Measuring Misallocation with Experiments*

David Hughes† Jeremy Majerovitz ‡

January 2025

Abstract

Misallocation of inputs across firms has been proposed as a reason for low levels of development in some countries. However, existing work has largely relied on strong assumptions about production functions in order to estimate the cost of misallocation. We show that, for arbitrary production functions, the cost of misallocation can be expressed as a function of the variance of marginal products. Using an RCT that gave grants to microenterprises, we estimate heterogeneous returns to capital by baseline characteristics, and provide a lower bound on the total variance of returns to capital. This lower bound is a nonlinear function of the parameters from a linear IV model, so that standard methods (e.g. the delta method or projection) have poor performance. We provide novel econometric tools that provide uniformly valid confidence intervals for nonlinear functions of parameters. We find evidence for sizable losses from misallocation of inputs across the firms we study, although the magnitude depends critically on which inputs we allow to be reallocated. We estimate that optimally reallocating capital would increase output by 22%, while optimally reallocating all inputs would increase output by 301%.

Keywords: Misallocation, Returns to Capital, Randomized Controlled Trials, Testing Nonlinear Restrictions
JEL: C12, C13, D24, D61, E1, E23, O11, O12, O4

*We are grateful to David Atkin, Abhijit Banerjee, Paco Buera, Joel David, Dave Donaldson, Ben Olken, Sergio Ocampo, David McKenzie, Jacob Moscona, Michael Peters, Martin Rotemberg, Karthik Sastry, Pari Sastry, Yongs Shin, Rob Townsend, and Mitchell VanVuren for helpful comments. We also thank seminar participants at Boston University, the Federal Reserve Bank of St. Louis, Harvard Kennedy School, MIT, the IGC/Yale Firms, Trade and Development Conference, NEUDC, PACDEV, Oxford CSAE Annual Conference, MWIEDC, the PSE-CEPR Development Economics Symposium, the North American Econometric Society Summer Meeting, the SED, the RIDGE Forum, and the STEG Annual Conference for thoughtful questions and feedback. This research was supported by the University of Notre Dame's Building Inclusive Growth (BIG) Lab. We thank Zeina Shalaby for excellent research assistance.

†Department of Economics, Boston College; Email: dw.hughes@bc.edu
‡University of Notre Dame; Email: jeremy.majerovitz@gmail.com

1 Introduction

In the absence of distortions, competitive markets allocate inputs across firms to their efficient use. Deviations from this efficient benchmark can lower aggregate productivity substantially. An extensive literature in macroeconomics and development has found large losses in output due to misallocation, especially in less developed economies. This has led many economists to view misallocation as "our best candidate answer to the question of why are some countries so much richer than others" (Jones, 2016).

However, an important shortcoming in this literature has been a heavy reliance on restrictive assumptions about firm production functions. Thus, most prior estimates of the cost of misallocation are implicitly a joint test of market efficiency and of the strong auxiliary assumptions that underlie these calculations. In the cases where these methods have apparently found large losses from misallocation, it is not always obvious whether this suggests a rejection of efficient markets or a rejection of the auxiliary assumptions.

In this paper, we show how to measure the cost of misallocation without relying on restrictive assumptions about production functions. To measure misallocation, we need to connect the cost of misallocation back to something that we can estimate in the data. We start from our question — what is the cost of misallocation of inputs — and work backwards.

We first provide an aggregation result, showing, for arbitrary production functions that misallocation is a function of the variance of the log marginal revenue product of capital (MRPK). We then show how to measure marginal products using an RCT, exploiting a randomized controlled trial by de Mel et al. (2008) that randomly assigned grants to microentrepreneurs in Sri Lanka. We use this experiment to estimate heterogeneous returns to capital by baseline characteristics, which provides a lower bound on the total variance of log MRPK. Finally, since standard methods cannot be used to conduct correct inference on this object, we provide new econometric tools to construct confidence intervals for nonlinear functions of parameters.

We find substantial dispersion in the marginal revenue product of capital among Sri Lankan microentrepreneurs. We estimate that the standard deviation of monthly returns to capital is 9.8%, or 1.23 times the mean; our 90% confidence interval rules out a standard deviation below 4% and a standard-deviation-to-mean ratio below 0.47. In our preferred calibration, this implies that optimally reallocating capital would increase output by 22%, while optimally reallocating all inputs would increase output by 301%.

Our results connect a macroeconomic question (what is the cost of misallocation?) to a microeconomic question (what is the variance of marginal products across firms?), and then to an econometric question (how do we measure this variance in an instrumental vari-

ables setting, and construct valid confidence intervals?). In doing so, we also draw connections between literatures on the microeconomics and macroeconomics of development. Our methodology shows how to correctly aggregate microeconomic evidence of dispersed marginal products into an aggregate cost of misallocation. Equivalently, we show how to use experimental or quasi-experimental variation to provide rigorous empirical microfoundations for macroeconomic models of misallocation.

Macro to Micro: Measuring Misallocation in Terms of Marginal Products. We begin by connecting misallocation to the distribution of marginal products. In an efficient economy, the marginal product of capital should be equalized across all firms. If firms produce heterogeneous products and households are price takers, then this condition can instead be expressed in terms of the "value of the marginal product." Focusing on capital, the "VMPK" is the price of the firm's output times the marginal product of capital. In an efficient economy, the VMPK must be the same across firms.

We consider a horizontal economy in which firms use a single input, capital, to produce differentiated products, which are then aggregated into a final good. We allow for arbitrary smooth production functions at the firm level. We do not make any assumptions about firm conduct, except that the household, is a price taker. We focus on counterfactuals that hold the aggregate supply of inputs fixed, in order to home in on the idea of misallocation of inputs *across* firms. In the first-best, VMPK is equalized across firms, but we allow for reduced-form "wedges" that represent deviations from the planner's efficient first-order condition.

In this economy, we show that when households have CES demand, the cost of misallocation is given by

$$\mathcal{L} \approx \frac{1}{2}\mathcal{E}\text{Var}\left(\log \text{VMPK}_i\right)$$

where $\mathcal{E}$ is the (negative) elasticity of firm output with respect to the wedge, and $\mathcal{L}$ is the potential gains, in terms of log aggregate output, from optimally reallocating inputs. This result is exact for Cobb-Douglas production functions with lognormally distributed productivity and wedges, and is a second-order approximation for arbitrary production functions.[1] The magnitude of $\mathcal{E}$ depends on both the CES parameter and on returns to scale in the production function. Thus, the potential gains from optimally reallocating inputs will depend critically on which inputs are being reallocated. If all inputs can be reallocated, then a constant-returns-to-scale production function implies that $\mathcal{E}$ equals the CES parameter; if only capital can be reallocated, then attempts to reallocate inputs will quickly run into

[1]The second-order approximation replaces the elasticity $\mathcal{E}$ with a sales-weighted average of firm-specific elasticities $\mathcal{E}_i$, and $\text{Var}\left(\log \text{VMPK}_i\right)$ with a sales-times-elasticity weighted variance.

decreasing returns to scale, dampening potential gains.

Finally, since we will not have separate data on prices and quantities, we show that the assumption of CES demand also allows us to re-express misallocation in terms of the variance of the log marginal revenue product of capital. Whereas the VMPK measures the price of output times the marginal product of capital, the MRPK measures the derivative of revenue with respect to capital. In general, the MRPK will be lower than the VMPK because an increase in capital will raise output and thus lower prices. However, under CES demand, the MRPK will equal to the VMPK times a constant, and so the variance of log VMPK and the variance of log MRPK will be the same.

Micro to Metrics: Measuring Marginal Products with an IV Regression. Our next step is to develop a strategy to estimate the variance of log MRPK across firms. To do this, we note two challenges. First, we must identify the causal effect of changes in capital, but variation in capital is in general endogenous: firms choose their capital as a function of productivity, so we would expect changes in capital to be correlated with changes in productivity. We solve this problem by using data from an RCT by de Mel et al. (2008). This experiment, conducted on a sample of microenterprises in Sri Lanka, randomized grants to firms in order to estimate the returns to capital. We use the grant as an instrument for capital, in order to identify the MRPK.

The second challenge is that we must identify not just the average returns to capital, but the variance of returns to capital across firms. In general, this is not possible without additional assumptions: the variance of treatment effects is not identified. However, we can provide an informative lower bound by projecting the returns to capital onto observable baseline characteristics. By the law of total variance, the total variance of MRPK will be equal to the variance of expected MRPK given baseline characteristics, plus the expected variance of MRPK conditional on those characteristics. Thus, the variance of the conditional average treatment effects provides an estimable lower bound on the total variance of treatment effects.

Targeting the predictable component of the variance of MRPK, rather than the total variance, also has attractive features from an economic perspective. In principle, dispersion in returns to capital *ex post* can result from misallocation or from risk: some investments are good ideas *ex ante* but do not pay off. Instead, dispersion in *ex ante* returns to capital reflects true misallocation. By focusing on the predictable component of the variance of returns, we ensure that we are measuring true misallocation.

To implement this, we express the returns to capital using a linear IV model, with capital entering both directly and interacted with baseline covariates. To simplify our formulas and to improve variable selection, we use principal components analysis to recast the baseline

characteristics as orthogonal variables with mean zero and standard deviation one. This orthonormal basis provides us with a simple expression for the variance of log expected returns to capital, as a nonlinear function of the parameters of a linear IV model.

Inference for Nonlinear Functions of Parameters. Given that the variance of log MRPK is a nonlinear function of parameters, our final step is to conduct valid inference on this function. The standard methods to construct confidence intervals for functions of parameters are the delta method and the projection method . The delta method approximates the distribution of a function of parameter estimates using a first order linear approximation. The function that we study is highly nonlinear, so that this method can perform poorly in practice – in fact, the function has zero derivative at the point where misallocation is equal to zero, and has infinite derivative at the point where the average returns to capital are zero and so fails completely in these cases. The projection method is valid regardless of the degree of nonlinearity, but it is generally conservative (and increasingly so as the number of parameters increases). Simulations calibrated to our empirical setting suggest that the performance of the delta method is sensitive to the true value of parameters, while the projection method results in confidence intervals that are too wide to be useful.We thus develop novel econometric tools in order to construct uniformly valid confidence intervals for functions of parameters, in settings where the delta method fails. To test a given null hypothesis, our method uses the inverse-variance-weighted distance between the estimated parameter and the constraint imposed by the null. We obtain critical values for this test statistic by treating the underlying parameter estimates as Gaussian and then simulating the distribution of the test statistic. We show in simulations that our method delivers correct size, even when other methods fail.

Results: Estimates of Misallocation for Sri Lankan Microenterprises. Finally, having developed a methodology to measure the cost of misallocation, we put these tools to work. Our estimates suggest that the variance of log MRPK across firms is sizable. Our preferred point estimates suggest that the average monthly returns to capital is 8.0%, and the standard deviation of returns is 9.8%. This implies a variance of log MRPK of 93 log points, with the 90% confidence interval ruling out values below 20 log points. If we had instead assumed a homogeneous returns-to-scale Cobb-Douglas production function as in Hsieh and Klenow (2009), we would have inferred an average monthly return of 7.8% and a variance of log MRPK of 130 log points. Our confidence intervals cannot rule out the Cobb-Douglas estimates. However, the advantage of our approach is that our estimates of the variance of marginal products are valid regardless of whether firms truly produce with a homogeneous Cobb-Douglas production function.

We then combine our main estimates with a standard calibration for $\mathcal{E}$, in order to back out the cost of misallocation. We estimate that optimally reallocating capital would increase output by 22%, while optimally reallocating all inputs would increase output by 301%. This suggests a potentially important role for misallocation, although also highlights the importance of firm returns-to-scale in determining the extent of misallocation.

Related Literature

We contribute to a large literature on the cost of misallocation. After the seminal contributions of Restuccia and Rogerson (2008) and Hsieh and Klenow (2009), many authors have worked on estimating and better understanding the costs of misallocation. This literature is summarized in Hopenhayn (2014) and in Restuccia and Rogerson (2017). Recent work (Baqaee and Farhi, 2020; Bigio and La'O, 2020; Dávila and Schaab, 2023; Liu, 2019) on aggregation has elucidated the connection between changes in aggregate output (and aggregate welfare) and individual marginal products and marginal utilities. By integrating along a path from the distorted equilibrium to an undistorted equilibrium, this line of research has also provided insights into the measurement of misallocation. This work informs our own paper, which highlights the connection between misallocation and the distribution of marginal products. We generalize previous results to allow for arbitrary firm production functions, though we still impose CES demand.

Our paper also connects to a literature in development microeconomics that finds high and dispersed returns to capital, and interprets this as evidence of misallocation. An influential paper by Banerjee and Duflo (2005) summarizes much of this evidence; since then, more work has found evidence that returns to capital are high (de Mel et al., 2008; Fafchamps et al., 2014; McKenzie, 2017) and vary substantially across firms (Hussam et al., 2022; Beaman et al., 2023; Crépon et al., 2023). We view our paper as providing a bridge between these related literatures in development microeconomics and macroeconomics. Our methods show how to correctly aggregate this rigorous microeconomic evidence, in order to provide estimates of the cost of misallocation.

A number of authors have noted challenges in the measurement of misallocation. Bils et al. (2021) highlight the problem presented by measurement error, and present a methodology to use panel data to separate misallocation from measurement error. Rotemberg and White (2021) also focus on measurement error, showing how differential data-cleaning methods by the statistical agencies in different countries can make apparent misallocation look very different across countries. Our methodology is robust to measurement error: a byproduct of using an instrumental variables regression is that (classical) measurement error does not bias our estimates. Gollin and Udry (2021) address both measurement error and risk

(*ex post* shocks to productivity), using data on farmers who produce on multiple plots and exploiting the assumption that allocations are undistorted within-farm across-plot. By projecting returns onto baseline observables, our method allows us to isolate misallocation from risk. Haltiwanger et al. (2018) highlight the strong assumptions required by the standard approach to measuring misallocation: in particular, isoelastic demand and homogeneous, constant-returns-to-scale production. Our approach relaxes these assumptions to allow for arbitrary production functions, although we will still require isoelastic demand (CES).

Most closely related to our work is a contemporaneous paper by Carrillo et al. (2023). Like ours, their paper studies misallocation, and uses random shocks (demand shocks from procurement lotteries, instead of capital supply shocks from an RCT) to identify moments of the distribution of marginal products.

We view both papers as complementary, and together providing a useful toolkit for future applications. Our paper differs from theirs in a few important ways. First, we target a different variance: the variance of expected returns, rather than the total variance of returns. Thus, our estimates provide a lower bound on misallocation, while their estimates provide an upper bound. Since we target different variances, the econometric method of our paper is also different from theirs. Their paper uses a correlated-random-coefficients model (Masten and Torgovitsky, 2016) to estimate the variance of marginal products across firms, relying on the linearity of the model. In contrast, we project marginal products onto baseline characteristics, in order to derive a lower bound on the variance of MRPK.

These different methods have different data requirements. Their method requires that the instrument be fully independent of the residual (as opposed to just uncorrelated), and also requires at least three points of support for the instrument. This does not rely on any assumptions beyond the typical ones for linear IV models with interaction effects. In practice, we find that the Carrillo et al. (2023) method produces uninformative confidence intervals in our setting, suggesting that our method may provide more statistical power in some settings.

Finally, and perhaps most importantly, we study a different setting and get different results: Carrillo et al. (2023) find a very small cost of misallocation for construction companies in Ecuador, while we find a more sizable cost of misallocation among microenterprises in Sri Lanka. Taken together, our results suggest that the degree of misallocation may vary across sectors and countries.

Comparison to Standard Approach. Our approach to measuring misallocation shares some elements in common with the standard approach, pioneered by Hsieh and Klenow (2009). The aggregation assumptions behind our approach are the same as those in the standard approach: we rely on CES demand to aggregate differentiated products across

firms. In the lognormal case, our aggregation is identical to that in the standard approach.[2] More generally we use a second-order approximation to misallocation, which should yield very similar results to the standard approach.

However, our approach differs from the standard approach in that we do not rely on assumptions about the functional form of the firm-level production function. Recasting the standard approach into our own framework, the standard approach assumes a particular production function (homogeneous loglinear) so that the average product is proportional to the marginal product. This approach will fail in settings where the production function does not take the assumed functional form (e.g. setting with fixed costs), or in which the production function is loglinear but the slope parameters are heterogeneous across firms. In contrast, we use an RCT to that provides exogenous variation in capital, allowing us to estimate marginal products directly. This is the critical distinction between our approach and the standard approach: we measure marginal products with variation in inputs *on the margin*, rather than inferring them from average products.

Outline. Section 2 shows how the cost of misallocation can be measured as a function of the distribution of marginal products across firms. Section 3 shows how to measure heterogeneous marginal products using an RCT, and provides a lower bound on the total variance of log MRPK as a nonlinear function of the parameters from a linear IV model. Section 4 explains the econometrics of nonlinear functions of parameters, such as our lower bound, and provides novel tools to provide valid inference in this setting. Section 5 uses the tools we develop to estimate the cost of misallocation. Each section begins with a less technical summary, so readers who wish to skip some sections can understand later sections without too much loss. Section 6 concludes.

2 Measuring Misallocation with Marginal Products

Summary

We begin by showing how the cost of misallocation depends on the distribution of marginal products across firms. In doing so, we recast a macroeconomic question ("What is the cost of misallocation?") as a microeconomic question ("What is the variance of log MRPK across firms?").

We start by highlighting that allocative efficiency requires the equation of marginal prod-

[2]We focus on a single sector version of the model, motivated by a desire for clarity and the fact that the microenterprises we study operate in relatively few sectors. However, extending our results to multiple sectors would be straightforward, and would yield extremely similar results.

ucts across firms. In a horizontal economy with heterogeneous products and price-taking consumers, this can be expressed in terms of the value of the marginal product: the marginal product times the price of output. Focusing on capital, equating VMPK is a necessary condition for productive efficiency, and is a sufficient condition under concavity.

Our first main result expresses misallocation as a function of the variance of log VMPK. Under CES aggregation, the cost of misallocation is given by

$$\mathcal{L} \approx \frac{1}{2}\mathcal{E}\text{Var}\left(\log \text{VMPK}_i\right)$$

where $\mathcal{E}$ is the (negative) elasticity of firm output with respect to the wedge, and $\mathcal{L}$ is the potential gains, in terms of log aggregate output, from optimally reallocating inputs. We show that this result is exact for log-linear production functions with log-normally distributed productivity and wedges, and holds more generally as a second-order approximation for arbitrary production functions. We also highlight that $\mathcal{E}$ depends critically on what inputs can be reallocated. If only capital can be reallocated, then decreasing returns to scale will make $\mathcal{E}$ small. If other inputs can also be reallocated, then $\mathcal{E}$ will be larger,[3] and thus the gains from reallocating inputs will also be larger.

Although production efficiency depends on the the distribution of VMPK across firms, in practice we typically do not observe separate data on prices and quantities. Thus, the best we can hope to do is to estimate MRPK: the derivative of revenue with respect to capital. In general, MRPK will be less than VMPK because an increase in capital increases output and thus decreases the price of the firm's output. Fortunately, we show that under CES aggregation, VMPK and MRPK are proportional to each other. Under CES demand the variance of log VMPK is thus the same as variance of log MRPK, and so we can focus on measuring the latter.

2.1 Setup

We begin by describing a fairly general production economy. We will focus throughout on horizontal economies: many firms produce intermediate goods, drawing from a common pool of inputs and supplying intermediates to an aggregator that creates the final good. We focus on this economy because it is the benchmark economy in the literature, and because it is a fairly accurate description of our setting, in which microenterprises produce similar products for final consumption).[4] We will focus on single product firms, and we will consider a single

[3] This is a consequence of Le Chatelier's principle: the elasticity of output to the wedge is larger when all inputs can adjust.

[4] Different network structures of production will in general imply different levels of misallocation (see Baqaee and Farhi, 2020).

input (we call this input capital) unless otherwise noted.[5]

There is a unit mass of firms indexed by $i \in [0, 1]$. Each firm has an individual production function:

$$y_i = f_i(k_i) \tag{1}$$

The final good, Y, is aggregated by an aggregator:

$$Y = Y\left(\{y_i\}_{i\in[0,1]}\right) \tag{2}$$

The final good aggregator can be viewed as the production function of a final good producer or as the utility function of a representative household: both formulations are mathematically identical. We will assume that the individual production functions, as well as the aggregator, are smooth. Moreover, we will assume throughout that any expectation we take exists: for example, we will assume that the variance of marginal products exists and is finite.

There is also an aggregate supply of the homogeneous input, capital. We define aggregate capital as:

$$K := \int_0^1 k_i di = \mathbb{E}[k_i] \tag{3}$$

Following the literature on misallocation, we will focus on counterfactuals in which aggregate inputs are held fixed. This allows us to focus on the production side of the economy: modeling an elastic input supply would require a model of household's preferences to supply that input.[6] Focusing on the losses from misallocation under fixed aggregate inputs will provide us a lower bound on the full cost of misallocation: the welfare gains from optimally reallocating inputs under the constraint that aggregate capital is held fixed must be less than or equal to the gains from selecting the unconstrained optimum allocation.

2.2 Marginal Products Are Equalized Across Firms In Efficient Economies

To study efficiency in this setting, we set up the planner's problem. The planner allocates capital among the firms to maximize the quantity of the final good, subject to the supply

[5]Treating capital as the only input implies that we are holding other inputs fixed: implicitly, fixed inputs are treated as part of productivity.

[6]This has the potential to be especially complicated for capital, since capital is accumulated over time and would require a dynamic theory of investment and savings.

constraint:

$$\max_{\{k_i\}_{i\in[0,1]}} Y\left(\{f_i(k)\}_{i\in[0,1]}\right) \quad \text{s.t.} \mathbb{E}[k_i] = \bar{K} \tag{4}$$

The planner's problem yields the first order condition:

$$\frac{dY}{dy_i} \cdot \frac{dy_i}{dk_i} = r \ \forall i \tag{5}$$

where r is the Lagrange multiplier on the supply constraint. The above is a necessary condition for efficiency. It also implies that $\frac{dY}{dy_i} \cdot \frac{dy_i}{dk_i} = \frac{dY}{dy_j} \cdot \frac{dy_j}{dk_j}$, for all i and j.

To build intuition, consider the case where firms produce homogeneous products. In this case, the aggregator is simply $Y = \int_0^1 y_i di$. It is well known that in this setting, efficiency requires equalizing the marginal product of capital (MPK) across firms. If firm i had a higher MPK than firm j, then a planner could increase output, without changing inputs, by taking a small amount of capital from j and giving it to i. Equalization of marginal products is a necessary condition for efficiency in the homogeneous-products setting, and becomes a sufficient condition for efficiency (conditional on a level of aggregate capital) if production functions are concave.

Introducing Prices and the Value of the Marginal Product of Capital (VMPK). We can simplify this condition by introducing prices. Let P be the price of the final good, and p_i be the price of the good produced by firm i. If the aggregator is a profit-maximizing firm, then its objective function is given by $PY - \mathbb{E}[p_i y_i]$. If the aggregator is a representative consumer, then it maximizes consumption, Y, subject to a budget constraint $\mathbb{E}[p_i y_i] \leq W$. These problems are of course the same, and yield equivalent first-order conditions.

Suppose that the aggregator takes prices as given. Then, from the first-order condition, we can show that $p_i = P \cdot \frac{dY}{dy_i}$. Define the value of the marginal product of capital (VMPK) as the price times MPK. That is,

$$\text{VMPK}_i := p \cdot \frac{dy_i}{dk_i} = P \cdot \frac{dY}{dy_i} \cdot \frac{dy_i}{dk_i} \tag{6}$$

It follows that equalization of VMPK across firms is a necessary condition for efficiency. Under appropriate concavity assumptions and along with the supply constraint, equalization of VMPK across firms would also be sufficient for efficiency.

This analysis is simple, but reveals a fundamental fact about the nature of misallocation.

Marginal products are equalized across firms in efficient economies. In horizontal economies with a price-taking aggregator, this can be expressed precisely as requiring VMPK to be equalized across firms. It is thus natural to assume that the cost of misallocation will be a function of the dispersion of VMPK. We next turn to derive the relationship between the distribution of VMPK and the cost of misallocation.

Wedges Rationalize Deviations from Efficiency. To rationalize variation in VMPK, we will introduce the notion of a wedge, μ_i. The wedge is a distortion of the planner's first-order condition for the firm. Letting r denote the price of capital that clears the input market, this yields the distorted first-order condition:

$$\underbrace{p_i \cdot \frac{dy_i}{dk_i}}_{\text{VMPK}_i} = \underbrace{r \cdot \mu_i}_{\text{Distorted Marginal Cost}} \tag{7}$$

In a competitive market without distortions, $\mu_i = 1$. More generally, the first-order condition can be distorted by a variety of factors, such as market power, credit constraints, taxes, and other market imperfections.

A few points are worth special note. First, note that, by the first welfare theorem, the wedgeless economy is efficient, and achieves the highest possible Y given K.[7] Moreover, if we double all of the wedges and halve the interest rate r, then no allocations will change. $Y\left(\{\mu_i\}_{i\in[0,1]}\right)$ will be homogeneous of degree zero.

Second, note that although we will refer to p_i and r as prices, our analysis in this subsection does not actually depend on the existence of markets where prices can be observed. In fact, all of our aggregation results would be the same if we simply defined $p_i = \frac{dY}{dy_i}$ and defined r solely as the Lagrange multiplier that implements market clearing in the input market. Instead, we use this notation to highlight the connection between our aggregation results and markets, and to connect to our later measurement results.

Finally, note that the wedge is defined in Equation 7 as a distortion of the *planner's* first-order condition for the firm, rather than of the firm's profit-maximizing first-order condition. If firms charge markups, then that markup will be included in the wedge, and if markups vary across firms then that will be reflected as variation in wedges across firms. Our definition of wedges thus does not require us to make any assumption about firm's conduct: wedges could arise due to firms' market power, or could be a result of perfectly competitive firms

[7] This result is an immediate consequence of the first welfare theorem because we have defined wedges in terms of deviations from the (planner's) efficient first-order condition. Some authors instead define wedges in terms of the firm's first-order condition under monopolistic competition, which will also incorporate the effects of market power. In this case, the wedgeless economy is still efficient, but only in the case of CES aggregation, since CES induces constant multiplicative markups across firms (Dhingra and Morrow, 2019).

facing credit constraints. If two sets of market imperfections implement the same allocation of inputs, then they will imply the same wedges (up to scale). Moreover, under appropriate concavity assumptions, a set of wedges will implement a unique allocation and prices. Thus, our wedges (along with technologies and the capital supply constraint) provide a complete description of the economy, without specifying firm conduct.

2.3 The Cost of Misallocation Depends on the Variance of log VMPK

We have shown so far that misallocation arises when there is dispersion across firms in the VMPK. We will now show how to map the distribution of VMPK across firms into a cost of misallocation. In particular, we will show that the cost of misallocation depends on the variance of log wedges, times one half times the elasticity of output with respect to the wedge.

For the remainder of our analysis, we will specialize slightly to CES demand. We will assume that the final good is produced by a constant-elasticity-of-substitution (CES) aggregator:

$$Y = \left(\int_0^1 y_i^{\frac{\theta-1}{\theta}} di \right)^{\frac{\theta}{\theta-1}} \tag{8}$$

where θ is the elasticity of substitution across varieties. This aggregator is the standard in the literature, and ensures that the demand for each firm's output can be expressed as a log-linear function of the firm's price, p_i, and aggregate output, Y. Normalizing the price of the final good, P, to one, we have:

$$\log y_i = -\theta \log p_i + \log Y \tag{9}$$

We can characterize the firm's behavior as a function of supply and demand: Equation 9 charactizes demand for the firm's output, and the firm's first order condition (Equation 7) determines the firm's supply curve. We show this graphically in Figure 1. We plot the demand curve and the undistorted supply curve in blue, and the response to a wedge in red. The interest rate and wedge shift the firm's supply curve, while the firm's demand curve is shifted by aggregate output, Y.

The slope of the demand and supply curves determine how much firm output responds to these shifters. The demand curve will be inelastic if the elasticity of substitution across varieties, θ, is low. The slope of the supply curve depends on the firms's (physical) returns to scale. If the firm faces decreasing returns to scale, then its MPK will fall as it gets larger; equivalently, its marginal cost rises. The slope of supply thus depends on ϕ_i, the elasticity of MPK with respect to output.

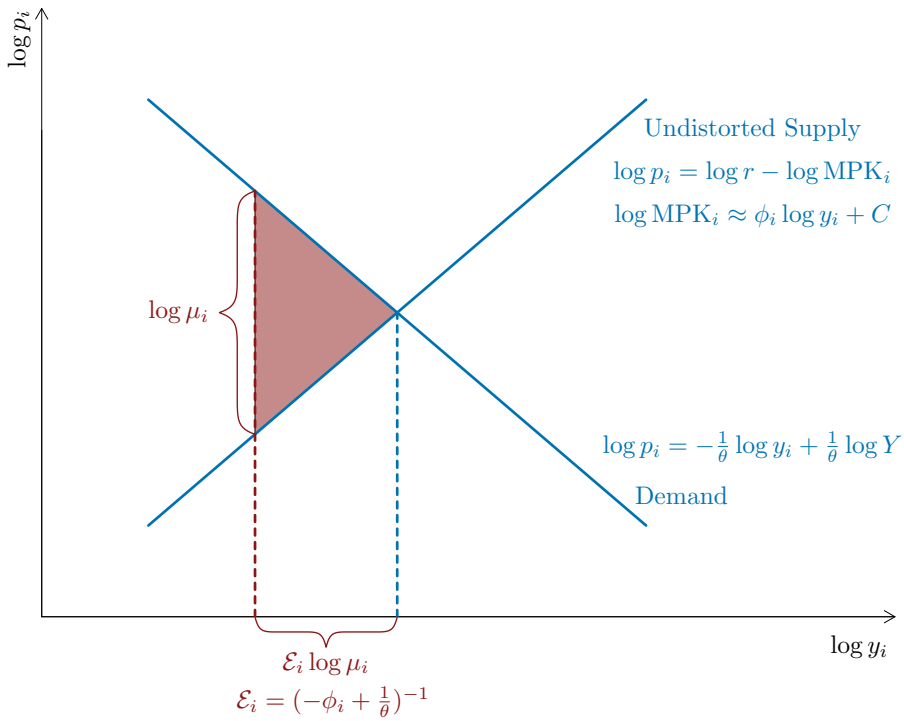


Figure 1: Firm Behavior and Deadweight Loss from Wedges

In order to ensure that the firm's first-order condition is sufficient to pin down a unique solution to the firm's problem, we will impose the following assumption:

Assumption 1. *The slope of the firm's supply curve is everywhere greater than the slope of the demand curve. That is,* $\phi_i(y_i) < \frac{1}{\theta}$, $\forall i$, $\forall y_i$, *where* $\phi_i := \frac{y_i \cdot f_i''}{(f_i')^2}$.

With this assumption, we can characterize the firm's behavior on the margin with the following Lemma:

Lemma 1 (Firm Behavior on the Margin)**.** *Assume the firm faces CES demand. The firm's behavior on the margin is described by*

$$d \log y_i = \underbrace{-\mathcal{E}_i d \log \mu_i}_{\textit{Wedge}} \underbrace{- \mathcal{E}_i d \log r}_{\textit{Input Cost}} + \underbrace{\frac{\mathcal{E}_i}{\theta} d \log Y}_{\textit{Demand}} \tag{10}$$

where $\mathcal{E}_i := \left(-\phi_i + \frac{1}{\theta}\right)^{-1}$ *is the firm-specific (negative) elasticity of output with respect to the wedge, and* $\phi_i := \frac{y_i \cdot f_i''}{(f_i')^2}$ *is the firm-specific elasticity of MPK with respect to output.*

As we can see graphically in Figure 1, the (negative) elasticity of output with respect to the wedge, $\mathcal{E}_i$, depends on the slopes of demand and supply. If the firm faces inelastic

demand (low θ) and low returns to scale (very negative ϕ_i), then firm output will not change much in response to the wedge. Later in this section, we will see that these same forces govern the scope for increasing aggregate output through reallocation of inputs.

The Harberger Triangle. In partial equilibrium, the deadweight loss from misallocation can also be read off of Figure 1. The pink triangle denotes the deadweight loss arising from the wedge. This is often known as the Harberger triangle, recognizing the contributions of Harberger (1954), who analyzed the welfare costs of market power in the United States.[8] This graphical intuition immediately suggests two things. First, the losses from misallocation will be quadratic in the size of the wedges. Small distortions to an initially undistorted economy will have approximately zero effect on aggregate output, but the welfare losses can grow large quickly as the distortions grow. Second, the magnitude of the losses will depend on $\mathcal{E}_i$, and thus depends on the slopes of supply and demand.

However, the simple graphical analysis in Figure 1 is incomplete, because it only analyzes the firm's problem in isolation. To solve for the welfare losses in general equilibrium, we must also keep track of the interest rate (which affects the firm's supply curve) and of the price of the final good (which affects demand). To do this, we will adapt and extend methods from Baqaee and Farhi (2020) to show how distortions affect aggregate output in distorted economies.

Welfare Losses in General Equilibrium. We can combine Lemma 1 with the input market clearing condition (Equation 3) and the aggregator (8) to solve for the marginal effect of wedges on the final good, Y. We will adopt the notation of Baqaee and Farhi (2020): they prove a version of our results under constant-returns-to-scale production.[9][10] We next derive how changes in wedges affect aggregate output, Y.

Proposition 1 (Effect of Wedges on Output in the General CES Case)**.** *Consider a horizontal economy with CES aggregation. The effect of a change in wedges on aggregate output is*

$$d \log Y = -\mathbb{E}\left[\mathcal{E}_i \lambda_i \hat{\mu} \cdot d \log \mu_i\right] \tag{11}$$

[8] See Hines (1999) for a history of this literature in economics, dating back to almost 200 years to Jules Dupuit in 1844.

[9] The results of Baqaee and Farhi (2020) are substantially more general in that they allow for arbitrary input-output structure. Their formulas can also be modified to capture decreasing returns to scale through a fixed-factors approach; our results are slightly more general than the fixed-factors approach in that we can allow for increasing returns to scale (as long as downward-sloping demand ensures that the firm's objective remains concave).

[10] Baqaee and Farhi (2020) distinguish between sales shares, revenue-based Domar weights, and cost-based Domar weights. Here, those notions are all equivalent because we focus on a horizontal economy, and so we simply define λ_i as the sales share.

where $\lambda_i := \frac{p_i y_i}{\int p_i y_i di}$ *denotes the sales share of firm* i*,* $\mathcal{E}_i$ *is the (negative) elasticity of* y_i *to the wedge* μ_i*, and* $\hat{\mu}_i := \frac{\mu_i - \tilde{\mu}}{\mu_i}$ *is the percent deviation of the wedge from the weighted harmonic average,* $\tilde{\mu} := \frac{\mathbb{E}[\lambda_i \mathcal{E}_i]}{\mathbb{E}\left[\lambda_i \mathcal{E}_i \mu_i^{-1}\right]}$*.*

To derive a formula for the cost of misallocation, we can integrate $d\log Y / d\log\mu$ along the path from the distorted to the undistorted economy, taking advantage of the fact that the wedgeless economy is efficient. Let $\mathcal{L} := \log Y^* - \log Y$ denote the losses from misallocation. Define $\log \check{\mu}(t) = t \cdot \log \mu$. With some abuse of notation, we have

$$\begin{aligned}
\mathcal{L} &= -\int_0^1 \frac{d\log Y\left(\check{\mu}(t)\right)}{d\log\mu} \cdot \frac{d\log\check{\mu}(t)}{dt} dt \\
&= -\int_0^1 \mathbb{E}\left[\frac{d\log Y\left(\check{\mu}(t)\right)}{d\log\mu_i} \cdot \frac{d\log\check{\mu}_i(t)}{dt}\right] dt \\
&= -\mathbb{E}\left[\left(\int_0^1 \frac{d\log Y\left(\check{\mu}(t)\right)}{d\log\mu_i} dt\right) \log\mu_i\right]
\end{aligned} \tag{12}$$

To approximate this integral up to second order, we can use the trapezoid rule. This tells us that the integral is approximated by the wedges, $\log\mu$, times the average of $\frac{d\log Y(\check{\mu}(t))}{d\log\mu}$ evaluated at $\check{\mu} = \mu$ and $\check{\mu} = 1$. As shown in Bigio and La'O (2020), the envelope theorem implies that the first-order effect of wedges on output (holding inputs fixed) is zero, so $\frac{d\log Y}{d\log\mu}$ is zero in the wedgeless economy. Thus, the losses from misallocation are given by:

$$\mathcal{L} \approx -\frac{1}{2} \cdot \mathbb{E}\left[\frac{d\log Y}{d\log\mu_i} \log\mu_i\right] \tag{13}$$

This leads to our main aggregation result.

Proposition 2 (Approximate Formula for the General CES Case)**.** *Consider a horizontal economy with CES aggregation. The cost of misallocation,* $\mathcal{L} := \log Y^* - \log Y$*, is given by*

$$\begin{aligned}
\mathcal{L} &\approx \frac{1}{2}\mathbb{E}\left[\mathcal{E}_i \lambda_i \hat{\mu} \log\mu_i\right] \\
&\approx \frac{1}{2}\mathbb{E}_{\lambda_i}\left[\mathcal{E}_i\right] \cdot Var_{\lambda_i \mathcal{E}_i}\left(\log\mu_i\right)
\end{aligned} \tag{14}$$

where $Var_{\lambda_i \mathcal{E}_i}(\log\mu_i)$ *is the sales-times-elasticity-weighted variance of the log wedges, and* $\mathbb{E}_{\lambda_i}[\mathcal{E}_i]$ *is the sales-weighted average* $\mathcal{E}_i$*.*

Misallocation depends on the (weighted) variance of log wedges, and on the (weighted average) elasticity of output with respect to the wedge. Equivalently, since $\log \text{VMPK}_i = \log r + \log\mu_i$, misallocation depends on the variance of log VMPK. The discussion earlier in

this section made clear that VMPK is equalized across firms in efficient economies. Proposition 2 further tightens the connection between dispersion in VMPK and misallocation, providing us with the relevant moment of the VMPK distribution and the formula to map that moment to the cost of misallocation. Note also that this weighted variance formula can be interpreted as the sum of Harberger triangles, with the wedges appropriately centered around the weighted average wedge.

Special Case: Log-Linear-Log-Normal. Our approximate formula will become exact in a special case. For this special case, we will specialize to a log-linear (Cobb-Douglas) production function

$$\log y_i = \log z_i + \alpha \log k_i \tag{15}$$

with all firms having the same elasticity of output with respect to capital, α. We will assume that wedges and productivity are jointly log-normal. That is, we assume that $(\log z_i, \log \mu_i)$ is multivariate normal.

We will define aggregate productivity as

$$\log Z := \log Y - \alpha \log K \tag{16}$$

This formulation is convenient because we will find that when aggregate productivity is defined this way, we can express aggregate productivity as depending only on the distribution of individual productivities and wedges, and not on the aggregate supply of capital. Thus, our results on the effect of wedges on $\log Z$ will also tell us how wedges affect Y, holding aggregate capital K fixed.

Exploiting the assumption of joint log-normality, as well as the log-linearity of the setup, we obtain the following formula through some manipulations:

$$\log Z = \mathbb{E}\left[\log z_i\right] - \frac{1}{2}\mathcal{E}\mathrm{Var}\left(\log \mu_i\right) + \frac{1}{2}\mathcal{E}\frac{1}{\alpha^2} \cdot \mathrm{Var}\left(\log z_i\right) - \frac{1}{2}\frac{1}{\alpha}\mathrm{Var}\left(\log z_i\right) \tag{17}$$

where $\mathcal{E} := \left(\frac{1-\alpha}{\alpha} + \frac{1}{\theta}\right)^{-1}$ is the (negative) elasticity of firm output with respect to the wedge. To derive the cost of misallocation, we simply compare Z under the economy with wedges to Z^*: aggregate productivity in the efficient, wedgeless (meaning $\mu_i = 1$) economy. This yields our aggregation result for the special case:

Proposition 3 (Exact Formula for the CES-Log-Linear-Log-Normal Case)**.** *Consider a horizontal economy with CES aggregation, log-linear production with a homogeneous elasticity of output with respect to capital, and log-normally distributed productivity and wedges. The*

cost of misallocation is given by

$$\underbrace{\log Z^* - \log Z}_{\textit{Losses from Capital Misallocation}} = \frac{1}{2}\mathcal{E} \cdot Var(\log \mu_i) \tag{18}$$

where $\mathcal{E} := \left(\frac{1-\alpha}{\alpha} + \frac{1}{\theta}\right)^{-1}$ *is the (negative) elasticity of output with respect to the wedge.*

This shows that our misallocation formula, which was a valid second-order approximation in the general case, is an exact result in the Cobb-Douglas log-normal case. Moreover, in the special case, weighted objects coincide with their unweighted counterparts.

Weighted vs. Unweighted Variance. Measuring the weighted variance in the fully general case requires observing the weights $\lambda_i \cdot \mathcal{E}_i$. This is not feasible in practice, since we do not observe each firm's $\mathcal{E}_i$. We can go part of the way towards a weighted variance by using sales weights, which we do in Appendix Table 9. However, our main results focus on estimating the unweighted variance of log VMPK, rather than the weighted variance.

Under appropriate statistical assumptions about the joint distribution of the weights and wedges, the weighted and unweighted variance of wedges will coincide (this is true, for example, in the log-normal special case). In these cases, it will typically be more efficient to estimate an unweighted variance than a weighted variance. More generally, we suspect that the difference between the weighted and unweighted variances is unlikely to be too large, especially compared to the statistical uncertainty of the estimates.

Simulation Results on Accuracy of Second-Order Approximation. Our formula for the cost of misallocation is exact for a special case, but in general is only accurate as a second-order approximation. How accurate is that approximation? We explore this question in Appendix Table 6. We assume a Cobb-Douglas production function, and a standard deviation of 1.2 for $\log z_i$. We take from Section 5 our preferred calibration of θ and α, and our preferred estimates of the standard deviation of $\log \mu_i$. We explore different distributions of $\log \mu_i$, including a normal distribution, uniform distribution, and two-point distribution (as in Restuccia and Rogerson, 2008). We also explore heterogeneous returns to scale. Throughout, we find that the simulated gains from optimally reallocating capital are similar to the analytic approximation.

Misallocation with Multiple Inputs. Our main analysis focuses on the case in which capital is the only input. We do this because we exploit an experiment which identifies the marginal revenue product of capital. What if there are other inputs, such as labor?

When there are multiple inputs, we can say two things. First, without any additional assumptions, we can say that our thought experiment identifies a lower bound on the losses

from misallocation. This is because the problem of optimally allocating capital can be thought of as part of a larger, nested optimization problem, in which the planner allocates all inputs. First, let us define the firm's production function, with only capital as an input, as the output of its multi-input production function in which other inputs (e.g. labor) are fixed:

$$f_i\left(k_i; l_i\right) := y_i\left(k_i, l_i\right) \tag{19}$$

We can then set up the planner's problem as a nested optimization problem:

$$\begin{aligned} Y^* &= \max_{\{k_i, l_i\}_{i\in[0,1]}} Y\left(\{y_i\left(k_i, l_i\right)\}_{i\in[0,1]}\right) \\ &= \max_{\{l_i\}_{i\in[0,1]}} \max_{\{k_i\}_{i\in[0,1]}} Y\left(\{f_i\left(k_i; l_i\right)\}_{i\in[0,1]}\right) \\ &= \max_{\{l_i\}_{i\in[0,1]}} Y^*\left(\{l_i\}_{i\in[0,1]}\right) \end{aligned} \tag{20}$$

In the context of Equation 20, our measure of misallocation focuses on the inner problem of capital allocation, measuring the benefits of optimally reallocating capital while holding other inputs (e.g. labor) fixed. These benefits must, necessarily, be at least as large as the benefits of reallocating not only capital but also other inputs.

Second, under stronger assumptions, we can use the dispersion in the marginal revenue product of capital to infer the cost of misallocation of all inputs. For this, we need two assumptions. First, we need that wedges affect all inputs symmetrically: this implies that distortions take the form of a markup or a "revenue tax." This allows us to infer the distortion to other inputs from the distortion facing capital. Second, we need the firm's production function to be Cobb-Douglas, and for all firms to have the same relative shares of each inputs. This second assumption is quite restrictive, but is necessary to safely extend our earlier results to a multiple input setting.

With these two assumptions, firms may differ in their returns to scale, but they will all have the same input mix: the revenue tax only distorts scale, and the assumption on production functions ensures that undistorted firms will all choose the same mix of inputs. The fact that the input mix is the same at all firms is essential. For example, if firm A was more labor intensive than firm B, then we could not simply reallocate inputs from A to B: this would require additional labor, otherwise we would violate the input supply constraint. By guaranteeing that the input mix is the same at all firms, we can safely reduce the various inputs that firms use to a single composite input.

We can then apply our previous results. Assume there are are M inputs, with each input denoted $x_{m,i}$, and constrain aggregate inputs to be inelastic, $\mathbb{E}\left[x_{m,i}\right] = \bar{X}_m$. This yields:

Proposition 4 (Approximate Formula with Multiple Inputs)**.** *Consider a horizontal economy with CES aggregation. Let each firm have a Cobb-Douglas production function* $\log y_i = \log z_i + \sum_M \alpha_{m,i} \log x_{m,i}$*, where* $\alpha_{m,i} = \tilde{\alpha}_m \cdot \sum_M \alpha_{m,i}$ *for all firms. Assume that wedges only take the form of a markup on the firm's output. The cost of misallocation,* $\mathcal{L} := \log Y^* - \log Y$*, is given by*

$$\mathcal{L} \approx \frac{1}{2}\mathbb{E}_{\lambda_i}\left[\mathcal{E}_i\right] \cdot Var_{\lambda_i \mathcal{E}_i}\left(\log \mu_i\right) \tag{21}$$

and the firm-specific elasticity $\mathcal{E}_i$ *is given by*

$$\mathcal{E}_i = \left(\frac{1 - \sum_M \alpha_{m,i}}{\sum_M \alpha_{m,i}} + \frac{1}{\theta}\right)^{-1}.$$

In our empirical analysis, we will focus on the benefits of reallocating only capital, keeping other inputs fixed. However, we will use the formula in Proposition 4 to provide an estimate of the benefits from optimally reallocating all inputs. As Proposition 4 makes clear, this requires stronger assumptions than our main results, but strictly weaker assumptions about production functions than in the standard method. The proposition imposes Cobb-Douglas production with constant shares, as in Hsieh and Klenow (2009), but our formula allows for arbitrary heterogeneity in returns to scale across firms.

Extending Results for Microenterprises to Whole Economy. We will exploit an experiment run on microenterprises, and thus our results will be informative only about the variance of log wedges for this subset of firms. What implications do these results have for the overall economy? One route is to model the macroeconomy in a way that allows us to separate the "microenterprise sector" from other sectors of the economy. For example, if household demand has a nested CES structure with the microenterprise sector as one of the nests, and if we hold each sector's aggregate capital fixed, then our previous results would tell us how much the aggregate output of the microenterprise sector could increase if we reallocated capital within sector.

What if we instead allowed for reallocation of capital across all firms, including those that are not necessarily in the same size category? Since our formula for misallocation depends in a simple way on the variance of log wedges, it is straightforward to see how the variance of log wedges for the whole economy depends on the moments of the wedge distribution for microenterprises. We will simplify by considering a setting where $\mathcal{E}_i$ is independent of (λ_i, μ_i), so we can focus solely on the sales-weighted variance of log wedges. Suppose there are two groups of firms, Small and Big. Define λ_g as the total share of sales by group g. By

the law of total variance, we have that:

$$\begin{aligned}\text{Var}_\lambda\left(\log \mu_i\right) = \lambda_{\text{Small}} \cdot \text{Var}_\lambda\left(\log \mu_i \mid \text{Small}\right) + \lambda_{\text{Big}} \cdot \text{Var}_\lambda\left(\log \mu_i \mid \text{Big}\right) \\ + \sum_{g \in \{\text{Small,Big}\}} \lambda_g \cdot \left(\mathbb{E}_\lambda\left[\log \mu_i \mid g\right] - \mathbb{E}_\lambda\left[\log \mu_i\right]\right)^2\end{aligned} \tag{22}$$

The overall level of misallocation in the economy thus depends on a weighted average of the weighted variance of log wedges within each group, plus the variance across groups of the conditional expectation of the log wedge. It follows that the variance of log wedges that we find among microenterprises provides a lower bound on the overall degree of misallocation: the overall variance must be at least $\lambda_{\text{Small}} \cdot \text{Var}_\lambda\left(\log \mu_i \mid \text{Small}\right)$. Moreover, if we could identify the moments of the wedge distribution for large firms in a separate analysis, then we could easily combine the two estimates to get a measure of the total misallocation in the economy.

2.4 Under CES Demand, $\text{Var}\left(\log \text{VMPK}_i\right) = \text{Var}\left(\log \text{MRPK}_i\right)$

We have so far shown that the cost of misallocation is a function of the variance of log VMPK. However, in practice we will not have access to separate data on prices and quantities, and thus we will measure the marginal revenue product of capital (MRPK) rather than VMPK. The MRPK is in general lower than VMPK, because the former includes a price effect: when capital increases, output rises, which lowers revenue. However, under CES demand, MRPK will be proportional to VMPK, because the price elasticity of demand is constant and the same across all goods. We have:

$$\begin{aligned}\text{MRPK}_i &= p_i \frac{dy_i}{dk_i} + y_i \frac{dp_i}{dy_i} \cdot \frac{dy_i}{dk_i} \\ &= \left(1 + \frac{d \log p_i}{d \log y_i}\right) \cdot p_i \frac{dy_i}{dk_i} \\ &= \frac{\theta - 1}{\theta} \cdot \text{VMPK}_i\end{aligned} \tag{23}$$

This shows us that under CES aggregation, $\log \text{MRPK}_i = \log \text{VMPK}_i + \log \frac{\theta-1}{\theta}$. By extension the variance of log VMPK and of log MRPK are the same. More broadly, this implies that the variance of log wedges and log MRPK are the same, given the firm's first order condition in Equation 7. Note that this relies solely on CES demand and on the final good producer being a price taker.

We summarize this the following proposition.

Proposition 5 (Variance of log VMPK and log MRPK Are the Same Under CES)**.** *Consider a horizontal economy with CES aggregation and a price-taking final good producer. In this economy,*

$$\begin{aligned} \text{Var}\left(\log \mu_i\right) &= \text{Var}\left(\log \text{VMPK}_i\right) \\ &= \text{Var}\left(\log \text{MRPK}_i\right) \end{aligned} \tag{24}$$

This result is convenient because it allows us to focus on estimating the variance of log MRPK, which is something we will show how to measure in the next section. Moreover, this result is closely connected to a special property of CES demand. Dhingra and Morrow (2019) show that in CES economies, the monopolistically competitive equilibrium (without additional distortions) is efficient, despite the fact that firms charge markups. A key reason for this is that firms charge homogeneous multiplicative markups, and thus equalization of MRPK implies equalization of VMPK. In CES economies with distortions, the above result shows that it does not matter whether the wedge is expressed as a distortion to the planner's first-order condition (VMPK deviates from the marginal cost) or as a distortion to the firm's first-order condition (MRPK deviates from marginal cost): the variance of log wedges is the same, and thus the implied cost of misallocation is the same.

Alternative with Homogeneous Products. Even without CES demand, we will also have that $\text{Var}\left(\log \text{VMPK}_i\right) \approx \text{Var}\left(\log \text{MRPK}_i\right)$ if goods are highly substitutable, that is, if $\frac{d \log p_i}{d \log y_i} \approx 0 \, \forall i$. This may be the case in a setting such as the one we study, in which many microenterprises produce similar products. This provides additional assurance that our estimates of $\text{Var}\left(\log \text{MRPK}_i\right)$ are likely to also be estimates of $\text{Var}\left(\log \text{VMPK}_i\right)$.

3 Measuring Marginal Products with an Experiment

Summary

In the previous section, we showed that the cost of misallocation can be expressed as the variance of log MRPK, times one half the elasticity of output with respect to the wedge. We will measure this variance in the data, and then calibrate the elasticity using standard parameter values.

In this section, we will show how to measure the variance of log MRPK using randomly assigned grants to microenterpreneurs. In doing so, we recast a a microeconomic question ("What is the variance of log MRPK across firms?") as an econometric question ("How can we conduct valid inference on a particular nonlinear function of parameters from a linear IV

model?").

We use the randomly assigned grants as an instrument for capital, solving the problem that capital is typically endogenous to productivity. We then project MRPK onto observable baseline characteristics, by using the grants instrument to estimate a linear IV model with heterogeneous treatment effects by baseline characteristics.

Projecting MRPK onto observables allows us to estimate the variance of the conditional expectation of MRPK. By the law of total variance, this provides us with a lower bound on the total variance of MRPK. Moreover, focusing on variation in returns to capital that can be predicted *ex ante* ensures that we are estimating misallocation, rather than risk.

Finally, we show how to use standardized principal components to construct an orthonormal basis for the baseline characteristics. We run the heterogeneous linear IV specification using these principal components as the heterogeneity variables. In addition to being useful for variable selection, this allows us to express the variance as a simple function of the coefficients from the IV model. In particular, we have that $\text{Var}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right) = \gamma'\gamma$, where X_i are the baseline characteristics and γ is the (vector-valued) coefficient on the interaction between X_i and capital. We also have that $\frac{\text{SD}(\mathbb{E}[\text{MRPK}_i|X_i])}{\mathbb{E}[\text{MRPK}_i]} = \frac{\sqrt{\gamma'\gamma}}{\beta}$, where β is the coefficient on capital. Using the log-normal approximation, this yields the formula $\text{Var}\left(\log \mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right) = \log\left(1 + \frac{\gamma'\gamma}{\beta^2}\right)$, which provides a lower bound on the total variance of log MRPK.

3.1 Solving the Identification Challenge with an Experiment

We wish to estimate the average MRPK, as well as moments of its distribution across firms. However, we face an identification challenge: capital is chosen endogenously, and so it will generally covary with productivity. In order to isolate the effect of capital, we need an instrument for capital. This instrument needs to affect capital, to be exogenous (e.g. it cannot be correlated with productivity), and to only affect the outcome through its effect on capital.

Using Grants as an Instrument. We will use an experiment by de Mel et al. (2008) to provide this instrument. They run an experiment among a sample of Sri Lankan microentreprises, in which they randomly offer grants to some microentrepreneurs in order to fund capital investment. In addition to the control group, their experiment has four treatment arms: participants could receive grants as cash or in-kind,[11] in the amount of 10,000 or 20,000 rupees. Importantly, the rollout of the treatment was staggered: in the first wave,

[11]The grant winner would tell the experimenter what inventory and equipment they wished to buy, up to the size of the grant, and then the research team would buy that capital on behalf of the entrepreneur.

no firms were treated and they did not have knowledge of the treatment, some firms were randomly treated between waves 1 and 2, some more firms were randomly treated between waves 3 and 4, and the control group received 2,500 rupees after wave 5, as a surprise gift and encouragement to stay in the study.

We will use the grant in this experiment as an instrument for capital. Following de Mel et al. (2008), we will pool the different arms of the treatment, and instead use the amount of the grant received as our instrument. The grant affects capital, and by design it is exogenous (uncorrelated with other shocks, such as productivity).

Using Profits to Isolate the MRPK. Importantly however, we also need our instrument to satisfy an exclusion restriction. The primary concern here is that the grant will also affect other inputs, besides capital. This will be a problem because those other inputs also affect revenue. More concretely, if we take the total derivative and linearize, we have:

$$p_i y_i = \text{MRPK}_i \cdot k_i + \text{MRPL}_i \cdot l_i + \text{MRPM}_i \cdot m_i \tag{25}$$

$$\implies p_i y_i - w l_i - c m_i = \text{MRPK}_i \cdot k_i + (\text{MRPL}_i - w) \cdot l_i + (\text{MRPM}_i - c) \cdot m_i \tag{26}$$

If our outcome is revenue, and the instrument affects other inputs like labor, l_i, or materials, m_i, then that would result in a violation of the exclusion restriction.

To resolve this issue, we follow de Mel et al. (2008) and use reported profits as the outcome. In practice, we believe that this means subtracting off the cost of labor and materials, but not subtracting off a cost of capital.[12] In accounting terms, we suspect that microentrepreneurs answer the profits question by giving their earnings before interest, depreciation, and amortization (EBIDA).

By using profits as the outcome, we attenuate the bias coming from changes in other inputs. If the marginal revenue product on other inputs is equal to the price of those inputs (that is, if $\text{MRPL}_i = w$ and $\text{MRPM}_i = c$), then this strategy will eliminate violations of the exclusion restriction. This assumption is common for materials: many authors, such as Hsieh and Klenow (2009), use a value-added production function that implicitly assumes materials are undistorted. In this setting, we have both a theoretical and empirical justification for this assumption. Based on a simple model of credit constraints, we would suspect that any distortions for materials are likely to be much smaller than those for capital, since materials are purchased in smaller amounts on an as-needed basis. Empirically, we can estimate the

[12] Moreover, the survey asks for profits before payments to the owner, so it is not accounting for any implicit wage for the owner. However, de Mel et al. (2008) find that attempting to adjust profits by subtracting off an implicit wage for the owner does not meaningfully affect estimates of returns.

average marginal revenue product of materials using the following regression:

$$\text{Revenue}_{it} = \beta \cdot m_{it} + \alpha_i + \delta_t + \varepsilon_{it}$$

where Revenue_{it} is reported real revenue at firm i in time t, m_{it} is materials (we compute materials as the difference between revenue and profits). As in de Mel et al. (2008), we use the cumulative amount of the grant received by firm i at time t as our instrument. This yields a point estimate of 1.18 for the marginal revenue product of materials, with a firm-clustered standard error of 0.07. Note that this estimate is biased upward: the instrument also increases capital, and so the true marginal revenue product of materials is lower than this estimate. Since an undistorted firm would have an MRPM of one (one rupee of materials increases profits by one rupee on the margin), this is strong evidence that any materials wedges are small in this setting.

For labor, we will instead rely on the fact that labor does not seem to respond much to the treatment. As a result, de Mel et al. (2008) find that accounting for the effect of the treatment on labor does not seem to meaningfully affect the estimated returns to capital, for plausible values of MRPL_i.

3.2 Projecting Onto Baseline Characteristics Provides a Lower Bound for the Total Variance

To estimate the returns to capital, de Mel et al. (2008) estimate the following linear IV model:

$$\text{Profit}_{it} = \beta k_{it} + \alpha_i + \delta_t + \varepsilon_{it} \tag{27}$$

where Profit_{it} is real profits,[13] k_{it} is capital, and the excluded instrument, Z_{it}, is the cumulative amount of the grant that the firm i has received by time t. Note that the time fixed effects are necessary for identification in this setting, since the treatment was staggered over time, and is thus correlated with the time fixed effect.

We modify this homogeneous model to estimate heterogeneous returns to capital based on the firm's baseline characteristics. Let X_i denote characteristics of the firm measured at baseline: these characteristics are measured before the treatment is announced, and thus are not affected by the treatment. We can estimate heterogeneous effects by interacting capital with these covariates. We estimate the following heterogeneous linear IV model:

$$\text{Profit}_{it} = \beta k_{it} + \gamma' X_i \times k_{it} + \alpha_i + \delta_t + \delta_t^{X\prime} X_i + \varepsilon_{it} \tag{28}$$

[13] We adjust profits for inflation, as in the original paper.

where the excluded instruments are now Z_{it} and $Z_{it} \times X_i$. Note that in order to ensure identification, we must now control for interacted time fixed effects, $\delta_t^{X\prime} X_i$. This is an extension of the earlier issue for the homogeneous model: the instrument is correlated with time, and therefore the instrument interacted with a baseline characteristic is correlated with that baseline characteristic interacted with time. Once we condition on these interacted fixed effects, Z_{it} and $Z_{it} \times X_i$ are uncorrelated with the residual ε_{it}.[14]

Once we know the parameters of the above model, we can estimate the distribution of $\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]$, the expected returns to capital given covariates X_i. For example, it is straightforward to compute the variance of expected returns to capital: $\text{Var}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right) = \text{Var}\left(\gamma' X_i\right) = \gamma' \text{Var}\left(X_i\right) \gamma$. In contrast, we cannot compute the distribution of MRPK_i: in general, it is not possible to compute the distribution of treatment effects without imposing additional assumptions.

Fortunately, we can use the variance of expected returns as a lower bound on the total variance. The law of total variance states

$$\text{Var}\left(\text{MRPK}_i\right) = \text{Var}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right) + \mathbb{E}\left[\text{Var}\left(\text{MRPK}_i\right) \mid X_i\right] \tag{29}$$

Since the expectation of the conditional variance, $\mathbb{E}\left[\text{Var}\left(\text{MRPK}_i\right) \mid X_i\right]$, cannot be negative, this implies that the variance of expected MRPK is a lower bound on the total variance of MRPK. We will focus on estimating this variance, and use it to provide a lower bound on the cost of misallocation. Our estimates are thus conservative, in the sense that we will only a capture a portion of the full dispersion in MRPK.

Although our estimates provide a lower bound on the variance of MRPK, our aggregation results are actually stated in terms of the variance of log MRPK. To estimate the variance of log MRPK, we will use an approximation based on the lognormal distribution. If MRPK is lognormally distributed, then we can back out the variance of log MRPK from the coefficient of variation for MRPK (the standard deviation divided by the mean). Then, we have $\text{Var}\left(\log \text{MRPK}_i\right) = \log\left(1 + \frac{\text{Var}(\text{MRPK}_i)}{\mathbb{E}[\text{MRPK}_i]^2}\right)$. This formula is convenient because we can replace $\text{Var}\left(\text{MRPK}_i\right)$ with $\text{Var}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)$, and still be sure that the formula gives us a lower bound on the total variance of log MRPK.[15]

[14] Whenever one estimates a model with an interaction with X_i, the model needs to include a main effect for X_i. Here, that main effect is absorbed by the interacted fixed effects, and also would be absorbed by the firm fixed effects.

[15] An alternative approach would have been to compute $\text{Var}\left(\log \mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)$ in sample, using the estimated β and γ. However, this approach has three problems. First, this approach does not necessarily recover a lower bound on $\text{Var}\left(\log \mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)$, since it is not the variance of the conditional expectation of log MRPK (it is the variance of the log of the conditional expectation). Second, for certain values of X_i, the estimated expected MRPK may be negative in practice: one cannot take the log of a negative. Finally, and relatedly, even if all the predicted values of MRPK are positive, this approach is likely to be

3.3 Standardized Principal Components Turns $\mathbf{Var}\left(\log \mathbf{MRPK}_i\right)$ into a Simple Function of IV Coefficients

Our strategy so far provides a formula for the variance of expected returns in terms of both model parameters and the distribution of covariates: $\text{Var}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right) = \gamma' \text{Var}\left(X_i\right) \gamma$. We can simplify this formula by re-expressing the covariates X_i using an orthonormal basis: a set of variables that spans the original X_i, but in which the new variables are orthogonal to each other and each have standard deviation one. Under this new basis, $\text{Var}\left(X_i\right)$ is simply an identity matrix, and so $\text{Var}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right) = \gamma'\gamma$.

We construct this orthonormal basis by using standardized principal components. Principal components gives us a set of orthogonal factors, ordered by how much of the variance of the variables they explain.[16] The ordered nature of the factors also has auxiliary benefits for variable selection: if we wish to instead use a subset of our factors, then principal components gives us a natural choice of which ones to use (if we want to only use K covariates, then we use the first K factors). By standardizing these components, we also ensure they have mean zero and standard deviation one.

With an orthonormal basis of mean zero variables, we obtain simple formulas for our objects of interest. We have the following formulas:

$$\text{Var}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right) = \gamma'\gamma \tag{30}$$

$$\frac{\text{SD}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)}{\mathbb{E}\left[\text{MRPK}_i\right]} = \frac{\sqrt{\gamma'\gamma}}{\beta} \tag{31}$$

Using the log-normal approximation, we get a formula for $\text{Var}\left(\log \mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)$, which also serves as a lower bound for $\text{Var}\left(\log \text{MRPK}_i\right)$.[17]

$$\text{Var}\left(\log \mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right) \approx \log\left(1 + \frac{\gamma'\gamma}{\beta^2}\right) \tag{32}$$

$$\text{Var}\left(\log \text{MRPK}_i\right) \approx \log\left(1 + \frac{\text{Var}\left(\text{MRPK}_i\right)}{\mathbb{E}\left[\text{MRPK}_i\right]^2}\right) \geq \log\left(1 + \frac{\gamma'\gamma}{\beta^2}\right) \tag{33}$$

very unstable when some firms have low predicted MRPK, and would be very sensitive to outliers in the X_i distribution.

[16] As is standard practice, we also standardize the raw variables before performing principal components.

[17] In general, $\text{Var}\left(\log \mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right) \neq \text{Var}\left(\mathbb{E}\left[\log \text{MRPK}_i \mid X_i\right]\right)$, and so an estimate of the former need not be a lower bound for the total variance, $\text{Var}\left(\log \text{MRPK}_i\right)$. We are relying, however, on the log-normal approximation, under which $\text{Var}\left(\log \text{MRPK}_i\right) = \log\left(1 + \frac{\text{Var}(\text{MRPK}_i)}{\mathbb{E}[\text{MRPK}_i]^2}\right)$. Since $\text{Var}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right) \leq \text{Var}\left(\text{MRPK}_i\right)$, we have a lower bound on the total variance of log MRPK.

3.4 Our Method Solves Many Measurement Challenges

Our method is distinct in two ways: we measure marginal products directly using exogenous variation in capital, and we project variation in marginal products onto observable characteristics. These distinctive features of our method solve many measurement challenges that have plagued prior work.

Comparison to Standard Approach. The standard approach to measuring misallocation, pioneered by Hsieh and Klenow (2009), assumes a Cobb-Douglas production function and CES demand, and infers marginal products from data on inputs and outputs. Implicitly, this methodology infers marginal products from average products. For a loglinear production function, $y = zk^{\alpha}$, the marginal product will in general be proportional to the average product: $\frac{dy}{dk} = \alpha z k^{\alpha-1} = \alpha \frac{y}{k}$. Note also that under CES demand, firms charge constant multiplicative markups, and so APK is proportional to $\frac{PY}{K}$, and VMPK is proportional to MRPK. Thus if all firms use a loglinear production function, with the same elasticity of output to capital, α, then the variance of log average products will be the same as the variance of log marginal products, and the standard approach will recover the correct variance.

However, since the standard approach relies on a homogeneous, loglinear production function, it will fail if the production function is not homogeneous or not loglinear. For example, suppose that firms have loglinear production functions with different elasticities, α_i. By our earlier derivation, we have that $\text{APK}_i = \frac{p_i y_i}{k_i} = \frac{1}{\alpha_i} \cdot \text{VMPK}_i$. Taking logs, we then have:

$$\text{Var}\left(\log \text{APK}_i\right) = \text{Var}\left(\log \text{VMPK}_i\right) + \text{Var}\left(\log \alpha_i\right) - 2 \cdot \text{Cov}\left(\log \alpha_i, \log \text{VMPK}_i\right) \tag{34}$$

It follows that $\text{Var}\left(\log \text{APK}_i\right)$ will generally not measure $\text{Var}\left(\log \text{VMPK}_i\right)$ in an environment with loglinear production functions that have different elasticities, since it will mix up true variation in VMPK with variation in α_i (for further discussion of this point, see also Haltiwanger et al. 2018 and Carrillo et al. 2023). In fact, in an allocatively efficient environment, there will be no variation in VMPK, and $\text{Var}\left(\log \text{APK}_i\right)$ will simply measure $\text{Var}\left(\log \alpha_i\right)$.

Deviations from loglinear production, such as fixed costs, will also cause the standard approach to fail. Suppose that we have loglinear production with a fixed cost, so $y_i = z_i k_i^{\alpha} - c$. Even if productivity z_i is the only part of the production function that varies across firms, this will cause the standard approach to fail. In this setting, $\text{VMPK}_i = p_i \alpha z_i k_i^{\alpha-1}$, and $\text{APK}_i = \frac{1}{\alpha} \cdot \text{VMPK}_i - p_i c / k_i = \frac{1}{\alpha} \cdot \text{VMPK}_i - c \cdot \left(\frac{\text{VMPK}_i}{\alpha z_i}\right)^{1/(1-\alpha)}$. In this case, average products are in general not proportional to marginal products. Moreover, like before, other sources of variation, besides wedges, will drive variation in average products. Variation in

productivity z_i will lead to variation in average products under fixed costs, even if VMPK is the same for all firms.

In contrast, our method sidesteps this issue because we measure marginal products directly. By using randomized grants as an instrument for capital, we isolate how a change in capital on the margin will affect output. We thus do not rely on any assumed relationship between average products and marginal products.

Flexibility vs. Heterogeneity. The standard approach imposes a homogeneous Cobb-Douglas production function on all firms. Deviations from these assumptions will cause it to infer misallocation even in an efficient setting. A natural question is whether the standard approach can be rescued by estimating a more flexible functional form, such as translog. Unfortunately, flexibility is in general insufficient to remedy the problem.

To see this, suppose firms use the production function $z_i \cdot f(k_i)$.[18] For expositional simplicity, we will assume $p_i = 1$ for all firms. Equalization of marginal products implies:

$$z_i = f'(k_i)/r \,\forall i \tag{35}$$

$$\implies z_i = z_j \text{ if } k_i = k_j \tag{36}$$

$$\implies y_i = y_j \text{ if } k_i = k_j \tag{37}$$

Imposing a fully flexible but homogeneous production function requires that if any two firms have the same level of inputs, they must also have the same level of output, and any deviation from this will be interpreted as misallocation. Thus, estimating a flexible production will in general be insufficient to ensure robust estimates of the cost of misallocation. The robustness of our method comes not just from its ability to handle flexible functional forms, but also from its ability to handle heterogeneous production functions.

Our Method is Robust to Measurement Error. The standard approach measures the variance of log average products, and is thus very sensitive to measurement error in inputs or in output. Prior work has shown that accounting for this measurement error has quantitatively important implications for the measurement of misallocation (Bils et al., 2021; Gollin and Udry, 2021; Rotemberg and White, 2021). In contrast, classical measurement error in inputs and outputs will in general have no effect on the consistency of IV estimates of MRPK. Thus, our method is completely robust to this form of measurement error.

[18] Here we are allowing limited heterogeneity by allowing for a Hicks-neutral productivity shifter. If we allowed arbitrary productivity shifters, then we would simply replicate the fully flexible and heterogeneous model $y_i = f_i(k_i)$.

Measurement error in our covariates, X_i, will in general lower the usefulness of these covariates in predicting MRPK. This will lower our estimate of $\text{Var}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)$, but only because it will actually lower the true variance of $\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]$. Regardless, our method will still provide a valid lower bound for the variance of MRPK.

We Measure Misallocation Rather Than Risk. Projecting returns onto baseline observables is useful econometrically, but it also clarifies the economic interpretation of our estimates. It is important to distinguish between *ex ante* and *ex post* variation in returns. If firms have different expected returns *ex ante* then we would interpret that as misallocation; if firms have the same expected returns but different returns *ex post*, then we would interpret that as risk rather than misallocation.

In general, economists define efficiency relative to what the social planner could implement. Since the planner cannot see the future, efficiency depends on equalizing *expected* marginal products based on the information available at the time of investment. If investment is reversible, then this means that efficiency depends on $\text{Var}\left(\mathbb{E}\left[\text{MRPK}_{it} \mid \Omega_{t-1}\right]\right)$, where Ω_{t-1} represents the planner's information set in $t-1$. We think that the baseline variables we observe as econometricians would also be reasonably be included in the planner's information set. Thus, by the law of total variance, our $\text{Var}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)$ will provide a lower bound on the misallocation-relevant variance of MRPK.

Recent work by David et al. (2022) has highlighted an alternative connection between risk and misallocation: firms whose returns are risky (in the sense of being correlated with aggregate risk) may have higher marginal products, reflecting a risk premium. This variation in expected returns need not reflect inefficiency, since the risk-adjusted returns could be the same across firms. In principle, our method could be extended to estimate a risk-adjusted return if we multiplied profits by the appropriate stochastic discount factor (e.g. we could infer the marginal utility of the representative Sri Lankan household data on aggregate consumption, and construct risk-adjusted profits using the implied marginal utilities). In practice, this would likely require a very large number of time periods to estimate consistently.

Our Estimates Are Not Likely to Be Driven by Adjustment Costs. Prior work on misallocation (Asker et al., 2014; David and Venkateswaran, 2019) has also studied how adjustment frictions may lead to dispersion in APK and MRPK. A critical distinction between our approach and prior approaches is that we measure MRPK, rather than APK. If adjustment costs show up in the data as reduced profits, then adjustment costs are just another component of the MRPK. In models where adjustment costs are a continuously differentiable function of the change in capital (i.e., Q-theory models of investment), the MRPK (including adjustment costs) will not vary across firms in the planner's solution.

Here, the critical difference between our approach and the standard approach is that we estimate MRPK directly, rather than inferring it from the output to capital ratio, and thus our estimates of MRPK should incorporate adjustment costs.

There are two complications to this story, which can theoretically generate (efficient) dispersion in MRPK. First, adjustment cost models where adjustment costs are not continously differentiable, such as a fixed cost of adjustment or partial irreversibility of investment will generate inaction regions. This can cause MRPK to differ across firms, even in the planner's solution.

Second, in models with adjustment costs, the appropriate measure of the MRPK (inclusive of adjustment costs) is the effect of capital on the net present value of output, and so returns to capital over the next month may exhibit some dispersion. For example, the planner may expect that high productivity firms will have lower productivity in the future, and thus today's MRPK may seem higher because the planner anticipates higher adjustment costs in the future.

However, in practice we would expect these differences to die out over the course of a two-year experiment. de Mel et al. (2008) find that there is low autocorrelation of profits among their sample of firms: this is a setting in which a theory based on adjustment costs would predict that returns should quickly revert to the mean.

In the data, we find that *ex ante* differences in returns are persistent. In Appendix Table 11, we estimate an alternative model that projects MRPK onto baseline APK, and allows the covariance of MRPK and baseline APK to vary between the first and second years of the survey. We find that the projection of returns onto baseline APK yields very similar coefficients in the first and second year, implying that differences in returns are persistent. This suggests that adjustment costs are unlikely to explain a large portion of our estimated dispersion in returns to capital.

Comparison to Approach in Carrillo et al. (2023). Our approach is most closely related to recent work by Carrillo et al. (2023). Our work differs from theirs in a number of ways: we study a different setting (Sri Lankan microenterprises vs. Ecuadorian construction firms), focus on a different type of shock (grants that shock capital vs. procurement lotteries that shock demand), and find a different result (we find a sizable cost of misallocation, while they find little misallocation). Methodologically, our approach differs from theirs in that we focus on projecting the wedges onto covariates, and estimating the variance of expected wedges. This produces a lower bound on misallocation, and also ensures that we are measuring misallocation as opposed to risk.

In contrast, Carrillo et al. (2023) target the total variance of the wedges. Economically, this does not distinguish between risk and misallocation, and should thus be viewed as an

upper bound. Since they find low levels of misallocation, an upper bound is useful in their setting. Since we find substantial misallocation, our lower bound approach is more useful in our setting.

Econometrically, Carrillo et al. (2023) estimate the total variance of wedges using an instrumental variable correlated random coefficients model (IV-CRC), following the method of Masten and Torgovitsky (2016). In order to identify not just the mean but also the variance of the treatment effects, they run the linear model and then also square the model, in order to identify $\mathbb{E}[\mu_i]$ and $\mathbb{E}[\mu_i^2]$. This method relies on the linearity of the model, and also requires the instrument to have multiple points of support: a binary instrument will be collinear with its square, and thus cannot separately identify the linear and quadratic endogenous regressors in their squared model. In contrast, our method will work even in the case of binary instruments. Relatedly, in order to identify $\mathbb{E}[\mu_i^2]$, the IV-CRC approach requires that the instrument is fully independent from the residual, rather than just mean-independent. Our method relies only on orthogonality (partialling out controls), as is standard for IV models with interaction effects.

In the setting we study, our method yields substantially more precise estimates than the IV-CRC approach. In Appendix Table 12, we provide estimates of the total variance of MRPK using the Carrillo et al. (2023) method in our setting. We find that the resulting confidence intervals are too wide to be informtive: they include both very large values of misallocation and zero misallocation. Whether our method is also more efficient in other settings will likely depend on how useful baseline covariates are in predicting MRPK.

Broadly, we view the two papers as complementary: Carrillo et al. (2023) provide a method to target the total variance of wedges, which provides an upper bound, while we provide a method to target a component of the variance of wedges that can be predicted by baseline covariates, which provides a lower bound. Future work may find it useful to use one or both methods, depending on the setting.

4 Inference for Nonlinear Functions of Parameters

Summary

We have shown that the variance of the log of expected MRPK can be expressed as a nonlinear function of the parameters of a linear IV model, estimated using an experiment that randomized grants to microenterprises. This provides a lower bound on the total variance of log MRPK and, combined with a calibration of the elasticity of output with respect to the wedge, provides an estimate of the cost of misallocation. In this section, we show how to

conduct valid inference on nonlinear functions of parameters, which allows us to construct confidence sets for our measure of the cost of misallocation using experimental data on firms.

Given asymptotic normality of the IV coefficients, i.e. $\sqrt{n}(\hat{\delta}-\delta_0) \xrightarrow{d} N(0,\Sigma)$, the standard approach to inference on a nonlinear function of these coefficients would be to use the delta method. For a continuous function $g(\cdot)$, the delta method approximates the distribution of $g\left(\hat{\delta}\right)$ as $\mathcal{N}\left(g(\delta_0), \nabla g' \Sigma \nabla g\right)$. In our setting however, where $g(\delta) = \sqrt{\gamma'\gamma}/\beta$, this derivative can be close to zero whenever there is little predictable heterogeneity in marginal products, and can be very large when the average returns to capital are low. This means that the delta method can provide a poor approximation to the distribution of our parameter of interest. In simulations calibrated to our setting, we show that tests based on the delta method suffer from severe size distortions.

An alternative approach is to construct a confidence set using the projection method, i.e. by first constructing a joint confidence set for δ, and then including in the confidence interval for g every value of $g(\delta)$ corresponding to a value of δ in the confidence set. In general this method will be conservative, resulting in overly wide confidence intervals, particularly as the dimension of δ grows.

We suggest an alternative simulation-based approach for generating critical values. The method performs very well in simulations calibrated to the data: it provides correct coverage across a range of true parameter values, while retaining good power properties. We first explain the procedure, and then apply it to construct confidence intervals for the cost of misallocation.

4.1 Inference Procedure

Our proposed method uses simulation to construct critical values for test statistics, rather than relying on the asymptotic approximations given by the delta method. To describe the method, suppose that we observe a vector of parameter estimates for which $\sqrt{n}(\hat{\delta} - \delta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$, and are interested in testing the null hypothesis $H_0 : g(\delta) = \tau_0$, for some continuous function g. Then, given some chosen test statistic $\widehat{T} = T(\widehat{\delta}, \tau_0)$ we can simulate its asymptotic distribution under the null by drawing $\delta^* \sim \mathcal{N}(\delta_0, \Sigma/n)$ and constructing the corresponding statistic $T(\delta^*, \tau_0)$. Quantiles of the simulated distribution can then be used to construct critical values for testing the null hypothesis.

The test statistic we propose is the minimum distance statistic

$$T(\widehat{\delta}, \tau_0) = \min_{\delta : g(\delta) = \tau_0} n(\delta - \widehat{\delta})' \widehat{\Sigma}^{-1} (\delta - \widehat{\delta}) \tag{38}$$

The statistic is intuitive in the sense that it measures the extent to which the data disagrees

with the null hypothesis, taking into account our relative uncertainty about each element of $\widehat{\delta}$. Implementation of the test requires knowledge of the true parameter vector δ_0 and its covariance matrix Σ. We replace these with estimated values that are consistent with the null hypothesis. A constrained estimator $\bar{\delta}$ is given by the solution to

$$\bar{\delta} = \arg \min_{\delta: g(\delta)=\tau_0} n(\delta - \widehat{\delta})' \widehat{\Sigma}^{-1} (\delta - \widehat{\delta}),$$

and a corresponding constrained covariance matrix $\bar{\Sigma}$ can be constructed using $\bar{\delta}$. Given the constrained estimates, we can simulate a p-value for the test statistic by taking some large number of draws $\delta_b^* \sim \mathcal{N}(\bar{\delta}, \widehat{\Sigma})$ and computing the proportion of simulated statistics $T(\delta_b^*, \tau_0)$ that exceed the observed test statistic, i.e.

$$\widehat{p}_{\tau_0} = \frac{1}{B} \sum_{b=1}^{B} 1\big\{T(\delta_b^*, \tau_0) \geq T(\widehat{\delta}, \tau_0)\big\}.$$

A corresponding confidence set is then easily constructed by inverting the resulting test, i.e. collecting the set of τ for which $\widehat{p}_\tau \geq \alpha$ so that we cannot reject the null hypothesis. The procedure is summarized in the following algorithm.

Algorithm 1. *Confidence interval for τ (non-uniform version)*

1. *Estimate the IV regression to obtain parameter estimates $\widehat{\delta} = (\widehat{\beta}, \widehat{\gamma}')'$ and variance matrix $\widehat{\Sigma}$*

2. *Set a null hypothesis $H_0 : \tau = \tau_0$:*

   (a) *compute the constrained parameter estimate $\bar{\delta}$ and the constrained variance matrix $\bar{\Sigma}$,*

   (b) *compute the test statistic*

   $$T(\widehat{\delta}, \tau_0) = \min_{\delta: g(\delta)=\tau_0} n(\delta - \widehat{\delta})' \bar{\Sigma}^{-1} (\delta - \widehat{\delta}),$$

   (c) *for $b = 1, \ldots, B$, simulate $\delta_b \sim N(\bar{\delta}, \bar{\Sigma})$ and compute the statistic*

   $$T_b^*(\delta_b, \tau_0) = \min_{\delta: g(\delta)=\tau_0} n(\delta - \delta_b)' \bar{\Sigma}^{-1} (\delta - \delta_b),$$

   and set the critical value $c_{1-\alpha}(\tau_0)$ as the $(1-\alpha)$-quantile of $T_b^*(\delta_b, \tau_0)$

   (d) *reject $H_0 : \tau = \tau_0$ if $T(\widehat{\delta}, \tau_0) > c_{1-\alpha}(\tau_0)$*

3. *Repeat step 2 for a grid of τ_0 values and collect the set of τ_0 for which the test does not reject*

$$\widehat{C}_{1-\alpha} = \{\tau : \widehat{p}_\tau \geq \alpha\}.$$

The following proposition establishes the asymptotic validity of the procedure for fixed δ_0 (see Appendix B for proof).[19]

Proposition 6. *Assume that $\sqrt{n}(\widehat{\delta} - \delta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$ where Σ is a positive definite matrix, and that $\widehat{\Sigma} \xrightarrow{p} \Sigma$ and $\bar{\Sigma} \xrightarrow{p} \Sigma$. Suppose that we wish to test either the null hypothesis $H_0 : \sqrt{\gamma'\gamma} = \tau_0$ or $H_0 : \sqrt{\gamma'\gamma}/\beta = \tau_0$, for any γ_0 and $\beta_0 \neq 0$. Then, letting $B \to \infty$ and for any $\alpha \in [0, 1]$, we have that $\Pr(\widehat{p}_{\tau_0} \leq \alpha) \to \alpha$ as $n \to \infty$.*

The test remains valid under two distinct regimes. When $\gamma_0 \neq 0$, the simulated distribution converges to a chi-squared distribution with one degree of freedom under the null and so is asymptotically equivalent to the delta method. In contrast, when $\gamma_0 = 0$ (and the delta method fails), the simulated distribution converges to the chi-squared distribution with $\dim(\gamma)$ degrees of freedom, and is equivalent to the Wald test for the null hypothesis $\gamma_0 = 0$. While the procedure is not uniformly valid over δ, we show in simulations calibrated to our empirical setting that the test has good size control across a range of parameter settings.[20] In particular, it performs well even for γ close to (or equal to) zero, where the delta method performs poorly. As another advantage, the test is invariant to parameterization of the null hypothesis, since it depends only on the restricted parameter space. In contrast, the delta method is known to be sensitive to parameterization and delivers different results for different choices, e.g. $H_0 : \sqrt{\gamma'\gamma} = \beta\tau_0$ versus $H_0 : \frac{\sqrt{\gamma'\gamma}}{\beta} = \tau_0$.

Remark 1. The method proposed here is distinct from an alternative simulation based approach that simulates $\delta^* \sim \mathcal{N}(\widehat{\delta}, \widehat{\Sigma})$, constructs the corresponding distribution for $\tau^* = g(\delta^*)$, and takes the $\alpha/2$ and $(1 - \alpha/2)$ quantiles of this distribution as a confidence set. Although straightforward, and perhaps deceptively intuitive, this approach does not deliver valid confidence sets in many settings, as highlighted by Ham and Woutersen (2013).[21] In fact, it can deliver zero coverage in some cases, for example when $g(\delta) = \delta'\delta$ and $\delta_0 = 0$ we have $g(\delta^*) > g(\delta_0) = 0$ with probability one so that the confidence set will have coverage zero. Instead, our method simulates the distribution *under the null hypothesis*, and constructs confidence sets by inverting the resulting test.

[19] For testing $H_0 : \sqrt{\gamma'\gamma}/\beta = \tau_0$ we exclude $\beta_0 = 0$ from the parameter space since the parameter of interest is not well-defined in this case.

[20] See Appendix B for a discussion on uniformly valid inference in this setting.

[21] Ham and Woutersen (2013) present a simulation based approach for confidence sets which essentially recovers the projection confidence set, which can be useful in settings in which the projection set is otherwise difficult to compute. As discussed above, this procedure is generally conservative. They also propose an adjustment based on linear approximation to the function g that can reduce conservativeness of the interval.

Remark 2. Our method for constructing critical values could be applied to alternative test statistics. For example, we might use the distance between our estimated parameter of interest and its null value $|\widehat{\tau} - \tau_0|$, rather than measuring distance in terms of the underlying parameter vector δ. We choose the distance metric statistic in order to improve the power of the test. For example, it can be the case that $T(\widehat{\delta}, \tau_0)$ is large even when $|\widehat{\tau} - \tau_0|$ is small since the distance $T(\cdot, \tau_0)$ takes into account the relative precision in which we estimate δ in different directions.

4.2 Simulation evidence

Here we provide the results of simulations that are calibrated to our empirical setting. We simulate an IV model with treatment D_{it} and outcome Y_{it} according to

$$\begin{aligned} Y_{it} &= \beta D_{it} + \gamma' D_{it} \times X_i + \sigma_Y(\rho e_{it} + \sqrt{1-\rho^2} u_{it}) \\ D_{it} &= \alpha Z_{it} + \pi' Z_{it} \times X_i + \sigma_D e_{it} \end{aligned}$$

where e_{it} and u_{it} are both independent standard normal variables. The instrument Z_{it} and covariates X_i are held fixed and taken from the empirical data – they are the grant and the first four principal components of firm baseline characteristics. All parameters in the model are set equal to their estimates using the empirical data, except where noted.

We run simulations under four settings, in which the model parameters are rescaled so that $\sqrt{\gamma'\gamma} = \{0.1, 0.01\}$ and $\beta = \{0.1, 0.03\}$. These choices are intended to represent settings in which the nonlinearities from either the quadratic form in γ or from division by β are weak or strong. Table 1 reports the mean, median, 5th percentile and 95th percentile of the point estimates, relative to their true values. The distributions of point estimates are in general positively biased and skewed – this is particularly true for the ratio parameter when β is small.

Table 1: Simulated point estimates

		$\sqrt{\widehat{\gamma}'\widehat{\gamma}} - \sqrt{\gamma_0'\gamma_0}$				$\sqrt{\widehat{\gamma}'\widehat{\gamma}}/\widehat{\beta} - \sqrt{\gamma_0'\gamma_0}/\beta_0$			
$\sqrt{\gamma_0'\gamma_0}$	β_0	Mean	Median	5%	95%	Mean	Median	5%	95%
0.1	0.1	0.01	0.01	-0.03	0.04	0.15	0.07	-0.28	0.58
0.1	0.03	0.01	0.01	-0.03	0.04	0.47	-0.17	-10.17	14.38
0.01	0.1	0.04	0.03	0.01	0.07	0.38	0.26	0.06	0.73
0.01	0.03	0.04	0.03	0.01	0.07	1.50	0.79	-3.73	6.07

Notes: This table shows the mean, median and 5th and 95th percentiles of the simulated distribution of point estimates relative to their true value. Each row corresponds to a different calibration of the model's true parameters. Results are for 1000 simulations.

Table 2 reports the rejection rates for tests under the null hypothesis for each of the four simulation settings. Three tests are compared: a Wald test for $H_0 : \frac{\sqrt{\gamma'\gamma}}{\beta} = \tau_0$; a Wald test for the null $H_0 : \sqrt{\gamma'\gamma} = \beta\tau_0$; and our proposed simulation-based test in Algorithm 1.[22] In the first row, both $\sqrt{\gamma'\gamma}$ and β are well separated from zero, so that all tests have size at or below the nominal level, although W_1 appears to under-reject. In the second row, β is close to zero so that the first Wald statistic (which divides by β) over-rejects. In the final two rows, $\sqrt{\gamma'\gamma}$ is close to zero. In this case both W_1 and W_2 have poor coverage, with rejection rates around twice the nominal level. The simulated statistic T has approximately correct size in all four cases, highlighting its robustness to settings in which the delta method fails.

Table 2: Simulated rejection rates

		10 % rejection			5 % rejection		
$\sqrt{\gamma'\gamma}$	β	W_1	W_2	T	W_1	W_2	T
0.1	0.1	0.058	0.082	0.098	0.021	0.045	0.052
0.1	0.03	0.150	0.077	0.094	0.102	0.041	0.050
0.01	0.1	0.198	0.186	0.109	0.109	0.099	0.053
0.01	0.03	0.068	0.202	0.104	0.037	0.101	0.045

Notes: This table shows simulated coverage rates for $g(\beta, \gamma) = \frac{\sqrt{\gamma'\gamma}}{\beta}$. Each row corresponds to a different calibration of the model's true parameters. The cells show the share of simulations in which the test statistic (falsely) rejected the null, for a nominal 10% test and for a nominal 5% test. The columns labeled W_1 correspond to the Wald statistic for the null hypothesis $H_0 : \frac{\sqrt{\gamma'\gamma}}{\beta} = \tau_0$. The columns labeled W_2 correspond to the Wald statistic for the null hypothesis $H_0 : \sqrt{\gamma'\gamma} = \beta\tau_0$. The columns labeled T correspond to our proposed simulation-based test.

5 Empirical Estimates of the Cost of Misallocation

Summary

In the preceding sections, we developed a methodology to measure the cost of misallocation, exploiting experiments in order to measure the variance of log MRPK. We now put those tools to work.

Our estimates suggest, for a sample of Sri Lankan microenterprises, the variance of log MRPK across firms is substantial. Our point estimates suggest a (lower bound) variance of log MRPK of roughly 93 log points. Using our novel econometric tools, we find that 90% confidence intervals rule out values below roughly 20 log points, while 95% confidence intervals rule out values below roughly 16 log points.

To feed these estimates into our aggregation formulas, we select a standard calibration for the CES parameter, $\theta = 3$, and provide two calibrations for the elasticity of output to

[22] We use $B = 1000$ draws from the simulated distribution to compute critical values

capital, $\alpha = \frac{1}{3}$ and $\alpha = 1$. We validate these values using parametric estimates of firms' returns to scale. The first calibration corresponds to a standard value for the capital share, and is useful for a thought experiment in which capital can be reallocated but other inputs are fixed. The second calibration corresponds to a constant-returns-to-scale production function, and is useful for a thought experiment in which all inputs can be reallocated. Focusing on the point estimates, we find that optimally reallocating capital only would increase output by 22%, while optimally reallocating all inputs would increase output by 301%. These estimates are sizable, implying that misallocation plays an important role in determining aggregate productivity, and that input markets are meaningfully inefficient in this setting.

5.1 Estimates of Heterogeneous MRPK

We begin by estimating heterogeneous MRPK across different firms. For our vector of baseline covariates, X_i, we use seven variables, all measured in the baseline: capital, profit, business age, owner's education, owner's hours worked, average product of capital, and the log of the average product of capital. Throughout, we use standard errors and confidence intervals that cluster at the firm level.

We begin by estimating the homogeneous linear IV model in Equation 27, replicating the main results in de Mel et al. (2008). This provides us with a homogeneous estimate of the MRPK for all firms, which under appropriate assumptions will be the average MRPK.[23] The results are in Table 3, in the first row of Panel A. The homogeneous linear IV model yields an average *monthly* return to capital of 6%.

In the rest of Table 3, we estimate the heterogeneous linear IV model in Equation 28. Each column uses a single covariate for X_i. In Panel A, we compute $\mathbb{E}[\text{MRPK}_i]$ as $\hat{\beta}+\hat{\gamma}\mathbb{E}[X_i]$. Our estimates range from 6-8% monthly returns across specifications.

In Panel B, we compute the standard deviation of expected MRPK, or $\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])$. Note that in the single covariate setting, this is equal (up to sign) to the standardized coefficient $\gamma \cdot \text{SD}(X_i)$, which represents how a one standard deviation change in the covariate affects the MRPK. We thus denote the sign of the interaction effect by including a plus or minus sign before the estimate. We provide 90% confidence intervals for Panels B and C, computed using Algorithm 1 from Section 4.

The strongest predictor of MRPK is the average product of capital at baseline. This is somewhat expected: under a homogeneous Cobb-Douglas production function, the MRPK is proportional to APK. However, the fact that the APK in wave 1 is a useful predictor of the

[23] In general, this IV model will identify a local average treatment effect, which may differ from the average treatment effect to the extent that the first stage (the effect of the grant on capital) covaries with the firm's MRPK. de Mel et al. (2008) argue that, in this setting, the LATE and ATE are likely to be similar.

Table 3: Estimates of Heterogeneous MRPK by Baseline Covariates

Panel A, without Covariates: $\mathbb{E}[\text{MRPK}_i]$							
Estimate	0.061	SE	(0.024)				
Covariate	Capital	Age	Education	Profit	Hours	APK	log(APK)
Panel A, with Covariates: $\mathbb{E}[\text{MRPK}_i]$							
Estimate	0.062	0.061	0.060	0.063	0.072	0.085	0.069
SE	(0.025)	(0.024)	(0.027)	(0.025)	(0.030)	(0.029)	(0.027)
Panel B: $SD\big(\mathbb{E}[\text{MRPK}_i \vert X_i]\big)$*, With Sign of Interaction Effect*							
Estimate	-0.070	+0.018	+0.044	-0.011	-0.023	+0.128	+0.052
90% CI	$[0.03, 0.64]$	$[0.00, 0.83]$	$[0.02, \infty]$	$[0.00, 0.13]$	$[0.00, 1.06]$	$[0.06, 0.22]$	$[0.02, 0.11]$
Panel C: $SD\big(\mathbb{E}[\text{MRPK}_i \vert X_i]\big)/\mathbb{E}[\text{MRPK}_i]$							
Estimate	1.121	0.300	0.723	0.171	0.314	1.505	0.747
90% CI	$[0.41, \infty]$	$[0.00, 0.90]$	$[0.30, 1.76]$	$[0.00, 3.49]$	$[0.00, 0.63]$	$[0.84, 2.48]$	$[0.38, 1.35]$

Notes: This table shows estimates of heterogeneous models of MRPK. All standard errors and confidence intervals are clustered at the firm level. The first row in Panel A shows estimates from Equation 27; a homogeneous model without covariates. The rest of the table show estimates from the heterogeneous model described in Equation 28; each column uses one covariate, which is measured at baseline. The second part of Panel A shows the $\mathbb{E}[\text{MRPK}_i]$ implied by these heterogeneous models, which is computed as $\hat{\beta} + \hat{\gamma}\mathbb{E}[X_i]$. Panel B shows estimates of $\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])$, as well as 90% confidence intervals computed using Algorithm 1. The sign of the interaction term is indicated by a plus or minus sign in front of the estimate; however, the confidence interval is for the unsigned standard deviation. Panel C shows the implied estimate of $\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])/\mathbb{E}[\text{MRPK}_i]$, as well as 90% confidence intervals computed using Algorithm 1. Where the confidence intervals have an upper bound of infinity, this indicates that the largest null tested (2 for Panel B and 5 for Panel C) could not be rejected.

Table 4: Estimates of Variance of MRPK

	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$
Panel A: $\mathbb{E}[\text{MRPK}_i] = \beta$							
Estimate	0.073	0.075	0.077	0.084	0.080	0.105	0.105
SE	(0.026)	(0.025)	(0.026)	(0.027)	(0.051)	(0.054)	(0.115)
Panel B: $SD\big(\mathbb{E}[\text{MRPK}_i \mid X_i]\big) = \sqrt{\gamma'\gamma}$							
Estimate	0.066	0.063	0.109	0.107	0.098	0.131	0.128
90% CI	$[0.03, 0.11]$	$[0.02, 0.11]$	$[0.04, 0.70]$	$[0.04, 0.53]$	$[0.04, \infty]$	$[0.08, \infty]$	$[0.05, \infty]$
Panel C: $SD\big(\mathbb{E}[\text{MRPK}_i \mid X_i]\big)/\mathbb{E}[\text{MRPK}_i] = \sqrt{\gamma'\gamma}/\beta$							
Estimate	0.913	0.840	1.415	1.275	1.234	1.247	1.213
90% CI	$[0.46, 1.80]$	$[0.21, 1.82]$	$[0.56, \infty]$	$[0.52, 3.97]$	$[0.47, \infty]$	$[0.78, \infty]$	$[0.56, \infty]$

Notes: This table shows estimates of heterogeneous models of MRPK. All standard errors and confidence intervals are clustered at the firm level. In Panel A, each column shows estimates from the heterogeneous model described in Equation 28. Each column uses the first K principal components of our vector of covariates. Panel A shows estimates of $\mathbb{E}[\text{MRPK}_i] = \beta$, along with standard errors. Panel B shows estimates of $\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i]) = \sqrt{\gamma'\gamma}$, as well as 90% confidence intervals computed using Algorithm 1. Panel C shows the implied estimate of $\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])/\mathbb{E}[\text{MRPK}_i] = \sqrt{\gamma'\gamma}/\beta$, as well as 90% confidence intervals computed using Algorithm 1. Where the confidence intervals have an upper bound of infinity, this indicates that the largest null tested (2 for Panel B and 5 for Panel C) could not be rejected.

MRPK in later waves also suggests that some component of the firm's MRPK is persistent over time.

In Panel C, we compute the implied estimates of $\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])/\mathbb{E}[\text{MRPK}_i]$. Note that these are not our main estimates: in the next subsection, we will use multiple covariates to predict MRPK, and combine them using principal components. Although the estimates based on APK are informative, many of these single-covariate confidence intervals cannot rule out zero misallocation. This highlights the importance of selecting the correct covariates, and/or incorporating multiple covariates in order to get a more precise estimate of misallocation, as we do next.

5.2 Estimates of Var $(\log \text{MRPK}_i)$

We now implement our main methodology for estimating the variance of log MRPK. We estimate Equation 28 using standardized principal components as our covariates X_i. Our results are in Table 4. Each column corresponds to our estimates using a different number of factors K for the X_i (e.g. the $K = 4$ row uses the first four standardized principal components of the baseline covariates). Each panel follows the same structure as the previous table.

Panel A shows estimates of $\mathbb{E}[\text{MRPK}_i]$: these are similar to previous estimates, with average monthly returns ranging from 7-10%. Panel B shows estimates of $\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])$,

or $\sqrt{\gamma'\gamma}$. The point estimates range from a standard deviation of 6% to a standard deviation of 13%.

Our main focus is Panel C, where we provide estimates of $\text{SD}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)/\mathbb{E}\left[\text{MRPK}_i\right]$, which is computed as $\frac{\sqrt{\gamma'\gamma}}{\beta}$. In Section 5.5, we will use these estimates to measure the cost of misallocation. The point estimates are fairly high, and for $K > 2$ they are all above one. This implies very sizable dispersion: according to these estimates, a firm that is one standard deviation below the mean has negative expected returns.

For Panels B and C, we provide 90% confidence intervals based on Algorithm 1. The lower bound of the 90% confidence interval is fairly high: in Panel C, it is roughly 0.5 for most values of K. This is of course lower than the point estimates, but still sizable. A firm that is two standard deviations below the mean would have near zero returns under these estimates.

We provide additional estimates and confidence intervals in the Appendix. In Appendix Table 7, we provide 95% confidence intervals, again based on Algorithm 1. These intervals are wider, but still imply substantial dispersion in returns, except for $K = 2$. We also compute uniformly valid confidence intervals using Algorithm 2. We show these intervals in Appendix Table 8. Although these intervals are modestly wider, they still rule out low values of $\text{SD}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)$ and $\frac{\text{SD}(\mathbb{E}[\text{MRPK}_i|X_i])}{\mathbb{E}[\text{MRPK}_i]}$, despite the interval for the latter being somewhat conservative due to projection. The uniformly valid 90% confidence intervals for $K = 5$, which are representative of the rest of the estimates, suggest that the standard deviation of monthly returns is at least 3.1%, and the ratio of the standard deviation over the mean is at least 0.387.

We also provide estimates of the weighted variance, using each firm's baseline profits as weights. To implement this, we construct our factors using weighted PCA, and standardize them based on the weighted mean and weighted variance. This provides us with a set of covariates whose weighted mean is zero, weighted variance is one, and whose weighted covariance with each other is zero. We then follow the same procedure as before, but use these weighted covariates to compute estimates and confidence intervals of β, $\sqrt{\gamma'\gamma}$, and $\sqrt{\gamma'\gamma}/\beta$. Note that although we use the weights to construct the covariates, we still run an unweighted IV regression. Because the weights are quite skewed, a weighted regression would be very noisily estimated. Moreover, note that the baseline profit weights are not quite the same weights as in Proposition 2: that result called for the sales-times-elasticity weights. In this setting it is not feasible for us to estimate firm-specific elasticities, and so we cannot use them as weights.

The results are in Appendix Table 9. The weighted variance estimates and confidence intervals are broadly similar to our main results for the unweighted variance, although they

are somewhat noisier for certain values of K. For all values of K, we are able to rule out low dispersion. Even for the most unfavorable confidence intervals ($K = 2$), we can rule out values of $\sqrt{\gamma'\gamma}/\beta$ below 0.35 with 90% confidence.

5.3 Comparison to Other Approaches

In this subsection, we compare our estimates to those from two othe approaches: the "standard approach" using a homogeneous Cobb-Douglas production function as in Hsieh and Klenow (2009), and the IV-CRC approach of Carrillo et al. (2023).

Comparison to Standard Approach. We first compare our results to the standard approach. To do this, we compute the MRPK under the assumption of CES demand and Cobb-Douglas production, as in Hsieh and Klenow (2009). Under these assumptions, we can observe MRPK directly from the average product of capital, through the formula $\text{MRPK}_i = \alpha\frac{\theta-1}{\theta}\text{APK}_i$. We use standard values of α and θ: we calibrate $\alpha = \frac{1}{3}$ to match the capital share, and we use $\theta = 3$, following Hsieh and Klenow (2009).[24] Throughout, we exclude MRPK data from the first wave, in order to make them more comparable to our IV estimates. We do this to make our estimates correspond more closely to the MRPK identified by the grant instrument: in the first wave there is no variation in the grant, and therefore our IV results were identified only off of later waves.

We begin by computing statistics for the unconditional distribution of MRPK, based on this Cobb-Douglas calibration. The results are in Panel A of Appendix Table 10. The panel shows $\mathbb{E}[\text{MRPK}_i]$, $\text{SD}(\text{MRPK}_i)/\mathbb{E}[\text{MRPK}_i]$, and $\text{SD}(\log \text{MRPK}_i)$, which we compute using their sample counterparts. The Cobb-Douglas calibration yields a mean monthly return of 7.8%, which is similar to our IV estimates. For our unconditional estimates of $\text{SD}(\text{MRPK}_i)$ and $\text{SD}(\log \text{MRPK}_i)$, we first partial out wave fixed effects, reflecting the idea that we are interested in misallocation across firms within the same time period, rather than varying returns over time. However, this has a trivial effect on our estimates.[25][26] The unconditional dispersion is extremely large: the standard deviation is roughly twice the mean. However, note that the unconditional distribution of returns mixes both ex ante (misallocation) and ex post (risk) differences in returns, and thus should be viewed as an upper bound on misallocation.

[24] Three is typically considered a low value of θ, and was used by Hsieh and Klenow (2009) because it gave a conservative estimate of misallocation. The exercise we conduct in this subsection is about measuring MRPK rather than misallocation per se, and so is less sensitive to the value of θ. A calibration of $\theta = 3$ yields a scaling factor of $\frac{\theta-1}{\theta} = \frac{2}{3}$, while a calibration where $\theta \to \infty$ has a scaling factor of one.

[25] The standard deviation of MRPK is 16.18% if we control for wave fixed effects, and 16.24% if we do not.

[26] Note also that the decision to exclude the first wave has little effect on these unconditional results: including the first wave yields a mean return of 8.2% and a standard deviation of 16.8%.

We then project MRPK onto covariates, so that we can compute $\text{SD}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)$. For each covariate, we estimate a regression analogous to our IV analysis:

$$\text{MRPK}_{it} = \gamma' X_i + \delta_t + \varepsilon_{it} \tag{39}$$

where δ_t is a wave fixed effect. As before, we exclude data from the first wave to maintain comparability to our IV results. Excluding the first wave also ensures that our results are not just mechanical. For example, it must be the case that baseline APK is highly predictive of MRPK in the first wave, since MRPK was computed as proportional to APK. However, the fact that baseline APK predicts future MRPK reflects that there is a persistent component to these variables.

Using our estimates from Equation 39, we compute $\text{SD}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)$, as well as $\text{SD}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)/\mathbb{E}\left[\text{MRPK}_i\right]$.[27] We show results for individual covariates in Panels B and C of Appendix Table 10. As before, we indicate the sign of the interaction effect in Panel B, since the point estimate is also the effect of a one standard deviation change in the covariate on the MRPK. In Panels D and E, we show results using standardized principal components as our covariates. For Panels B through E, we provide 90% confidence intervals computed using Algorithm 1.

These estimates highlight that the unconditional variance of returns is substantially larger than the predictable component of that variance. The unconditional estimate of the ratio $\text{SD}\left(\text{MRPK}_i\right)/\mathbb{E}\left[\text{MRPK}_i\right]$ is 2.082. In contrast, the estimates in Panel E suggest that $\text{SD}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)/\mathbb{E}\left[\text{MRPK}_i\right]$ is less than half as large, at roughly 0.8. Our confidence intervals here are tight, suggesting these differences are not driven by sampling error.

Unsurprisingly, our estimates based on a Cobb-Douglas production function are more precise than those based on an IV regression. Under the strong assumption of a homogeneous Cobb-Douglas production, MRPK can be observed directly rather than estimated, yielding smaller standard errors.

However, in part due to significant uncertainty in our IV estimates, we cannot reject equality between any cell in Table 10 and the corresponding cell in Tables 3 and 4, with the exception of the cell projecting MRPK onto baseline APK (the Cobb-Douglas point estimate however is just outside the 90% confidence interval). At the same time, we also cannot rule out large differences. For example, the IV regression without covariates yields a point estimate of 6.1% monthly returns. The 95% confidence interval from this regression includes the 7.8% average monthly return implied by our Cobb-Douglas calibration, but it

[27] For this calculation, we compute $\mathbb{E}\left[\text{MRPK}_i\right]$ using the subset of firms for which the covariates are non-missing. Due to a few non-missing firms, this is slightly lower than the $\mathbb{E}\left[\text{MRPK}_i\right]$ in Panel A (7.9% rather than 8.2%).

also includes values as high as 10.8% and as low as 1.4%.

Comparison to Carrillo et al. (2023) Approach. In Appendix Table 12, we implement the method of Carrillo et al. (2023) for comparison. We implement three versions of their estimator: one controlling for the expected size of the grant in a given wave, a second controlling for both the first and second moment of the grant size in that wave, and one controlling for wave fixed effects. The three estimates produce similar results, although the precise point estimate is somewhat sensitive to the controls.

In principle, Carrillo et al. (2023) target the total variance rather than the predictable component of the variance, and so their estimate provides an upper bound on misallocation rather than a lower bound. In practice however, their method produces confidence intervals in this setting that are too wide to be informative. The point estimate for the variance of MRPK is 0.224 (controlling for the expected first and second moment of grant size), implying a standard deviation of monthly returns of 47%. Yet the clustered bootstrap standard error is even larger than the point estimate, and thus the confidence interval also includes zero misallocation. At least in this setting, our approach provides a substantial improvement in statistical precision.

5.4 Selecting $\mathcal{E}$

Our formula for misallocation tells us that we can measure the gains from optimally reallocating inputs with the formula $\log Y^* - \log Y \approx \frac{1}{2}\mathcal{E} \cdot \text{Var}(\log \text{MRPK}_i)$. The elasticity $\mathcal{E}$, depends on the elasticity of substitution across varieties, θ, and on the elasticity of MPK with respect to output, ϕ. There are many estimates of θ in the literature. Following the calibration in Hsieh and Klenow (2009), we will focus on $\theta = 3$; this is a relatively conservative calibration, and larger values of θ would imply larger levels of misallocation.

The elasticity ϕ measures the curvature of the firm's production function and can, in principle, also be measured using the same experimental variation in capital that we exploit. In practice however, the experiment we study is not well-powered to pick up non-linearity in the returns to capital.[28] To make progress, we base our calibration of ϕ on standard values for a Cobb-Douglas production function, which we also show are consistent with estimates from the data. In particular, if we assume a Cobb-Douglas production function $f(k) = z \cdot k^{\alpha}$, then the elasticity ϕ is given by $\phi = -\frac{1-\alpha}{\alpha}$. For a standard capital share calibration of $\alpha = \frac{1}{3}$, this yields $\phi = -2$. We will also explore an alternative constant-returns-to-scale calibration of $\phi = 0$: we use this calibration for our thought experiment in which all inputs can be

[28] This fact is also noted by de Mel et al. (2008) in their original paper. When we include a quadratic term in our specification, we cannot reject zero.

reallocated.

We can also estimate returns to scale parametrically in the data, exploiting the randomly assigned grant. Using the amount of the grant as an instrument, we run the regression:

$$\log \text{Profit}_{it} = \beta \cdot \log k_{it} + \alpha_i + \delta_t + \varepsilon_{it}$$

where profit and capital are defined as before. We can use the regression coefficient β to identify the returns to scale parameter α in the Cobb-Douglas production function.[29] In particular, we have $\beta = \frac{d \log py}{d \log k} = \left(1 + \frac{d \log p}{d \log y}\right) \cdot \frac{d \log y}{d \log k} = \left(\frac{\theta - 1}{\theta}\right) \cdot \alpha$. In Appendix Table 13, we show the results of this regression. We also show the results of a similar regression replacing $\log k_{it}$ with $\left(\frac{1}{3} \log k_{it} + \frac{2}{3} \log l_{it}\right)$, measuring l_{it} as total hours worked, which we use to measure returns to scale for our calibration in which all inputs can be reallocated.

The results of these estimates suggest our calibration is reasonable, and if anything somewhat conservative. Our point estimates of β are 0.38 (in the only-capital case) and 1.05 (in the capital-and-labor case), implying values of α of 0.57 and 1.58 under our assumed value of θ. The confidence intervals contain our preferred values, and the point estimates are in fact somewhat larger than the our calibrated values. We also run sales-weighted versions of these regressions; the confidence intervals for the weighted versions are wider, but again contain our preferred values.[30]

Using the formula $\mathcal{E} = \left(-\phi + \frac{1}{\theta}\right)^{-1}$, our main calibration yields $\mathcal{E} = \frac{3}{7}$, while our "reallocate all inputs" calibration yields $\mathcal{E} = \theta = 3$. The difference in these elasticities reflects the difference between the two thought experiments. The gains from optimally reallocating all inputs will in general be much larger than the gains from reallocating only capital. Reallocating only capital quickly runs into diminishing returns on the production side, while the benefits from reallocating all inputs (under constant returns to scale) are held back only by downward sloping demand.

We have aimed to be somewhat conservative in calibrating $\mathcal{E}$, e.g. our value of θ is at the low end of the estimates in the literature, and our values for returns to scale are very standard. A feature of our analysis, however, is that our formula for misallocation is a simple and transparent function of $\mathcal{E}$ and $\text{Var}\left(\log \text{MRPK}_i\right)$. Thus, a reader can easily substitute her own preferred values of θ, ϕ, and/or α, and solve for $\mathcal{E}$ by hand. This can easily be coupled with our estimates of $\text{Var}\left(\log \text{MRPK}_i\right)$ to compute the cost of misallocation under the reader's preferred value of $\mathcal{E}$.

[29] We are assuming here that the production function takes a value-added form, which allows us to treat profit (which in this context is equivalent to value added) as output.

[30] Our weighted regressions use firm's baseline profits as the weights.

Table 5: Estimated Cost of Misallocation ($K = 5$)

	Point Estimate	90% CI	95% CI	Cobb-Douglas
$\frac{\sqrt{\gamma'\gamma}}{\beta}$	1.234	0.470	0.419	—
$\text{Var}(\log \text{MRPK}_i)$	0.93	0.20	0.16	1.30
$\log Z^* - \log Z$ $(\mathcal{E} = \frac{3}{7})$	0.20	0.04	0.03	0.28
$Z^*/Z - 1$ $(\mathcal{E} = \frac{3}{7})$	0.22	0.04	0.04	0.32
$\log Z^* - \log Z$ $(\mathcal{E} = 3)$	1.39	0.30	0.24	1.95
$Z^*/Z - 1$ $(\mathcal{E} = 3)$	3.01	0.35	0.27	6.03

Notes: This table summarizes point estimates and confidence interval lower bounds for $\frac{\sqrt{\gamma'\gamma}}{\beta}$, the variance of log MRPK, and the implied cost of misallocation. We focus on results for $K = 5$; results for other values of K are similar. The first row summarizes results for $\frac{\sqrt{\gamma'\gamma}}{\beta}$, replicating results in Panel A of Table 4. The second row provides a (lower bound) estimate of $\text{Var}(\log \text{MRPK}_i)$, using the formula $\log\left(1 + \frac{\gamma'\gamma}{\beta^2}\right)$. The remaining rows provide estimates of the cost of misallocation in log points, using the formula $\log Z^* - \log Z = \frac{1}{2}\mathcal{E} \cdot \text{Var}(\log \text{MRPK}_i)$. The third and fourth rows are calibrated to reflect the gains from optimally reallocating capital, while the fifth and sixth rows are calibrated to reflect the gains from optimally reallocating all inputs, assuming a constant-returns-to-scale production function. For comparison, we also show estimates based on the variance of log MRPK, computed under the assumption of a Cobb-Douglas production function. As in Table 4, our estimate of the variance of log MRPK is computed after partialling out wave fixed effects, in order to focus on within-period misallocation, and uses only data after the first wave, to correspond with the returns captured by the instrument.

5.5 Implied Estimates of the Cost of Misallocation

Finally, we use our estimates of the variance of log MRPK, along with our calibration of $\mathcal{E}$, to back out estimates of the cost of misallocation. We summarize the results in Table 5. Since we generated a range of estimates for $\frac{\sqrt{\gamma'\gamma}}{\beta}$, based on different numbers of factors, we focus on results for $K = 5$, which is fairly representative of the broader set of estimates. We then use the formula $\log\left(1 + \frac{\gamma'\gamma}{\beta^2}\right)$, to provide a lower bound estimate of $\text{Var}(\log \text{MRPK}_i)$, as discussed in Section 3. This gives a point estimate of 93 log points, while the lower bound of the 90% confidence interval is 20 log points, and the lower bound of the 95% confidence interval is 16 log points.

Focusing on the point estimates, we find that optimally reallocating capital would increase output by 20 log points, or 22%. Under the stronger assumptions of Proposition 4, we find that optimally reallocating all inputs would increase output by 139 log points, or 301%. Our confidence intervals rule out low values for the gains from reallocating all inputs, although combining the lower bound of the confidence interval with a low elasticity of $\mathcal{E} = \frac{3}{7}$ does yield modest estimates. Overall, we interpret these estimates as suggesting sizable losses from misallocation of inputs, at least for our sample of Sri Lankan microenterprises.

Our point estimates are large, but not as large as the misallocation implied by the Cobb-Douglas benchmark. Under the assumption of a homogeneous Cobb-Douglas production

function, the variance of log MRPK is 130 log points.[31] This implies that optimally reallocating capital would increase output by 32%, and optimally reallocating all inputs would increase output by 603%.

Our results do not imply a level of misallocation this large, but they also do not necessarily rule it out. The confidence interval on our estimates include high values of misallocation. Moreover, we focus on the predictable component of the variance of log MRPK, thus our estimates are a lower bound on the total variance. This is somewhat beneficial: we would not want to label unpredictable variation in MRPK as "misallocation." However, some variation in returns may be predictable ex ante, but not captured by our seven covariates. When we instead compute the predictable component of the variance of MRPK under the Cobb-Douglas assumption, the estimates are not far from our IV estimates, and in fact are somewhat smaller (see Panels D and E of Appendix Table 10).

Our results thus do not provide decisive evidence on whether a homogeneous Cobb-Douglas production function fits the data well. Instead, our results provide robust evidence for sizable misallocation in this setting, that does not depend on strong auxiliary assumptions about the production function.

6 Conclusion

The misallocation of inputs across firms has been an important area of study in macroeconomics and development. Although some prior work has found large potential gains from reallocating inputs, the literature has typically relied on strong assumptions about the functional form of production, and other papers have suggested that estimates of misallocation are sensitive to these assumptions. Understanding the extent to which misallocation of inputs lowers aggregate productivity may be crucial for understanding large cross-country differences in output per capita; moreover, the degree to which inputs are misallocated is fundamental to our understanding of whether markets are efficient in practice.

In this paper, we show how to use experiments to measure misallocation in a credible way. We show that misallocation can be expressed as a function of the variance of log marginal products. We then show how to use data from a randomized controlled trial, which randomized grants to microenterprises, to measure an ex-ante-predictable component of the variance of log MRPK as a function of the parameters of a heterogeneous linear IV model. We develop new econometric tools to construct uniformly valid confidence intervals for this function of parameters. Finally, we apply the tools we develop to estimate the

[31] As in Appendix Table 10, these estimates are for the data after the first wave, and partial out wave fixed effects.

cost of misallocation for a sample of Sri Lankan microenterprises. We find that optimally reallocating capital would raise output by 22%, while optimally reallocating all inputs would raise output by 301%.

Our results highlight the potentially important role played by misallocation in holding back aggregate productivity. However, our estimates focus on misallocation of inputs among a sample of microenterprises in Sri Lanka. It is not obvious how these estimates compare to those for other countries and sectors. Moreover, our design does not capture misallocation between microenterprises and other firms. If the average MRPK is different for other firms than it is for microenterprises, this would imply further misallocation. The methodology we develop can be flexibly applied in other settings where there is exogenous variation in inputs: future work can use the techniques we develop to deepen our understanding of misallocation across a range of settings.

References

Andrews, I. and A. Mikusheva (2016): "A Geometric Approach to Nonlinear Econometric Models," *Econometrica*, 84, 1249–1264.

Asker, J., A. Collard-Wexler, and J. D. Loecker (2014): "Dynamic Inputs and Resource (Mis)Allocation," *Journal of Political Economy*, 122, 1013–1063.

Banerjee, A. V. and E. Duflo (2005): "Chapter 7 Growth Theory through the Lens of Development Economics," in *Handbook of Economic Growth*, Elsevier, 473–552.

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

Beaman, L., D. Karlan, B. Thuysbaert, and C. Udry (2023): "Selection Into Credit Markets: Evidence From Agriculture in Mali," *Econometrica*, 91, 1595–1627.

Bigio, S. and J. La'O (2020): "Distortions in Production Networks," *The Quarterly Journal of Economics*, 135, 2187–2253.

Bils, M., P. J. Klenow, and C. Ruane (2021): "Misallocation or Mismeasurement?" *Journal of Monetary Economics*, 124, S39–S56.

Carrillo, P., D. Donaldson, D. Pomeranz, and M. Singhal (2023): "Misallocation in Firm Production: A Nonparametric Analysis Using Procurement Lotteries," Tech. rep.

Crépon, B., M. E. Komi, and A. Osman (2023): "Is It Who You Are or What You Get? Comparing the Impacts of Loans and Grants for Microenterprise Development," *American Economic Journal: Applied Economics*.

David, J. M., L. Schmid, and D. Zeke (2022): "Risk-adjusted capital allocation and misallocation," *Journal of Financial Economics*, 145, 684–705.

David, J. M. and V. Venkateswaran (2019): "The Sources of Capital Misallocation," *American Economic Review*, 109, 2531–2567.

Dávila, E. and A. Schaab (2023): "Welfare Accounting," Tech. rep.

de Mel, S., D. McKenzie, and C. Woodruff (2008): "Returns to Capital in Microenterprises: Evidence from a Field Experiment," *Quarterly Journal of Economics*, 123, 1329–1372.

Dhingra, S. and J. Morrow (2019): "Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity," *Journal of Political Economy*, 127, 196–232.

Fafchamps, M., D. McKenzie, S. Quinn, and C. Woodruff (2014): "Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in Ghana," *Journal of Development Economics*, 106, 211–226.

Gollin, D. and C. Udry (2021): "Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture," *Journal of Political Economy*, 129, 1–80.

Haltiwanger, J., R. Kulick, and C. Syverson (2018): "Misallocation Measures: The Distortion That Ate the Residual," Tech. rep.

Ham, J. C. and T. Woutersen (2013): "Calculating Confidence Intervals for Continuous and Discontinuous Functions of Estimated Parameters," Tech. rep.

Harberger, A. C. (1954): "Monopoly and Resource Allocation," *The American Economic Review*, 44, 77–87.

Hines, J. R. (1999): "Three Sides of Harberger Triangles," *Journal of Economic Perspectives*, 13, 167–188.

Hopenhayn, H. A. (2014): "Firms, Misallocation, and Aggregate Productivity: A Review," *Annual Review of Economics*, 6, 735–770.

Hsieh, C.-T. and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124, 1403–1448.

Hussam, R., N. Rigol, and B. N. Roth (2022): "Targeting High Ability Entrepreneurs Using Community Information: Mechanism Design in the Field," *American Economic Review*, 112, 861–898.

Jones, C. (2016): "The Facts of Economic Growth," in *Handbook of Macroeconomics*, Elsevier, 3–69.

Liu, E. (2019): "Industrial Policies in Production Networks," *The Quarterly Journal of Economics*, 134, 1883–1948.

Masten, M. A. and A. Torgovitsky (2016): "Identification of Instrumental Variable Correlated Random Coefficients Models," *Review of Economics and Statistics*, 98, 1001–1005.

McKenzie, D. (2017): "Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition," *American Economic Review*, 107, 2278–2307.

Restuccia, D. and R. Rogerson (2008): "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic Dynamics*, 11, 707–720.

——— (2017): "The Causes and Costs of Misallocation," *Journal of Economic Perspectives*, 31, 151–174.

Rotemberg, M. and K. White (2021): "Plant-to-Table(s and Figures): Processed Manufacturing Data and Measured Misallocation," Tech. rep.

A Proofs for Section 2

A.1 Proof of Lemma 1

Proof. We solve for the firm's behavior using the firm's FOC, the firm production function, and the demand curve faced by the firm. We begin by log-differentiating the firm FOC from Equation 7. This yields

$$d\log p_i = d\log \mu_i + d\log r - d\log \underbrace{f_i'(k_i)}_{\text{MPK}} \tag{40}$$

To obtain an expression for MPK, we next log-differentiate the production function, twice.

$$d\log y_i = (f_i') \cdot \frac{k_i}{y_i} d\log k_i \tag{41}$$

$$\implies \frac{y_i}{f_i'} \cdot d\log y_i = k_i d\log k_i \tag{42}$$

$$d\log f_i' = \frac{d\log f_i'}{d\log k_i} d\log k_i \tag{43}$$

$$= f_i'' \cdot \frac{k_i}{f_i'} d\log k_i \tag{44}$$

$$= \phi_i d\log y_i \tag{45}$$

where $\phi_i := \frac{y_i \cdot f_i''}{(f_i')^2}$ is the elasticity of MPK with respect to output. Plugging back into the firm FOC yields:

$$d\log p_i = d\log \mu_i + d\log r - \phi_i d\log y_i \tag{46}$$

We then plug in the demand curve from Equation 9, combining the firm-level demand and firm-level supply curves:

$$\underbrace{\frac{1}{\theta} d\log Y - \frac{1}{\theta} d\log y_i}_{\text{Firm-Level Demand}} = \underbrace{d\log \mu_i + d\log r - \phi_i d\log y_i}_{\text{Firm-Level Supply}} \tag{47}$$

This yields:

$$d\log y_i = -\underbrace{\mathcal{E}_i d\log \mu_i}_{\text{Wedge}} - \underbrace{\mathcal{E}_i d\log r}_{\text{Input Cost}} + \underbrace{\frac{\mathcal{E}_i}{\theta} d\log Y}_{\text{Demand}} \tag{48}$$

where $\mathcal{E}_i := \left(-\phi_i + \frac{1}{\theta_i}\right)^{-1}$ is the negative elasticity of output with respect to the wedge. □

A.2 Proof of Proposition 1

Proof. With Lemma 1 describing firm behavior, we close the system of equations using input market clearing and the aggregator. First, input market clearing with a fixed supply of capital requires

$$\mathbb{E}\left[k_i d\log k_i\right] = 0 \tag{49}$$

Using the firm's production function, and then the firm's FOC (to substitute $f_i' = \frac{r_i\mu_i}{p_i}$), we have:

$$\begin{aligned} k_i d\log k_i &= \frac{y_i}{f_i'} d\log y_i && (50)\\ &= \frac{p_i y_i}{r_i \mu_i} d\log y_i && (51) \end{aligned}$$

Substituting into our original expression, and multiplying both sides by $\frac{r}{\mathbb{E}[p_i y_i]}$, this yields:

$$\mathbb{E}\left[\frac{\lambda_i}{\mu_i} d\log y_i\right] = 0 \tag{52}$$

where λ_i is the sales share of firm i.

Next, we use our constant-returns-to-scale aggregator to get an expression for $d\log Y$. Normalizing $P = 1$, we have that $p_i = \frac{dY}{dy_i}$. Then, using Euler's homogeneous function theorem, we have

$$\mathbb{E}\left[p_i y_i\right] = \mathbb{E}\left[\frac{dY}{dy_i} y_i\right] = Y \tag{53}$$

We can then log-differentiate the aggregator, and then plug this in, which gives us

$$\begin{aligned} d\log Y &= \mathbb{E}\left[\frac{dY}{dy_i} \cdot \frac{y_i}{Y} d\log y_i\right] && (54)\\ &= \mathbb{E}\left[\frac{p_i y_i}{Y} d\log y_i\right] && (55)\\ &= \mathbb{E}\left[\frac{p_i y_i}{\mathbb{E}\left[p_i y_i\right]} d\log y_i\right] && (56)\\ \implies d\log Y &= \mathbb{E}\left[\lambda_i d\log y_i\right] && (57) \end{aligned}$$

Finally, we can combine input market clearing (Equation 52) and aggregation (Equation 57), along with firm behavior from Lemma 1, in a way that r falls out. Take Equation 57

and subtract off C times Equation 52, where C is some constant. We have:

$$\mathbb{E}\left[\lambda_i d\log y_i\right] - C\cdot\mathbb{E}\left[\frac{\lambda_i}{\mu_i}d\log y_i\right] = d\log Y \tag{58}$$

$$\mathbb{E}\left[\left(\lambda_i - C\cdot\frac{\lambda_i}{\mu_i}\right)\cdot\left(-\mathcal{E}_i d\log\mu_i - \mathcal{E}_i d\log r + \frac{\mathcal{E}_i}{\theta}d\log Y\right)\right] = d\log Y \tag{59}$$

To ensure that the interest rate falls out, we must select a C such that $\mathbb{E}\left[\left(\lambda_i - C\cdot\frac{\lambda_i}{\mu_i}\right)\cdot\mathcal{E}_i\right] = 0$. To do this, we select $C = \frac{\mathbb{E}[\lambda_i\mathcal{E}_i]}{\mathbb{E}\left[\lambda_i\mathcal{E}_i\mu_i^{-1}\right]}$. Note also that this C is the weighted harmonic average of the wedges, which we will denote $\tilde{\mu} := \frac{\mathbb{E}[\lambda_i\mathcal{E}_i]}{\mathbb{E}\left[\lambda_i\mathcal{E}_i\mu_i^{-1}\right]}$. Let $\hat{\mu}_i := \frac{\mu_i - \tilde{\mu}}{\mu_i}$. We then have:

$$d\log Y = \mathbb{E}\left[\left(\lambda_i - \tilde{\mu}\cdot\frac{\lambda_i}{\mu_i}\right)\cdot\left(-\mathcal{E}_i d\log\mu_i - \mathcal{E}_i d\log r + \frac{\mathcal{E}_i}{\theta}d\log Y\right)\right] \tag{60}$$

$$= \mathbb{E}\left[\lambda_i\hat{\mu}_i\cdot\left(\frac{\mathcal{E}_i}{\theta}d\log Y - \mathcal{E}_i d\log\mu_i\right)\right] \tag{61}$$

$$\implies \left(1 - \mathbb{E}\left[\frac{\mathcal{E}_i\lambda_i\hat{\mu}_i}{\theta}\right]\right)d\log Y = \mathbb{E}\left[-\mathcal{E}_i\lambda_i\hat{\mu}_i d\log\mu_i\right] \tag{62}$$

Finally, we will show that $\mathbb{E}\left[\frac{\mathcal{E}_i\lambda_i\hat{\mu}_i}{\theta}\right] = 0$. We have:

$$\mathbb{E}\left[\frac{\mathcal{E}_i\lambda_i\hat{\mu}_i}{\theta}\right] = \frac{1}{\theta}\mathbb{E}\left[\mathcal{E}_i\lambda_i\frac{\mu_i - \tilde{\mu}}{\mu_i}\right] \tag{63}$$

$$= \frac{1}{\theta}\mathbb{E}\left[\lambda_i\mathcal{E}_i\left(1 - \mu_i^{-1}\frac{\mathbb{E}\left[\lambda_i\mathcal{E}_i\right]}{\mathbb{E}\left[\lambda_i\mathcal{E}_i\mu_i^{-1}\right]}\right)\right] \tag{64}$$

$$= \frac{1}{\theta}\mathbb{E}\left[\lambda_i\mathcal{E}_i - \frac{\lambda_i\mathcal{E}_i\mu_i^{-1}\mathbb{E}\left[\lambda_i\mathcal{E}_i\right]}{\mathbb{E}\left[\lambda_i\mathcal{E}_i\mu_i^{-1}\right]}\right] \tag{65}$$

$$= \frac{1}{\theta}\left(\mathbb{E}\left[\lambda_i\mathcal{E}_i\right] - \frac{\mathbb{E}\left[\lambda_i\mathcal{E}_i\mu_i^{-1}\right]}{\mathbb{E}\left[\lambda_i\mathcal{E}_i\mu_i^{-1}\right]}\mathbb{E}\left[\lambda_i\mathcal{E}_i\right]\right) \tag{66}$$

$$= 0 \tag{67}$$

Plugging back into our earlier expression, this yields our desired result:

$$d\log Y = -\mathbb{E}\left[\mathcal{E}_i\lambda_i\hat{\mu}_i d\log\mu_i\right] \tag{68}$$

□

A.3 Proof of Proposition 2

Proof. As described in the main text, we integrate $\frac{d\log Y}{d\log\mu}$ along a path from the distorted to the wedgeless economy. We use the trapezoid rule to get an approximation that is accurate up to second-order:

$$\mathcal{L} \approx -\frac{1}{2}\cdot\mathbb{E}\left[\left(\frac{d\log Y\left(\mu=1\right)}{d\log\mu_i}+\frac{d\log Y\left(\mu=\mu\right)}{d\log\mu_i}\right)\log\mu_i\right] \tag{69}$$

$$= -\frac{1}{2}\cdot\mathbb{E}\left[\frac{d\log Y}{d\log\mu_i}\log\mu_i\right] \tag{70}$$

where the second line takes advantage of the fact that, thanks to the envelope theorem, $\frac{d\log Y}{d\log\mu_i}=0$ around the undistorted (efficient) economy. Plugging in our formula from Proposition 1, we have

$$\mathcal{L} = \frac{1}{2}\mathbb{E}\left[\mathcal{E}_i\lambda_i\hat{\mu}\log\mu_i\right] \tag{71}$$

We can turn $\mathbb{E}\left[\mathcal{E}_i\lambda_i\hat{\mu}\log\mu_i\right]$ into a more intuitive expression using some additional approximations. First, we will use a first-order Taylor approximation to convert $\hat{\mu}$ into a function of log wedges.

$$\log\mu_i-\log\tilde{\mu}\approx\frac{1}{\mu_i}\left(\mu_i-\tilde{\mu}\right) \tag{72}$$

$$=\hat{\mu} \tag{73}$$

$$\implies \mathbb{E}\left[\mathcal{E}_i\lambda_i\hat{\mu}\log\mu_i\right]\approx\mathbb{E}\left[\mathcal{E}_i\lambda_i\left(\log\mu_i-\log\tilde{\mu}\right)\log\mu_i\right] \tag{74}$$

Note that since $\hat{\mu}$ was a valid first-order approximation to $\log\mu_i-\log\tilde{\mu}$, and we are then multiplying by $\log\mu_i$, our new approximation is equivalent to the old one up to second-order.

Next, we will replace the weighted harmonic average, $\tilde{\mu}$, with a geometric average that uses the same weights. We define:

$$\log\bar{\mu}=\frac{\mathbb{E}\left[\mathcal{E}_i\lambda_i\log\mu_i\right]}{\mathbb{E}\left[\mathcal{E}_i\lambda_i\right]} \tag{75}$$

Substituting this into our old expression yields:

$$\mathbb{E}\left[\mathcal{E}_i\lambda_i\hat{\mu}\log\mu_i\right]\approx\mathbb{E}\left[\mathcal{E}_i\lambda_i\left(\log\mu_i-\log\bar{\mu}\right)\log\mu_i\right] \tag{76}$$

$$=\mathbb{E}\left[\mathcal{E}_i\lambda_i\right]\cdot\mathbb{E}_{\mathcal{E}_i\lambda_i}\left[\left(\log\mu_i-\log\bar{\mu}\right)\log\mu_i\right] \tag{77}$$

$$=\mathbb{E}_{\lambda_i}\left[\mathcal{E}_i\right]\cdot\mathrm{Var}_{\mathcal{E}_i\lambda_i}\left(\log\mu_i\right) \tag{78}$$

where the last line uses the fact that $\mathbb{E}\left[\lambda_i\right]=\mathbb{E}\left[\frac{p_iy_i}{\mathbb{E}[p_iy_i]}\right]=1$. We can then plug this back

into our full expression, to obtain our desired expression:

$$\mathcal{L} \approx \frac{1}{2}\mathbb{E}_{\lambda_i}\left[\mathcal{E}_i\right] \cdot \mathrm{Var}_{\mathcal{E}_i\lambda_i}\left(\log \mu_i\right) \tag{79}$$

□

A.4 Proof of Proposition 3

Proof. Normalizing the price of the final good to one ($P = 1$), CES demand yields the demand curve:

$$\implies p_i = Y^{\frac{1}{\theta}} y_i^{\frac{-1}{\theta}} \tag{80}$$

We can then solve for the firm's optimal level of capital, using the distorted first order condition and then plugging in demand and the firm production function:

$$\text{FOC: } \log p_i = \log \mu_i + \log r - \log \frac{dy_i}{dk_i} \tag{81}$$

$$\text{Demand: } \frac{1}{\theta}\log Y - \frac{1}{\theta}\log y_i = \log \mu_i + \log r - \log\left(\frac{d\log y_i}{d\log k_i} \cdot \frac{y_i}{k_i}\right) \tag{82}$$

$$= \log \mu_i + \log r - \log \alpha - \log y_i + \log k_i \tag{83}$$

$$= \log \mu_i + \log r - \log \alpha - \log y_i + \frac{1}{\alpha}\left(\log y_i - \log z_i\right) \tag{84}$$

$$\implies \left(\frac{1-\alpha}{\alpha} + \frac{1}{\theta}\right)\log y_i = \frac{1}{\alpha}\log z_i - \log \mu_i + \frac{1}{\theta}\log Y - \log r + \log \alpha \tag{85}$$

$$\implies \log y_i = \mathcal{E} \cdot \left[\frac{1}{\alpha}\log z_i - \log \mu_i + \underbrace{\frac{1}{\theta}\log Y - \log r + \log \alpha}_{C}\right] \tag{86}$$

where $\mathcal{E} := \left(\frac{1-\alpha}{\alpha} + \frac{1}{\theta}\right)^{-1}$ is the elasticity of output with respect to the wedge, and C is a constant that will fall out.

Next, we need to solve for $\log Z := \log Y - \alpha \log K$. To do this, we can exploit the joint lognormality of z_i and μ_i. Since $\log z_i$ and $\log \mu_i$ are multivariate normal, and since $\log y_i$ is a linear function of $\log z_i$ and $\log \mu_i$, we have that y_i is jointly lognormal with z_i and μ_i, and

by extension so is k_i ($\log k_i = \log y_i - \alpha \log k_i$). We thus have

$$\log Y = \frac{\theta}{\theta - 1} \cdot \log \mathbb{E}\left[y_i^{\frac{\theta-1}{\theta}}\right] \tag{87}$$

$$= \frac{\theta}{\theta - 1} \cdot \left(\frac{\theta - 1}{\theta} \cdot \mathbb{E}\left[\log y_i\right] + \left(\frac{\theta - 1}{\theta}\right)^2 \cdot \frac{1}{2}\mathrm{Var}\left(\log y_i\right)\right) \tag{88}$$

$$= \mathbb{E}\left[\log y_i\right] + \left(\frac{\theta - 1}{\theta}\right) \cdot \frac{1}{2}\mathrm{Var}\left(\log y_i\right) \tag{89}$$

and similarly

$$\log K := \log \mathbb{E}\left[k_i\right] \tag{90}$$

$$= \log \mathbb{E}\left[\left(\frac{y_i}{z_i}\right)^{\frac{1}{\alpha}}\right] \tag{91}$$

$$= \frac{1}{\alpha}\mathbb{E}\left[\log y_i - \log z_i\right] + \frac{1}{2}\frac{1}{\alpha^2}\mathrm{Var}\left(\log y_i - \log z_i\right) \tag{92}$$

$$= \frac{1}{\alpha}\mathbb{E}\left[\log y_i - \log z_i\right] + \frac{1}{2}\frac{1}{2\alpha^2}\mathrm{Var}\left(\log y_i\right) + \frac{1}{2}\frac{1}{\alpha^2}\mathrm{Var}\left(\log z_i\right) - \frac{1}{\alpha^2}\mathrm{Cov}\left(\log y_i, \log z_i\right) \tag{93}$$

We now combine these two expressions to solve for $\log Z$. We have:

$$\log Z := \log Y - \alpha \log K \tag{94}$$

$$= \mathbb{E}\left[\log y_i\right] + \left(\frac{\theta - 1}{\theta}\right) \cdot \frac{1}{2}\mathrm{Var}\left(\log y_i\right) - \mathbb{E}\left[\log y_i - \log z_i\right] - \frac{1}{2}\frac{1}{\alpha}\mathrm{Var}\left(\log y_i\right) - \frac{1}{2}\frac{1}{\alpha}\mathrm{Var}\left(\log z_i\right) + \frac{1}{\alpha}\mathrm{Cov}\left(\log y_i, \log z_i\right) \tag{95}$$

$$= \mathbb{E}\left[\log z_i\right] - \frac{1}{2}\frac{1}{\alpha}\mathrm{Var}\left(\log z_i\right) + \frac{1}{2} \cdot \left(\frac{\theta - 1}{\theta} - \frac{1}{\alpha}\right)\mathrm{Var}\left(\log y_i\right) + \frac{1}{\alpha}\mathrm{Cov}\left(\log y_i, \log z_i\right) \tag{96}$$

Solving just for $\frac{1}{2} \cdot \left(\frac{\theta-1}{\theta} - \frac{1}{\alpha}\right)\mathrm{Var}\left(\log y_i\right) + \frac{1}{\alpha}\mathrm{Cov}\left(\log y_i, \log z_i\right)$, and noting that $\left(\frac{\theta-1}{\theta} - \frac{1}{\alpha}\right) =$

$-\mathcal{E}^{-1}$,[32] we have:

$$\begin{aligned}
&\frac{1}{2}\cdot\left(\frac{\theta-1}{\theta}-\frac{1}{\alpha}\right)\text{Var}\left(\log y_i\right)-\frac{1}{\alpha}\text{Cov}\left(\log y_i,\log z_i\right)\\
&=-\frac{1}{2}\cdot\mathcal{E}^{-1}\text{Var}\left(\mathcal{E}\cdot\left[\frac{1}{\alpha}\log z_i-\log\mu_i\right]\right)+\frac{1}{\alpha}\text{Cov}\left(\mathcal{E}\cdot\left[\frac{1}{\alpha}\log z_i-\log\mu_i\right],\log z_i\right) \qquad (97)\\
&=-\frac{1}{2}\cdot\mathcal{E}\left(\frac{1}{\alpha^2}\text{Var}\left(\log z_i\right)+\text{Var}\left(\log\mu_i\right)-2\cdot\frac{1}{\alpha}\text{Cov}\left(\log z_i,\log\mu_i\right)\right)\\
&\quad+\frac{1}{\alpha}\mathcal{E}\left(\frac{1}{\alpha}\text{Var}\left(\log z_i\right)-\text{Cov}\left(\log\mu_i,\log z_i\right)\right) \qquad (98)\\
&=\frac{1}{2}\cdot\mathcal{E}\frac{1}{\alpha^2}\text{Var}\left(\log z_i\right)-\frac{1}{2}\cdot\mathcal{E}\text{Var}\left(\log\mu_i\right) \qquad (99)
\end{aligned}$$

Plugging this back in, we obtain the formula:

$$\log Z=\mathbb{E}\left[\log z_i\right]-\frac{1}{2}\cdot\mathcal{E}\text{Var}\left(\log\mu_i\right)+\frac{1}{2}\cdot\mathcal{E}\frac{1}{\alpha^2}\text{Var}\left(\log z_i\right)-\frac{1}{2}\frac{1}{\alpha}\text{Var}\left(\log z_i\right) \qquad (100)$$

which is Equation 17 from the main text. From here, it is immediate that this is maximized when the variance of the log wedges is zero. Thus, we have

$$\log Z^*-\log Z=\frac{1}{2}\cdot\mathcal{E}\text{Var}\left(\log\mu_i\right) \qquad (101)$$

which completes the proof. □

A.5 Proof of Proposition 4

Proof. As described in the main text, the assumption of a Cobb-Douglas production function where $\alpha_{m,i}=\tilde{\alpha}_m\cdot\sum_M\alpha_{m,i}$, along with the assumption that wedges only appear as a revenue tax, ensures that each firm will demand the same mix of inputs. In particular, we solving the firm's cost minimization problem yields the well-known result for cost shares under Cobb-Douglas production:

$$\frac{r_m x_m}{\sum_M r_m x_m}=\frac{\alpha_{m,i}}{\sum_M\alpha_{m,i}}=\tilde{\alpha}_m \qquad (102)$$

where r_m denotes the cost of input m. Since all firms consume inputs in the same proportion, it is straightforward for us to show how to reduce the problem to be equivalent to our single input case. First, note that market clearing pins down the ratio of input prices to not depend on the wedges

$$\frac{r_m X_m}{r_n X_n}=\frac{\tilde{\alpha}_m}{\tilde{\alpha}_n}\implies\frac{r_m}{r_n}=\frac{\tilde{\alpha}_m}{\tilde{\alpha}_n}\frac{X_n}{X_m}\forall m,n \qquad (103)$$

[32] To see this, observe that: $\left(\frac{\theta-1}{\theta}-\frac{1}{\alpha}\right)=\frac{\alpha\theta-\alpha-\theta}{\alpha\theta}=-\left(\frac{\theta-\alpha\theta+\alpha}{\alpha\theta}\right)=-\left(\frac{1-\alpha}{\alpha}+\frac{1}{\theta}\right)=-\mathcal{E}^{-1}$

Since the ratio of input prices is pinned down by the (inelastic) aggregate supply of those inputs (as well as by the production parameters $\tilde{\alpha}_m$ and $\tilde{\alpha}_n$), it follows that relative input prices will not change, and thus that input shares will also not change.

We can thus, without loss, reduce our firm's production function to take a composite input, with the shares pinned down as described above. We then have the composite production function:

$$\begin{aligned}\log y_i &= \log z_i + \sum_M \alpha_{m,i} \log x_{m,i} \\ &= \log z_i + \left(\sum_M \alpha_{m,i}\right) \log x_i\end{aligned}$$

where x_i is a composite input defined such that $x_{m,i} \propto x_i$.

Having defined our composite production function, the remainder of the proposition is immediate from our previous results. The formula for misallocation is simply an application of 2, while the formula for $\mathcal{E}_i$ is simply an application of our result in 3. □

A.6 Proof of Proposition 5

Proof. The proposition has two components. The first is that $\text{Var}(\log \mu_i) = \text{Var}(\log \text{VMPK}_i)$. This is an immediate result of the efficient first-order condition in Equation 7, which implies

$$\log \text{VMPK}_i = \log r + \log \mu_i$$

Since r is the same across firms by definition, this implies that the variance of log wedges and log VMPK is the same. Note that this implicitly relies on the final good producer being a price taker, so that $p_i = P \cdot \frac{dY}{dy_i}$; since VMPK is defined in terms of the observed price, while the wedges are defined as distortions that lead to deviations from efficient solution to the planner's problem.

The second component is that $\text{Var}(\log \text{VMPK}_i) = \text{Var}(\log \text{MRPK}_i)$. As discussed in the text, under CES demand $\log \text{MRPK}_i = \log \text{VMPK}_i + \log \frac{\theta-1}{\theta}$. Thus, their variance is the same. □

B Proofs for Section 4

B.1 Proof of Proposition 6

Proof. First, assume that $\gamma_0 \neq 0$ in which case we also have $\tau_0 \neq 0$ so that $G = \nabla g(\delta_0)$ is a non-zero vector for both of the null hypotheses considered in the lemma. From the assumptions of the proposition we have

$$\sqrt{n}\Sigma^{-1/2}(\widehat{\delta} - \delta_0) = \mathcal{Z} + o_p(1)$$

where $\mathcal{Z} \sim \mathcal{N}(0, 1)$. We can then show using standard methods, see for example Newey and McFadden (1994), that

$$\sqrt{n}(\widehat{\delta} - \bar{\delta}) = \Sigma G(G'\Sigma G)^{-1}G'\Sigma^{1/2}\mathcal{Z} + o_p(1)$$

Since $\widehat{\Sigma} \xrightarrow{p} \Sigma$, it then follows that

$$\begin{aligned} T(\widehat{\delta}, \tau_0) &= \sqrt{n}(\widehat{\delta} - \bar{\delta})\widehat{\Sigma}^{-1}\sqrt{n}(\widehat{\delta} - \bar{\delta}) \\ &= \mathcal{Z}\Sigma^{1/2}G(G'\Sigma G)^{-1}G'\Sigma^{1/2}\mathcal{Z} + o_p(1) \\ &\xrightarrow{d} \chi^2(1), \end{aligned}$$

where $\chi^2(1)$ is a chi-squared distributed variable with one degree of freedom. Since we simulate draws of the parameter vector from $\delta^* \sim \mathcal{N}(\bar{\delta}, \bar{\Sigma})$, and $\bar{\Sigma} \xrightarrow{p} \Sigma$, identical steps show that

$$\sqrt{n}(\delta^* - \bar{\delta}^*) = \Sigma G(G'\Sigma G)^{-1}G'\Sigma^{1/2}\mathcal{Z} + o_p(1)$$

where $\bar{\delta}^*$ is the constrained minimizer of $T(\delta^*, \tau_0)$, and hence the simulated test statistic also converges in distribution to $\chi^2(1)$. Let $F_n(t) = P(T(\delta^*, \tau_0) \leq t)$ and $F(t) = P(\chi^2(1) \leq t)$. Then convergence in distribution implies that $\sup_t |F_n(t) - F(t)| \to 0$. An extended continuous mapping theorem (e.g. 1.11.1 in van der Vaart and Wellner, 1996) then gives

$$F_n\big(T(\widehat{\delta}, \tau_0)\big) \xrightarrow{d} F\big(\chi^2(1)\big)$$

which is a uniform random variable.

For the case in which $\gamma_0 = 0$ and hence $\tau_0 = 0$, then we must have that $\bar{\delta} = (\bar{\beta}, 0)$. In this case the test statistic $T(\widehat{\delta}, 0)$ is equivalent to a standard Wald test of the null hypothesis $H_0 : \gamma = 0$. Standard results give $T(\widehat{\delta}, 0) \xrightarrow{d} \chi^2(p)$. Similarly, the simulated test statistic is

also simply

$$T(\delta^*, 0) = \gamma^*(\bar{\Sigma}_{\gamma\gamma'})^{-1}\gamma^*,$$

where $\bar{\Sigma}_{\gamma\gamma'}$ is the block of the variance matrix corresponding to $\widehat{\gamma}$, and since $\gamma^* \sim \mathcal{N}(0, \bar{\Sigma}_{\gamma\gamma'})$ we have that $T(\delta^*, 0) \sim \chi^2(p)$ exactly. □

B.2 Uniformly valid inference

Loosely speaking, a test is *uniformly valid* if it has correct asymptotic size under any sequence of true models. Uniform procedures provide guarantees that the test will behave well at any value of the true parameters. Here we show that the infeasible simulation based test that simulates δ^* from a normal distribution with mean δ_0 is uniformly valid.

Let $\delta^* \sim \mathcal{N}(\delta_0, \widehat{\Sigma})$ be a simulated draw of the parameter vector δ, and $F_{\delta_0}(t) = P\big(T(\delta^*, \tau_0) \leq t\big)$ be the corresponding CDF of the simulated test statistic. A p-value for the test statistic (38) is given by $\widehat{p}(\delta_0) = 1 - F_{\delta_0}\big(T(\widehat{\delta}, \tau_0)\big)$. In practice, the CDF F_{δ_0} could be approximated with arbitrary accuracy via simulation. We first demonstrate that under some straightforward conditions the p-value $\widehat{p}(\delta_0)$ converges uniformly to a uniformly distributed variable.

Assumption 2. *Let the data be drawn from some distribution indexed by the possibly infinite dimensional parameter $\lambda \in \Lambda$. We assume that:*

(i) Uniformly consistent variance estimator: $\widehat{\Sigma}$ is a uniformly consistent estimator of the symmetric positive definite variance matrix $\Sigma(\lambda)$, i.e.

$$\sup_{\lambda \in \Lambda} P_\lambda\big(\|\widehat{\Sigma} - \Sigma(\lambda)\| > \varepsilon\big) \to 0,$$

where $\lambda_{\min}(\Sigma(\lambda)) \geq c > 0$ for some constant c for all $\lambda \in \Lambda$.

(ii) Uniform convergence of parameter estimates: $\sqrt{n}(\widehat{\delta} - \delta_0)$ converges uniformly in distribution to $\mathcal{Z}(\lambda) \sim N(0, \Sigma(\lambda))$, i.e.

$$\sup_{\lambda \in \Lambda} d_{BL}^{\lambda}\big(\sqrt{n}(\widehat{\delta} - \delta_0), \mathcal{Z}(\lambda)\big) \to 0,$$

where d_{BL} is the bounded Lipschitz metric (e.g. see Kasy, 2018).

Assumption 2 requires uniform consistency of the variance estimator $\widehat{\Sigma}$ along with uniform convergence of the parameter estimate $\widehat{\delta}$. This will hold in many standard settings; for the instrumental variables estimators used in this paper, uniform convergence of the IV estimates will require an assumption of strong identification.

Lemma 2. *Let $F_{\delta_0}(t) = P\big(T(\delta^*, g(\delta_0)) \leq t\big)$ be the CDF of the statistic $T(\delta^*, \tau_0)$ for $\tau_0 = g(\delta_0)$, where $\delta^* \sim \mathcal{N}(\delta_0, \widehat{\Sigma}/n)$. Define the p-value of the test statistic $T(\widehat{\delta}, \tau_0)$ as $\widehat{p}(\delta_0) = 1 - F_{\delta_0}\big(T(\widehat{\delta}, g(\delta_0))\big)$. Then, under Assumption 2, $\widehat{p}(\delta_0)$ converges uniformly in distribution to a uniform random variable $\mathcal{U}$*

$$\sup_{\lambda \in \Lambda} d_{BL}^{\lambda}\big(\widehat{p}(\delta_0), \mathcal{U}\big) \to 0.$$

Proof. Consider the test statistic $S_n(\widehat{\delta}_n, \delta_0) = T(\widehat{\delta}, g(\delta_0))^{1/2}$

$$S_n(\widehat{\delta}_n, \delta_0) = \min_{d: g(\delta_0 + \frac{1}{\sqrt{n}}\widehat{\Sigma}^{1/2} d) = 0} \sqrt{(d - \widehat{\delta}_n)'(d - \widehat{\delta}_n)},$$

where $\widehat{\delta}_n = \sqrt{n}\widehat{\Sigma}^{-1/2}(\widehat{\delta} - \delta_0)$. We first show that the statistic $S_n(\delta, \delta_0)$ is Lipschitz continuous in its first argument. Let

$$\bar{d} = \arg \min_{d: g(\delta_0 + \frac{1}{\sqrt{n}}\widehat{\Sigma}^{1/2} d) = 0} \sqrt{(d - \delta)'(d - \delta)}$$

be the constrained minimizer associated with $S_n(\delta, \delta_0)$. Similarly, let $\tilde{d}$ be the constrained minimizer corresponding to $S_n(\tilde{\delta}, \delta_0)$. Using the fact that $\tilde{d}$ is a minimizer, and applying the triangle inequality, we find

$$\begin{aligned} S_n(\tilde{\delta}, \delta_0) &\leq \sqrt{(\bar{d} - \tilde{\delta})'(\bar{d} - \tilde{\delta})} \\ &\leq \sqrt{(\bar{d} - \delta)'(\bar{d} - \delta)} + \sqrt{(\delta - \tilde{\delta})'(\delta - \tilde{\delta})} \\ &= S_n(\delta, \delta_0) + \|\delta - \tilde{\delta}\| \end{aligned}$$

Similarly, we have $S_n(\delta, \delta_0) \leq S_n(\tilde{\delta}, \delta_0) + \|\delta - \tilde{\delta}\|$, and hence

$$|S_n(\delta, \delta_0) - S_n(\tilde{\delta}, \delta_0)| \leq \|\delta - \tilde{\delta}\|,$$

and hence $S_n(\delta, \delta_0)$ is Lipschitz continuous in its first argument. Since $\sqrt{n}(\widehat{\delta} - \delta_0)$ converges uniformly in distribution to $N(0, \Sigma)$ and the variance estimator is uniformly consistent, we have that $\widehat{\delta}_n = \sqrt{n}\widehat{\Sigma}^{-1/2}(\widehat{\delta} - \delta_0)$ converges uniformly to $\mathcal{Z} \sim N(0, I)$. We can then apply Theorem 1 of Kasy (2018) to find that $S_n(\widehat{\delta}_n, \delta_0)$ converges uniformly in distribution to $S_n(\mathcal{Z}, \delta_0)$.[33] Then, since the CDF $F_\delta(t) = P\big(S_n(\mathcal{Z}, \delta) \leq t\big)$ is also a Lipschitz con-

[33] The theorem is stated for a fixed function ψ, while our function depends on the sample size n and the

tinuous function we also have that $F_{\delta_0}\big(S_n(\widehat{\delta}_n, \delta_0)\big)$ converges uniformly in distribution to $F_{\delta_0}\big(S_n(\mathcal{Z}, \delta_0)\big) \sim U[0,1]$. Letting $G_\delta(t) = P\big(T(\delta^*, g(\delta)) \leq t\big) = F_\delta(\sqrt{t})$ we have that $F_{\delta_0}\big(S_n(\widehat{\delta}_n, \delta_0)\big) = G_{\delta_0}\big(T_n(\widehat{\delta}, g(\delta_0))\big)$ and so

$$\widehat{p}(\delta_0) = 1 - G_{\delta_0}\big(T_n(\widehat{\delta}, g(\delta_0))\big)$$

converges uniformly to $1 - F_{\delta_0}\big(S_n(\mathcal{Z}, \delta_0)\big) \sim U[0,1]$. □

The test described in the paper, which replaces the unknown δ_0 with an estimate $\bar{\delta}$, is no longer uniformly valid. We now briefly describe a feasible and uniformly valid version of the test. Let $\widehat{p}_{\tau_0}(\delta)$ be the p-value associated with the test that simulates δ^* from a normal distribution with mean δ. The 'worst case' p-value is given by

$$\tilde{p}_{\tau_0} = \sup_{\delta: g(\delta) = \tau_0} \widehat{p}_{\tau_0}(\delta),$$

that is, the largest p-value over all parameter vectors δ that are consistent with the null hypothesis. Construction of this p-value is feasible since it does not require knowledge of the true δ_0. Also, since $\tilde{p}_{\tau_0}$ is no smaller than the p-value constructed using δ_0, it is also uniformly valid as shown in the following proposition.

Proposition 7. *Let Assumption 2 hold and let $\widehat{p}_\tau = \sup_{\delta: g(\delta)=\tau} \widehat{p}(\delta)$ be the largest p-value over all δ satisfying the null hypothesis. Then the confidence set*

$$\widehat{C}_{1-\alpha} = \big\{\tau : \widehat{p}_\tau \geq \alpha\big\}$$

is uniformly valid, in the sense that

$$\lim_{n\to\infty} \sup_{\lambda \in \Lambda} P_\lambda\big(\tau(\lambda) \in \widehat{C}_{1-\alpha}\big) \geq 1 - \alpha$$

Proof. Since $\widehat{p}_{\tau(\lambda)} = \sup_{\delta: g(\delta)=\tau} \widehat{p}(\delta) \geq \widehat{p}(\delta_\lambda)$, we have

$$P_\lambda\big(\tau(\lambda) \in \widehat{C}_{1-\alpha}\big) = P_\lambda\big(\widehat{p}_{\tau(\lambda)} \geq \alpha\big) \geq P_\lambda\big(\widehat{p}(\delta_\lambda) \geq \alpha\big)$$

and hence

$$\lim_{n\to\infty} \sup_{\lambda \in \Lambda} P_\lambda\big(\tau(\lambda) \in \widehat{C}_{1-\alpha}\big) \geq \lim_{n\to\infty} \sup_{\lambda \in \Lambda} P_\lambda\big(\widehat{p}(\delta_\lambda) \geq \alpha\big) = 1 - \alpha$$

variance matrix $\widehat{\Sigma}$. Inspection of the proof indicates that the result may still be applied so long as the Lipschitz constant is fixed, which is true in this case (since it is one).

by the fact that $\widehat{p}(\delta_\lambda)$ converges uniformly to uniformly distributed variable, which is equivalent to uniform convergence of its CDF at all continuity points (which includes the point α since the uniform CDF is continuous on $(0,1)$). □

B.3 A feasible uniformly valid test

In practice searching over all δ satisfying the null hypothesis for the worst case critical values is likely to be computationally demanding, particularly when the dimension of δ is not small. In some settings it is possible to show that the distribution F_δ depends only on $\tau = g(\delta)$ so that valid critical values can be simulated using any δ satisfying the null hypothesis. This is the case for example when $g(\cdot)$ is linear in δ, or when $g(\delta) = \delta'\delta$ and $\Sigma = I$. Andrews and Mikusheva (2016) provide a procedure for generating critical values for the distance metric statistic that is valid for any δ satisfying the null, although their method is conservative in general.

In other cases it may be possible to identify the worst case choice of δ directly. In the case that we are interested in a one-sided hypothesis on the parameter $\tau = \sqrt{\gamma'\gamma}$, we conjecture that the worst case value of γ is related the a particular eigenvector of the variance matrix Σ.[34]

Conjecture 1. *Let Σ_γ be the variance matrix associated with $\widehat{\gamma}$, and let $\Sigma_\gamma = VDV'$ be its eigen-decomposition, where $D = diag(d_1, \ldots, d_p)$ is a diagonal matrix of eigenvalues in decreasing order $d_1 \geq d_2 \geq \cdots \geq d_p$ and V is an orthonormal matrix of eigenvectors. The worst case γ for testing the null hypothesis $H_0 : \sqrt{\gamma'\gamma} \leq \tau_0$ is given by*

$$\gamma_{worst} = \tau_0 v_p \tag{104}$$

where v_p is the eigenvector associated with the smallest eigenvalue of Σ_γ.

Assuming the conjecture to be true, this would allow us to test the null hypothesis $H_0 : \sqrt{\gamma'\gamma} \leq \tau_0$ by simulating draws $\gamma_b^* \sim \mathcal{N}(\gamma_{worst}, \widehat{\Sigma})$ computing the corresponding quantiles of the test statistic. We could similarly construct a confidence set for $\tau = \sqrt{\gamma'\gamma}/\beta$ by using a worst case value of $\delta = (\beta, \gamma)$. However, the worst case distribution is likely to be particularly bad for values of β close to zero and so this method may be overly conservative. Instead, we suggest constructing a joint confidence set for $(\beta, \sqrt{\gamma'\gamma})$ by testing the null hypothesis $H_0 : \beta = \beta_0, \sqrt{\gamma'\gamma} \leq S_0$. Critical values for this joint null are then simulated from $\delta = (\beta_0, \gamma_{worst})$. We can then use the projection method to construct a confidence set for τ from this joint confidence set by finding the minimum value of $\tau_0 = S_0/\beta_0$ across all

[34] Although we are currently unable to prove the conjecture, we have verified it in a range of simulations.

(β_0, S_0) that cannot be rejected. Although this uses the projection method, it does so only for the two-dimensional parameter $(\beta, \sqrt{\gamma'\gamma})$, rather than the full parameter vector δ, and so it much less conservative. The full procedure is summarized in the following algorithm.

Algorithm 2. *A one-sided uniformly valid confidence set for* $\tau = \sqrt{\gamma'\gamma}/\beta$

1. *Estimate the IV regression to obtain parameter estimates* $\widehat{\delta} = (\widehat{\beta}, \widehat{\gamma}')'$ *and variance matrix* $\widehat{\Sigma}$

2. *For the joint null hypothesis* $H_0 : \beta = \beta_0, \sqrt{\gamma'\gamma} \leq S_0$*:*

   (a) *compute the worst-case* γ *value,* $\gamma_{worst} = S_0 v_p$ *as in (104), and constrained variance matrix* $\bar{\Sigma}$

   (b) *compute the test statistic*

$$T(\widehat{\delta}, S_0, \beta_0) = \min_{\substack{\delta: \sqrt{\gamma'\gamma} = \leq S, \\ \beta = \beta_0}} n(\delta - \widehat{\delta})' \bar{\Sigma}^{-1} (\delta - \widehat{\delta}),$$

   (c) *for* $b = 1, \ldots, B$*, simulate* $\delta_b \sim N(\delta_{worst}, \bar{\Sigma})$*, and compute the statistic*

$$T_b^*(\delta_b, S_0, \beta_0) = \min_{\substack{\delta: \sqrt{\gamma'\gamma} \leq S_0, \\ \beta = \beta_0}} n(\delta - \delta_b)' \bar{\Sigma}^{-1} (\delta - \delta_b),$$

   and set the critical value $c_{1-\alpha}(S_0, \beta_0)$ *as the* $(1-\alpha)$*-quantile of* $T_b^*(\delta_b, S_0, \beta_0)$

   (d) *reject* $H_0 : \beta = \beta_0, \sqrt{\gamma'\gamma} \leq S_0$ *if* $T(\widehat{\delta}, \tau_0) > c_{1-\alpha}(\tau_0)$

3. *Repeat step 2 for a grid of* (β_0, S_0) *values to construct a joint confidence set for* (β, S)*,* $\widehat{C}_{1-\alpha}(\beta, S)$*. Then compute a one-sided confidence set for* $\tau = S/\beta$ *as*

$$\widehat{C}_{1-\alpha} = \Big(\min_{(\beta,S) \in \widehat{C}_{1-\alpha}(\beta,S)} \frac{S}{\beta}, \infty \Big)$$

C Additional Tables and Figures

Table 6: Accuracy of Second-Order Approximation: Simulation Results

	Log-Normal	Log-Uniform	Two-Point	Heterogeneous RTS
True Gains: Y^*/Y-1	22%	18%	17%	26%
$\exp\left(\frac{1}{2}\mathbb{E}_{\lambda_i}\left[\mathcal{E}\right]\cdot\mathrm{Var}\left(\log\mu_i\right)\right)-1$	22%	22%	22%	25%
$\exp\left(\frac{1}{2}\mathbb{E}_{\lambda_i}\left[\mathcal{E}\right]\cdot\mathrm{Var}_{\lambda_i}\left(\log\mu_i\right)\right)-1$	22%	21%	20%	26%
$\exp\left(\frac{1}{2}\mathbb{E}_{\lambda_i}\left[\mathcal{E}\right]\cdot\mathrm{Var}_{\lambda_i\mathcal{E}_i}\left(\log\mu_i\right)\right)-1$	22%	21%	20%	26%
$\exp\left(\frac{1}{2}\mathbb{E}_{\lambda_i}\left[\mathcal{E}\right]\cdot\log\left(1+\frac{\mathrm{Var}(\mu_i)}{\mathbb{E}[\mu_i]^2}\right)\right)-1$	22%	13%	10%	25%

Notes: This table compares the true benefits of reallocating capital to various approximate formula, under different simulations. In all simulations, we assume that firms use a Cobb-Douglas production function: $\log y_i = \log z_i + \alpha \log k_i$, face CES demand with elasticity of substitution $\theta = 3$, and that there is an aggregate capital constraint $\mathbb{E}[k_i] = \bar{K}$. We also assume throughout that $\log z_i \sim N\left(0, \sigma_z^2\right)$, where $\sigma_z = 1.2$. Following our preferred point estimates, we assume that the variance of $\log \mu_i$ is 0.93. We draw μ independently from z. In the first column, we simulate from a lognormal distribution for μ_i. In the second column, we simulate $\log \mu_i$ from a uniform distribution, $\log \mu_i \sim [-a, a]$, where a is selected such that the variance of $\log \mu_i$ is 0.93. In the third column, we simulate $\log \mu$ from an equal-probability two-point distribution, $\log \mu_i = \{-\sigma_\mu, \sigma_\mu\}$, which again imposes that the variance of $\log \mu_i$ is 0.93. For the first three columns, we assume $\alpha = \frac{1}{3}$. For the fourth column, we return to drawing μ_i from a log-normal distribution, but we now draw heterogeneous α_i, drawing $\alpha_i = \left\{\frac{1}{6}, \frac{1}{3}, \frac{1}{2}\right\}$ with equal probability, and drawing α_i independently from (μ, z).
The first row shows the simulated welfare benefits from optimally reallocating capital across firms. The second row shows a second-order approximation to these gains, using the unweighted variance of log wedges. The third row shows the same approximation using the sales-weighted variance, and the fourth row shows the same approximation using the elasticity-times-sales-weighted variance, as in Proposition 2. The fifth row shows the same formula using $\log\left(1 + \mathrm{Var}\left(\mu_i\right)/\mathbb{E}\left[\mu_i\right]^2\right)$ in place of the variance of $\log \mu$, as discussed in Section 3.

Table 7: Estimates of Variance of MRPK: 95% Confidence Intervals

	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$
Panel A: $SD\big(\mathbb{E}[\mathrm{MRPK}_i\|X_i]\big) = \sqrt{\gamma'\gamma}$							
Estimate	0.066	0.063	0.109	0.107	0.098	0.131	0.128
95% CI	$[0.02, 0.13]$	$[0.00, 0.12]$	$[0.03, \infty]$	$[0.03, \infty]$	$[0.03, \infty]$	$[0.07, \infty]$	$[0.03, \infty]$
Panel B: $SD\big(\mathbb{E}[\mathrm{MRPK}_i\|X_i]\big)/\mathbb{E}[\mathrm{MRPK}_i] = \sqrt{\gamma'\gamma}/\beta$							
Estimate	0.913	0.840	1.415	1.275	1.234	1.247	1.213
95% CI	$[0.34, 2.08]$	$[0.00, 2.30]$	$[0.33, \infty]$	$[0.41, 4.72]$	$[0.38, \infty]$	$[0.71, \infty]$	$[0.45, \infty]$

Notes: This table shows estimates of heterogeneous models of MRPK. All standard errors and confidence intervals are clustered at the firm level. In Panel A, each column shows estimates from the heterogeneous model described in Equation 28. Each column uses the first K principal components of our vector of covariates. Panel A shows estimates of $\mathrm{SD}\left(\mathbb{E}\left[\mathrm{MRPK}_i \mid X_i\right]\right) = \sqrt{\gamma'\gamma}$, as well as 95% confidence intervals computed using Algorithm 1. Panel B shows the implied estimate of $\mathrm{SD}\left(\mathbb{E}\left[\mathrm{MRPK}_i \mid X_i\right]\right)/\mathbb{E}\left[\mathrm{MRPK}_i\right] = \sqrt{\gamma'\gamma}/\beta$, as well as 95% confidence intervals computed using Algorithm 1. Where the confidence intervals have an upper bound of infinity, this indicates that the largest null tested (2 for Panel A and 5 for Panel B) could not be rejected.

Table 8: Estimates of Variance of MRPK: Uniformly Valid Confidence Intervals

	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$
Panel A: $SD\big(\mathbb{E}[\text{MRPK}_i\|X_i]\big) = \sqrt{\gamma'\gamma}$							
Estimate	0.066	0.063	0.109	0.107	0.098	0.131	0.128
90% CI	$[0.03, 0.11]$	$[0.02, 0.11]$	$[0.03, 1.05]$	$[0.04, 2.00]$	$[0.03, \infty]$	$[0.06, \infty]$	$[0.03, \infty]$
Panel B: $SD\big(\mathbb{E}[\text{MRPK}_i\|X_i]\big)/\mathbb{E}[\text{MRPK}_i] = \sqrt{\gamma'\gamma}/\beta$							
Estimate	0.913	0.840	1.415	1.275	1.234	1.247	1.213
90% CI	$[0.21, \infty]$	$[0.00, \infty]$	$[0.21, \infty]$	$[0.39, \infty]$	$[0.39, \infty]$	$[0.67, \infty]$	$[0.44, \infty]$

Notes: This table shows estimates of heterogeneous models of MRPK, with uniformly valid confidence intervals based on Algorithm 2. All confidence intervals are clustered at the firm level. Each column shows estimates from the heterogeneous model described in Equation 28. Each column uses the first K principal components of our vector of covariates. Panel A shows estimates of $\text{SD}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right) = \sqrt{\gamma'\gamma}$, as well as 90% confidence intervals computed using Algorithm 2. Panel B shows the implied estimate of $\text{SD}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)/\mathbb{E}\left[\text{MRPK}_i\right] = \sqrt{\gamma'\gamma}/\beta$, as well as 90% confidence intervals computed using Algorithm 2. Where the confidence intervals have an upper bound of infinity, this indicates that the largest null tested (2 for Panel A and 5 for Panel B) could not be rejected.

Table 9: Estimates of Variance of MRPK: Weighted Variance

	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$
Panel A: $\mathbb{E}[\text{MRPK}_i] = \beta$							
Estimate	0.069	0.053	0.020	0.019	0.047	0.084	0.095
SE	(0.026)	(0.042)	(0.070)	(0.083)	(0.113)	(0.104)	(0.244)
Panel B: $SD\big(\mathbb{E}[\text{MRPK}_i\|X_i]\big) = \sqrt{\gamma'\gamma}$							
Estimate	0.060	0.070	0.139	0.123	0.109	0.130	0.121
90% CI	$[0.03, 0.11]$	$[0.03, 0.15]$	$[0.06, \infty]$	$[0.04, \infty]$	$[0.04, \infty]$	$[0.08, \infty]$	$[0.04, \infty]$
Panel C: $SD\big(\mathbb{E}[\text{MRPK}_i\|X_i]\big)/\mathbb{E}[\text{MRPK}_i] = \sqrt{\gamma'\gamma}/\beta$							
Estimate	0.866	1.335	7.045	6.550	2.315	1.551	1.272
90% CI	$[0.41, 1.73]$	$[0.35, \infty]$	$[0.88, \infty]$	$[0.43, 3.97]$	$[0.44, \infty]$	$[0.66, \infty]$	$[0.39, \infty]$

Notes: This table shows estimates of heterogeneous models of MRPK, using covariates that target the weighted variance of returns. The weights are firm profits at baseline. All standard errors and confidence intervals are clustered at the firm level. In Panel A, each column shows estimates from the heterogeneous model described in Equation 28. Each column uses the first K principal components of our vector of covariates, with the principal components constructed based on the weighted variance matrix, and standardized to ensure that the standardized factors have weighted mean zero and a weighted variance of one. Panel A shows estimates of $\mathbb{E}\left[\text{MRPK}_i\right] = \beta$, along with standard errors. Panel B shows estimates of $\text{SD}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right) = \sqrt{\gamma'\gamma}$, as well as 90% confidence intervals computed using Algorithm 1. Panel C shows the implied estimate of $\text{SD}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)/\mathbb{E}\left[\text{MRPK}_i\right] = \sqrt{\gamma'\gamma}/\beta$, as well as 90% confidence intervals computed using Algorithm 1. Where the confidence intervals have an upper bound of infinity, this indicates that the largest null tested (2 for Panel B and 5 for Panel C) could not be rejected.

Table 10: Cobb-Douglas Estimates of MRPK

Panel A: Unconditional Estimates							
$\mathbb{E}[\text{MRPK}_i]$			$\text{SD}(\text{MRPK}_i)/\mathbb{E}[\text{MRPK}_i]$			$\text{SD}(\log \text{MRPK}_i)$	
0.078			2.082			1.141	
Covariate	Capital	Age	Education	Profit	Hours	APK	log(APK)
Panel B: $SD\big(\mathbb{E}[\text{MRPK}_i\|X_i]\big)$							
Estimate	-0.044	+0.015	+0.029	-0.005	-0.023	+0.053	+0.060
90% CI	[0.04, 0.05]	[0.01, 0.02]	[0.02, 0.04]	[0.002, 0.01]	[0.02, 0.03]	[0.04, 0.06]	[0.05, 0.07]
Panel C: $SD\big(\mathbb{E}[\text{MRPK}_i\|X_i]\big)/\mathbb{E}[\text{MRPK}_i]$							
Estimate	0.568	0.187	0.370	0.067	0.295	0.666	0.770
90% CI	[0.53, 0.60]	[0.12, 0.26]	[0.30, 0.44]	[0.02, 0.12]	[0.25, 0.35]	[0.56, 0.78]	[0.71, 0.81]
	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$
Panel D: $SD\big(\mathbb{E}[\text{MRPK}_i\|X_i]\big)$							
Estimate	0.061	0.061	0.062	0.062	0.062	0.062	0.064
90% CI	[0.05, 0.07]	[0.05, 0.07]	[0.05, 0.07]	[0.06, 0.07]	[0.06, 0.07]	[0.06, 0.07]	[0.06, 0.07]
Panel E: $SD\big(\mathbb{E}[\text{MRPK}_i\|X_i]\big)/\mathbb{E}[\text{MRPK}_i]$							
Estimate	0.784	0.790	0.794	0.799	0.800	0.802	0.821
90% CI	[0.71, 0.83]	[0.72, 0.84]	[0.72, 0.84]	[0.72, 0.85]	[0.73, 0.85]	[0.73, 0.86]	[0.75, 0.88]

Notes: This table shows estimates of MRPK, under the assumption of Cobb-Douglas production and CES demand. In this case we can compute $\text{MRPK}_i = \alpha\frac{\theta-1}{\theta}\text{APK}_i$. We calibrate $\alpha = \frac{1}{3}$ and $\theta = 3$. All confidence intervals are clustered at the firm level. We exclude MRPK estimates from the first wave, in order to maintain comparability with our IV estimates, which were estimated using an instrument that only varies after the first wave. Panel A shows estimates from the unconditional distribution of MRPK; the variance estimates are residualized on wave fixed effects. Panels B and C show estimates from a regression of the MRPK on the covariate, with wave fixed effects, as described in Equation 39. Panels D and E shows estimates from Equation 39 using the first K standardized principal components as covariates. Panels B through E include confidence intervals based on Algorithm 1. Where the confidence intervals have an upper bound of infinity, this indicates that the largest null tested (2 for Panel B and 5 for Panel C) could not be rejected. Note that Panels C and E use a mean MRPK, $\mathbb{E}[\text{MRPK}_i]$, that is computed for the subset of firms with no missing covariates, and is thus slightly different from the $\mathbb{E}[\text{MRPK}_i]$ in Panel A.

Table 11: Estimates of Persistence of MRPK Differences by Baseline APK

	(1)	(2)
β	0.086	
	(0.029)	
$\beta_{\text{First Year}}$		0.104
		(0.034)
$\beta_{\text{Second Year}}$		0.061
		(0.032)
$\gamma_{\text{First Year}}$	0.133	0.127
	(0.066)	(0.067)
$\gamma_{\text{Second Year}}$	0.142	0.115
	(0.053)	(0.061)

Notes: This table shows estimates from a heterogeneous model of MRPK, where we allow the heterogeneity in MRPK to vary by year. Standard errors are clustered at the firm level. We estimate the model:

$$\text{Profit}_{it} = \beta k_{it} + \gamma_{\text{First Year}} X_i \times \mathbf{1}_{t\leq 5} \times k_{it} + \gamma_{\text{Second Year}} X_i \times \mathbf{1}_{t>5} \times k_{it} + \alpha_i + \delta_t + \delta_t^{X\prime} X_i + \varepsilon_{it}$$

where X_i is APK measured at baseline and demeaned (this allows us to interpret β as $\mathbb{E}[\text{MRPK}_i]$), $\mathbf{1}_{t\leq 5}$ is an indicator equal to one in the first five waves, and $\mathbf{1}_{t>5}$ is an indicator equal to one in the last four waves. Capital and its interaction effects are instrumented for using the grant amount, Z_{it}, and the corresponding interactions. Each column shows estimates from the model described above: the first column shows estimates that assume a common average MRPK for both years, while the second column shows estimates that allow the average MRPK to be different in the first year and second year.

Table 12: Estimates of Variance of MRPK: Carrillo et al. (2023) Method

	(1)	(2)	(3)
$\mathbb{E}[\text{MRPK}_i]$	0.131	0.129	0.127
	(0.044)	(0.046)	(0.041)
$\mathbb{E}\left[(\text{MRPK}_i)^2\right]$	0.110	0.241	0.209
	(0.236)	(0.292)	(0.249)
$\text{Var}(\text{MRPK}_i)$	0.093	0.224	0.193
	(0.234)	(0.290)	(0.245)
Controls:			
$\mathbb{E}[\text{Amount}_{it} \mid t]$	Yes	Yes	—
$\mathbb{E}\left[(\text{Amount}_{it})^2 \mid t\right]$	No	Yes	—
Wave Fixed Effects	No	No	Yes

Notes: This table shows estimates of the mean and variance of MRPK, using the approach proposed in Carrillo et al. (2023), based on the IV-CRC model studied in Masten and Torgovitsky (2016). Standard errors are based on a firm-clustered bootstrap with 100 bootstrap draws. The first column controls for the expected value of the instrument (amount of grant) in that wave. The second column add in a control for the expected value of the instrument squared in that wave, while the third column controls for wave fixed effects.

Table 13: Parametric Estimates of Returns to Scale

	(1)	(2)	(3)	(4)
β	0.381	1.050	0.290	0.438
	(0.122)	(0.311)	(0.293)	(0.471)
Weighted	No	No	Yes	Yes

Notes: This table shows estimates of returns to scale based on loglinear models. Standard errors are clustered at the firm level. In the first and third columns, we estimate the model:

$$\log \text{Profit}_{it} = \beta \log k_{it} + \alpha_i + \delta_t + \varepsilon_{it}$$

while in the second and fourth columns, we estimate the model:

$$\log \text{Profit}_{it} = \beta \cdot \left(\frac{1}{3} \log k_{it} + \frac{2}{3} \log l_{it} \right) + \alpha_i + \delta_t + \varepsilon_{it}$$

where k_{it} is capital and l_{it} is total hours worked at the firm in the past week, summing up the hours of the owner, their family members, and any other workers. In either model, we use the grant amount, Z_{it}, as an instrument for the endogenous variable. The first and second columns are unweighted, while the third and fourth columns use the firms profits at baseline as weights.