Measuring Misallocation with Experiments

David Hughes (Boston College) and Jeremy Majerovitz (St. Louis Fed)

CEPR/PSE Development Symposium June 2024

Misallocation and Development

> *"There is a large literature on misallocation and development — this is our best candidate answer to the question of why are some countries so much richer than others." (Jones, 2016)*

Misallocation is when inputs are not allocated efficiently across firms

Literature finds very large costs: Indian manufacturing TFP would rise 40-60% if it raised allocative efficiency to US level (Hsieh and Klenow 2009)

Misallocation drastically changes what type of development policies we consider

In efficient economies: Policies like credit subsidies are disortionary

When inputs are misallocated: Role for policies that reallocate inputs

- Policy can raise output if reallocates from low-return to high-return firms
- Subsidized credit, microfinance, credit guarantees, other financial interventions

How Should We Measure Misallocation?

Misallocation is a very influential theory of development, but also controversial

Existing work has largely relied on strong assumptions about production functions in order to estimate the cost of misallocation

Where these methods have found large losses, not obvious whether this suggests a rejection of efficient markets or a rejection of the auxiliary assumptions

This talk: What are the costs of capital misallocation? How can we measure misallocation with minimal assumptions?

Measuring Misallocation with Experiments

Parallel literature in micro-development has found high and dispersed returns to capital (Banerjee and Duflo 2005)

Influential experiment by de Mel, McKenzie, and Woodruff (2008):

- Gave cash or in-kind grants to Sri Lankan microenterprises
- Using grant as instrument for capital, finds *monthly* returns of $\approx$6%

We show how to also estimate the variance of expected monthly returns

- Sidestep functional form assumptions by estimating marginal products directly

We use (and extend) recent advances in the macroeconomics of aggregation to show how misallocation depends on this variance, under arbitrary firm production functions

Key Idea: Measure with (micro) experiment, aggregate with (macro) model

Outline of the Talk

Macro $\Rightarrow$ Micro $\Rightarrow$ Metrics $\Rightarrow$ Data

Methodology connects macro question (misallocation) to the microdata:

1. Represent misallocation as function of variance of marginal products
   - Extend Baqaee and Farhi (2020) to arbitrary firm production functions
2. Recover lower bound on variance of marginal products with a linear IV model
   - Extend de Mel et al. (2008) to estimate (variance of) heterogeneous returns
3. Provide new tools for valid inference on nonlinear functions of parameters
4. Bring tools to the data and estimate misallocation

Preview of Results

Provide new methodology: aggregation, measurement, and inference

- Allows arbitrary and heterogeneous production functions
- Separates misallocation from risk
- Robust to measurement error

Using de Mel et al. (2008) grant RCT in Sri Lanka

- Standard deviation of monthly returns of 9.8%; 90% CI rules out below 4%
- Standard deviation over mean is 1.23, 90% CI $= [0.47, \infty]$
- Cobb-Douglas estimates actually in the right ballpark (with some caveats)

Optimally reallocating capital increases aggregate output by 22%

Literature Review

Misallocation (Macro): Restuccia and Rogerson (2008, 2017); Hsieh and Klenow (2009); Bartelsman et al. (2013); Hopenhayn (2014)

High and Dispersed Returns to Capital (Micro): Banerjee and Duflo (2005); de Mel et al. (2008); Fafchamps et al. (2014); McKenzie, (2017); Hussam et al. (2022); Beaman et al. (2023); Crépon et al. (2023)

We bridge these literatures and show how to aggregate the micro evidence

Measuring Misallocation: Bils et al. (2021); Rotemberg and White (2021); Gollin and Udry (2021); Haltiwanger et al. (2018); Carrillo et al. (2023)

We measure without Cobb-Douglas and robust to risk and measurement error

Aggregation: Liu (2019); Baqaee and Farhi (2020); Bigio and La'O (2020)

We use these tools to measure misallocation with arbitrary production functions

Measuring Misallocation in Terms of Marginal Products

Environment (Horizontal Economy)

Preferences: Household aggregates differentiated products into final utility

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

Technologies: Unit mass of firms $i \in [0,1]$. Each has individual production function:

$$y_i = f_i\left(k_i\right)$$

Endowments: Aggregate capital constraint

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

Horizontal economy is benchmark in literature and not bad description of setting

Planner's Problem and Efficiency

Planner maximizes household utility, given endowment and technologies

Planner solves

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

Yields first order condition

$$\underbrace{\frac{dY}{dy_i}}_{\text{Marginal Utility of Good } i} \cdot \underbrace{\frac{dy_i}{dk_i}}_{\text{MPK of Firm } i} = r \; \forall i$$

where "r" is Lagrange multiplier on capital constraint

Introducing Prices and VMPK

Marginal utility is unobservable, so for measurement we need prices

Assume a **price-taking** household that optimizes consumption bundle

Normalizing $P = 1$, household's FOC gives us

$$p_i = \frac{dY}{dy_i}$$

For efficiency, need to equalize $\frac{dY}{dy_i} \cdot \frac{dy_i}{dk_i}$ across firms.

So, define "Value of the Marginal Product of Capital"

$$\text{VMPK}_i := p_i \cdot \text{MPK}_i = \frac{dY}{dy_i} \cdot \frac{dy_i}{dk_i}$$

In efficient economies, VMPK is equalized across firms

“Wedges” Rationalize Deviations from Efficiency

When VMPK is not equalized, what is the cost of misallocation?

To talk about misallocation, we need to rationalize deviations from efficiency

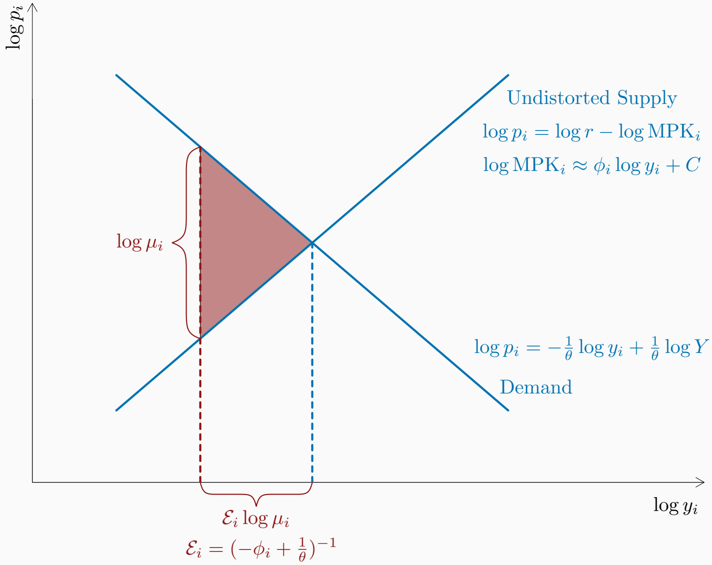
The planner’s first-order condition of the firm is distorted by some “wedge,” μ_i:

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

Wedge could come from markups, taxes, credit constraints, etc.

Note: No assumption about firm conduct, just price-taking consumer

Harberger's Triangle: Deadweight Loss from a Wedge



Misallocation is a Heap of Harberger Triangles

Misallocation is sum of Harberger triangles, even in GE Proof Sketch

Proposition

Consider a horizontal economy with CES aggregation. The cost of misallocation is given by

$$\underbrace{\log Y^* - \log Y}_{\text{Cost of Capital Misallocation}} \approx \frac{1}{2} \cdot \underbrace{\mathbb{E}_{\lambda_i}\left[\mathcal{E}_i\right]}_{\text{Weighted Average Elasticity}} \cdot \underbrace{Var_{\mathcal{E}_i \lambda_i}\left(\log \mu_i\right)}_{\text{Weighted Variance of Log Wedges}}$$

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

Extends Baqaee and Farhi (2020) to allow arbitrary firm production functions

Misallocation depends on the variance of log wedges

More Aggregation Results

Cobb-Douglas Log-Normal Special Case:

- With homogeneous Cobb-Douglas + log-normally distributed (μ, z), second-order approximation becomes exact and weights fall out More

Accuracy of Second-Order Approximation

- Simulations suggest second-order approximation is fairly accurate More

Misallocation with Multiple Inputs

- Gains from reallocating capital lower bound on reallocating all inputs More
- Formula for cost of misallocation (requires stronger assumptions) More

Measuring VMPK Without Data on Quantities

With price-taking household:

$$\mathrm{Var}\left(\log \mu_i\right) = \mathrm{Var}\left(\log\left(\frac{dY}{dy_i}\frac{dy_i}{dk_i}\right)\right) = \mathrm{Var}\left(\log\left(p_i\frac{dy_i}{dk_i}\right)\right) = \mathrm{Var}\left(\log\left(\mathrm{VMPK}_i\right)\right)$$

Ideally we would measure VMPK...

...*in practice* can only measure MRPK (no data on physical quantities)

Marginal revenue product of capital measures effect of capital on revenue

Under CES demand, MRPK is proportional to VMPK:

$$\mathrm{MRPK}_i = p_i\frac{dy_i}{dk_i} + y_i\frac{dp_i}{dy_i}\cdot\frac{dy_i}{dk_i} = \left(1 + \frac{d\log p_i}{d\log y_i}\right)\cdot p_i\frac{dy_i}{dk_i} = \frac{\theta-1}{\theta}\cdot \mathrm{VMPK}_i$$

So, under CES, $\mathrm{Var}\left(\log \mathrm{VMPK}_i\right) = \mathrm{Var}\left(\log \mathrm{MRPK}_i\right)$ Beyond CES

Measuring Misallocation with MRPK

Proposition

Consider a horizontal economy with CES aggregation and a price-taking consumer. In this economy,

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

Price-taking household gives us the first equality ($\mu_i \propto \mathrm{VMPK}_i$)

CES gives us the second equality ($\mathrm{VMPK}_i \propto \mathrm{MRPK}_i$)

Can also get the second equality if markets are competitive, so $\mathrm{VMPK}_i = \mathrm{MRPK}_i$

Combined with earlier result, tells us losses depend on $\mathrm{Var}(\log \mathrm{MRPK}_i)$

From Micro to Macro

Theory gives us a mapping from the distribution of marginal products to the equilibrium cost of misallocation

In particular, cost of misallocation is $\mathcal{L} \approx \frac{1}{2}\mathcal{E} \cdot \mathrm{Var}\left(\log \mathrm{MRPK}_i\right)$

- Sufficient statistics approach makes the empirics transparent
- Later, we will use estimates of CES parameter and returns-to-scale to calibrate $\mathcal{E}$
- To test productive efficiency, key is to identify $\mathrm{Var}\left(\log \mathrm{MRPK}_i\right)$

Next up: How do we measure MRPK?

Measuring Marginal Products with an IV Regression

Grants Experiment (de Mel, McKenzie, and Woodruff 2008)

Setting is three districts of Sri Lanka: sample of microenterprises

RCT randomly offered grants to fund capital

- Four treatment arms (10,000 vs. 20,000 rupees; cash vs. in-kind) and control
- Staggered rollout of treatment over quarterly waves:
  - None treated in wave 1
  - Some treated between wave 1 and 2
  - Rest treated betwen wave 3 and 4
  - Control gets 2,500 rupees after wave 5

Estimate returns to capital, using grant as instrument:

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

Roll treatments into one: Z_{it} = Cumulative Grant Amount Received

Identifying Returns to Capital with Randomized Grants

Consider heterogeneous, (locally) linear model of profits:

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

Three challenges:

- **Exogeneity:** Need instrument that provides exogenous variation in k_i

  Solution: Use grants RCT from de Mel et al. (2008)
- **Excludability:** Need to isolate MRPK from other marginal products

  Solution: Profits regression isolates MRPK More
- **Heterogeneity:** Need moments of the distribution of β_i

  Solution: Project β_i onto ex ante characteristics

IV Regression with Heterogeneous Treatment Effects

Convert original model to a model with heterogeneous returns to capital

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

Heterogeneity by baseline characteristics, X_i

Instruments are Z_{it} and $Z_{it} \times X_i$

Need interacted time fixed effects ($\delta_t^X \times X_i$) to ensure that $Z_{it} \times X_i \perp \varepsilon_{it}$

- Since Z_{it} correlated with time, need δ_t to control for time trends
- Need $\delta_t^X \times X_i$ to control for differential trends by baseline X_i

With estimate of γ, can back out $\text{Var}\left(\mathbb{E}\left[\text{MRPK}_i \mid X_i\right]\right) = \text{Var}\left(\gamma' X_i\right)$

Isolate the Predictable Component of Returns

Can decompose the variance of returns to capital into two components, based on law of total variance:

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

Our strategy, which projects returns to capital onto ex ante characteristics, will provide a lower bound on total variance Comparison to Carrillo et al.

Focusing on variance of ex ante expected returns is a feature, not a bug

Need to distinguish misallocation from risk
(David and Venkateswaran 2019; Gollin and Udry 2021)

- *Ex ante* differences in returns are misallocation
- *Ex post* differences in returns are misallocation + risk

Our method isolates misallocation, as opposed to risk

Comparison to Standard Approach (Hsieh and Klenow 2009)

We relax assumptions relative to "standard approach" (e.g. Hsieh and Klenow 2009)

Hsieh and Klenow assume homogeneous Cobb-Douglas: implies MPK $\propto$ APK

For $y_i = z_i k_i^{\alpha}$, we have:

$$\text{APK}_i := \frac{y_i}{k_i} = z_i k_i^{\alpha-1}$$

$$\text{MPK}_i := \frac{dy_i}{dk_i} = \alpha z_i k_i^{\alpha-1}$$

$$\implies \text{VMPK}_i := p_i \cdot \text{MPK}_i = \alpha \frac{p_i y_i}{k_i}$$

In general, cannot infer MPK from APK (Haltiwanger, Kulick, and Syverson 2018)

- Example: Fixed costs and/or heterogeneous α_i (returns to scale)

We measure marginal products directly with an RCT

Our Method is Robust to Measurement Error

Measurement error shows up as misallocation in standard approach
(Bils, Klenow, and Ruane 2021; Gollin and Udry 2021; Rotemberg and White 2021)

Under homogeneous Cobb-Douglas:

$$\mathrm{Var}\left(\log \mu_i\right) = \mathrm{Var}\left(\log\left(\alpha \frac{p_i y_i}{k_i}\right)\right) = \mathrm{Var}\left(\log\left(p_i y_i\right) - \log k_i\right)$$

Classical measurement error will inflate this expression

Our method is doubly-robust to measurement error

- Classical measurement error has no effect on IV estimates
- Measurement error in inputs/output does not affect projection onto observables

Inference for Nonlinear Functions of Parameters

The *Shape* of Misallocation

What is the exact function of parameters we are interested in?

Convenient change of basis: Original covariates $\Rightarrow$ Standardized principal components

- Orthonormal basis: New covariates have $\text{Var}(X_i) = I$

To measure lower bound on $\text{Var}(\text{MRPK}_i)$:

- $\text{Var}(\mathbb{E}[\text{MRPK}_i \mid X_i]) = \gamma' \text{Var}(X_i)\gamma = \gamma'\gamma = \sum_j {\gamma_j}^2$
- **Misallocation is the radius of the circle**

To back out $\text{Var}(\log \text{MRPK}_i)$ with log-normal approximation:

- $\text{Var}(\log \mathbb{E}[\text{MRPK}_i \mid X_i]) = \log\left(1 + \left(\frac{\sqrt{\gamma'\gamma}}{\beta}\right)^2\right)$
- Geometrically, null hypothesis $\frac{\gamma'\gamma}{\beta^2} = g_0$ is a (hyper-)cone
- **Misallocation is the angle of the cone**

How do we do inference for this object?

Inference for Nonlinear Functions of Parameters

We are interested in a high-dimensional, highly nonlinear function of parameters $g(\beta, \gamma)$

Standard methods perform poorly in this setting

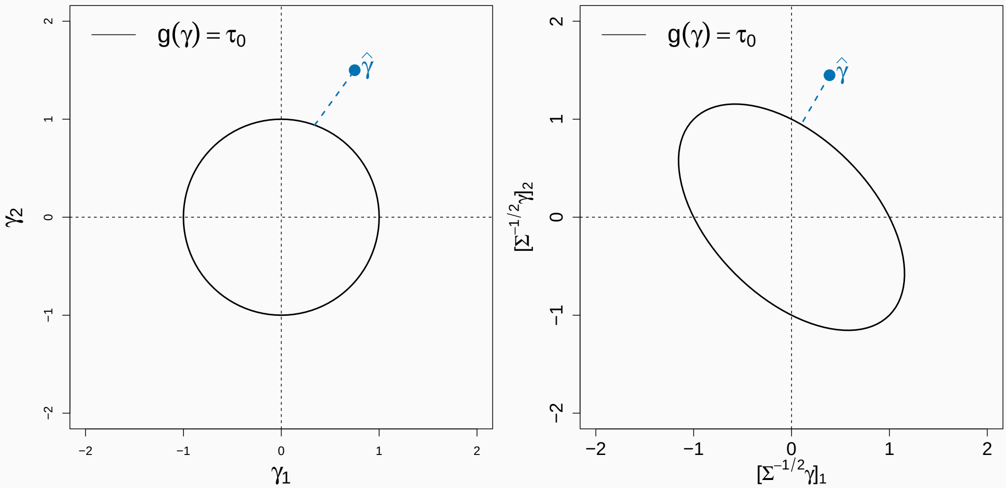
- Delta method takes linear approximation to nonlinear function
  - Fails due to severe nonlinearity
- Projection method uses confidence set for (β, γ) to create CI for $g(\beta, \gamma)$
  - Very conservative in high dimensions

We provide new method to construct valid confidence intervals

- Null hypothesis: $g(\beta, \gamma) = g_0$
- Test statistic: inverse-variance-weighted distance from $\left(\hat{\beta}, \hat{\gamma}\right)$ to constraint
- Critical values: obtained from Gaussian simulation

Prove asymptotic validity, and excellent performance in Monte Carlo

Graphical Summary of the Distance Metric



Distance-Metric Test Works Well in Simulation

Simulation with Gaussian errors, calibrated to the actual data ($K = 4$)

Each row shows probability of rejecting a true null, varying β and γ

Table 1: Simulated Rejection Rates

		10 % Rejection			5 % Rejection		
$\sqrt{\gamma'\gamma}$	β	Wald	Projection	Distance	Wald	Projection	Distance
0.10	0.10	0.058	0.004	0.098	0.021	0.003	0.052
0.10	0.03	0.150	0.000	0.094	0.102	0.000	0.050
0.01	0.10	0.198	0.003	0.109	0.109	0.000	0.053
0.01	0.03	0.068	0.004	0.104	0.037	0.000	0.045

Wald test performs terribly: sometimes over-rejects, sometimes under-rejects

Projection method is extremely conservative, as expected in high dimensions

Our distance-metric test gives correct size in each case! Point Estimator Simulations

Empirical Estimates of the Cost of Misallocation

Summary of Results

Estimate returns heterogeneity with single baseline covariate

- Some covariates not very informative, although baseline APK is useful

Estimate returns heterogeneity with first K principal components

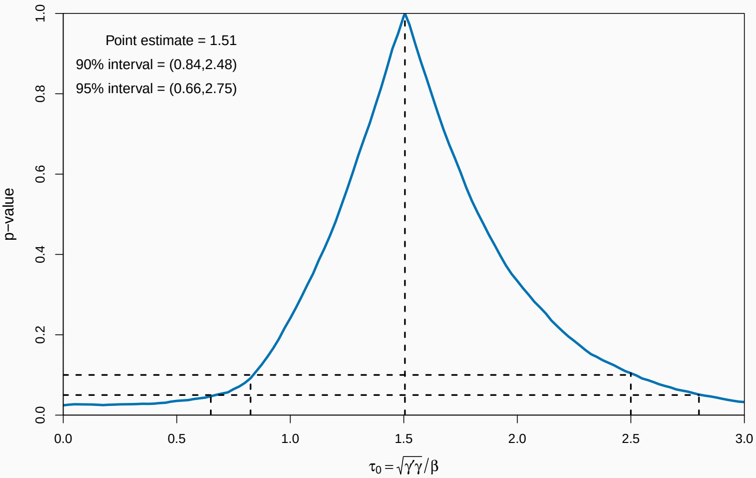
- For $K = 5$: Mean 8% (monthly) vs. SD of 9.8%; SD/Mean = 1.23
- 90% CI rules out SD below 4% or ratio below 0.47

In *principle*, can estimate $\mathcal{E}$ nonparametrically; in *practice* too noisy

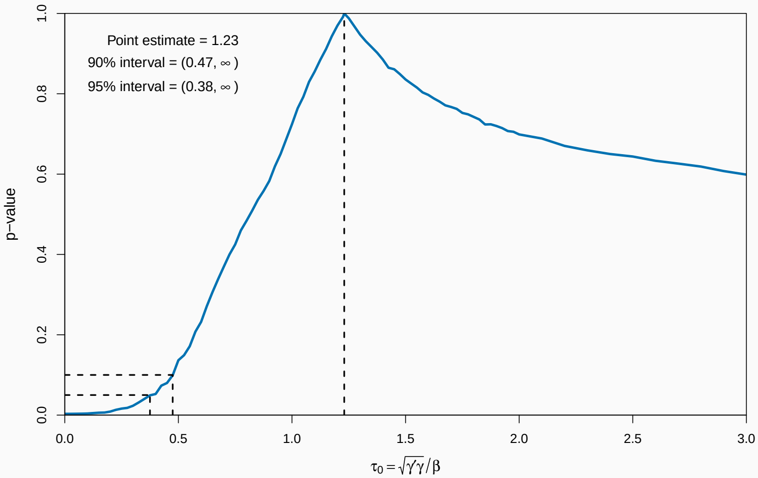
Use standard calibrations, as in Hsieh and Klenow (2009):

- Standard calibration $\alpha = \frac{1}{3}, \theta = 3$; also consistent with (noisily) estimated RTS
- Implies elasticity of output w.r.t. the wedge of $\mathcal{E} = \frac{3}{7}$

Confidence Interval for Baseline APK Covariate



Confidence Interval for $K = 5$



The Cost of Misallocation

Table 2: Estimated Cost of Misallocation ($K = 5$)

	Point Estimate	90% CI	95% CI	Cobb-Douglas
$\mathrm{SD}(\mathrm{MRPK}_i)/\mathbb{E}[\mathrm{MRPK}_i]$	1.234	0.470	0.384	—
$\mathrm{Var}(\log \mathrm{MRPK}_i)$	0.93	0.20	0.14	1.35
$Z^*/Z - 1$ ($\mathcal{E} = \frac{3}{7}$)	22%	4%	3%	34%

Point estimates imply large potential gains from reallocating capital

Cobb-Douglas estimates are larger, but not so different

Robustness and Extensions

Other estimates/confidence intervals are similar

- Uniformly valid ("worst case") intervals More
- Weighted by firm profits More

Results do not seem to be driven by:

- Adjustment costs More
- Differential exposure to aggregate risk More

Misallocation with multiple inputs

- Main results are a lower bound on gains from reallocating all inputs
- Can estimate this too, at the price of stronger assumptions More

Extension: Returns Heterogeneity in Other RCTs

Apply our method to estimate SD/Mean return in other experiments

Hussam, Rigol, and Roth (2022): Peer rankings predict high returns to grant

Beaman, Karlan, Thuysbaert, and Udry (2023): Loan take-up predicts high return

Note: Can only estimate "return to grant" (reduced form), not return to capital (IV)

	Covariate	Ratio Point Estimate	90% CI
De Mel et al. (2008)	$K = 5$	1.23	$[0.47, \infty]$
	APK	1.51	$[0.84, 2.48]$
	log (APK)	0.75	$[0.38, 1.35]$
Hussam et al. (2022)	Peer Rank	2.07	$[0.64, \infty]$
	APK	0.70	$[0, 7.68]$
	log (APK)	0.51	$[0.14, 3.45]$
Beaman et al. (2023)	Loan Status	1.73	$[0.58, 4.51]$

Find heterogeneity in returns in many settings!

Conclusion

Conclusion

Show how to measure misallocation without auxiliary assumptions

Measure misallocation as a function of the distribution of marginal products

Derive formula for misallocation under arbitrary production functions:

$$\log Y^* - \log Y \approx \frac{1}{2}\mathcal{E} \cdot \mathrm{Var}\left(\log \mathrm{MRPK}_i\right)$$

Use RCT to identify heterogeneous *ex ante* returns to capital

Develop novel econometric tools to handle cases where traditional methods fail

We find substantial dispersion of MRPK

Potential gains large: Optimally reallocating capital increases output 22%

Appendix

Proof Sketch for Main Result

Combine firm behavior with input market clearing and aggregation to show:

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

where $\lambda_i := \frac{p_i y_i}{\int p_i y_i di}$ is sales share, $\hat{\mu}_i := \frac{\mu_i - \tilde{\mu}}{\mu_i}$ is "wedge deviation", and $\tilde{\mu} = \frac{\mathbb{E}[\lambda_i \mathcal{E}_i]}{\mathbb{E}[\lambda_i \mathcal{E}_i \mu_i^{-1}]}$

Then integrate along a path from μ to wedgeless equilibrium. Let $\log \check{\mu}(t) = t \cdot \log \mu$

$$\mathcal{L} := \log Y^* - \log Y = \int_0^1 \frac{d \log Y(\check{\mu}(t))}{d \log \mu} \cdot \frac{d \log \check{\mu}(t)}{dt} dt$$

Trapezoid rule + Envelope Theorem at $\mu = 1$: Trapezoid $\implies$ (Harberger) triangle:

$$\mathcal{L} \approx \frac{1}{2}\mathbb{E}\left[\frac{d \log Y(\mu)}{d \mu_i} \log \mu_i\right] = \frac{1}{2}\mathbb{E}\left[\mathcal{E}_i \lambda_i \hat{\mu} \log \mu_i\right] \approx \frac{1}{2}\mathbb{E}_{\lambda_i}\left[\mathcal{E}_i\right] \cdot \mathrm{Var}_{\mathcal{E}_i \lambda_i}\left(\log \mu_i\right)$$

Back

Cost of Misallocation (Exact Formula for Special Case)

Assume $\log y_i = \log z_i + \alpha \log k_i$, and $(\log z_i, \log \mu_i)$ distributed multivariate normal.

Proposition

Consider a horizontal economy with CES aggregation, log-linear production, and lognormally distributed productivity and wedges. The cost of misallocation is given by

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

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

Formula is exact, and weights fall out!

Back

Benefits of Reallocating Capital vs. All Inputs

Why are the benefits from reallocating capital a lower bound on the benefits from reallocating all inputs?

Think about planner's problem as nested optimization:

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

Capital allocation is just the inner problem $\implies$ Allocation where only capital optimized must be no better than allocation where all inputs optimized

$$\log Y^* - \log Y \geq \log Y^*\left(\{l_i\}_{i\in[0,1]}\right) - \log Y$$

Back

Cost of Misallocation (Multiple Inputs)

Our experiment only measures MRPK...

...how can we get cost of misallocation for all inputs?

Need additional assumptions!

In particular, need to be able to ignore the input mix

Horizontal CES economy (as before) with M inputs supplied inelastically

Cobb-Douglas production: $\log y_i = \log z_i + \sum_M \alpha_{m,i} \log x_{m,i}$

Same input shares for all firms ($\alpha_{m,i} = \tilde{\alpha}_m \cdot \sum_M \alpha_{m,i}$ for all i)

Wedges are markups on output: only distort scale, not input mix

Cost of Misallocation (Multiple Inputs)

Under these assumptions, input mix same at all firms, and pinned down by input supply

Lets us to aggregate to composite input, and then apply our old results

Proposition

Under the assumptions of the previous slide, the cost of misallocation is given by

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

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

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

More general than Hsieh-Klenow (we allow heterogeneous returns to scale), but stricter than our results for just capital Back to Main Slides Back to Empirical Appendix

Accuracy of Second-Order Approximation

Draw $\log z_i \sim N\left(0, \sigma_z^2\right)$, $\sigma_z = 1.2$. Set $\text{Var}\left(\log \mu_i\right) = 0.93$. Set $\theta = 3$.

Cobb-Douglas with $\alpha = \frac{1}{3}$. Last column draw $\alpha_i = \left\{\frac{1}{6}, \frac{1}{3}, \frac{1}{2}\right\}$ equal prob.

Table 3: Accuracy of Second-Order Approximation: Simulation Results

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

Back

The Variance of log VMPK Outside of CES

Outside CES, demand elasticity differs for different size firms. We have:

$$\log \mathrm{VMPK}_i = \log \underbrace{\mathrm{MRPK}_i}_{\text{Other Frictions}} + \underbrace{\log \left(1 \Big/ \left(1 + \frac{d \log p_i}{d \log y_i}\right)\right)}_{\text{Endogenous Markups}}$$

Under monopolistic competition, endogenous markup is $1/\left(1 + \frac{d \log p_i}{d \log y_i}\right)$, while all variation in MRPK is due to other frictions

Standard estimates suggests $\mathrm{Cov}\left(\log \mathrm{MRPK}_i, \log \left(1/\left(1 + \frac{d \log p_i}{d \log y_i}\right)\right)\right) \geq 0$

- Endogenous markups depend only on firm size, and are larger at large firms
- "Correlated distortions" thought to be positively correlated with size

Implies that $\mathrm{Var}\left(\log \mathrm{VMPK}_i\right) \geq \mathrm{Var}\left(\log \mathrm{MRPK}_i\right)$, so we measure a lower bound

Back

Excludability: Using Profits to Isolate the MRPK

RCT provides exogenous instrument which affects capital

But we also need excludability: instrument affects outcome only through capital

What if the instrument affects other inputs?

Solution (de Mel et al. 2008) is to use profits (in practice, really EBIDA)

Take total derivative and linearize:

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

$$\implies \Delta \underbrace{(p_i y_i - w l_i - c m_i)}_{\text{“Profits” or EBIDA}} = \mathrm{MRPK}_i \cdot \Delta k_i + (\mathrm{MRPL}_i - w) \cdot \Delta l_i + (\mathrm{MRPM}_i - c) \cdot \Delta m_i$$

Get excludability either if

- Input not very affected by instrument (empirically true for labor) or
- Wedge on that input is small (appears true for materials) Back

Comparison to Carrillo, Donaldson, Pomeranz, Singhal (2023)

Carrillo, Donaldson, Pomeranz, Singhal (2023) use procurement lotteries as demand shocks to estimate dispersion in markups. A few key differences from our paper:

Different variance, different econometrics

- They target *total* variance: upper bound on misallocation
- They use an IV-CRC model and run regression + squared regression
- Need Z *independent* ε, with 3+ points of support, rules out some nonlinearities
- In our setting, yields very wide confidence intervals CDPS Intervals

Different setting, different results

- They find little misallocation for Ecuadorian construction firms
- We find large misallocation for Sri Lankan microenterprises
- Suggests importance of studying many settings!

Papers are complementary + provide a toolkit for future applications Back

CDPS Method Estimates

Table 4: Estimates of Variance of MRPK: Carrillo et al. 2023 Method

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

Back

The Delta Method

Typical Wald test relies on delta method

- Null hypothesis is $H_0 : \; g(\beta, \gamma) = \frac{\sqrt{\gamma'\gamma}}{\beta} = g_0$
- Given $\sqrt{n}\left((\hat{\beta}, \hat{\gamma}) - (\beta_0, \gamma_0)\right) \Rightarrow N(0, \Sigma)$, the delta method implies

$$\sqrt{n}(g\left(\hat{\beta}, \hat{\gamma}\right) - g_0) \overset{a}{\sim} N(0, A'\Sigma A),$$

$$A = \nabla g = (-\frac{\sqrt{\gamma'\gamma}}{\beta^2}, \frac{\gamma'}{\beta\sqrt{\gamma'\gamma}})'$$

- Yields Wald statistic: $W = n\frac{(\widehat{g}-g_0)^2}{\hat{A}'\hat{\Sigma}\hat{A}} \overset{a}{\sim} \chi^2(1)$

Requires $\nabla g \neq 0$ and $\nabla g \neq \infty$

Approximation is poor when $\gamma_0 \approx 0$ and/or $\beta_0 \approx 0$

But $g_0 = 0$ (no misallocation) requires $\gamma_0 = 0$!

Back

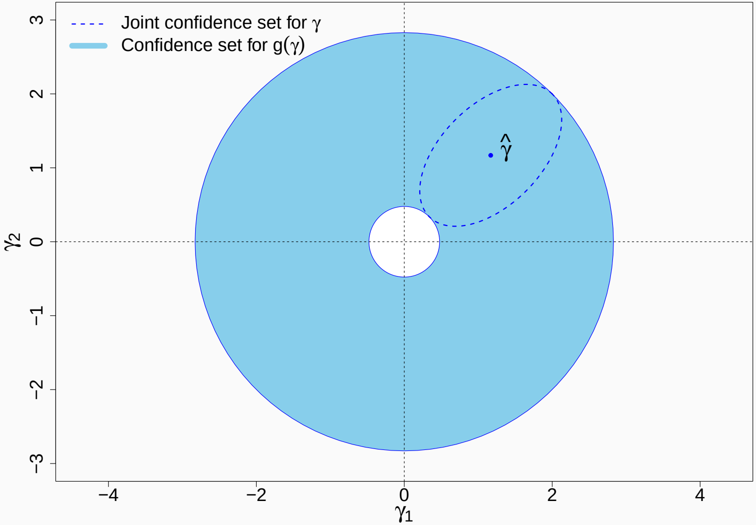
The Projection Method

A common alternative to the delta method is the projection method:

- Let $CI(\beta, \gamma)$ be a joint confidence set for (β, γ)
- The projection confidence set contains $g_0 = g(\beta, \gamma)$ for every $(\beta, \gamma) \in CI(\beta, \gamma)$

Problem: The projection method is conservative, especially in high dimensions

Projection Method Confidence Sets Are Too Large



Back

Algorithm

1. Estimate the IV regression to obtain $\widehat{\delta}$ and $\widehat{\Sigma}$, where $\delta := (\beta, \gamma)$
2. Set a null hypothesis $g(\delta) = g_0$:
   2.1 compute the constrained parameter estimates $\bar{\delta}$ and $\bar{\Sigma}$
   2.2 compute the test statistic
   $$\widehat{DM}(g_0) = \min_{\delta: g(\delta) = g_0} n(\delta - \widehat{\delta})' \bar{\Sigma}^{-1} (\delta - \widehat{\delta})$$
   2.3 for $b = 1, \ldots, B$, simulate $\delta_b \sim N(\bar{\delta}, \bar{\Sigma})$ and compute
   $$DM_b(g_0) = \min_{\delta: g(\delta) = g_0} n(\delta - \delta_b)' \bar{\Sigma}^{-1} (\delta - \delta_b)$$
   and set the critical value $c_{1-\alpha}(g_0)$ as the $(1-\alpha)$-quantile of $DM_b(g_0)$.
   2.4 reject $H_0 : g = g_0$ if $\widehat{DM}(g_0) > c_{1-\alpha}(g_0)$
3. Repeat step 2 for a range of g_0 values to construct the confidence interval
$$CI_{1-\alpha}(g) = \{g : \widehat{DM}(g) \leq c_{1-\alpha}(g)\}$$

Back

Extension: Uniformly Valid Confidence Intervals

Our approach can be adjusted to provide uniformly valid inference. More

- It is possible show that using critical values constructed by simulation $\delta \sim N(\delta_0, \Sigma)$ provides a uniformly valid test
- However, we do not know δ_0. Replacing δ_0 with $\bar{\delta}$ gives a test that is asymptotically correct, but not uniform

Idea is to replace the unknown δ_0 with a 'worst case' value, i.e. the value of δ that:

1. satisfies the null hypothesis $g(\delta) = g_0$
2. leads to the largest $1 - \alpha$ critical value

This provides a uniformly valid, although slightly conservative test. Back

Uniformly valid inference

A confidence set $CI_{1-\alpha,n}$ is said to have correct coverage asymptotically if

$$P\left(g_0 \in CI_{1-\alpha,n}\right) \to 1 - \alpha$$

A stronger condition is that the confidence set has *uniformly* correct coverage asymptotically

$$P^{\lambda_n}\left(g(\lambda_n) \in CI_{1-\alpha,n}\right) \to 1 - \alpha, \qquad \text{for any sequence } \lambda_n$$

where λ_n indexes some sequence of true DGPs.

- the uniformity condition provides some protection against settings in which asymptotic approximations may be poor for some values of parameters
- when confidence sets are not uniform, even in very large samples, there exists some value of the true parameters for which coverage may be poor

Back

Point Estimator

Same simulation as for our rejection rates, with 1000 simulation draws

Compute mean and median bias (although technically IV has no mean)

Table 5: Simulated Bias of Point Estimator

		Standard Deviation $(\sqrt{\gamma'\gamma})$		Ratio $(\sqrt{\gamma'\gamma}/\beta)$		
$\sqrt{\gamma'\gamma}$	β	Mean Bias	Median Bias	$\sqrt{\gamma'\gamma}/\beta$	Mean Bias	Median Bias
0.10	0.10	0.001	0.000	1.0	0.011	0.010
0.10	0.03	0.001	0.000	$3.\bar{3}$	0.298	0.087
0.01	0.10	0.008	0.007	0.1	0.076	0.071
0.01	0.03	0.008	0.007	$0.\bar{3}$	0.293	0.244

The standard deviation point estimator shows little bias in practice

Ratio shows bias if β is near zero: result of small and noisy denominator

Back

Results (Single Covariate)

Table 6: Estimates of Heterogeneous MRPK by Baseline Covariates

Covariate	Capital	Age	Education	Profit	Hours	APK	log(APK)
Panel A, with Covariates: $\mathbb{E}[\text{MRPK}_i]$							
Estimate	0.062	0.061	0.060	0.063	0.072	0.085	0.069
SE	(0.025)	(0.024)	(0.027)	(0.025)	(0.030)	(0.029)	(0.027)
Panel B: $SD\left(\mathbb{E}[\text{MRPK}_i \mid X_i]\right)$, *With Sign of Interaction Effect*							
Estimate	-0.070	+0.018	+0.044	-0.011	-0.023	+0.128	+0.052
90% CI	$[0.03, 0.64]$	$[0.00, 0.83]$	$[0.02, \infty]$	$[0.00, 0.13]$	$[0.00, 1.06]$	$[0.06, 0.22]$	$[0.02, 0.11]$
Panel C: $SD\left(\mathbb{E}[\text{MRPK}_i \mid X_i]\right)/\mathbb{E}[\text{MRPK}_i]$							
Estimate	1.121	0.300	0.723	0.171	0.314	1.505	0.747
90% CI	$[0.41, \infty]$	$[0.00, 0.90]$	$[0.30, 1.76]$	$[0.00, 3.49]$	$[0.00, 0.63]$	$[0.84, 2.48]$	$[0.38, 1.35]$

Back

95% Confidence Intervals

Table 7: Estimates of Heterogeneous MRPK by Baseline Covariates

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

Back

Uniformly Valid ("Worst Case") Confidence Intervals

Table 8: Estimates of Heterogeneous MRPK: Many Covariates

	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$	$K=7$
Main Estimates							
$\frac{\mathrm{SD}(\mathbb{E}[\mathrm{MRPK}_i \mid X_i])}{\mathbb{E}[\mathrm{MRPK}_i]}$	0.913	0.840	1.415	1.275	1.234	1.247	1.213
90% CI Bound	0.456	0.210	0.557	0.519	0.470	0.784	0.555
Worst Case Estimates							
90% CI Bound	0.205	0.000	0.205	0.392	0.387	0.672	0.436

Back

Results (Weighted Variance)

Construct weighted standardized principal components such that $\text{Var}_{\lambda_i}(\gamma' X_i) = \gamma'\gamma$

Table 9: Estimates of Variance of MRPK: Weighted Variance

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

Standard deviation similar to unweighted; ratio point estimates are unstable Back

Our Results Are Not Driven by Adjustment Costs

Adjustment costs can generate dispersion in MRPK, even in planner's problem

- Adjustment costs should be picked up as part of MRPK, but...
- Short-run MRPK can be dispersed because firms expect future adjustment costs
- Also get MRPK dispersion if adjustment costs generate inaction regions

Differences in MRPK due to adjustment costs should not be persistent

- Eventually, the firm adjusts and/or productivity reverts to the mean

However, differences in MRPK in the data are very persistent

Project MRPK onto baseline APK, and estimate separate coefficients for years 1 and 2

- Same coefficient in year 2 as year 1, suggesting little mean reversion

Our Results Are Not Driven by Adjustment Costs

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

Table 10: Estimates of Persistence of MRPK Differences by Baseline APK

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

Back

Our Results Are Not Driven by Differential Exposure to Aggregate Risk

Average returns can (efficiently) differ if firms have different exposure to aggregate risk

Capital Asset Pricing Model (CAPM):

$$r_{it} - r_t^{\text{risk-free}} = \alpha_i + \beta_i \cdot \left(r_t^{\text{market}} - r_t^{\text{risk-free}} \right)$$

Variation in α_i is true misallocation, variation in β_i is compensation for risk

How much can risk compensation alone explain?

If $\alpha_i = \alpha \ \forall i$, then $SD\left(r_{it}\right) = SD\left(\beta_i\right) \cdot \left(r_t^{\text{market}} - r_t^{\text{risk-free}}\right)$

Equity premium, $\mathbb{E}\left[r_t^{\text{market}} - r_t^{\text{risk-free}}\right]$, is 6% *annual*, or 0.5% monthly

$SD\left(\beta_i\right)$ is setting-dependent, but very likely < 1

Implies $SD\left(r_{it}\right)$ of $< 0.5\%$ monthly, but $SD\left(\text{MRPK}_i\right)$ is roughly 10% monthly!

Back

The Gains from Reallocating Multiple Inputs

The experiment we study only identifies the returns to *capital*

Need stronger assumptions for multiple-inputs case Proposition

- Wedges show up as a markup/tax on output (distort scale not input mix)
- Cobb-Douglas production with homogeneous input-mix
- Still allows heterogeneous returns to scale

Under these assumptions, can use same formula as before

New returns to scale parameter is for all inputs, not just capital

Under constant returns to scale, $\alpha = 1$ and $\mathcal{E} = \theta = 3$

Gains from optimally reallocating all inputs are 301%

Back