function. The table below lists industry-level estimation results of both returns to scale, $\beta_k + \beta_l + \beta_m + \beta_e$, and the estimated dispersion of log-total factor productivity, $StD(\omega_{it} + \varepsilon_{it})$.

Table 6: Olley-Pakes Estimates by Industry

NAICS-4	β_k	β_l	β_m	β_e	Returns to scale	TFP Disp. (StD)
2361	0.1057*	0.2868***	0.4742***	0.0823***	0.9490	0.4211
	(0.0730)	(0.0067)	(0.0067)	(0.0093)		
2362	0.0357***	0.3294***	0.416***	0.1318***	0.9129	0.5692
	(0.0151)	(0.0057)	(0.0069)	(0.0062)		
2371	0.089***	0.4021***	0.2874***	0.1228***	0.9013	0.4263
	(0.0202)	(0.0097)	(0.0059)	(0.0056)		
2373	0.0878***	0.3560***	0.3377***	0.1078***	0.8893	0.4454
	(0.0126)	(0.0096)	(0.0109)	(0.0054)		
2381	0.0409**	0.3917***	0.457***	0.0388***	0.9284	0.4027
	(0.0183)	(0.0073)	(0.0077)	(0.0049)		
2382	0.0123	0.4908***	0.3493***	0.0698***	0.9222	0.4540
	(0.0102)	(0.0069)	(0.0047)	(0.0039)		
2383	0.0394***	0.3900***	0.4399***	0.0490***	0.9183	0.3907
	(0.0091)	(0.0036)	(0.0033)	(0.0028)		
2389	0.0629***	0.4333***	0.3609***	0.0769***	0.9340	0.4454
	(0.0082)	(0.0042)	(0.0042)	(0.0019)		
3111	0.1461***	0.0924***	0.6231***	0.1702***	1.0318	0.2763
	(0.0376)	(0.0106)	(0.0164)	(0.0220)		
3112	0.1976***	0.2023***	0.5529***	0.1018***	1.0546	0.3519
	(0.0320)	(0.0205)	(0.0243)	(0.0131)		
3113	0.1048***	0.2341***	0.6660***	-0.0005	1.0044	0.2751
	(0.0309)	(0.0157)	(0.0426)	(0.0254)		
3114	0.0743*	0.1095***	0.6506***	0.0899***	0.9243	0.3251
	(0.0464)	(0.0184)	(0.0498)	(0.0295)		
3115	0.1339***	0.0777***	0.5851***	0.1629***	0.9596	0.2758
	(0.0209)	(0.0085)	(0.0348)	(0.0197)		
3116	0.0879***	0.1492***	0.6900***	0.065***	0.9921	0.3059
	(0.0197)	(0.0146)	(0.0270)	(0.0153)		

continued ...

NAICS-4	β_k	β_l	β_m	β_e	Returns to scale	TFP Disp. (StD)
3117	0.2002***	0.1554***	0.5570***	0.0793**	0.9919	0.3286
	(0.0439)	(0.0178)	(0.0572)	(0.0371)		
3118	0.1137***	0.1996***	0.3815***	0.2145***	0.9093	0.3523
	(0.0196)	(0.0088)	(0.0132)	(0.0127)		
3119	0.1687***	0.0634***	0.5996***	0.0582***	0.8899	0.3633
	(0.0168)	(0.0099)	(0.0267)	(0.0150)		
3131	0.0624**	0.3759***	0.4147***	0.1338***	0.9868	0.3035
	(0.0297)	(0.0202)	(0.0250)	(0.0312)		
3132	0.0868***	0.3778***	0.4284***	0.0695***	0.9625	0.3194
	(0.0151)	(0.0137)	(0.0143)	(0.0135)		
3133	0.1286***	0.3159***	0.3993***	0.0981***	0.9419	0.3344
	(0.0161)	(0.0074)	(0.0125)	(0.0107)		
3141	0.1077***	0.2750***	0.3914***	0.1325***	0.9066	0.3490
	(0.0319)	(0.0109)	(0.0141)	(0.0125)		
3149	0.1615***	0.2988***	0.5051***	-0.0201**	0.9453	0.4338
	(0.0207)	(0.0122)	(0.0094)	(0.0121)		
3152	0.1702***	0.3129***	0.325***	0.1201***	0.9282	0.4441
	(0.0091)	(0.0058)	(0.0021)	(0.0048)		
3159	0.1797***	0.2034***	0.3787***	0.1251***	0.8869	0.3569
	(0.0537)	(0.0191)	(0.0426)	(0.0291)		
3161	0.1808***	0.2876***	0.4855***	0.0768***	1.0307	0.3184
	(0.0430)	(0.0249)	(0.0296)	(0.0308)		
3162	0.1030***	0.2709***	0.4543***	0.1671***	0.9953	0.3273
	(0.0378)	(0.0322)	(0.0362)	(0.0309)		
3169	0.1162***	0.2305***	0.4245***	0.1741***	0.9453	0.2808
	(0.0281)	(0.0167)	(0.0333)	(0.0217)		
3211	0.1971***	0.1981***	0.3934***	0.1399***	0.9285	0.3497
	(0.0123)	(0.0063)	(0.0124)	(0.0094)		
3212	0.1165***	0.2118***	0.5552***	0.0726***	0.9561	0.2848
	(0.0213)	(0.0116)	(0.0289)	(0.0114)		
3219	0.1270***	0.2236***	0.5805***	0.0101	0.9412	0.3304
	(0.0192)	(0.0057)	(0.0125)	(0.0102)		
3221	0.0638**	0.2504***	0.4690***	0.2514***	1.0346	0.3738
	(0.0305)	(0.0289)	(0.0396)	(0.0368)		
3222	0.0755***	0.2923***	0.5206***	0.0522***	0.9406	0.2966
	(0.0104)	(0.0053)	(0.0105)	(0.0049)		
3231	0.1916***	0.1996***	0.3671***	0.1207***	0.8790	0.3359
	(0.0054)	(0.0018)	(0.0034)	(0.0035)		

continued ...

NAICS-4	β_k	β_l	β_m	β_e	Returns to scale	TFP Disp. (StD)
3241	0.0471*	0.1504***	0.7253***	0.0123	0.9351	0.3443
	(0.0289)	(0.0036)	(0.0172)	(0.0096)		
3251	0.1891***	0.1915***	0.3679***	0.1542***	0.9027	0.5410
	(0.0207)	(0.0178)	(0.0174)	(0.0134)		
3252	0.1421***	0.1039***	0.6855***	0.0301	0.9616	0.3524
	(0.0354)	(0.0136)	(0.0485)	(0.0379)		
3253	0.2089***	0.1489***	0.5065***	0.0871***	0.9514	0.4078
	(0.0310)	(0.0185)	(0.0265)	(0.0141)		
3254	0.2511***	0.2020***	0.5810***	-0.0753***	0.9588	0.4907
	(0.0476)	(0.0135)	(0.0235)	(0.0179)		
3255	0.1467***	0.1528***	0.6316***	0.0543**	0.9854	0.2581
	(0.0184)	(0.0116)	(0.0385)	(0.0238)		
3256	0.139***	0.2187***	0.6071***	-0.0467***	0.9181	0.3774
	(0.0446)	(0.0082)	(0.0326)	(0.0156)		
3259	0.1502***	0.1636***	0.5967***	0.0376***	0.9481	0.3351
	(0.0129)	(0.0108)	(0.0196)	(0.0054)		
3261	0.1510***	0.2028***	0.5057***	0.0739***	0.9334	0.3044
	(0.0093)	(0.0040)	(0.0086)	(0.0054)		
3262	0.2075***	0.2706***	0.444***	0.0211	0.9432	0.3561
	(0.0367)	(0.0312)	(0.0672)	(0.0334)		
3271	0.1758***	0.3467***	0.3224***	0.0437***	0.8886	0.4450
	(0.0278)	(0.0161)	(0.0177)	(0.0143)		
3272	0.1229***	0.2460***	0.5324***	0.0493***	0.9506	0.3156
	(0.0162)	(0.0122)	(0.0140)	(0.0081)		
3273	0.1655***	0.1636***	0.5138***	0.0823***	0.9252	0.2952
	(0.0119)	(0.0048)	(0.0116)	(0.0057)		
3274	-0.1620**	0.4129***	0.4904***	0.1142***	0.8555	0.4779
	(0.0824)	(0.0521)	(0.0310)	(0.0319)		
3279	0.1765***	0.2431***	0.4237***	0.0663***	0.9096	0.3739
	(0.0360)	(0.0111)	(0.0128)	(0.0112)		
3312	0.0909**	0.1998***	0.6180***	0.1053***	1.0140	0.2845
	(0.0506)	(0.0177)	(0.0345)	(0.0270)		
3313	0.096***	0.2427***	0.6009***	0.0666***	1.0062	0.3200
	(0.0282)	(0.0139)	(0.0194)	(0.0138)		
3314	0.3847***	0.2209***	0.6099***	-0.0695**	1.1460	0.5640
	(0.0704)	(0.0272)	(0.0340)	(0.0312)		
3315	0.1041***	0.2761***	0.5015***	0.1136***	0.9953	0.3026
	(0.0234)	(0.0134)	(0.0184)	(0.0098)		

continued . . .

NAICS-4	β_k	β_l	β_m	β_e	Returns to scale	TFP Disp. (StD)
3321	0.1594*** (0.0117)	0.2788*** (0.0098)	0.5170*** (0.0123)	-0.0031 (0.0083)	0.9521	0.2938
3322	0.2036*** (0.0209)	0.2148*** (0.0067)	0.3769*** (0.0110)	0.1612*** (0.0104)	0.9565	0.3014
3323	0.1638*** (0.0105)	0.2250*** (0.0054)	0.4605*** (0.0096)	0.0785*** (0.0081)	0.9278	0.3148
3324	0.1482*** (0.0119)	0.2516*** (0.0063)	0.4619*** (0.0102)	0.1138*** (0.0115)	0.9755	0.2763
3325	0.1198*** (0.0321)	0.2113*** (0.0183)	0.5065*** (0.0253)	0.1176*** (0.0149)	0.9552	0.3332
3326	0.1383*** (0.0254)	0.2213*** (0.0153)	0.5278*** (0.0182)	0.0485*** (0.0139)	0.9359	0.3203
3327	0.2229*** (0.0423)	0.2261*** (0.0101)	0.3493*** (0.0216)	0.1127*** (0.0181)	0.9110	0.3181
3328	0.2079*** (0.0238)	0.3258*** (0.0080)	0.2375*** (0.0089)	0.1203*** (0.0063)	0.8915	0.3869
3329	0.1718*** (0.0126)	0.2519*** (0.0094)	0.4684*** (0.0064)	0.0823*** (0.0046)	0.9744	0.3433
3331	0.1558*** (0.0214)	0.1924*** (0.0167)	0.469*** (0.0208)	0.1471*** (0.0152)	0.9643	0.3635
3332	0.1607*** (0.0239)	0.2593*** (0.0138)	0.4471*** (0.0190)	0.0861*** (0.0119)	0.9532	0.3418
3333	0.1158*** (0.0205)	0.1886*** (0.0204)	0.5876*** (0.0238)	0.1072*** (0.0208)	0.9992	0.4116
3334	0.1306*** (0.0182)	0.2155*** (0.0170)	0.5250*** (0.0245)	0.1116*** (0.0138)	0.9827	0.2894
3335	0.1736*** (0.0121)	0.2268*** (0.0055)	0.3933*** (0.0109)	0.1185*** (0.0097)	0.9122	0.3113
3336	0.1199*** (0.0240)	0.3092*** (0.0189)	0.4614*** (0.0150)	0.0930*** (0.0161)	0.9835	0.3194
3339	0.1504*** (0.0104)	0.2596*** (0.0127)	0.4087*** (0.0093)	0.1453*** (0.0117)	0.9640	0.3402
3341	0.0862** (0.0488)	0.204*** (0.0261)	0.6827*** (0.0404)	0.1333*** (0.0297)	1.1062	1.0316
3342	0.1693*** (0.0119)	0.2654*** (0.0096)	0.4213*** (0.0130)	0.1054*** (0.0119)	0.9614	0.3963
3343	0.2015*** (0.0524)	0.2062*** (0.0155)	0.5625*** (0.0345)	0.0628** (0.0271)	1.0330	0.2710

continued ...

NAICS-4	β_k	β_l	β_m	β_e	Returns to scale	TFP Disp. (StD)
3344	-0.4473*** (0.0920)	1.0016*** (0.0342)	0.9931*** (0.0510)	-0.2446*** (0.0748)	1.3028	1.5409
3345	0.2138*** (0.0154)	0.2355*** (0.0099)	0.3742*** (0.0087)	0.1483*** (0.0110)	0.9718	0.3540
3346	0.1039* (0.0694)	0.1322*** (0.0179)	0.6073*** (0.0574)	0.0002 (0.0320)	0.8436	0.4158
3351	0.0950*** (0.0181)	0.2368*** (0.0167)	0.5308*** (0.0295)	0.1177*** (0.0114)	0.9803	0.3090
3352	0.0741* (0.0537)	0.1777*** (0.0325)	0.6624*** (0.0453)	0.1020** (0.0492)	1.0162	0.2731
3353	0.1535*** (0.0222)	0.2676*** (0.0103)	0.4725*** (0.0162)	0.0702*** (0.0159)	0.9638	0.3376
3359	0.1812*** (0.0201)	0.3039*** (0.0098)	0.4153*** (0.0214)	0.0642*** (0.0177)	0.9646	0.3703
3361	0.0966** (0.0478)	0.1557*** (0.0225)	0.6563*** (0.0750)	0.0456 (0.0563)	0.9542	0.3157
3362	0.0626*** (0.0181)	0.1829*** (0.0203)	0.5785*** (0.0247)	0.1368*** (0.0201)	0.9608	0.2888
3363	0.2645*** (0.0072)	0.2663*** (0.0034)	0.2794*** (0.0061)	0.0959*** (0.0053)	0.9061	0.3131
3364	0.2164*** (0.0241)	0.2265*** (0.0155)	0.4228*** (0.0197)	0.1274*** (0.0185)	0.9931	0.3590
3365	0.1045*** (0.0424)	0.2160*** (0.0440)	0.5563*** (0.0407)	0.1288*** (0.0438)	1.0056	0.2966
3366	0.2123*** (0.0316)	0.1513*** (0.0151)	0.4737*** (0.0302)	0.1215*** (0.0195)	0.9588	0.3103
3369	0.1088*** (0.0362)	0.1830*** (0.0135)	0.5995*** (0.0274)	0.0845*** (0.0151)	0.9758	0.2939
3371	0.0987*** (0.0057)	0.224*** (0.0065)	0.5117*** (0.0163)	0.1162*** (0.0089)	0.9506	0.2660
3372	0.2000*** (0.0207)	0.2288*** (0.0086)	0.4505*** (0.0179)	0.0805*** (0.0144)	0.9598	0.3047
3379	0.0846*** (0.0257)	0.2146*** (0.0150)	0.5162*** (0.0312)	0.1050*** (0.0143)	0.9204	0.2566
3391	0.1183*** (0.0169)	0.2364*** (0.0127)	0.5491*** (0.0172)	0.0143* (0.0109)	0.9181	0.3479
3399	0.1246*** (0.0076)	0.2462*** (0.0054)	0.5187*** (0.0080)	0.0447*** (0.0029)	0.9342	0.3033

Note: Note: This table reports estimation results for each NAICS-4 industry using the method proposed by Olley and Pakes (1996). Output, capital, material, and energy inputs are measured in year-2005 dollar values, labour in hours. *, **, *** significantly different from 0 at the 10%, 5%, 1% level, respectively.
Some industries had to be omitted because they are too small or too concentrated to satisfy Census' disclosure rules. The estimation procedure in those industries is typically fragile as few observations mean less information in the non-parametric inversion of the investment policy rule.