Measuring Misallocation with Experiments*

(Job Market Paper)

David Hughes[†] Jeremy Majerovitz[‡]

October 11, 2023

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, and we show that standard methods (e.g. the delta method or projection) fail in this setting. 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, Jacob Moscona, Michael Peters, Karthik Sastry, Pari Sastry, Yongs Shin, Rob Townsend, and to seminar participants at MIT and the Federal Reserve Bank of St. Louis for helpful comments. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

[†]Department of Economics, Boston College; Email: dw.hughes@bc.edu

[‡]Federal Reserve Bank of St. Louis; 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 underly 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 infer the cost of misallocation from the data, we need to be able to measure the dispersion of marginal products and to aggregate that dispersion into an implied loss in output.

We show that, for arbitrary firm production functions, misallocation is a function of the variance of marginal products and an elasticity that depends on returns to scale and the slope of the demand curve. Our aggregation result is a non-parametric counterpart to more parametric results in the existing literature, and suggests that these parametric assumptions were not crucial for aggregation.

We then show how to estimate the variance of marginal products, using experimental variation to measure marginal products directly. We exploit a randomized controlled trial by de Mel et al. (2008) that randomly assigned grants to microentrepreneurs in Sri Lanka. We estimate heterogeneity in returns to capital given baseline covariates, which provides us with a lower bound on the total variance of the marginal revenue product of capital. By directly estimating marginal products using (experimentally induced) changes in capital, we sidestep the need for restrictive assumptions about production functions.

The variance of expected returns is a highly nonlinear function of the parameters from a linear IV model, and we find that traditional methods fail to provide accurate inference for this function. We thus develop a new econometric tools to construct uniformly valid confidence intervals for nonlinear functions of parameters. In a simulation calibrated to the data, we show that whereas traditional methods fail, our method delivers correct coverage across a range of true parameters.

We find substantial dispersion in the marginal revenue product of capital among Sri Lankan microentrepreneurs. We estimate that the variance of the log marginal revenue product of capital is 93 log points, with the 90% confidence interval ruling out values below 20 log points, and the 95% confidence interval ruling out values below 16 log points. 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 variables 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.

Methodology: Three Steps from Data to the Cost of Misallocation

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, connecting misallocation to the variance of the log marginal revenue product of capital (MRPK). We then show how to measure marginal products using an RCT, and show how to estimate heterogeneous returns to capital by baseline characteristics. This provides a lower bound on the total variance of log MRPK, which is a nonlinear function of the parameters of the linear IV model. Finally, since standard methods cannot be used to conduct correct inference on this object, we provide new econometric tools to construct uniformly valid confidence intervals for nonlinear functions of parameters.

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 hone 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 under CES aggregation the cost of misallocation is given by

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

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 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 VMPK and MRPK will differ if demand is downward-sloping: 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 price elasticity of demand is constant for all firms, and so $\mathrm{MRPK}_i = \frac{\theta-1}{\theta} \cdot \mathrm{VMPK}_i$, where θ is the CES parameter. Since this multiplier is the same across all firms, 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

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

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 (succinctly, $\mathrm{Var}\left(\mathrm{MRPK}_i\right) = \mathrm{Var}\left(\mathbb{E}\left[\mathrm{MRPK}_i \mid X_i\right]\right) + \mathbb{E}\left[\mathrm{Var}\left(\mathrm{MRPK}_i \mid X_i\right)\right]$). Thus, the variance of the conditional average treatment effects provides an estimatable 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 orthonornmal 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 projection method and the delta method. The projection method — construct a confidence set for the parameters and then project this confidence set to create a confidence interval for the function — will in general yield confidence intervals that are too large. The delta method in principle would yield confidence intervals with correct size asymptotically, but the delta method requires that the derivative of the function be finite and non-zero. However, the function we study 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. This makes the delta method fail at these points. More broadly a high degree of nonlinearity will make the delta method perform poorly. Our simulations suggest that the projection method is extremely conservative, while the delta method either rejects too often or not enough, depending on parameters.

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 simulation 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 8.2% and a variance of log MRPK of 135 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 higlights 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. 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

[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.

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 in Terms of 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 products 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

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

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 loglinear production functions with lognormally 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, 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.[3] We will focus on single product firms, and we will consider a single input (we call this input capital) unless otherwise noted.

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

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

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

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

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 mathemtically identical. We will assume that the individual production functions, as well as the aggregator, are smooth.

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

[3]Different network structures of production will in general imply different levels of misallocation (see Baqaee and Farhi, 2020). We focus on horizontal economies because these are the benchmark economy in the literature, and the simplest economy that allows for multiple firms that produce consumption goods. The other simplest model, a vertical economy, is unappealing because it cannot capture misallocation of inputs.

capital as:

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

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.[4] 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 constraint:

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

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

[4] 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.

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 efficient first-order condition of 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 efficient 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.[5] 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 firm's *efficient* first-order condition, rather than of the firm's profit-maximizing first-order condition. If firms charge markups, then that 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 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

With our economy fully specified, we can now characterize how deviations from the efficient, wedgeless economy affect welfare. We will first show that in a special case, which is the leading model in the literature on misallocation, the cost of misallocation is exactly equal to the variance of the log wedges, times one half times the elasticity of firm-level output with respect to the wedge. This elasticity is a function of the elasticity of demand and of the decreasing returns to scale of the firm. Results of this form are well-known for this special case, but we will show that they apply much more generally. We will show how to characterize misallocation in horizontal economies for arbitrary smooth production functions, without distributional assumptions about wedges or productivity. We will find that CES demand is sufficient to ensure that the same result we derived in the special case

[5] 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).

is in fact valid as a second-order approximation in the general case. This will put us on firm theoretical footing: our strategy to measure the variance of log MRPK will not require any assumptions about production, and thus neither should our aggregation results.

Special Case: CES-Loglinear-Lognormal. We will begin by focusing on a special case. Consider a horizontal economy with 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. We will specialize to a loglinear production function

$$\log y_i = \log z_i + \alpha \log k_i \tag{9}$$

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

We will define aggregate productivity as

$$\log Z := \log Y - \alpha \log K \tag{10}$$

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 loglinearity of the setup, we obtain the following formula through some manipulations:

$$\log Z = \mathbb{E}\left[\log z_i\right] - \frac{1}{2}\mathcal{E}\mathrm{Var}\left(\log \mu_i\right) + \frac{1}{2}\mathcal{E}\frac{1}{\alpha^2} \cdot \mathrm{Var}\left(\log z_i\right) - \frac{1}{2}\frac{1}{\alpha}\mathrm{Var}\left(\log z_i\right) \tag{11}$$

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 first aggregation result:

Proposition 1 (Exact Formula for the CES-Loglinear-Lognormal Case)**.** *Consider a horizontal economy with CES aggregation, loglinear production with a homogeneous elasticity of output with respect to capital, and lognormally distributed productivity and wedges. The cost* *of misallocation is given by*

$$\underbrace{\log Z^* - \log Z}_{\text{Losses from Capital Misallocation}} = \frac{1}{2}\mathcal{E} \cdot Var(\log \mu_i) \tag{12}$$

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

Misallocation depends on the variance of log wedges, and on the elasticity of output with respect to the wedge. Equivalently, since $\log \mathrm{VMPK}_i = \log r + \log \mu_i$, 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 1 further tightens the connection between dispersion in VMPK and misallocation, providing us with the relevant moment of the VMPK distribution (the variance of log VMPK) and the formula to map that moment to the cost of misallocation.

General Case: Horizontal Economies. The result in Proposition 1 is exact, but it relies on strong simplifying assumptions: in particular, CES demand, a loglinear production function with homogeneous α, and a lognormal distribution of z_i and μ_i. The goal of this paper is to measure misallocation without these simplifying assumptions, to the extent possible. Our next result thus generalizes the special case to allow arbitrary firm-level production functions and distributions of wedges, as a second-order approximation to the cost of misallocation. We will still require that the aggregator be CES: this aggregator is the standard in the literature, and ensures that the demand for each firm's output can be expressed as a loglinear function of the firm's price, p_i, and aggregate output, Y. Normalizing the price of the final good, P, to one, we have:

$$\log y_i = -\theta \log p_i + \log Y \tag{13}$$

We can combine the firm's production function, the firm's first order condition (Equation 7), and the firm's demand curve (Equation 13) to obtain an equation that characterizes the firm's behavior on the margin.

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

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

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* $\phi_i := \frac{y_i \cdot f_i''}{\left(f_i'\right)^2}$ *is the firm-specific elasticity of MPK with respect to output.*

This lemma summmarizes the solution to the firm's problem, and generalizes our earlier (parametric) notion of $\mathcal{E}$ to its firm-specific, non-parametric counterpart. The negative elasticity of output with respect to the wedge, $\mathcal{E}_i$, depends the elasticity of the firm's demand curve (governed by θ), and on physical returns to scale (governed by ϕ_i). 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.

We can combine this lemma with the input market clearing condition and the aggregator to solve for the 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.[6] In the spirit of Baqaee and Farhi (2020), we next derive how changes in wedges affect aggregate output, Y.

Proposition 2 (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*

$$d \log Y = -\mathbb{E}\left[\mathcal{E}_i \lambda_i \hat{\mu} \cdot d \log \mu_i\right] \tag{14}$$

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.

[6] 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).

Define $\log \check{\mu}(t) = t \cdot \log \mu$. With some abuse of notation, we have

$$\begin{aligned}
\mathcal{L} &= -\int_0^1 \frac{d \log Y(\check{\mu}(t))}{d \log \mu} \cdot \frac{d \log \check{\mu}(t)}{dt} dt \\
&= -\int_0^1 \mathbb{E}\left[\frac{d \log Y(\check{\mu}(t))}{d \log \mu_i} \cdot \frac{d \log \check{\mu}_i(t)}{dt}\right] dt \\
&= -\mathbb{E}\left[\left(\int_0^1 \frac{d \log Y(\check{\mu}(t))}{d \log \mu_i} dt\right) \log \mu_i\right]
\end{aligned} \tag{15}$$

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:

$$\mathcal{L} \approx -\frac{1}{2} \cdot \mathbb{E}\left[\frac{d \log Y}{d \log \mu_i} \log \mu_i\right] \tag{16}$$

This leads to our main aggregation result.

Proposition 3 (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*

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

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 $\mathcal{E}_i$.

Thus, our exact result for the CES-loglinear-lognormal case extends as an approximate result for the more general CES case. The cost of misallocation can be measured as a function of the (weighted) variance of log VMPK. Note also that this weighted variance formula can be interpreted as the sum of Harberger triangles, in the spirit of a long literature dating back to Harberger (1954).[7]

In practice, we will measure the unweighted variance of log VMPK, rather than the weighted variance. Measuring the weighted variance would require observing the weights, which in the general case would require observing $\mathcal{E}_i$, which is not feasible in practice. However, under appropriate statistical assumptions about the joint distribution of the weights

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

and wedges, the weighted and unweighted variance of wedges will coincide (this was the case in the lognormal special case).[8] 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.

Selecting $\mathcal{E}$. Although we will focus on measuring $\mathrm{Var}\left(\log \mathrm{VMPK}_i\right)$, the elasticity $\mathcal{E}$ is also an important input into our formula for misallocation. In principle, this parameter can also be estimated, and the literature contains estimates of both the elasticity of substitution across goods θ and of the returns to scale in firm production. We will select this elasticity through calibration, based on standard values for CES demand and for the capital share. Following the calibration in Hsieh and Klenow (2009), we will focus on $\theta = 3$ as our value for the CES parameter; this is a relatively conservative calibration, and larger values of θ would imply larger levels of misallocation.

We consider two values of α. One calibration is $\alpha = \frac{1}{3}$, matching the capital share. This calibration corresponds to a thought experiment in which only capital can be reallocated, and results in a relatively elasticity $\mathcal{E} = \frac{3}{7}$. We also consider $\alpha = 1$, which implies constant-returns-to-scale production. This corresponds to a thought experiment in which all inputs can be reallocated, rather than just capital. Under $\alpha = 1$, we get a much higher elasticity, $\mathcal{E} = \theta = 3$.

The gains from optimally reallocating all inputs will in general be much larger than the gains from reallocating capital only. This is because reallocating capital only quickly runs into diminishing returns on the production side, while the benefits from reallocating all inputs are held back only by downward sloping demand. Implicitly, the thought experiment in which we reallocate all inputs assumes that the variance of log VMPK also captures the distortions on other inputs. This will be exactly true in a case where wedges are on revenue; this is the case we will focus on in our results. If wedges vary across inputs, then a precise result would require measuring wedges for various inputs and aggregating appropriately.

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

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

[8] When log wedges are normally distributed, it will in general be more efficient to estimate an unweighted variance than a weighted variance.

will be proportional to VMPK, because the price elasticity of demand is constant and the same across all goods. We have:

$$\begin{aligned} \text{MRPK}_i &= p_i \frac{dy_i}{dk_i} + y_i \frac{dp_i}{dy_i} \cdot \frac{dy_i}{dk_i} \\ &= \left(1 + \frac{d\log p_i}{d\log y_i}\right) \cdot p_i \frac{dy_i}{dk_i} \\ &= \frac{\theta - 1}{\theta} \cdot \text{VMPK}_i \end{aligned} \tag{17}$$

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

We summarize this the following proposition.

Proposition 4 (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,*

$$\begin{aligned} Var(\log \mu_i) &= Var(\log VMPK_i) \\ &= Var(\log MRPK_i) \end{aligned}$$

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 efficient 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.

3 Measuring Heterogeneous Marginal Products with an IV Regression

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 $\mathrm{Var}\left(\mathbb{E}\left[\mathrm{MRPK}_i \mid X_i\right]\right) = \gamma'\gamma$, 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 $\frac{\mathrm{SD}(\mathbb{E}[\mathrm{MRPK}_i|X_i])}{\mathbb{E}[\mathrm{MRPK}_i]} = \frac{\sqrt{\gamma'\gamma}}{\beta}$, where β is the coefficient on capital. Using the log-normal approximation, this yields the formula $\mathrm{Var}\left(\log \mathbb{E}\left[\mathrm{MRPK}_i \mid X_i\right]\right) = \log\left(1 + \frac{\gamma'\gamma}{\beta^2}\right)$, 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 be exogenous (e.g. it cannot be correlated with productivity), to affect capital, 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 microenteprises, 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 [9], in the amount of 10,000 or 20,000 rupees. Importantly, the rollout of the treatment was staggered: in the first wave, 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 an incentive 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:

$$p_i y_i = \mathrm{MRPK}_i \cdot k_i + \mathrm{MRPL}_i \cdot l_i + \mathrm{MRPM}_i \cdot m_i \tag{18}$$

$$\implies p_i y_i - w l_i - c m_i = \mathrm{MRPK}_i \cdot k_i + (\mathrm{MRPL}_i - w) \cdot l_i + (\mathrm{MRPM}_i - c) \cdot m_i \tag{19}$$

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.[10] In accounting terms, we suspect

[9]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.

[10]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.

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. We 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. For labor, we will instead rely on the fact that labor does not seem to respond much to the treatment. Thus, 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:

$$y_{it} = \beta k_{it} + \alpha_i + \delta_t + \varepsilon_{it} \tag{20}$$

where y_{it} is profits, 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:

$$y_{it} = \beta k_{it} + \gamma' X_i \times k_{it} + \alpha_i + \delta_t + \delta_t^{X\prime} X_i + \varepsilon_{it} \tag{21}$$

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}.[11]

Once we know the parameters of the above model, we can estimate the distribution of $\mathbb{E}[\text{MRPK}_i \mid X_i]$, the expected returns to capital given covariates X_i. For example, it is straightforward to compute the variance of expected returns to capital: $\text{Var}(\mathbb{E}[\text{MRPK}_i \mid X_i]) = \text{Var}(\gamma' X_i) = \gamma' \text{Var}(X_i)\gamma$. 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

$$\text{Var}(\text{MRPK}_i) = \text{Var}(\mathbb{E}[\text{MRPK}_i \mid X_i]) + \mathbb{E}[\text{Var}(\text{MRPK}_i) \mid X_i] \tag{22}$$

Since the expectation of the conditional variance, $\mathbb{E}[\text{Var}(\text{MRPK}_i) \mid X_i]$, 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 $\text{Var}(\log \text{MRPK}_i) = \log\left(1 + \frac{\text{Var}(\text{MRPK}_i)}{\mathbb{E}[\text{MRPK}_i]^2}\right)$. This formula is convenient because we can replace $\text{Var}(\text{MRPK}_i)$ with $\text{Var}(\mathbb{E}[\text{MRPK}_i \mid X_i])$, and still be sure that the formula gives us a lower bound on the total variance of log MRPK.[12]

[11] 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.

[12] An alternative approach would have been to compute $\text{Var}(\log \mathbb{E}[\text{MRPK}_i \mid X_i])$ 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 $\text{Var}(\log \mathbb{E}[\text{MRPK}_i \mid X_i])$, 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 very unstable when some firms have low predicted MRPK, and would be very sensitive to outliers in the X_i distribution.

3.3 Standardized Principal Components Turns $\mathbf{Var}\,(\log \mathbf{MRPK}_i)$ 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: $\mathrm{Var}\,(\mathbb{E}\,[\mathrm{MRPK}_i \mid X_i]) = \gamma'\mathrm{Var}\,(X_i)\,\gamma$. 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, $\mathrm{Var}\,(X_i)$ is simply an identity matrix, and so $\mathrm{Var}\,(\mathbb{E}\,[\mathrm{MRPK}_i \mid X_i]) = \gamma'\gamma$.

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.[13] 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:

$$\mathrm{Var}\,(\mathbb{E}\,[\mathrm{MRPK}_i \mid X_i]) = \gamma'\gamma \tag{23}$$

$$\frac{\mathrm{SD}\,(\mathbb{E}\,[\mathrm{MRPK}_i \mid X_i])}{\mathbb{E}\,[\mathrm{MRPK}_i]} = \frac{\sqrt{\gamma'\gamma}}{\beta} \tag{24}$$

Using the log-normal approximation, we get a formula for $\mathrm{Var}\,(\log \mathbb{E}\,[\mathrm{MRPK}_i \mid X_i])$, which also serves as a lower bound for $\mathrm{Var}\,(\log \mathrm{MRPK}_i)$.[14]

$$\mathrm{Var}\,(\log \mathbb{E}\,[\mathrm{MRPK}_i \mid X_i]) \approx \log\left(1 + \frac{\gamma'\gamma}{\beta^2}\right) \tag{25}$$

$$\mathrm{Var}\,(\log \mathrm{MRPK}_i) \approx \log\left(1 + \frac{\mathrm{Var}\,(\mathrm{MRPK}_i)}{\mathbb{E}\,[\mathrm{MRPK}_i]^2}\right) \geq \log\left(1 + \frac{\gamma'\gamma}{\beta^2}\right) \tag{26}$$

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

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

$$\text{Var}\left(\log \text{APK}_i\right) = \text{Var}\left(\log \text{VMPK}_i\right) + \text{Var}\left(\log \alpha_i\right) - 2 \cdot \text{Cov}\left(\log \alpha_i, \log \text{VMPK}_i\right) \tag{27}$$

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, $\text{VMPK}_i = \alpha z_i k_i^{\alpha-1}$, and $\text{APK}_i = \frac{1}{\alpha} \cdot \text{VMPK}_i - c/k_i = \frac{1}{\alpha} \cdot \text{VMPK}_i - c \cdot \left(\frac{\text{VMPK}_i}{\alpha z_i}\right)^{1/(1-\alpha)}$. 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.

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; 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.

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 $\mathrm{Var}\left(\mathbb{E}\left[\mathrm{MRPK}_i \mid X_i\right]\right)$, but only because it will actually lower the true variance of $\mathbb{E}\left[\mathrm{MRPK}_i \mid X_i\right]$. 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 is 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 $\mathrm{Var}\left(\mathbb{E}\left[\mathrm{MRPK}_{it} \mid \Omega_{t-1}\right]\right)$, 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 $\mathrm{Var}\left(\mathbb{E}\left[\mathrm{MRPK}_i \mid X_i\right]\right)$ will provide a lower bound on the misallocation-relevant variance of MRPK.

Prior work on misallocation (Asker et al., 2014; David and Venkateswaran, 2019) has also studied how adjustment frictions may lead to dispersion in MRPK. In some ways, this is a dynamic version of the above argument: a planner, limited to today's information, cannot avoid the fact that the firm may be hit with shocks after the investment that lead to dispersion in MRPK *ex post*, since adjusting capital after the fact is costly. This dispersion in marginal products is not necessarily misallocation, since a planner could not undo it.

However, theories in which dispersion in marginal products is driven by adjustment costs are unlikely to explain the variance in returns across firms with different covariates at baseline. If adjustment costs show up in the data as reduced profits, then the estimated expected return given baseline covariates should be the same across firms. These covariates are measured at baseline, and the instrument only affects investment in later waves. Thus, the heterogeneous MRPK we identify is based on information available at the time of investment, and thus would reflect misallocation. If adjustment costs are utility costs that do not show up in the data, then this could cause MRPK to differ across firms. However, we would expect these differences to die out over the course of a multi-year experiment. Moreover, 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.

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.

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, 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. Moreover, our method may have efficiency advantages in some settings (we attempted to implement their method in our setting, but found that the resulting confidence intervals were extremely wide, perhaps in part because the grant instrument only takes on three values).

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 Estimation and Inference for Nonlinear Functions of Parameters

Summary

In the previous section, we showed that we could express the variance of the log of expected MRPK 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. However, that result was a statement about population parameters: if we knew the true parameters of the linear IV model, then a nonlinear function of those parameters would tell us what we want to know about the distribution of MRPK across firms.

In this section, we show how to conduct valid inference on nonlinear functions of parameters. This provides the final piece of the puzzle for our methodology to measure the cost of misallocation using experimental data on firms.

We begin by reviewing the two most prominent methods for conducting inference on nonlinear functions of parameters: the projection method and the delta method. Unfortunately, both methods fail to deliver valid confidence intervals. Suppose that δ is the vector of true parameters from our IV model, with estimator $\hat{\delta}$, and our function of interest is $g(\delta)$. The projection method constructs a confidence interval for $g(\delta)$ by first constructing a confidence set for δ, and then including in the confidence interval for g every value of $g(\delta)$ where δ is in the confidence set. In general this method will be conservative: the confidence intervals will be wider than necessary, and nominal 95% confidence intervals will include the true parameter more than 95% of the time.

The delta method uses the fact that if $\hat{\delta}$ is asymptotically normal ($\sqrt{N} \cdot \left(\hat{\delta} - \delta\right) \xrightarrow{d} N(0, \Sigma)$), then $g\left(\hat{\delta}\right)$ is also asymptotically normal ($\sqrt{N} \cdot \left(g\left(\hat{\delta}\right) - g(\delta)\right) \xrightarrow{d} N(0, \nabla g' \Sigma \nabla g)$), as long as the derivative of g with respect to δ is finite and non-zero. In our setting however, this derivative will be zero whenever misallocation is zero (there is no predictable heterogeneity in marginal products), and will be infinite when the average returns to capital are zero. These points at which the derivative conditions fail mean that, in practice, tests based on the delta method perform poorly. In simulation, delta-method-based tests suffer from severe size distortions.

We thus build new econometric tools in order to construct uniformly valid confidence intervals for functions of Gaussian parameters. Our approach combines a standard test statistic — the inverse-variance-weighted distance between the unconstrained estimator and the constrained estimator — with a simulation-based approach for generating critical values. We show that our 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. This novel method allows us to construct reliable confidence intervals for the cost of misallocation.

4.1 Standard Methods Do Not Provide Correct Size

In order to understand how the delta method can fail to provide correct inference for certain functions of parameters, consider a simple case in which we estimate a two-dimensional parameter $\widehat{\delta} = (\widehat{\gamma}, \widehat{\beta})$ that is asymptotically normally distributed, i.e. $\sqrt{n}(\hat{\delta} - \delta) \xrightarrow{d} \mathcal{N}(0, \Sigma)$. First, suppose that we are interested in performing inference on γ^2. In this case, a simple Taylor expansion gives

$$\sqrt{n}(\widehat{\gamma}^2 - \gamma^2) = 2\gamma \cdot \sqrt{n}(\widehat{\gamma} - \gamma) + \frac{2}{\sqrt{n}}\left(\sqrt{n}(\widehat{\gamma} - \gamma)\right)^2,$$

which is the sum of two components, an asymptotically normal variable and an asymptotically chi-squared variable. Standard asymptotic analysis would ignore the second term, on the basis that it converges to zero as $n \to \infty$, and base inference on the first, normally distributed term. However, when γ is small, the second term can be equally important, or even dominate the first in finite samples. For $\gamma = 0$, the first term disappears altogether and $n(\widehat{\gamma}^2 - \gamma^2)$ is asymptotically chi-squared. The failure of the delta-method occurs here when the derivative of γ^2 is (close to) zero.

As a second example, consider inference for the ratio $\frac{\gamma}{\beta}$. In this case, the derivative of the ratio parameter with respect to (γ, β) diverges to infinity as β gets closer to zero and the delta method again fails since the remainder term in the linear approximation can be large, even in large samples. This is exactly the setting of weakly identified instrumental variables, a well known case in which standard asymptotic inference fails. Both of these cases are examples of a failure in uniform convergence, which implies the existence of certain parameter values for which the delta method approximation can be arbitrarily bad, even in large samples (see Kasy 2019 for a detailed discussion on uniformity issues with the delta method).

Our parameter of interest, $\theta = \sqrt{\gamma'\gamma}/\beta$ is an example of both of the above cases – we can expect the delta method to provide poor inference whenever $\gamma \approx 0$, so that there is limited heterogeneity in marginal products across firms, as well as when $\beta \approx 0$, so that average returns to capital are low. For similar reasons, the bootstrap will also provide poor coverage in these cases: the validity of the bootstrap typically relies on similar arguments to the delta method, and will fail where the delta method fails.

One simple solution to this problem is to construct confidence sets using the projection method. The projection confidence interval for θ contains all values of θ for which there exists a corresponding $\delta_0 = (\gamma_0', \beta_0)'$ for which we cannot reject the null hypothesis $H_0 : \delta = \delta_0$. In our setting, this would correspond to the confidence set

$$CI_{1-\alpha} = \{\theta = \sqrt{\gamma'\gamma}/\beta : (\delta - \widehat{\delta})'\widehat{\Sigma}^{-1}(\delta - \widehat{\delta}) \leq \chi^2_{1-\alpha}(\dim \delta)\}.$$

This method is in general conservative, particularly when the dimension of δ is large. For example, returning to our simple two-dimensional parameter $\widehat{\delta} = (\widehat{\gamma}, \widehat{\beta})$, suppose that $\Sigma = I_2$ so that the estimators are asymptotically uncorrelated with unit variance. A standard 95% confidence set for γ would be given by $(\widehat{\gamma} - 1.96, \widehat{\gamma} + 1.96)$, while the projection method would result in the interval $(\widehat{\gamma} - 2.45, \widehat{\gamma} + 2.45)$.[15] As an alternative, we propose a method that delivers asymptotically valid confidence sets that are robust to failures of the delta method,

[15] The value 2.45 is the square-root of the 95th percentile of a chi-squared distribution with two degrees of freedom.

without being conservative.

4.2 A Uniformly Valid 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 $\sqrt{n}(\hat{\delta} - \delta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$ uniformly, and are interested in testing the null hypothesis $H_0 : g(\delta) = \tau_0$, for some function g. Then, given some chosen test statistic $\widehat{T} = T(\widehat{\delta}, \tau_0)$ we can simulate its asymptotic distribution by drawing δ^* from $\mathcal{N}(\delta_0, \Sigma/n)$ and constructing the corresponding statistic $T(\delta^*, \tau_0)$. For a suitably chosen test statistic we will have uniform convergence of $T(\widehat{\delta}, \tau_0)$ to the simulated distribution. Quantiles of the simulated distribution can the be used to construct critical values for testing the null hypothesis.

The test statistic we propose is a measure of the distance between the unconstrained estimator $\widehat{\delta}$ and the constraint set $\{\delta : g(\delta) = \tau_0\}$, i.e. the set of δ that satisfy the null hypothesis. Specifically, we will use

$$T(\widehat{\delta}, \tau_0) = \min_{\delta : g(\delta) = \tau_0} n(\delta - \widehat{\delta})' \widehat{\Sigma}^{-1} (\delta - \widehat{\delta}) \tag{28}$$

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 $\widehat{\delta}$. Under conditions in which the delta method is in fact applicable, the statistic is asymptotically equivalent to the standard Wald statistic and converges to a chi-squared distribution with degrees of freedom equal to the rank of $D = \nabla g(\delta)$ (see for example, Newey and McFadden 1994). However, in finite sample settings where the delta method may work poorly, for example when D is either close to zero or unbounded, the distribution of the statistic is no longer well approximated by a chi-squared, but we may approximate it via simulation.

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 (28) 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 1. *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.*

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

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

Assumption 1 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 1,* $\widehat{p}(\delta_0)$ *converges uniformly in distribution to a uniform random variable* $\mathcal{U}$

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

Lemma 2 shows that critical values from the CDF F_{δ_0} could be used to construct a uniformly valid test. This is of course infeasible since δ_0 is not fully specified under the null hypothesis $H_0 : g(\delta_0) = \tau_0$. Here we discuss two feasible alternatives. The first method replaces the unknown δ_0 with a constrained estimator $\bar{\delta}$ and gives an test that valid although non-uniform test. The second method is based on finding the worst case δ under the null, and provides uniformly valid confidence sets.

4.2.1 An alternative test

Replacing δ_0 with a constrained parameter estimator $\bar{\delta}$, 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 that exceed the test statistic, i.e.

$$\widehat{p}(\bar{\delta}) = \frac{1}{B} \sum_b 1\big\{T(\delta^*_b, \tau_0) \geq T(\widehat{\delta}, \tau_0)\big\}.$$

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 constrained estimate is easy to compute and is available even in cases where the worst case δ is unknown. For example, we could find $\bar{\delta}$ by solving a constrained version of the original estimation procedure, or as the solution to

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

A description of the process for computing the confidence interval using this procedure is given below. The following proposition establishes asymptotic validity of the test.

Proposition 5. *Let Assumption 1 hold for $\delta = (\beta, \gamma)$. Suppose that we wish to test either the null hypothesis $H_0 : \sqrt{\gamma'\gamma} = \tau_0$ or $H_0 : \sqrt{\gamma'\gamma}/\beta = \tau_0$. Then for any $\delta_0 = (\beta_0, \gamma_0)$ with $\beta_0 \neq 0$, and for the number of simulation draws $B \to \infty$, we have that $\widehat{p}(\bar{\delta})$ converges in distribution to a uniform random variable as $n \to \infty$.*

The proposition establishes that the procedure is asymptotically correct for fixed values of δ_0; however, it is no longer uniformly valid. 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. When $\gamma_0 \neq 0$ the delta method is asymptotically valid for testing both of the null hypotheses in Proposition 5. In this case, the simulated distribution $F_{\bar{\delta}}(t)$ converges asymptotically to the chi-squared distribution with one degree of freedom under the null and so is asymptotically equivalent to the delta method. When $\gamma_0 = 0$ the delta method fails since the derivative of $\sqrt{\gamma'\gamma}$ with respect to γ is zero. In this setting we have $\bar{\gamma} = 0$ (the only value of γ consistent with the null hypothesis) and the simulated distribution $F_{\bar{\delta}}(t)$ converges asymptotically to the chi-squared distribution with $p = \dim(\gamma)$ degrees of freedom.

While the procedure is not uniformly valid over δ, it is likely to outperform the delta method when γ is small. When γ is close to zero the linear approximation used by the delta method provides a very poor approximation to the true distribution of the parameter of interest. In contrast, the difference between $F_{\bar{\delta}}$ and F_{δ_0} is likely to introduce much less size distortion. In simulations calibrated to our empirical setting we show that the test has good size control across a range of parameter settings, even where the delta method performs poorly.

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 $\delta_b \sim N(\bar{\delta}, \bar{\Sigma})$, 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

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

4.2.2 A feasible and uniformly valid procedure

In order to construct confidence sets with uniform size control, we must choose critical values that provide correct coverage for all δ satisfying the null hypothesis. The following proposition establishes uniform validity of confidence sets constructed using critical values that are based on the 'worst case' distribution of the statistic under the null hypothesis.

Proposition 6. *Let Assumption 1 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

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

The critical values used in Proposition 6 are feasible to compute since they depend only on the hypothesized value for τ. However, 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.

Remark 1. The methods proposed here are distinct from an alternative simulation based approach that simulates $\delta^* \sim \mathcal{N}(\widehat{\delta}, \widehat{\Sigma})$ and then constructs the corresponding distribution for $\tau^* = g(\delta^*)$. The confidence set for τ is then taken as the $\alpha/2$ and $(1-\alpha/2)$ quantiles of this distribution. Although straightforward, and perhaps deceptively intuitive, this approach does not deliver valid confidence sets in many settings, as highlighted by Ham and Woutersen (2013).[16] 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.

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 it 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.

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 $\Sigma = I$. 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.

Conjecture 1. *Let Σ_γ be the variance matrix associated with $\widehat{\gamma}$, and let $\Sigma_\gamma = VDV'$ be its eigendecomposition, 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*

$$\gamma_{worst} = \tau_0 v_p \tag{29}$$

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 $\gamma_b^* \sim \mathcal{N}(\gamma_{worst}, \widehat{\Sigma})$ for $b = 1, \ldots, B$, and computing the

[16] 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 produces conservative coverage levels. They also propose an adjustment based on linear approximation to the function g that can reduce conservativeness of the interval.

corresponding test statistic $T(\gamma_b^*, \tau_0)$. The p-value for the test would then be given by

$$\widehat{p}_{\tau_0} = \frac{1}{B}\sum_b 1\{T(\gamma_b^*, \tau_0) \geq T(\widehat{\gamma}, \tau_0)\}.$$

A $(1-\alpha)$-level confidence set for $\tau = \sqrt{\gamma'\gamma}$ is then given by $\widehat{C}_{1-\alpha} = (\tau_{min}, \infty)$, where $\tau_{min} = \min_\tau\{\tau : \widehat{p}_\tau \geq \alpha\}$.

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 construct 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 (β_0, S_0) that cannot be rejected. We summarize this process in Algorithm 2.

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 (29), 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*

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

4.3 Simulation evidence

Here we provide the results of simulations that are calibrated to our empirical setting. The simulated treatment D_{it} and outcome Y_{it} are generated from

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

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 adjusted so that

$$\begin{aligned} \sqrt{\gamma'\gamma} &= \{0.1, 0.01\}, \\ \beta &= \{0.1, 0.03\}. \end{aligned}$$

These settings 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. I rescale β and γ to match the above settings, keeping all other parts of the model the same. Three tests are compared:

1. W_1 - the Wald statistic for $H_0 : \frac{\sqrt{\gamma'\gamma}}{\beta} = \theta_0$

$$W_1 = n\frac{(\frac{\sqrt{\widehat{\gamma}'\widehat{\gamma}}}{\widehat{\beta}} - \theta_0)^2}{\widehat{A}'\widehat{\Sigma}\widehat{A}}, \qquad \widehat{A}' = (-\frac{\sqrt{\widehat{\gamma}'\widehat{\gamma}}}{\widehat{\beta}^2}, \frac{\widehat{\gamma}'}{\widehat{\beta}\sqrt{\widehat{\gamma}'\widehat{\gamma}}})$$

2. W_2 - the Wald statistic for $H_0 : \sqrt{\gamma'\gamma} = \beta\theta_0$

$$W_2 = n\frac{(\sqrt{\widehat{\gamma}'\widehat{\gamma}} - \widehat{\beta}\theta_0)^2}{\widehat{D}'\widehat{\Sigma}\widehat{D}}, \qquad \widehat{D}' = (-\theta_0, \frac{\widehat{\gamma}'}{\sqrt{\widehat{\gamma}'\widehat{\gamma}}})$$

3. T - the proposed simulation based test

Table 1: 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 $H_0 : \frac{\sqrt{\gamma'\gamma}}{\beta} = \theta_0$. The columns labeled W_2 correspond to the Wald statistic for the null hypothesis $H_0 : \sqrt{\gamma'\gamma} = \beta\theta_0$. The columns labeled T correspond to our proposed simulation-based test.

The two Wald statistics are compared to the critical value from a chi-squared distribution with one degree of freedom. We perform 1000 simulations, and use 1000 simulations to compute the critical value for the simulated test statistic.

Table 1 reports the rejection rates under the null hypothesis for each of the four simulation settings. 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 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.

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 120 log points. Using our novel econometric tools, we find that 90% confidence interval rule out values below roughly 50 log points, while 95% confidence intervals rule out values below roughly 40 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 capital, $\alpha = \frac{1}{3}$ and $\alpha = 1$. 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; it implicitly assumes that different inputs face the same wedges. 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

Before we estimate the variance of log 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 that cluster at the firm level.

We begin by estimating the homogeneous linear IV model in Equation 20, replicating the original paper by 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.[17] The results are in Table 2, in the first column of Panel A. The homogeneous linear IV model yields an average *monthly* return to capital of 6%.

In the other columns of Panel A, we estimate the heterogeneous linear IV model in Equation 21. Each column uses a single covariate for X_i. We compute $\mathbb{E}\left[\text{MRPK}_i\right]$ as $\hat{\beta} + \hat{\gamma}\mathbb{E}\left[X_i\right]$; our estimates range from 6-8% monthly returns across specifications. In the second row, we compute $\gamma \cdot \text{SD}\left(X_i\right)$, where γ is the interaction effect between the covariate and capital. This represents how a one standard deviation change in the covariate affects the MRPK. Moreover, this is equal to the standard deviation of expected MRPK, or $\text{SD}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)$, up to sign. This varies substantially across covariates, although the average product of capital at baseline appears to be the strongest predictor of MRPK, and it is a statistically significant predictor. This is somewhat expected: under a homogeneous log-linear production function (Cobb-Douglas), the MRPK is proportional to APK. However, the fact that the APK in wave 1 is a useful predictor of the MRPK in later waves also suggests that some component of the firm's MRPK is persistent over time.

[17] 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.

In Panel B, we compute the same statistics 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 $\theta = 3$, following Hsieh and Klenow (2009).[18] The first row shows $\mathbb{E}[\text{MRPK}_i]$, which we compute as the sample mean.

For each covariate, we then estimate a regression analogous to our IV analysis:

$$\text{MRPK}_{it} = \gamma' X_i + \delta_t + \varepsilon_{it} \tag{30}$$

where δ_t is a wave fixed effect. To estimate this regression, we exclude data from the first wave. We do this to make our estimates correspond more closely to the variation in 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 MRPK in later waves. 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 30, we compute $\gamma \cdot \text{SD}(X_i)$, as in Panel A. Unsurprisingly, our estimates in Panel B are more precise than those in Panel A. Under the strong assumption of Cobb-Douglas production, MRPK can be observed directly rather than estimated, yielding smaller standard errors. In part due to the large standard errors in Panel A, we cannot reject equality between any cell in Panel A and the corresponding cell in Panel B. At the same time, we also cannot rule out large differences. For example, in the first column, the 95% confidence interval for $\mathbb{E}[\text{MRPK}_i]$ includes the 8.2% average monthly return implied by our Cobb-Douglas calibration, but it also includes values as high as 10.8% and as low as 1.4%.

Finally, in the third row of Panel A and the third row of Panel B, we show the implied estimate of $\frac{\text{SD}(\mathbb{E}[\text{MRPK}_i|X_i])}{\mathbb{E}[\text{MRPK}_i]}$, using only the single covariate as our estimate. We also provide the lower bound of the 90% confidence interval, computed using our algorithm from Section 4. 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. Instead,

[18] 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 here 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.

Table 2: Estimates of Heterogeneous MRPK by Baseline Covariates

Panel A: IV Estimates

$\mathbb{E}[\text{MRPK}_i]$	0.061	0.062	0.061	0.060	0.063	0.072	0.085	0.069
	(0.024)	(0.071)	(0.063)	(0.058)	(0.134)	(0.034)	(0.032)	(0.032)
$\gamma \cdot \text{SD}(X_i)$	—	-0.070	0.018	0.044	-0.011	-0.023	0.128	0.052
		(0.263)	(0.128)	(0.135)	(0.372)	(0.029)	(0.060)	(0.036)
90% CI Lower Bound		0.028	0.000	0.019	0.000	0.000	0.064	0.024
$\frac{\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])}{\mathbb{E}[\text{MRPK}_i]}$	—	1.121	0.300	0.723	0.171	0.314	1.505	0.747
90% CI Lower Bound		0.405	0.000	0.302	0.000	0.000	0.835	0.383
Covariate (X_i)	None	Capital	Age	Education	Profit	Hours	APK	$\log(\text{APK})$

Panel B: Cobb-Douglas

$\mathbb{E}[\text{MRPK}_i]$	0.082	0.082	0.082	0.082	0.082	0.082	0.082	0.082
$\gamma \cdot \text{SD}(X_i)$	—	-0.044	0.015	0.029	-0.005	-0.023	0.052	0.060
		(0.028)	(0.013)	(0.013)	(0.008)	(0.019)	(0.021)	(0.023)
$\frac{\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])}{\mathbb{E}[\text{MRPK}_i]}$	—	0.540	0.177	0.351	0.064	0.281	0.633	0.731
Covariate (X_i)	None	Capital	Age	Education	Profit	Hours	APK	$\log(\text{APK})$

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, the first column shows estimates from Equation 20; a homogeneous model without covariates. The other columns show estimates from the heterogeneous model described in Equation 21; these columns each use one covariate, which is measured at baseline. The first row shows the $\mathbb{E}[\text{MRPK}_i]$ implied by these models, which is computed as $\hat{\beta} + \hat{\gamma}\mathbb{E}[X_i]$. The second row shows estimates of $\gamma \cdot \text{SD}(X_i)$, which represents how MRPK changes when the covariate X_i increases by one standard deviation. Note that this is also equal to the standard deviation of the conditional average MRPK, or $\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])$, up to sign. The third row shows the implied estimate of $\frac{\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])}{\mathbb{E}[\text{MRPK}_i]}$, and the lower bound of the 90% confidence interval, computed using our algorithm. Panel B shows corresponding estimates under the assumption of Cobb-Douglas production and CES demand, in which case we can compute $\text{MRPK}_i = \alpha \frac{\theta-1}{\theta} \text{APK}_i$. We calibrate $\alpha = \frac{1}{3}$ and $\theta = 3$. The first row shows $\mathbb{E}[\text{MRPK}_i]$, which we compute simply as the sample mean. The second row shows estimates of $\gamma \cdot \text{SD}(X_i)$, coming from a regression of the MRPK on the covariate, with wave fixed effects, as described in Equation 30. We estimate this regression excluding the first wave, reflecting the fact that grants were not distributed until after the first wave. The third row shows the implied estimate of $\frac{\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])}{\mathbb{E}[\text{MRPK}_i]}$.

these estimates serve to highlight the importance of selecting the correct covariates, and/or incorporating multiple covariates in order to get a more precise estimate of misallocation. Although the estimates based on APK, most of these single-covariate confidence intervals cannot rule out zero misallocation.

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

We now turn to implement our methodology, described in Sections 3 and 4, for estimating the variance of log MRPK. We estimate Equation 21, now using standardized principal components as our covariates X_i. Our results are in Table 3. We focus on estimates of $\frac{\text{SD}(\mathbb{E}[\text{MRPK}_i|X_i])}{\mathbb{E}[\text{MRPK}_i]}$, which is computed as $\frac{\sqrt{\gamma'\gamma}}{\beta}$. We will use this to obtain a lower bound on $\text{Var}\left(\log \text{MRPK}_i\right)$, and to provide estimates of the cost of misallocation. 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).

Panel A follows the same structure as the previous table. The first row shows estimates of $\mathbb{E}\left[\text{MRPK}_i\right]$, which is computed simply as $\hat{\beta}$. The estimates are similar to previous estimates, with average monthly returns ranging from 7-10%. The second row shows estimates of $\text{SD}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right)$, which is computed as $\sqrt{\hat{\gamma}'\hat{\gamma}}$. The point estimates range from a standard deviation of 6% to a standard deviation of 13%.

The third row shows our estimates of $\frac{\text{SD}(\mathbb{E}[\text{MRPK}_i|X_i])}{\mathbb{E}[\text{MRPK}_i]}$. The point estimates are computed as $\frac{\sqrt{\hat{\gamma}'\hat{\gamma}}}{\hat{\beta}}$. We also show the lower bound of confidence intervals based on our algorithm from the previous section. The point estimates are fairly high, and for $K > 2$ they are all above one, suggesting that the standard deviation of returns is larger than the mean. This implies very sizable dispersion: a firm that is one standard deviation below the mean has negative expected returns, according to these point estimates. Similarly, once we include more than two factors into the analysis, the lower bound of the confidence interval is fairly high: roughly 0.5 for most rows in the 90% confidence interval, and roughly 0.4 for most rows in the 95% confidence interval. These are more modest than the point estimates, but still sizable: a firm with returns two standard deviations below the mean would have near zero returns under these estimates.

In Panel B, we benchmark these results against estimates computed under the assumption of a Cobb-Douglas production function, as in 2. The first and second rows show results from estimating Equation 30, using the standardized principal components as covariates. We also report unconditional estimates of $\mathbb{E}\left[\text{MRPK}_i\right]$, $\frac{\text{SD}(\text{MRPK}_i)}{\mathbb{E}[\text{MRPK}_i]}$, and $\text{SD}\left(\log \text{MRPK}_i\right)$. For our unconditional estimates $\text{SD}\left(\text{MRPK}_i\right)$ and $\text{SD}\left(\log \text{MRPK}_i\right)$, we first partial out wave fixed

Table 3: Estimates of Variance of MRPK

Panel A: IV Estimates	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$
$\mathbb{E}[\text{MRPK}_i]=\beta$	0.073	0.075	0.077	0.084	0.080	0.105	0.105
	(0.026)	(0.086)	(0.118)	(0.154)	(0.219)	(0.235)	(0.185)
$\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])=\sqrt{\gamma'\gamma}$	0.066	0.063	0.109	0.107	0.098	0.131	0.128
90% CI Lower Bound	0.033	0.019	0.044	0.043	0.037	0.079	0.050
$\frac{\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])}{\mathbb{E}[\text{MRPK}_i]}=\frac{\sqrt{\gamma'\gamma}}{\beta}$	0.913	0.840	1.415	1.275	1.234	1.247	1.213
90% CI Lower Bound	0.456	0.210	0.557	0.519	0.470	0.784	0.555

Panel B: Cobb-Douglas	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$
$\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])$	0.061	0.061	0.062	0.062	0.062	0.062	0.064
$\frac{\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])}{\mathbb{E}[\text{MRPK}_i]}$	0.745	0.751	0.755	0.759	0.760	0.762	0.780

Unconditional Estimates	
$\mathbb{E}[\text{MRPK}_i]$	0.082
$\frac{\text{SD}(\text{MRPK}_i)}{\mathbb{E}[\text{MRPK}_i]}$	2.053
$\text{SD}(\log \text{MRPK}_i)$	1.161

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 21. Each column uses the first K principal components of our vector of covariates. The first row shows the $\mathbb{E}[\text{MRPK}_i]$ implied by these models, which is simply $\hat{\beta}$. The second row shows estimates of the standard deviation of the conditional average MRPK, or $\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])$. This is computed as $\sqrt{\hat{\gamma}'\hat{\gamma}}$. The third row shows the implied estimate of $\frac{\text{SD}(\mathbb{E}[\text{MRPK}_i|X_i])}{\mathbb{E}[\text{MRPK}_i]}$, and the lower bound of the 90% confidence interval, computed using our algorithm. Panel B shows corresponding estimates under the assumption of Cobb-Douglas production and CES demand, in which case we can compute $\text{MRPK}_i=\alpha\frac{\theta-1}{\theta}\text{APK}_i$. We calibrate $\alpha=\frac{1}{3}$ and $\theta=3$. The first row row shows estimates of $\text{SD}(\mathbb{E}[\text{MRPK}_i \mid X_i])$, coming from a regression of the MRPK on the covariates X_i, with wave fixed effects, as described in Equation 30. We estimate this regression excluding the first wave, reflecting the fact that grants were not distributed until after the first wave. The second row shows the implied estimate of $\frac{\text{SD}(\mathbb{E}[\text{MRPK}_i|X_i])}{\mathbb{E}[\text{MRPK}_i]}$. We also compute $\mathbb{E}[\text{MRPK}_i]$ as the sample mean, $\text{SD}(\text{MRPK}_i)$ as the sample standard deviation, and $\text{SD}(\log \text{MRPK}_i)$ as the sample standard deviation of log MRPK. To compute $\text{SD}(\text{MRPK}_i)$ and $\text{SD}(\log \text{MRPK}_i)$, we first partial out wave fixed effects, in order to focus on within-period misallocation.

Table 4: 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	—
$\mathrm{Var}\left(\log \mathrm{MRPK}_i\right)$	0.93	0.20	0.16	1.35
$\log Z^* - \log Z$ $(\mathcal{E} = \frac{3}{7})$	0.20	0.04	0.03	0.29
$Z^*/Z - 1$ $(\mathcal{E} = \frac{3}{7})$	0.22	0.04	0.04	0.34
$\log Z^* - \log Z$ $(\mathcal{E} = 3)$	1.39	0.30	0.24	2.02
$Z^*/Z - 1$ $(\mathcal{E} = 3)$	3.01	0.35	0.27	6.56

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 3. The second row provides a (lower bound) estimate of $\mathrm{Var}\left(\log \mathrm{MRPK}_i\right)$, 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 $\log Z^* - \log Z = \frac{1}{2}\mathcal{E} \cdot \mathrm{Var}\left(\log \mathrm{MRPK}_i\right)$. The third and fourth rows are calibrated to reflect the gains from optimally reallocating capital, while the fourth row is 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 3, our estimate of the variance of log MRPK is computed after partialling out wave fixed effects, in order to focus on within-period misallocation.

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: the standard deviation of MRPK is 16.8% if we control for wave fixed effects, and 16.9% if we do not.

5.3 Implied Estimates of the Cost of Misallocation

We next use our estimates of the variance of log MRPK to back out estimates of the cost of misallocation, using our earlier formula. We summarize the results of this exercise in Table 4. 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 $\mathrm{Var}\left(\log \mathrm{MRPK}_i\right)$, as discussed in Section 3. This gives a point estimate of 93 log points, while the lower bound of the 90% confidence interval is 22 log points, and the lower bound of the 95% confidence interval is 16 log points.

Our formula for misallocation tells us that we can measure the gains from optimally reallocating inputs, $\log Z^* - \log Z$, with the formula $\frac{1}{2}\mathcal{E} \cdot \mathrm{Var}\left(\log \mathrm{MRPK}_i\right)$. To obtain an appropriate value of $\mathcal{E}$, we will calibrate to standard values of θ and α. We use $\theta = 3$, reflecting a standard value in the misallocation literature (Hsieh and Klenow, 2009). We use two values of α. One calibration, $\alpha = \frac{1}{3}$, reflects a standard value of the capital share. We interpret this calibration as giving us the gains from optimally reallocating capital only.

An alternative calibration, $\alpha = 1$, reflects a constant-returns-to-scale production function. We interpret this calibration as giving us the gains from optimally reallocating all inputs, although we note that assuming constant returns to scale may somewhat overstate the scope for reallocation of inputs. These calibrations give an elasticity of output to the wedge of $\mathcal{E} = \frac{3}{7}$ and $\mathcal{E} = 3$, respectively.

Focusing on the point estimates, we find that optimally reallocating capital would increase output by 20 log points, or 22%. 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 small 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 125 log points. Under the elasticity $\mathcal{E} = 3$, this implies that optimally reallocating all inputs would increase output by 601%.

Our results do not imply a level of misallocation this large, but they also do not necessarily rule it out. First, we focus on the component of the variance of log MRPK that can be predicted with a set of baseline covariates, thus our estimates are a lower bound on the total variance, and thus a lower bound on misallocation. This is somewhat beneficial: we would not want to label unpredictable variation in MRPK as "misallocation." However, there may be some variation in returns that is predictable ex ante, but which is not captured by our seven covariates. Second, the confidence interval on our estimates is fairly wide: we cannot rule out high values of misallocation. Our results provide robust evidence for sizable misallocation in this setting, that does not depend on strong auxiliary assumptions about the production function. But they do not provide decisive evidence on whether a homogeneous Cobb-Douglas production function fits the data well.

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 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

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.

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.

Kasy, M. (2019): "Uniformity and the Delta Method," *Journal of Econometric Methods*, 8.

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.

Newey, W. K. and D. McFadden (1994): "Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, Elsevier, 2111–2245.

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 Omitted Proofs

A.1 Proof of Proposition 1

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

$$\implies p_i = Y^{\frac{1}{\theta}} y_i^{\frac{-1}{\theta}} \tag{31}$$

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:

$$\text{FOC: } \log p_i = \log \mu_i + \log r - \log \frac{dy_i}{dk_i} \tag{32}$$

$$\text{Demand: } \frac{1}{\theta}\log Y - \frac{1}{\theta}\log y_i = \log \mu_i + \log r - \log\left(\frac{d\log y_i}{d\log k_i}\cdot\frac{y_i}{k_i}\right) \tag{33}$$

$$= \log \mu_i + \log r - \log \alpha - \log y_i + \log k_i \tag{34}$$

$$= \log \mu_i + \log r - \log \alpha - \log y_i + \frac{1}{\alpha}\left(\log y_i - \log z_i\right) \tag{35}$$

$$\implies \left(\frac{1-\alpha}{\alpha} + \frac{1}{\theta}\right)\log y_i = \frac{1}{\alpha}\log z_i - \log \mu_i + \frac{1}{\theta}\log Y - \log r + \log \alpha \tag{36}$$

$$\implies \log y_i = \mathcal{E}\cdot\left[\frac{1}{\alpha}\log z_i - \log \mu_i + \underbrace{\frac{1}{\theta}\log Y - \log r + \log \alpha}_{C}\right] \tag{37}$$

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

$$\log Y = \frac{\theta}{\theta-1}\cdot\log \mathbb{E}\left[y_i^{\frac{\theta-1}{\theta}}\right] \tag{38}$$

$$= \frac{\theta}{\theta-1}\cdot\left(\frac{\theta-1}{\theta}\cdot\mathbb{E}\left[\log y_i\right] + \left(\frac{\theta-1}{\theta}\right)^2\cdot\frac{1}{2}\mathrm{Var}\left(\log y_i\right)\right) \tag{39}$$

$$= \mathbb{E}\left[\log y_i\right] + \left(\frac{\theta-1}{\theta}\right)\cdot\frac{1}{2}\mathrm{Var}\left(\log y_i\right) \tag{40}$$

and similarly

$$\log K := \log \mathbb{E}\left[k_i\right] \tag{41}$$

$$= \log \mathbb{E}\left[\left(\frac{y_i}{z_i}\right)^{\frac{1}{\alpha}}\right] \tag{42}$$

$$= \frac{1}{\alpha}\mathbb{E}\left[\log y_i - \log z_i\right] + \frac{1}{2}\frac{1}{\alpha^2}\mathrm{Var}\left(\log y_i - \log z_i\right) \tag{43}$$

$$= \frac{1}{\alpha}\mathbb{E}\left[\log y_i - \log z_i\right] + \frac{1}{2}\frac{1}{2\alpha^2}\mathrm{Var}\left(\log y_i\right)$$
$$+ \frac{1}{2}\frac{1}{\alpha^2}\mathrm{Var}\left(\log z_i\right) - \frac{1}{\alpha^2}\mathrm{Cov}\left(\log y_i, \log z_i\right) \tag{44}$$

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

$$\log Z := \log Y - \alpha \log K \tag{45}$$

$$= \mathbb{E}\left[\log y_i\right] + \left(\frac{\theta-1}{\theta}\right) \cdot \frac{1}{2}\mathrm{Var}\left(\log y_i\right)$$
$$- \mathbb{E}\left[\log y_i - \log z_i\right] - \frac{1}{2}\frac{1}{\alpha}\mathrm{Var}\left(\log y_i\right) - \frac{1}{2}\frac{1}{\alpha}\mathrm{Var}\left(\log z_i\right) + \frac{1}{\alpha}\mathrm{Cov}\left(\log y_i, \log z_i\right) \tag{46}$$

$$= \mathbb{E}\left[\log z_i\right] - \frac{1}{2}\frac{1}{\alpha}\mathrm{Var}\left(\log z_i\right)$$
$$+ \frac{1}{2} \cdot \left(\frac{\theta-1}{\theta} - \frac{1}{\alpha}\right)\mathrm{Var}\left(\log y_i\right) + \frac{1}{\alpha}\mathrm{Cov}\left(\log y_i, \log z_i\right) \tag{47}$$

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 $\left(\frac{\theta-1}{\theta} - \frac{1}{\alpha}\right) = -\mathcal{E}^{-1}$,[19] we have:

$$\frac{1}{2} \cdot \left(\frac{\theta-1}{\theta} - \frac{1}{\alpha}\right)\mathrm{Var}\left(\log y_i\right) - \frac{1}{\alpha}\mathrm{Cov}\left(\log y_i, \log z_i\right)$$
$$= -\frac{1}{2} \cdot \mathcal{E}^{-1}\mathrm{Var}\left(\mathcal{E} \cdot \left[\frac{1}{\alpha}\log z_i - \log \mu_i\right]\right) + \frac{1}{\alpha}\mathrm{Cov}\left(\mathcal{E} \cdot \left[\frac{1}{\alpha}\log z_i - \log \mu_i\right], \log z_i\right) \tag{48}$$

$$= -\frac{1}{2} \cdot \mathcal{E}\left(\frac{1}{\alpha^2}\mathrm{Var}\left(\log z_i\right) + \mathrm{Var}\left(\log \mu_i\right) - 2 \cdot \frac{1}{\alpha}\mathrm{Cov}\left(\log z_i, \log \mu_i\right)\right)$$
$$+ \frac{1}{\alpha}\mathcal{E}\left(\frac{1}{\alpha}\mathrm{Var}\left(\log z_i\right) - \mathrm{Cov}\left(\log \mu_i, \log z_i\right)\right) \tag{49}$$

$$= \frac{1}{2} \cdot \mathcal{E}\frac{1}{\alpha^2}\mathrm{Var}\left(\log z_i\right) - \frac{1}{2} \cdot \mathcal{E}\mathrm{Var}\left(\log \mu_i\right) \tag{50}$$

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

Plugging this back in, we obtain the formula:

$$\log Z = \mathbb{E}\left[\log z_i\right] - \frac{1}{2} \cdot \mathcal{E}\mathrm{Var}\left(\log \mu_i\right) + \frac{1}{2} \cdot \mathcal{E}\frac{1}{\alpha^2}\mathrm{Var}\left(\log z_i\right) - \frac{1}{2}\frac{1}{\alpha}\mathrm{Var}\left(\log z_i\right) \tag{51}$$

which is Equation 11 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

$$\log Z^* - \log Z = \frac{1}{2} \cdot \mathcal{E}\mathrm{Var}\left(\log \mu_i\right) \tag{52}$$

which completes the proof. □

A.2 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

$$d \log p_i = d \log \mu_i + d \log r - d \log \underbrace{f_i'\left(k_i\right)}_{\text{MPK}} \tag{53}$$

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

$$d \log y_i = \left(f_i'\right) \cdot \frac{k_i}{y_i} d \log k_i \tag{54}$$

$$\implies \frac{y_i}{f_i'} \cdot d \log y_i = k_i d \log k_i \tag{55}$$

$$d \log f_i' = \frac{d \log f_i'}{d \log k_i} d \log k_i \tag{56}$$

$$= f_i'' \cdot \frac{k_i}{f_i'} d \log k_i \tag{57}$$

$$= \phi_i d \log y_i \tag{58}$$

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

$$d \log p_i = d \log \mu_i + d \log r - \phi_i d \log y_i \tag{59}$$

We then plug in the demand curve from Equation 13, combining the firm-level demand

and firm-level supply curves:

$$\underbrace{\frac{1}{\theta}d\log Y - \frac{1}{\theta}d\log y_i}_{\text{Firm-Level Demand}} = \underbrace{d\log \mu_i + d\log r - \phi_i d\log y_i}_{\text{Firm-Level Supply}} \tag{60}$$

This yields:

$$d\log y_i = -\underbrace{\mathcal{E}_i d\log \mu_i}_{\text{Wedge}} - \underbrace{\mathcal{E}_i d\log r}_{\text{Input Cost}} + \underbrace{\frac{\mathcal{E}_i}{\theta} d\log Y}_{\text{Demand}} \tag{61}$$

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

A.3 Proof of Proposition 2

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

$$\mathbb{E}\left[k_i d\log k_i\right] = 0 \tag{62}$$

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:

$$k_i d\log k_i = \frac{y_i}{f_i'} d\log y_i \tag{63}$$

$$= \frac{p_i y_i}{r_i \mu_i} d\log y_i \tag{64}$$

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

$$\mathbb{E}\left[\frac{\lambda_i}{\mu_i} d\log y_i\right] = 0 \tag{65}$$

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

$$\mathbb{E}\left[p_i y_i\right] = \mathbb{E}\left[\frac{dY}{dy_i} y_i\right] = Y \tag{66}$$

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

$$d\log Y = \mathbb{E}\left[\frac{dY}{dy_i}\cdot\frac{y_i}{Y}d\log y_i\right] \tag{67}$$

$$= \mathbb{E}\left[\frac{p_i y_i}{Y}d\log y_i\right] \tag{68}$$

$$= \mathbb{E}\left[\frac{p_i y_i}{\mathbb{E}\left[p_i y_i\right]}d\log y_i\right] \tag{69}$$

$$\Longrightarrow \quad d\log Y = \mathbb{E}\left[\lambda_i d\log y_i\right] \tag{70}$$

Finally, we can combine input market clearing (Equation 65) and aggregation (Equation 70), along with firm behavior from Lemma 1, in a way that r falls out. Take Equation 70 and subtract off C times Equation 65, where C is some constant. We have:

$$\mathbb{E}\left[\lambda_i d\log y_i\right] - C\cdot\mathbb{E}\left[\frac{\lambda_i}{\mu_i}d\log y_i\right] = d\log Y \tag{71}$$

$$\mathbb{E}\left[\left(\lambda_i - C\cdot\frac{\lambda_i}{\mu_i}\right)\cdot\left(-\mathcal{E}_i d\log\mu_i - \mathcal{E}_i d\log r + \frac{\mathcal{E}_i}{\theta}d\log Y\right)\right] = d\log Y \tag{72}$$

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:

$$d\log Y = \mathbb{E}\left[\left(\lambda_i - \tilde{\mu}\cdot\frac{\lambda_i}{\mu_i}\right)\cdot\left(-\mathcal{E}_i d\log\mu_i - \mathcal{E}_i d\log r + \frac{\mathcal{E}_i}{\theta}d\log Y\right)\right] \tag{73}$$

$$= \mathbb{E}\left[\lambda_i\hat{\mu}_i\cdot\left(\frac{\mathcal{E}_i}{\theta}d\log Y - \mathcal{E}_i d\log\mu_i\right)\right] \tag{74}$$

$$\Longrightarrow \quad \left(1 - \mathbb{E}\left[\frac{\mathcal{E}_i\lambda_i\hat{\mu}_i}{\theta}\right]\right)d\log Y = \mathbb{E}\left[-\mathcal{E}_i\lambda_i\hat{\mu}_i d\log\mu_i\right] \tag{75}$$

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

$$\mathbb{E}\left[\frac{\mathcal{E}_i\lambda_i\hat{\mu}_i}{\theta}\right] = \frac{1}{\theta}\mathbb{E}\left[\mathcal{E}_i\lambda_i\frac{\mu_i-\tilde{\mu}}{\mu_i}\right] \tag{76}$$

$$= \frac{1}{\theta}\mathbb{E}\left[\lambda_i\mathcal{E}_i\left(1-\mu_i^{-1}\frac{\mathbb{E}\left[\lambda_i\mathcal{E}_i\right]}{\mathbb{E}\left[\lambda_i\mathcal{E}_i\mu_i^{-1}\right]}\right)\right] \tag{77}$$

$$= \frac{1}{\theta}\mathbb{E}\left[\lambda_i\mathcal{E}_i-\frac{\lambda_i\mathcal{E}_i\mu_i^{-1}\mathbb{E}\left[\lambda_i\mathcal{E}_i\right]}{\mathbb{E}\left[\lambda_i\mathcal{E}_i\mu_i^{-1}\right]}\right] \tag{78}$$

$$= \frac{1}{\theta}\left(\mathbb{E}\left[\lambda_i\mathcal{E}_i\right]-\frac{\mathbb{E}\left[\lambda_i\mathcal{E}_i\mu_i^{-1}\right]}{\mathbb{E}\left[\lambda_i\mathcal{E}_i\mu_i^{-1}\right]}\mathbb{E}\left[\lambda_i\mathcal{E}_i\right]\right) \tag{79}$$

$$= 0 \tag{80}$$

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

$$d\log Y = -\mathbb{E}\left[\mathcal{E}_i\lambda_i\hat{\mu}_i d\log\mu_i\right] \tag{81}$$

□

A.4 Proof of Proposition 3

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:

$$\mathcal{L} \approx -\frac{1}{2}\cdot\mathbb{E}\left[\left(\frac{d\log Y\left(\mu=1\right)}{d\log\mu_i}+\frac{d\log Y\left(\mu=\mu\right)}{d\log\mu_i}\right)\log\mu_i\right] \tag{82}$$

$$= -\frac{1}{2}\cdot\mathbb{E}\left[\frac{d\log Y}{d\log\mu_i}\log\mu_i\right] \tag{83}$$

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 2, we have

$$\mathcal{L} = \frac{1}{2}\mathbb{E}\left[\mathcal{E}_i\lambda_i\hat{\mu}\log\mu_i\right] \tag{84}$$

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.

$$\log \mu_i - \log \tilde{\mu} \approx \frac{1}{\mu_i} (\mu_i - \tilde{\mu}) \tag{85}$$

$$= \hat{\mu} \tag{86}$$

$$\implies \mathbb{E}[\mathcal{E}_i \lambda_i \hat{\mu} \log \mu_i] \approx \mathbb{E}[\mathcal{E}_i \lambda_i (\log \mu_i - \log \tilde{\mu}) \log \mu_i] \tag{87}$$

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:

$$\log \bar{\mu} = \frac{\mathbb{E}[\mathcal{E}_i \lambda_i \log \mu_i]}{\mathbb{E}[\mathcal{E}_i \lambda_i]} \tag{88}$$

Substituting this into our old expression yields:

$$\mathbb{E}[\mathcal{E}_i \lambda_i \hat{\mu} \log \mu_i] \approx \mathbb{E}[\mathcal{E}_i \lambda_i (\log \mu_i - \log \bar{\mu}) \log \mu_i] \tag{89}$$

$$= \mathbb{E}[\mathcal{E}_i \lambda_i] \cdot \mathbb{E}_{\mathcal{E}_i \lambda_i}[(\log \mu_i - \log \bar{\mu}) \log \mu_i] \tag{90}$$

$$= \mathbb{E}_{\lambda_i}[\mathcal{E}_i] \cdot \mathrm{Var}_{\mathcal{E}_i \lambda_i}(\log \mu_i) \tag{91}$$

where the last line uses the fact that $\mathbb{E}[\lambda_i] = \mathbb{E}\left[\frac{p_i y_i}{\mathbb{E}[p_i y_i]}\right] = 1$. We can then plug this back into our full expression, to obtain our desired expression:

$$\mathcal{L} \approx \frac{1}{2} \mathbb{E}_{\lambda_i}[\mathcal{E}_i] \cdot \mathrm{Var}_{\mathcal{E}_i \lambda_i}(\log \mu_i) \tag{92}$$

□

A.5 Proof of Proposition 4

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

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

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

A.6 Proof of Lemma 2

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 $S_n(\mathcal{Z}, \delta_0)$.[20] Then, since the CDF $F_\delta(t) = P\big(S_n(\mathcal{Z}, \delta) \leq t\big)$ is also a Lipschitz con-

[20]The theorem is stated for a fixed function ψ, while our function depends on the sample size n and the variance matrix $\widehat{\Sigma}$. Inspection of the proof indicates that the result may still be applied so long as 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 $1 - F_{\delta_0}\big(S_n(\mathcal{Z},\delta_0)\big) \sim U[0,1]$. □

A.7 Proof of Proposition 5

Proof. First, consider the case of $\gamma_0 \neq 0$ in which we also have $\tau_0 \neq 0$. In this case we have $G = \nabla g(\delta_0)$ is a non-zero vector for both of the null hypotheses considered in the lemma. We can then show using standard methods, see for example Newey and McFadden (1994), that

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

where $\mathcal{Z} \sim \mathcal{N}(0,1)$. It then follows that

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

where $\chi^2(1)$ is a chi-squared distributed variable with one degree of freedom. Since we simulated draws of the parameter vector from $\delta^* \sim \mathcal{N}(\bar{\delta}, \widehat{\Sigma})$, 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) then gives

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

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

Lipschitz constant is fixed, which is true in this case (since it is one).

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 $T(\widehat{\delta}, 0) \Rightarrow \chi^2(p)$. Similarly, the simulated test statistic is also simply

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

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

A.8 Proof of Proposition 6

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

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

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 $(0, 1)$). $\square$