

Misallocation in Firm Production: A Nonparametric Analysis Using Procurement Lotteries*

Paul Carrillo[†]
Dave Donaldson[‡]
Dina Pomeranz[§]
Monica Singhal[¶]

November 2023

Abstract

This paper develops new tools to study misallocation that do not require assumptions about the heterogeneity of firms' technologies. We show how features of the distribution of marginal products can be identified from exogenous variation in firms' input use and used both to test for misallocation and to quantify its resulting welfare losses. We apply this method to a setting with exogenous demand shocks from public procurement contracts for construction services in Ecuador. Our results reject the null of efficiency but our estimates of the resulting welfare losses from misallocation are small.

*We thank our team of outstanding, dedicated research assistants, as well as David Baqaee, Mark Bills, Richard Blundell, Kirill Borusyak, Arnaud Costinot, Pete Klenow, Jeremy Majerovitz, Matt Masten, Virgiliu Midrigan, Emi Nakamura, Karthik Narayan, Whitney Newey, Michael Peters, Richard Rogerson, Cian Ruane, Jesse Shapiro, John Sturm, Alex Torgovitsky, Ivan Werning, and numerous seminar participants for helpful comments. We are grateful to the Centro de Estudios Fiscales and the Departamento de Control of the Ecuadorian Tax Authority for outstanding collaboration and to Innovations for Poverty Action (IPA), International Growth Centre (IGC) and the Stanford Institute for Innovation in Development Economies (SEED) for generous research support. This paper has also benefited from funding from CEPR and UK Department for International Development (DFID) (under the Private Enterprise Development in Low-Income Countries program, reference MRG004 3834), the European Research Council (grant reference 758984), and the Swiss National Science Foundation (grant 100018 192588). The views expressed are not necessarily those of CEPR, DFID, ERC or SNF.

[†]George Washington University, pcarrill@gwu.edu

[‡]Massachusetts Institute of Technology and NBER, ddonald@mit.edu

[§]University of Zurich and CEPR, dina.pomeranz@uzh.ch

[¶]University of California, Davis, and NBER, msinghal@ucdavis.edu

1 Introduction

There is no shortage of reasons to suspect that the economies around us produce goods and services in an inefficient manner. Many firms seem to enjoy market power, many contracts look incomplete, and many policy actions (be they taxes and subsidies, regulations, or even corruption) appear to sacrifice production efficiency in pursuit of other goals. These sources of market failure result in misallocation because they cause the (value-adjusted) marginal product of an input to differ across its uses in an economy. But how large are such differences and their misallocative consequences?

In this paper we develop and apply new techniques for assessing the extent of misallocation among any given set of firms. Our methods leverage exogenous shocks to firms’ input use (from shocks to either output demand or input supply) in order to quantify features of the distribution of firms’ marginal products, for each input. This allows us to test for the presence of misallocation and to estimate its welfare costs. Importantly, this is possible without the need to assume that firms use technologies that share common features, an assumption that is crucial for existing methods and may bias estimates of true marginal product heterogeneity. We apply these new procedures to Ecuador’s construction industry, in which a lottery component of the public procurement system generates random sources of firm-level demand. Using such variation, we find that misallocation in this context appears to be small (causing costs of 1.6%) and several times smaller than existing methods would imply.

We begin in Section 2 by describing the economic environment that motivates our empirical procedures. An econometrician observes a set of potentially multi-product firms producing in a given initial cross-section. The goal is to test for and quantify misallocation—or equivalently, departures from allocative efficiency of production (AEP)—in this cross-section. To encompass prior work, we distinguish two notions of AEP. The first, which we term unconstrained AEP (henceforth “U-AEP”), is generically necessary for Pareto efficiency. Such an allocation requires that, for any firm and produced good or service, the ratio of the value marginal product of any input to the input’s price—which we define as the “wedge” for that firm-product-input—is equal to one no matter where the input is in use. The second notion

is analogous but conditions on the aggregate amount of each input type that can be used by the firms in question, as standard definitions of aggregate productivity do. An allocation consistent with this notion of constrained AEP (C-AEP) would feature wedges that are equal to a common value across firms and products for each input type, but that value is not necessarily one.

Assessing departures from either constrained or unconstrained AEP would be simple if marginal products, and hence wedges, were observed. But a firm’s marginal products, by definition, depend on its production function, which cannot be identified in the cross-section if firms’ technologies differ in unknown respects. Existing methods overcome this challenge by assuming that firms’ technologies share common features (such as their elasticity of scale). But doing so may increase the risk of attributing observed differences in firm behavior to apparent heterogeneity in wedges, rather than underlying differences in technology, affecting the validity of inferences about the extent of misallocation.

By contrast, the methods we develop proceed by placing minimal restrictions on the technologies that each firm uses, the demand or supply relations it faces, the extent of its optimizing behavior, or the underlying sources of market failure that cause misallocation.¹ This flexibility is possible for two reasons. First, we do not aim to identify each firm’s wedge on each input and product. Instead, we focus on features of misallocation—such as its existence or its welfare consequences—that are functions of the (arbitrarily weighted) wedge *distribution*. Second, we observe that any wedge is simply the appropriately price-adjusted “treatment effect” of a change in a firm’s input on its output. This implies that exogenous variation in firm input use can be used to identify features of the distribution of such treatment effects (and hence wedges), drawing on recent advances in the literature on treatment effects estimation due to Masten and Torgovitsky (2016).

Intuitively, if a given allocation features C-AEP, for example, then, starting from this allocation, the treatment effect of an exogenous change in any firm’s input use will be the same anywhere in this economy. And if it features U-AEP, then these treatment effects will be equal to one. Our procedures simply invert this intuition and ask what amount of

¹The main substantive assumption we do require is that each firm uses a technology that is differentiable at the point at which it is operating.

misallocation is implied by the distribution of treatment effects that arises due to exogenous input variation.

The techniques we develop can be applied to many contexts with different sources of exogenous variation in firms’ input use. This can stem from demand shocks that firms face, be it from the government, from other firms, or from consumers. For example, some studies isolate quasi-experimental features of firms’ access to procurement contracts (e.g., Ferraz et al. (2015)), as we do in our empirical application.² Equivalently, our method can be applied using shifts in residual demand due to changes in the competitive environment—from domestic entry, domestic expansion, or the arrival of competition from abroad.³ In addition to demand-related shocks, exogenous variation in the supply of firms’ inputs can also be used, such as (effective) subsidies to capital, labor, material inputs, and management services.⁴ Finally, the method can also be applied using variation that derives from assumptions about the timing of firms’ information about changes in productivity, as used in the literature on production function estimation.

We apply our new approach for assessing misallocation to the empirical context of Ecuador. As detailed in Section 3, Ecuador’s public procurement system allocates certain construction contracts (for specific projects at specific prices) by lottery, generating exogenous variation in demand. We study the approximately 9,000 firms that took part in the over 18,000 multi-participant procurement lotteries in our sample period (2008-2015). We trace the response of both outputs and inputs to these random demand shocks by merging the lottery contracts data with administrative tax records containing information on sales and costs, firm-to-firm transaction values, as well as employer-employee matches. Compar-

²See also Lee (2017), Kroft et al. (2022), and Hvide and Meling (2023). Numerous studies have drawn on other forms of exogenous variation in private demand, stemming from either domestic buyers—e.g. Bai et al. (2022) and Amiti et al. (2023)—or foreign ones—e.g. Hummels et al. (2014), Atkin et al. (2017), Alfaro-Ureña et al. (2022), and Demir et al. (forthcoming). Finally, Egger et al. (2022) use exogenous changes in households’ incomes to study the behavior of nearby firms.

³For example, Goolsbee and Syverson (2008), Jensen and Miller (2018), and Busso and Galiani (2019) study exogenous sources of domestic entry (both threatened and executed), and Amiti et al. (2019) and Bergquist and Dinerstein (2020) do the same for expansions of domestic competitors. Similar variation can derive from reductions in trade barriers, as exploited by e.g., De Loecker et al. (2016) and Felix (2021).

⁴Such shocks have been studied, for example, by de Mel et al. (2008), Kaboski and Townsend (2011), Banerjee and Duflo (2014), Zwick and Mahon (2017), and Lane (forthcoming) for capital, by de Mel et al. (2019) and Beerli et al. (2021) for labor, by Goldberg et al. (2010) and Boehm and Oberfield (2020) for materials, and by Bloom et al. (2013) and Giorcelli (2019) for management services.

ing these responses allows us to use the techniques described above to test for both forms of allocative efficiency, and to estimate the cost of misallocation.

Our first set of results, in Section 4, begins with an event-study specification that reveals the average time-path of output and input responses to a unit of randomly-determined procurement lottery winnings. Firms rapidly scale up sales as well as labor and non-labor inputs, with effects peaking about 6 months after the lottery and dissipating by 14 months.⁵ Cumulative sales responses are larger than cost responses by a factor of 1.15. Turning to the heterogeneity in these responses, we observe little dispersion in either output or input responses when grouping firms according to a variety of pre-determined firm characteristics, such as baseline sales, profitability, or number of employees, despite the fact that firms are highly heterogeneous along such dimensions. These descriptive findings are therefore consistent with a setting in which wedges are larger than one (in violation of U-AEP) but display limited dispersion across firms (consistent with C-AEP). However, these conclusions could still mask firm-level heterogeneity in unobserved respects.

Section 5 therefore turns to a formal test for both forms of AEP that does not rely on the ability to specify the sources of heterogeneity across firms and inputs. We implement these tests via randomization inference to avoid relying on asymptotic inferential assumptions. Results are consistent with the aforementioned descriptive analysis. The test for U-AEP—a test for the null that all wedges equal one—resoundingly rejects, with a p -value less than 0.001. On the other hand, the test for C-AEP—which asks whether wedges take a common value for each firm, within each input type—does not reject ($p = 0.35$).

In Section 6 we extend this analysis to estimate the cost of misallocation. We use weighted moments of the wedge distribution to provide a second-order approximation to the cost of misallocation, following Baqaee and Farhi (2020). Our results show a sales-weighted mean of wedges across firms of 1.126 and a corresponding variance of wedges of 0.014. Even at conservative values for elasticities of output demand and input supply, these point estimates imply that the total cost of misallocation in this context is just 1.6%.

It may seem surprising that the estimated cost of misallocation in this context is so much

⁵These findings are generally consistent with Fadic (2020), who examines the Ecuadorian lottery system studied here and documents temporary effects on revenues, labor payments, and assets of winning firms.

smaller than the typically substantial costs estimated in the existing literature (discussed below).⁶ However, as argued above, by placing only minimal restrictions on firms’ technologies, it is natural that our procedure may arrive at a lower estimated amount of wedge dispersion. To shed light on this, we can ask what our data would imply if we were to impose the assumption that all firms have technologies that feature common rates of returns-to-scale, as is common in the existing literature on misallocation. Applying this assumption to the same data would lead to the conclusion that the firms in our setting exhibit substantially greater wedge dispersion and a cost of misallocation that is generally an order of magnitude larger (between 4 and 54 times larger across specifications).⁷ An interpretation of our findings is, therefore, that the firms in our context truly have heterogeneous technologies—displaying, at a minimum, heterogeneous degrees of returns-to-scale—and that procedures that restrict such heterogeneity by assumption would overstate the extent of misallocation.

Prior work has focused on quantifying the allocative efficiency of inputs across firms for some time. For example, Restuccia and Rogerson (2008) and Hsieh and Klenow (2009) offer seminal contributions and Hopenhayn (2014) and Restuccia and Rogerson (2017) provide reviews. As discussed above, the typical approach in existing work has been to leverage parametric assumptions about firms’ production functions in order to estimate marginal products, and hence wedges. Such an approach is invaluable if one hopes to identify a separate wedge for every firm and input. However, it may lead to biased wedge estimates if the specification of firms’ technologies is incorrect (for example, in terms of assumed returns-to-scale as in Haltiwanger et al. (2018)) or understates cross-firm technological heterogeneity (as emphasized in Gollin and Udry (2021), for example). Our approach focuses instead on identifying features of the wedge distribution that are sufficient for our questions of interest: whether all wedges equal one (in order to test for U-AEP), whether the wedge distribution is degenerate for each input type (to test for C-AEP), and how large various weighted first and second moments of the wedge distribution are (to quantify the cost of misallocation).

⁶For example, in their meta-analysis of 72 studies of misallocation, Bun et al. (2023) estimate that the average cost of misallocation across studies deploying the production function-based approach referred to here is 66%.

⁷These conclusions are similar across a range of robustness checks. They are also similar across subsamples of firms that differ in the extent of their participation in procurement lotteries, which mitigates a potential concern that wedges on firms’ lottery sales may be smaller or less variable than those on firms’ other activities.

To do so, we leverage insights from the Hall (1988, 2018) method of estimating markups. Hall (2018), for example, uses a time-series regression of a U.S. sector’s output on a weighted bundle of its inputs, with instrumental variables based on military purchases and oil price fluctuations, to identify the sector’s markup under the assumption that the markup is constant over time. We extend this idea in several directions. First, our techniques estimate weighted moments of the wedge distribution within a single cross-section of firms, and hence require no restrictions on markup heterogeneity within the sample.⁸ Second, we apply the methods in Baqaee and Farhi (2020) to provide a mapping from such weighted moments into estimates of the cost of misallocation. And third, consistent with the misallocation literature, our method allows for arbitrary wedges on each input rather than only one overall markup.

Our approach is complementary with that in recent work by Hughes and Majerovitz (2023), who estimate the marginal product of capital using instruments for capital supply and analyze heterogeneity via projections onto observable firm characteristics. This approach delivers a lower bound for the variance, but one that is informative (in terms of ruling out small costs of misallocation) in an application to the seminal de Mel et al. (2008) study, which randomized subsidized capital among Sri Lankan microenterprises.⁹

Finally, as detailed above, an extensive literature has exploited field and natural experiments to examine effects of demand and input supply shocks on firm outcomes. Our paper complements and contributes to this literature by formalizing a method through which firms’ responses to such shocks can be used to test for the existence of allocative inefficiency and to quantify its magnitude.

⁸As discussed below, the Masten and Torgovitsky (2016) estimator is suitable for models like ours, in which coefficients are arbitrarily heterogeneous and regressors are endogenous. This estimator has been previously used by Gollin and Udry (2021) when estimating averages of heterogeneous Cobb-Douglas production function parameters in a model of Tanzanian and Ugandan farms. Klette (1999) also uses a random coefficient estimator to estimate unweighted markup dispersion across firms and years in a panel of Norwegian firms, but under the assumption that regressors are exogenous.

⁹Our approach also relates to work (such as McCaig and Pavcnik (2018), Rotemberg (2019), Bau and Matray (2023), and Sraer and Thesmar (2023)) that has used quasi-experimental designs, combined with existing methods for estimating firms’ wedges (such as that due to Hsieh and Klenow (2009)), to ask the important but distinct question of whether policy changes *change* the extent of misallocation.

2 Theoretical Framework

2.1 Setup

We consider an economy with \mathcal{I} potential firms indexed by i . Each has a technology for converting inputs (indexed by m in the set \mathcal{M}) into products (indexed by j in the set \mathcal{J}). The output and input bundles of firm i are denoted by vectors $\mathbf{y}_i \geq 0$ and $\mathbf{x}_i \geq 0$, respectively. Our goal is to assess the efficiency of production of the allocation observed at some time “0”, and we use a bar over any variable to denote its value at that time.¹⁰

Each firm $i \in \mathcal{I}$ produces outputs \mathbf{y}_i from inputs \mathbf{x}_i via the transformation function

$$F^{(i)}(\mathbf{y}_i, \mathbf{x}_i) \leq 0, \quad (1)$$

that is (by convention) strictly increasing in $(\mathbf{y}_i, -\mathbf{x}_i)$. We make the following assumption about firms’ behavior and technologies.

Assumption 1 (Firms). *For all firms $i \in \bar{\mathcal{I}}$: (a) $F^{(i)}(\mathbf{y}_i, \mathbf{x}_i) = 0$; and (b) $F^{(i)}(\cdot)$ is differentiable at $(\bar{\mathbf{y}}_i, \bar{\mathbf{x}}_i)$ with respect to the outputs in $\bar{\mathcal{J}}(i)$ and the inputs in $\bar{\mathcal{M}}(i)$.*

Part (a) assumes that firms produce on their technological frontiers, but does not otherwise restrict firm behavior (including the extent to which it may be constrained or not profit-maximizing). And (b) rules out cases where a firm is producing at time-0 at a kink in its frontier. Beyond this relatively mild restriction, technologies are free to vary in arbitrary ways across firms and across each firm’s levels of production, including arbitrary extents of increasing or decreasing returns-to-scale, non-homotheticity, and input substitution.

For any firm i , and given any set of prices $(\mathbf{p}_i, \mathbf{w})$ and an allocation $(\mathbf{y}_i, \mathbf{x}_i)$, we define the *wedge* on any input m in firm i ’s production of product j as

$$\mu_{ij,m}(\mathbf{y}_i, \mathbf{x}_i, p_{ij}, w_m) \equiv -\frac{p_{ij}}{w_m} \frac{F_{x_{im}}^{(i)}(\mathbf{y}_i, \mathbf{x}_i)}{F_{y_{ij}}^{(i)}(\mathbf{y}_i, \mathbf{x}_i)}, \quad (2)$$

¹⁰The set of active firms (i.e., those with $\bar{y}_{ij} > 0$ for at least one product j) is $\bar{\mathcal{I}} \subseteq \mathcal{I}$, and the sets of active products and inputs (those produced and used in strictly positive amounts) for firm i are denoted by $\bar{\mathcal{J}}(i)$ and $\bar{\mathcal{M}}(i)$.

where $F_x^{(i)}(\cdot) \equiv \partial F^{(i)}(\cdot)/\partial x$, etc.¹¹ That is, a wedge is the ratio of the value marginal product of an input to the price of that input.¹² Of particular interest is the set of time-0 wedges obtained by applying this definition to the time-0 allocation and prices:

$$\bar{\mu}_{ij,m} \equiv \mu_{ij,m}(\bar{\mathbf{y}}_i, \bar{\mathbf{x}}_i, \bar{p}_{ij}, \bar{w}_m) = -\frac{\bar{p}_{ij}}{\bar{w}_m} \frac{F_{x_{im}}^{(i)}(\bar{\mathbf{y}}_i, \bar{\mathbf{x}}_i)}{F_{y_{ij}}^{(i)}(\bar{\mathbf{y}}_i, \bar{\mathbf{x}}_i)}. \quad (3)$$

Turning to preferences, we specify the utility that household $h \in \mathcal{H}$ derives from consuming firms' outputs and providing firms' inputs as

$$U^{(h)}(\{\mathbf{y}_i^h\}, \{\mathbf{x}_i^h\}), \quad (4)$$

where, for example, y_{ij}^h denotes the amount of y_{ij} consumed by household h .¹³ Feasibility implies that $\sum_h y_{ij}^h \leq y_{ij}$ for all i and j , etc. Households behave in a simple manner:

Assumption 2 (Households). *For each household $h \in \mathcal{H}$: (a) $U^{(h)}(\cdot)$ is differentiable at the point of consumption, $(\{\bar{\mathbf{y}}_i^h\}, \{\bar{\mathbf{x}}_i^h\})$; and (b) $(\{\mathbf{y}_i^h\}, \{\mathbf{x}_i^h\})$ maximizes $U^{(h)}(\cdot)$, subject to the budget constraint $\sum_{i \in \mathcal{I}, j \in \mathcal{J}} p_{ij} y_{ij}^h \leq \sum_{i \in \mathcal{I}, m \in \mathcal{M}} w_m x_{im}^h$, while taking output prices \mathbf{p}_i and input prices \mathbf{w} as given.*

As we shall see, the role of this assumption is to allow observed relative prices to reveal households' marginal rates of substitution.¹⁴

2.2 Allocative Efficiency of Production

We define two notions of allocative efficiency. The first, which we term *unconstrained allocative efficiency of production* (U-AEP), corresponds to the choices made by a hypothetical social planner who faces no aggregate resource constraint. A standard derivation implies

¹¹For expositional simplicity, we refer to derivatives of functions (such as $F^{(i)}(\cdot)$) that are not necessarily differentiable everywhere.

¹²For a single-product firm we can write its technology as $y_i = \tilde{F}^{(i)}(\mathbf{x}_i)$, so that $\mu_{i,m} = \frac{p_i}{w_m} \frac{\partial \tilde{F}^{(i)}(\mathbf{x}_i)}{\partial x_{im}}$.

¹³We use braces " $\{q_k\}$ " to denote a vector whose elements are q_k .

¹⁴Given our focus on allocative efficiency of *production*, Assumption 2(b) effectively normalizes to one any wedges that arise on the household side (ruling out, for example, taxes on consumption or income). More generally, in the case of inputs sourced from other firms rather than households, our analysis normalizes to one any wedges in upstream firms so as to focus on misallocation among the firms under study.

that, under Assumptions 1 and 2, any Pareto-efficient allocation featuring U-AEP (denoted with double stars) would display

$$\mu_{ij,m}(\mathbf{y}_i^{**}, \mathbf{x}_i^{**}, p_{ij}^{**}, w_m^{**}) \equiv \mu_{ij,m}^{**} = 1, \quad (5)$$

for any $y_{ij}^{**} > 0$ and $x_{im}^{**} > 0$.¹⁵

Under the second notion of efficiency, which we term *constrained AEP* (C-AEP), the planner cannot use any more of an input type than is observed in use at time-0. This can be written as:

$$\sum_{i \in \mathcal{I}} x_{im} \leq \sum_{i \in \mathcal{I}} \bar{x}_{im} \equiv \bar{X}_m \quad \text{for all } m \in \mathcal{M}. \quad (6)$$

It is common in the misallocation literature to study allocative efficiency in the presence of such constraints. One benefit of doing so is that misallocation can be decomposed into two sources: (a) misallocation of inputs conditional on \bar{X}_m ; and (b) misallocation in the total input level \bar{X}_m . Another motivation involves a connection to aggregate TFP, which holds aggregate inputs constant by definition.

Under Assumptions 1 and 2, as well as constraint (6), any interior Pareto-efficient allocation featuring C-AEP (denoted with single stars) displays

$$\mu_{ij,m}(\mathbf{y}_i^*, \mathbf{x}_i^*, p_{ij}^*, w_m^*) \equiv \mu_{ij,m}^* = \chi_m, \quad (7)$$

for some $\chi_m > 0$. In this case, the planner's wedges $\mu_{ij,m}^*$ are all equal to some common value χ_m , within the same input type, but this value is not necessarily one as in the U-AEP case (5). The essence of C-AEP is lack of dispersion in wedges rather than their level.

¹⁵This follows from the first-order conditions for an interior solution, which imply

$$\frac{U_{y_j}^{(h)}(\{\mathbf{y}_i^{**}\}, \{\mathbf{x}_i^{**}\}) F_{x_{im}}^{(i)}(\mathbf{y}_i^{**}, \mathbf{x}_i^{**})}{U_{x_{im}}^{(h)}(\{\mathbf{y}_i^{**}\}, \{\mathbf{x}_i^{**}\}) F_{y_{ij}}^{(i)}(\mathbf{y}_i^{**}, \mathbf{x}_i^{**})} = 1 \quad \text{for all } i \in \mathcal{I}, j \in \mathcal{J}, m \in \mathcal{M}, h \in \mathcal{H},$$

and the application of Assumption 2 and wedge definition (2). The prices $(\mathbf{p}_i^{**}, \mathbf{w}^{**})$ are those that would prevail in a decentralized equilibrium that corresponds to the planner's allocation.

2.3 Testing for Allocative Efficiency of Production

We next turn to a test for whether the observed time-0 allocation is allocatively efficient, in either sense. The two null hypotheses to be tested are:

$$H_0 \text{ (U-AEP)} : \bar{\mu}_{ij,m} = 1 \text{ for all } i \in \bar{\mathcal{I}}, j \in \bar{\mathcal{J}}(i), m \in \bar{\mathcal{M}}(i), \quad (8)$$

$$H_0 \text{ (C-AEP)} : \bar{\mu}_{ij,m} = \chi_m \text{ for some } \chi_m > 0 \text{ and for all } i \in \bar{\mathcal{I}}, j \in \bar{\mathcal{J}}(i), m \in \bar{\mathcal{M}}(i).$$

Testing these hypotheses would be straightforward if the wedges $\{\bar{\mu}_{ij,m}\}$ were observable. But each wedge depends, by definition (3), on the marginal product of input m for product j in firm i . This cannot be determined without knowledge of each firm's transformation function $F^{(i)}(\cdot)$.

We overcome this challenge by analyzing a change over time starting at time 0. Let $\Delta y_{ij} \equiv y_{ij,t=1} - \bar{y}_{ij}$ denote the change in any variable from time 0 to some later point in time ($t = 1$). Then, following Hall (1988), we apply a first-order Taylor expansion to the transformation function $F^{(i)}(\cdot)$ around the point $(\bar{\mathbf{y}}_i, \bar{\mathbf{x}}_i)$, as is valid under Assumption 1(b). Using the definition of $\bar{\mu}_{ij,m}$, changes in outputs therefore relate to changes inputs via

$$\sum_{j \in \bar{\mathcal{J}}(i)} \frac{\bar{\mu}_{ij_0, m_0}}{\bar{\mu}_{ij, m_0}} \bar{p}_{ij} \Delta y_{ij} = \sum_{m \in \bar{\mathcal{M}}(i)} \bar{\mu}_{ij_0, m} \bar{w}_m \Delta x_{im} + \varepsilon_i, \quad (9)$$

for any firm $i \in \bar{\mathcal{I}}$, where m_0 and j_0 denote arbitrarily chosen reference inputs and products.¹⁶ Here, ε_i captures any changes in the production function (such as a change in TFP), the consequences of any higher-order terms in the Taylor expansion and of new products $j \notin \bar{\mathcal{J}}(i)$ or inputs $m \notin \bar{\mathcal{M}}(i)$. Importantly, this expression is valid for any set of input changes Δx_{im} and does not take a stand on why (or whether) these changes occur.

As with any test, we seek a function of observables that could distinguish H_0 from an alternative hypothesis (in this case, the existence of misallocation). To do so, we define the

¹⁶For simplicity of notation, this expression assumes that $m_0 \in \bar{\mathcal{M}}(i)$ for all i .

following change in (fixed-price) revenues minus (fixed-price and $\boldsymbol{\chi}$ -adjusted) costs

$$\Delta\Pi_i(\boldsymbol{\chi}) \equiv \sum_{j \in \overline{\mathcal{J}}(i)} \bar{p}_{ij} \Delta y_{ij} - \sum_{m \in \overline{\mathcal{M}}(i)} \chi_m \bar{w}_m \Delta x_{im}. \quad (10)$$

$\Delta\Pi_i(\boldsymbol{\chi})$ is a function of data alone, given any candidate value of $\boldsymbol{\chi}$. Combining equations (9) and (10), this function reduces to $\Delta\Pi_i(\boldsymbol{\chi}) = \varepsilon_i$ under either AEP null.

Finally, we introduce an additional observable, an instrument Z_i , that is assumed to satisfy the following:

Assumption 3 (Exogeneity). *An instrumental variable (IV), denoted Z_i , is observable and satisfies statistical independence with respect to ε_i : $Z_i \perp\!\!\!\perp \varepsilon_i$.*

When exogeneity holds, then $\Delta\Pi_i(\boldsymbol{\chi}) \perp\!\!\!\perp Z_i$ under either AEP null. Since both Z_i and $\Delta\Pi_i(\boldsymbol{\chi})$ are observable under any value of $\boldsymbol{\chi}$, the hypothesis of statistical independence is testable. One method is to consider the following nonparametric relation among all firms $i \in \overline{\mathcal{I}}$

$$\Delta\Pi_i(\boldsymbol{\chi}) = f(Z_i) + \nu_i, \quad (11)$$

for a set of flexible functions $f(\cdot)$ and under the normalization that $f(Z_i)$ is independent of ν_i . Then the AEP hypotheses can be stated as:

$$H_0(\text{U-AEP}) : \text{when } \boldsymbol{\chi} = \mathbf{1}, f(\cdot) = 0, \quad (12)$$

$$H_0(\text{C-AEP}) : \text{for some } \boldsymbol{\chi} > \mathbf{0}, f(\cdot) = 0.$$

We summarize the discussion so far in the following proposition.

Proposition 1. *Suppose Assumptions 1-3 hold. Then a nonparametric test for U-AEP can be performed by estimating the function $f(\cdot)$ in equation (11), evaluating $\Delta\Pi_i(\boldsymbol{\chi})$ at $\boldsymbol{\chi} = \mathbf{1}$, and rejecting the null whenever $f(\cdot) \neq 0$. A nonparametric test for C-AEP can similarly be performed by estimating the function $f(\cdot)$ in equation (11), but while evaluating $\Delta\Pi_i(\boldsymbol{\chi})$ at all $\boldsymbol{\chi} > \mathbf{0}$, and rejecting the null whenever there exists no value of $\boldsymbol{\chi} > \mathbf{0}$ at which $f(\cdot) = 0$.*

The intuition behind the test in Proposition 1 is as follows. Suppose that Z_i is a random demand shock that affects a subset of firms, and that Z_i is positively correlated with $\Delta\Pi_i(\mathbf{1})$

which would correspond to a case in which one estimates an $f(\cdot) \neq 0$. This implies that the demand shock caused some firms' (time 0 price-valued) outputs to grow by more than their (time-0 price-valued) inputs. This is inconsistent with U-AEP because it implies that some wedges are larger than one—or equivalently, that random demand shocks have caused a systematic improvement in the allocation, which would be impossible under efficiency. However, suppose that there does exist a value of $\boldsymbol{\chi} > \mathbf{0}$ at which $f(\cdot) = 0$ for all candidate functions within some flexible set. This would be consistent with C-AEP because it implies that there exists a set of “shadow values” χ_m such that, when inputs are valued in a way that includes χ_m , output changes and input changes are aligned and no improvements to the allocation can be found.

In practice, there are a number of ways to choose functional forms for $f(\cdot)$ to carry out the test of $f(\cdot) = 0$ in equation (11). In Section 5 we follow Ding et al. (2016) and estimate a quantile regression relationship between $\Delta\Pi_i(\boldsymbol{\chi})$ and Z_i and then test for zero effects at all quantiles; this effectively tests that the distribution of $\Delta\Pi_i(\boldsymbol{\chi})$ does not vary across values of Z_i , a necessary condition for their independence. We also employ randomization inference, given that the null is sharp (at any value of the nuisance parameter $\boldsymbol{\chi}$) and the stochastic distribution of Z_i is known in our setting.

Concerning the choice of instruments, there are two considerations. First, Z_i must satisfy Assumption 3. Recall that ε_i consists of three components: (a) changes to the production technology (such as TFP shocks) between time 0 and time 1; (b) non-linear terms in the Taylor expansion of equation (9); and (c) the impact of new types of inputs or outputs. An instrument satisfies component (a) if it is unrelated to the firm's technology. As long as the set of varying inputs is comprehensive, valid examples could include the types of exogenous sources of variation investigated in the studies referred to in the Introduction. This includes changes in demand, in the competitive environment, in the supply of inputs, or in the use of inputs that firms are assumed to allocate before technology shocks are realized. Our application follows the first of these options, using randomized demand shocks from government procurement contracts. Concerns due to (b) can be probed through the use of different values of Z_i that drive larger input changes. For example, we compare effects of relatively small and large demand shocks below. Component (c) causes bias to the extent

that the new inputs or outputs caused by the instrument have large value marginal products but, as equation (7) makes clear, under the null of efficiency these effects are small.

The second consideration in choosing instruments is that the test has greater power against typical alternatives when Z_i has a stronger correlation with input changes $\bar{w}_m \Delta x_{im}$. This is akin to a “first-stage” relevance condition in standard IV settings and is empirically observable.

2.4 Measuring Moments of the Wedge Distribution

Proposition 1 develops nonparametric tests for the existence of misallocation in production. These tests amount to evaluating whether the distribution of wedges $\{\bar{\mu}_{ij,m}\}$ satisfies constancy across firm-products within input types (in the case of C-AEP) or degeneracy at one for all firms, products and inputs (in the case of U-AEP). Our final procedure complements such tests by providing point estimates of features of the wedge distribution, such as its weighted moments (with observed weights). These estimates are themselves of interest, but they also play a key role in estimating the aggregate costs of misallocation.

We begin by rearranging equation (9) as

$$\bar{p}_{ij_0} \Delta y_{ij_0} = - \sum_{j \in \bar{\mathcal{J}}(i)-j_0} \frac{\bar{\mu}_{ij_0,m_0}}{\bar{\mu}_{ij,m_0}} \bar{p}_{ij} \Delta y_{ij} + \sum_{m \in \bar{\mathcal{M}}(i)} \bar{\mu}_{ij_0,m} \bar{w}_m \Delta x_{im} + \varepsilon_i, \quad (13)$$

which holds for each firm $i \in \bar{\mathcal{I}}$. For simplicity, we first focus on firms that produce a single product at time 0 (and therefore drop subscript j temporarily) before returning to multi-product firms. In this case, equation (13) becomes

$$\bar{p}_i \Delta y_i = \sum_{m \in \bar{\mathcal{M}}} \bar{\mu}_{i,m} \bar{w}_m \Delta x_{im} + \varepsilon_i, \quad (14)$$

where we follow the convention that $\Delta x_{im} = 0$ if $\bar{x}_{im} = 0$.

This equation corresponds to a cross-firm regression model relating $\bar{p}_i \Delta y_i$ to a set of regressors given by $\bar{w}_m \Delta x_{im}$ for each $m \in \bar{\mathcal{M}}$. In particular, this model takes the form of an instrumental variables correlated random coefficients (IVCRC) model, for three reasons. First, each unit of observation i not only has its own unobserved intercept ε_i but also its

own unobserved coefficient $\bar{\mu}_{i,m}$ on each regressor. Second, the model features endogeneity, since we expect any regressor (i.e., the change in inputs of type m) may be correlated with the error term. This is because the error term captures changes in the firm’s technology, to which the firm’s input choices are likely to respond. And third, the coefficients $\bar{\mu}_{i,m}$ may be correlated with the error term ε_i , for example, if firms with high markups are more likely to receive productivity shocks. Masten and Torgovitsky (2016) develop tools to study such IVCRC models in cases with suitable instruments, as described in the following.

Assumption 4 (IV). *The econometrician has access to a vector of instruments Z_{im} , one for each $m \in \mathcal{M}$, that satisfies: (a) $Z_{im} \perp (\varepsilon_i, \bar{\mu}_{i,m})$; (b) $\bar{w}_m \Delta x_{im} = h_m(Z_{im}, V_{im})$ for some unknown $h_m(\cdot)$ and scalar V_{im} , with $\frac{\partial h_m}{\partial V_{im}} > 0$ for all $m \in \bar{\mathcal{M}}(i)$; (c) $h_m(Z_{im}, V_{im})$ is strictly monotonic in Z_{im} and there exists variation in Z_{im} at almost every V_{im} for all $m \in \bar{\mathcal{M}}(i)$; and (d) Z_{im} has at least $K + 1$ points of support.*

Parts (a) and (c) are analogous to standard IV requirements—exogeneity and relevance, respectively. We return to part (d) below. Condition (a) requires the instruments to be independent of both the residual ε_i and the wedge $\bar{\mu}_{i,m}$.¹⁷ One *a priori* concern about the validity of part (c) is the presence of adjustment costs, as discussed above. Just as such costs will reduce the power to detect misallocation of inputs that rarely adjust, they may also hinder the ability to estimate the distribution of wedges on such inputs because it can be challenging to find sufficiently strong instruments for them. However, condition (c) is testable (Masten and Torgovitsky, 2016).

Part (b) is unique to IVCRC. Masten and Torgovitsky (2016) refer to this as a “first-stage rank-invariance” condition. It requires that the ranking of firms in terms of their input changes (for any input type m) is the same when all firms (hypothetically) receive a low value of Z_{im} as when they receive a high value. Put differently, firms can respond to Z_{im} in heterogeneous ways, but not in a way that alters their rank in the conditional-on- Z_{im} distribution of changes of input m . This is more likely when firms’ heterogeneity in responses

¹⁷Our wedge estimation procedure is less vulnerable to concerns about the endogeneity of ε_i than the test in Proposition 1 because here it is straightforward to control for higher-order terms in $\bar{w}_m \Delta x_{im}$ (to the extent that separate instruments are available for small and large changes in inputs) as well as for new inputs.

to Z_{im} is small relative to all other reasons for adjusting inputs (i.e. those that are driven by V_{im}). We return to this point in the context of our application in Section 6.

Masten and Torgovitsky (2016) show that, under Assumption 4, a consistent estimator of $\mathbb{E}[\bar{\mu}_m]$ can be constructed for any $m \in \mathcal{M}$ in equation (14)—that is, for the expected value of firms’ wedges $\bar{\mu}_{i,m}$ on input m , with the expectation taken across firms i .¹⁸ We augment this procedure to obtain an estimate of the analogous expectation weighted by an arbitrary vector of firm-specific weights $\{\alpha_m\}$. We denote this expectation by $\mathbb{E}_{\alpha_m}[\bar{\mu}_m]$. This can be achieved by using the regressor $\frac{1}{N\alpha_{im}}\bar{w}_m\Delta x_{im}$ instead of $\bar{w}_m\Delta x_{im}$.¹⁹

Higher-order moments of the wedge distributions can also be estimated consistently by a simple extension. As Masten and Torgovitsky (2016) discuss, the square of equation (14) implies

$$\begin{aligned} (\bar{p}_i\Delta y_i)^2 &= \sum_{m \in \mathcal{M}} \bar{\mu}_{i,m}^2 (\bar{w}_m\Delta x_{im})^2 + 2 \sum_{\substack{m, m' \in \mathcal{M}, \\ m \neq m'}} \bar{\mu}_{i,m}\bar{\mu}_{i,m'}\bar{w}_m\Delta x_{im}\bar{w}_{m'}\Delta x_{im'} \\ &\quad + 2\varepsilon_i \sum_{m \in \mathcal{M}} \bar{\mu}_{i,m}\bar{w}_m\Delta x_{im} + \varepsilon_i^2. \end{aligned} \tag{15}$$

This, too, is an IVCRC model. That is, to the extent that the regressors in equation (15) are not perfectly collinear (a sufficient condition for which is Assumption 4, part (d)), the previous argument can be used to construct a consistent estimator of second-order moments, $\mathbb{E}[\bar{\mu}_m^2]$ and $\mathbb{E}[\bar{\mu}_m\bar{\mu}_{m'}]$.²⁰ By a similar argument to that stated above, weighted moments such as $\mathbb{E}_{\alpha_m}[\bar{\mu}_m^2]$ are also identified. Extensions to third- and higher-order moments are straightforward.

We summarize the preceding discussion in the next proposition:

Proposition 2. *Suppose Assumptions 1-4 hold. Then all leading K th-order weighted mo-*

¹⁸Since Assumption 4 invokes conditions about inputs $m \in \bar{\mathcal{M}}(i)$, the identification of $\mathbb{E}[\bar{\mu}_m]$ for any given $m \in \bar{\mathcal{M}}$ refers to the expectation over firms i for which $m \in \bar{\mathcal{M}}(i)$.

¹⁹In the case of one regressor, $\mathbb{E}_{\alpha}[\bar{\mu}]$ can be approximated by interacting the regressor with indicators for groups based on quantiles of the weighting variable α_i and then constructing the group-weighted average of group-specific estimates. This approximates $\mathbb{E}_{\alpha}[\bar{\mu}]$ increasingly well as the number of groups grows.

²⁰If the number of input types is M , equation (15) suggests that $M(M+1)$ instruments are required for identification. However, an attractive feature of the Masten and Torgovitsky (2016) procedure is that only M instruments are required. This is because the method incorporates the mechanical relationships between basic and derived endogenous regressors (i.e., that the first-stage relationship $\bar{w}_m\Delta x_{im} = h_m(Z_{im}, V_{im})$ is the same no matter how component $\bar{w}_m\Delta x_{im}$ appears in the regression).

ments of the distribution of wedges $\{\bar{\mu}_{i,m}\}$ are identified among $i \in \bar{\mathcal{I}}$ and $m \in \bar{\mathcal{M}}(i)$.

Finally, we return to the multi-product firm from equation (13). The new element here is the additional coefficients $\frac{\bar{\mu}_{ij_0,m_0}}{\bar{\mu}_{ij,m_0}}$ on the regressors $\bar{p}_{ij}\Delta y_{ij}$, with $\bar{J}(i) - 1$ coefficients when firm i produces $\bar{J}(i)$ products at time-0.²¹ These coefficients capture within-firm, cross-product dispersion in wedges, measured by the ratio of the wedge (for an arbitrarily chosen input m_0) on reference product j_0 relative to any other product j . Given suitable instruments, Proposition 2 can be applied to the moments of these distributions too. This leads to the identification of all moments of the distribution of all wedges across firms, within firms across products, and within firms across inputs.

2.5 Quantifying the Cost of Misallocation

Proposition 2 developed a procedure for estimating weighted moments of the distribution of wedges. We now show how such moments can be used to measure the aggregate cost of misallocation. In particular, we compare the U-AEP allocation $(\mathbf{y}^{**}, \mathbf{x}^{**})$ with its wedges of $\{1\}$ to the actual time-0 allocation $(\bar{\mathbf{y}}, \bar{\mathbf{x}})$ with its wedges of $\{\bar{\mu}_{ij,m}\}$. We also decompose this total cost into two terms: (i) the cost of misallocation due to how each input type is allocated across firms (holding aggregate input use constant at \bar{X}_m); and (ii) the cost of misallocation of the aggregate input amounts themselves. Throughout, we focus—à la Harberger—on the second-order expansion to an aggregate welfare function around the point $(\mathbf{y}^{**}, \mathbf{x}^{**})$.²²

Quantifying the total cost of misallocation, up to second-order, amounts to summing up a series of Harberger triangles whose heights are related to the change in wedges (i.e., to the vector $\{\bar{\mu}_{ij,m} - 1\}$) and whose bases are related to the change in quantities of each good produced as a result of the change in wedges. Baqaee and Farhi (2020) emphasize how the cost of misallocation can be written as a set of weighted first- and second-order moments of the wedge distribution, where underlying elasticities of supply and demand govern the importance of each moment. That is, following Proposition 2, one can calculate the value of

²¹This can be executed as follows. Arbitrarily order each firm's $\bar{J}(i)$ products such that one is chosen as the reference product, and hence populates the left-hand side of equation (13), and the remaining $\bar{J}(i) - 1$ products appear as regressors. For firms with less than $\bar{J} \equiv \max_{i \in \bar{\mathcal{I}}} \bar{J}(i)$ products, use the convention of $\Delta y_{ij} = 0$ to populate the missing $\bar{J} - \bar{J}(i)$ regressor values.

²²Since Proposition 2 ensures identification of arbitrary moments of the wedge distribution, our focus on second-order expansions is for simplicity only.

each moment without knowledge of any supply or demand elasticities. But converting such values into a measure of the cost of misallocation does require such knowledge.

Consider the following example. The economy consists of profit-maximizing firms, each using production functions that are arbitrary (and use an arbitrary set of inputs) as long as they satisfy Assumption 1 and display constant returns locally to the time-0 allocation.²³ The representative household has nested CES preferences (with elasticity θ and arbitrary preference weights) over the products of these firms. The household also supplies inputs x_{im} to all firms by deciding how much of its time endowment to consume as leisure or convert into inputs, with an elasticity of substitution η between leisure and the bundle of final consumption goods. Finally, each firm has the same wedge on each of its inputs, which we denote by $\bar{\mu}_i = \bar{\mu}_{i,m}$ for all m .²⁴

The tools in Baqaee and Farhi (2020) make it straightforward to calculate the total cost of misallocation due to wedges. Let $\bar{\lambda}_i \equiv \frac{\bar{p}_i \bar{y}_i}{\sum_{i'} \bar{p}_{i'} \bar{y}_{i'}}$ denote the share of firm i 's sales in total goods consumption and $\bar{\omega}_C$ denote the share of the household's virtual income spent on consumption goods.²⁵ Then, up to a second-order approximation, the increase in welfare (relative to the initial level) from eliminating the wedges is

$$\frac{\Delta W}{W} = \frac{1}{2} \bar{\omega}_C (1 - \bar{\omega}_C) \eta (\mathbb{E}_{\bar{\lambda}} [\bar{\mu}] - 1)^2 + \frac{1}{2} \bar{\omega}_C \theta \text{Var}_{\bar{\lambda}} [\bar{\mu}], \quad (16)$$

where $\mathbb{E}_a[b]$ is the expectation of vector b weighted by vector a and $\text{Var}_a[b] \equiv \mathbb{E}_a[b^2] - (\mathbb{E}_a[b])^2$.

The first term in this expression captures the effect of the average wedge on consumption goods (relative to leisure, which has no wedge), which causes misallocation of the amount of aggregate inputs \bar{X}_m supplied at time-0. It therefore scales with η , the elasticity of input supply, and the extent to which the weighted average wedge $\mathbb{E}_{\bar{\lambda}} [\bar{\mu}]$ among firms differs from the efficient level of one. The second term captures the effect of wedge dispersion across firms, which causes misallocation of any given input supply $\{\bar{X}_m\}_m$. It scales with the demand elasticity θ and the weighted variance of wedges across firms $\text{Var}_{\bar{\lambda}} [\bar{\mu}]$.

²³A natural technology with constant returns locally is one with constant marginal costs (at fixed input prices and wedges) but arbitrary overhead costs.

²⁴This is the case if, for example, the underlying cause of potential wedges is firms' output market power, and/or taxes and subsidies on firms' sales.

²⁵If the household's time endowment is \bar{T} and it earns the price \bar{w} for selling inputs then $\bar{\omega}_C \equiv \frac{\sum_i \bar{p}_i \bar{y}_i}{\bar{w} \bar{T}}$.

As this example makes clear, estimates of a small set of weighted moments of the wedge distribution—along with a small number of additional structural parameters—can be used to estimate the welfare cost of misallocation. Appendix B provides two further examples that illustrate extended versions of this logic. These extensions add: (i) within-firm wedge dispersion across two types of inputs (as in Hsieh and Klenow (2009)), which adds a term involving the weighted variance of within-firm wedges; and (ii) a household with nested CES preferences across sectors and firms (as in Atkeson and Burstein (2008)), which adds a term involving the cross-sector dispersion in the average wedge within each sector.

3 Background and Data

As discussed above, our procedures require an instrument that is correlated with changes in firms' input use but uncorrelated with changes in firms' technologies. We now turn to an application in which we can construct such an instrument based on demand shocks: randomized allocation of certain public procurement contracts in Ecuador. This section describes this procurement lottery process, the data and descriptive statistics regarding the firms that participate in these lotteries, and our empirical strategy.²⁶

3.1 Ecuador's Procurement Lottery System

Starting in 2009, contracts for public construction projects below a certain value were allocated through randomized lotteries among qualified suppliers.²⁷ Examples of such contracts include construction or maintenance of public buildings, small roads and town squares, schools, sewerage, and wells. Included are both physical construction and related services such as those of architects. Lottery contracts represent about 4% of total public procurement.

Government entities aiming to procure small construction contracts first set the parameters of the contract (service to be provided and fixed price) and conduct a screening process to determine which interested firms qualify.²⁸ Among these firms, an automatic and cen-

²⁶Appendix C contains further details on data construction.

²⁷The threshold value is 0.00007% of the central government's annual budget. This corresponds to \$134,176 in 2009 and \$240,100 in 2014.

²⁸Starting in 2013, one qualification criteria was that the total value of lottery contracts a firm enters at

tralized program determines the winner through random selection. Entry into any given lottery is therefore endogenous—the result of selection on both sides. Conditional on entry, however, the allocation of the contract is randomly determined.²⁹

We collected data for each procurement lottery in 2009-2014, including the value, start date and anticipated duration of the contract, the participating firms, and the winner (SERCOP, 2014). We discard contracts for which only one firm qualifies (and hence there is no random assignment). The remaining sample includes 18,474 lotteries with 9,393 firms that participate at least once. The first two panels of Table 1 report summary statistics of these lotteries and of firms' participation. Contracts have a mean value of about \$47,000 (median \$32,000) and usually have a short expected duration, with a mean (and median) of about 2 months. Lotteries have 10 participants on average (median 4). Participation is relatively frequent among lottery firms, averaging about 3.5 times a year (median once a year).

3.2 Firm Data

We match the procurement records to administrative data from Ecuador's tax authority including firm income tax filings, matched employer-employee social security data, and firm-to-firm transaction data from value added tax (VAT) filings. We have annual firm income tax returns for 2008 to 2015, including all line items such as wages, non-wage costs, revenues, and profits (SRI, 2015a). We use these line items to construct total sales and costs: total sales include domestic sales, exports, and other income (e.g., received professional fees); total costs comprise labor and non-labor costs (including capital costs, intermediate input costs, maintenance and repairs, and real estate rent).³⁰

We combine these annual income tax filings with two data sources that contain information on wages, costs and sales on a monthly basis. The first is matched employer-employee social security data, which includes information (available for 2007-2017) on earnings at the worker-firm level (SRI, 2017). And the second is VAT data, which contains third-party re-

any given time be below 0.0007% of the government budget.

²⁹Appendix C includes a more detailed step-by-step description of the lottery process.

³⁰All our sales and costs variables are gross of taxes. The sample includes both incorporated firms, which file the corporate income tax form (F101), and sole proprietorships, which file a combined business and individual income tax return (F102).

ported information about firms' sales as constructed from their clients' purchase annexes (SRI, 2015b). Such purchase annexes must be filed by all incorporated firms, large sole proprietorships, and government agencies as part of their VAT requirements.³¹ We use this data to calculate third-party reported sales by summing all purchases from a supplier across the purchase annexes of all its client entities.³² These monthly data for sales and labor costs, though less complete than the annual data, allow us to better illustrate the dynamic paths of the treatment effects of winning a lottery.

By their nature, administrative tax data reflect reported economic activity. We can mitigate concerns about potential misreporting in two ways. First, we validate our sales estimates, showing that estimated treatment effects on sales are almost identical when using self-reported or third-party reported sales. Second, we find that the treatment effect on total sales is almost entirely accounted for by sales to procuring entities. This is reassuring since these are public entities that have no incentive to misreport tax filings.

Our analysis focuses on the 9,393 firms that ever participate in a multi-participant lottery in 2009-2014.³³ The third panel of Table 1 reports summary statistics of these firms (for the first year they participate in a lottery). Sample firms are on average 11 years old and sell to 5 clients a year. About 18% are incorporated while the rest are sole-proprietorships. Mean annual third-party reported sales are around \$133,000 (with a median of \$48,000, reflecting the typical skewness of firm size distribution). Self-reported annual sales are similar, with a mean of \$141,000. Annual costs are \$124,000 on average, and profits \$17,000. The average number of employees is 4.4, with a wage bill of around \$7,400.

Figure 1 plots the distribution of firm sizes (in terms of sales) both for lottery participants and other firms in Ecuador. Panel 1(a) compares lottery participants to other economically active firms, whereas panel 1(b) compares them to those in the same industry (construction and engineering). The size distributions are broadly similar. Relative to both comparison

³¹Purchase annexes include the value, date and supplier ID of all purchases a firm makes.

³²This measure is a lower bound on total sales since not all client firms file purchase annexes and sales to final consumers are not included. In our sample, the two sales measures are very similar (see Table 1).

³³Firms that are not yet active in 2008 enter the sample when they first appear to be economically active, defined by when they first report any sales or costs or appear in any of the above data sets (as a lottery participant, an employer in the social security data, or as a supplier in another firm's purchase annex). Once a firm is economically active, we impute zeros for any future missing data, since filing nothing is legally equivalent to filing zero.

groups, the distribution of lottery participants is shifted to the right but with less mass in the upper and lower tails of the distribution.

3.3 Using Procurement Lotteries to Construct Demand Shocks

The randomized assignment of contracts provides a source of demand-side exogenous variation in the use of inputs. We use this variation to construct an instrumental variable Z_{it} for each firm i and time period t and then apply this instrument to the testing and estimation procedures described above. The instrument needs to satisfy the statistical independence assumption invoked in Assumptions 3 and 4(a). While any given lottery is akin to a simple randomized trial, pooling across all lotteries is more complex because selection into specific lottery entry is endogenous and may correlate with unobserved determinants of firm growth.³⁴ To address this, we draw on the ideas in Doran et al. (2022) and Borusyak and Hull (forthcoming) to aggregate over all lotteries.

Let k index all procurement lotteries, and \mathcal{K}_{it} denote the set of lotteries firm i enters in period t . Further, let A_k denote the contract value and N_k the number of participating firms in lottery k . The amount of winnings that firm i obtains from lottery k is then a random variable W_{ik} that is binomially distributed (equal to A_k with probability $1/N_k$, and zero otherwise). Hence, a firm’s total winnings in period t is a random variable $W_{it} \equiv \sum_{k \in \mathcal{K}_{it}} W_{ik}$ that is a weighted, binomially distributed random variable, with a p th central moment—which we denote by $M_p(\{A_k, N_k\}_{k \in \mathcal{K}_{it}})$.³⁵ When firms participate in lotteries at different points in time, they are therefore (potentially endogenously) exposed to different probability distributions of randomly-generated winnings. However, conditional on two firms participating in lotteries with the same distribution, the *realization* of the variable W_{it} differs in a purely random manner across the two firms. Put differently, conditional on all moments of $M_p(\{A_k, N_k\}_{k \in \mathcal{K}_{it}})$, W_{it} is independent of any pre-determined firm attributes.

One way to proceed is therefore to use W_{it} as our instrument while controlling for flexible functions of leading moments of $M_p(\{A_k, N_k\}_{k \in \mathcal{K}_{it}})$. However, in practice, we find that

³⁴Since firms can enter multiple lotteries per time period, we cannot use lottery fixed effects to isolate purely random variation.

³⁵For example, the first moment (or expected value) of firm i ’s winnings at time t is simply $M_1(\{A_k, N_k\}_{k \in \mathcal{K}_{it}}) \equiv \mathbb{E}[W_{it} | \{A_k, N_k\}_{k \in \mathcal{K}_{it}}] = \sum_{k \in \mathcal{K}_{it}} \frac{A_k}{N_k}$.

controlling for anything beyond the first moment is quantitatively inconsequential—both when applied to estimators that rely only on mean independence of instruments and those that rely on stronger forms of independence. In addition, as Borusyak and Hull (forthcoming) explain, a simpler procedure that is equivalent to controlling for the first moment is to exploit the fact that, even in the absence of controls, demeaned winnings

$$D_{it} \equiv W_{it} - \mathbb{E}[W_{it} \mid \{A_k, N_k\}_{k \in \mathcal{K}_{it}}], \quad (17)$$

is mean-independent of any firm characteristics, or potential outcomes. Intuitively, even though firms can control expected winnings by choosing which lotteries to enter, they cannot control the random deviations from expected winnings. We refer to D_{it} , firm i 's deviation from expected winnings in time t , as its *procurement winnings shock*.³⁶

We check for balance of randomization by regressing D_{it} on firms' pre-treatment characteristics. Table 2 shows that indeed, there is no statistically significant correlation between these characteristics and the procurement winnings shock.³⁷ In addition, we show below that there are no spurious pre-treatment “effects” for any of our outcomes.

4 Effects of Demand Shocks: Descriptive Results

This section presents descriptive evidence on firms' average responses to procurement demand shocks and explores response heterogeneity by firm characteristics. These results preview our formal tests and quantification of allocative efficiency in Sections 5 and 6.

4.1 Average Treatment Effects

We estimate effects of demand shocks on firm outcomes in an event-study framework by regressing firm outcomes on procurement winnings shocks D_{it} . Formally, for any outcome

³⁶In our analysis below, time periods t are either months or years. By the linearity of expectations, D_{it} at the annual level corresponds to the sum of the monthly D_{it} in each year.

³⁷This is consistent with Brugués et al. (2022), who show that, while political connections influence the distribution of regular procurement contracts in Ecuador, they do not affect contracts allocated by lottery.

Y_{it} , we estimate

$$Y_{it} = \alpha + \sum_{\tau=-T_{lead}}^{T_{lag}} \beta_{\tau} D_{i,t-\tau} + \epsilon_{it}, \quad (18)$$

where α is an intercept, β_{τ} is a coefficient that captures the effect of winnings shock $D_{i,t-\tau}$ (in thousands of USD) on outcome Y_{it} , and T_{lead} and T_{lag} denote the number of included lead and lag coefficients, respectively. Confidence intervals and standard errors are clustered at the firm level.

Sales

Figure 2 displays estimated treatment effects of demand shocks on monthly firm sales (using the third-party reported sales from VAT filings), based on (18). There is a flat pre-trend at zero during the six months prior to the lottery, lending further support to the validity of the lottery randomization. The small spike in sales one month after the lottery represents the fact that many lottery contracts stipulate some up-front payments. Monthly sales rise thereafter, peaking 5 months after the lottery, and dissipate by about 14 months. The temporary demand shock does not leave these firms permanently larger.

We next annualize these third-party reported sales in order to compare them to the self-reported sales from firms' annual income tax returns. Columns (1) and (2) of Table 3 report the effect of \$1,000 in procurement winnings shocks on sales, based on these two measures, in the year of the lottery and the year after.³⁸ The total impact on sales over the two years is \$708 for third-party reported sales and \$669 for self-reported sales.³⁹

Figure 3 examines whether treatment effects vary with contract size. Effects—per \$1000 of procurement winnings shocks—are remarkably similar for contracts of above- and below-median size. The same is true when examining how relatively small firms respond to relatively large contracts and vice versa (Appendix Figure A.2). And responses to demand shocks are very similar regardless of whether the firm has won or lost other lotteries in the recent

³⁸Appendix Figure A.1 shows these results graphically, starting two years before the lotteries. Again, there are no differential pre-trends.

³⁹As discussed below, we find virtually no crowd-out of sales to other clients. The fact that cumulative sales effects are lower than the winnings shock itself therefore derives from two sources: firms that win a contract are somewhat less likely to enter subsequent lotteries and contract values and scope are sometimes modified *ex post*. However, this does not affect our empirical strategy since it only relies on the existence of a random source of firm input changes (not the full pass through of contract winnings into sales growth).

past (Appendix Figure A.3). These findings all imply that responses to demand shocks are approximately linear, suggesting that adjustment costs and the higher-order terms in the Taylor expansion of equation (9) are unlikely to be large in our context.

Costs

We next analyze how firm inputs change given the increase in sales. Figure 4 shows the impact of procurement winnings shocks on labor inputs—employment and wage payments—from the monthly social security data. The time path of effects is similar to that of sales, though there is a minor longer-run effect for labor. Column (3) of Table 3 reports the impact on total costs from the annual firm income tax returns (which include both labor costs and all non-labor costs, such as materials and capital) in the year of the lottery and the year after, which together amount to an increase in total costs of \$583.⁴⁰ This cost increase is driven primarily by an increase in non-labor costs (columns 4 and 5). Finally, column (6) reports the corresponding treatment effects on profits (i.e., column 2 minus column 3).

Overall, these findings have two important implications. First, firms in our context cannot meet additional demand simply by using existing capacity. Second, the rapid scale-up and scale-down of different categories of inputs is again consistent with this being a context with limited adjustment costs.

Price versus quantity adjustments

The procedures from Section 2 rely on the ability to measure the initial period prices and changes in quantities of firm inputs and outputs. As is standard in administrative tax data, prices and quantities are not reported separately in our data (except for the social security data, which tracks individual employees and their wages). However, to the extent that the demand shocks have no effect on prices, our data on sales and costs would capture the appropriate response (initial price times change in quantity).

Figure 5 provides evidence consistent with this in regards to output prices, plotting impacts on four categories of clients that firms sell to.⁴¹ Importantly, there is no appreciable

⁴⁰Appendix Figure A.4 presents estimated dynamic effects on total costs. As with sales, there is a flat pre-trend followed by a sharp increase in the year of the lottery, which dissipates within two years.

⁴¹These mutually-exclusive categories are: (a) procuring entities for which firm i participated in at least one lottery; (b) other procuring entities (i.e., other entities that made at least one purchase through the lottery system) that firm i may sell to via non-lottery means; (c) other public entities; and (d) privately-owned

effect on sales to the private sector: we can reject monthly effects on private sector sales larger than $\pm \$10$ for each \$1,000 dollars in procurement winnings shocks at the 95% level in all periods. If the demand from private sector buyers is at all responsive to prices, as we would expect it to be, then the lack of any change in purchases from lottery winners suggests that the price at which procurement firms sell is not affected by lottery-based demand shocks.

Turning to input price adjustments, in the case of labor we can assess these directly in our data. Appendix Figure A.5 shows wage responses per worker—first for all workers in the firm in a given month (panel a) and second among continuing workers who remained employed at the firm for a spell running from 6 months before the lottery to 18 months thereafter (panel b).⁴² The wages paid to continuing workers are remarkably stable in response to the demand shock, consistent with the assumption of (input) price stability.

Implications for homogeneous wedges

The pioneering Hall (1988) study of US markups assumed: (a) no wedge dispersion within firms (across products or input types, i.e. $\bar{\mu}_{ij,m} = \bar{\mu}_i$ for all j and m), which is consistent with the wedge (i.e. the markup) arising purely from output market power; and (b) no markup dispersion across firms (in a single cross-section of producers) or across time periods (when tracking one producer over time). Under these homogeneity assumptions, the single markup can be estimated from an IV regression of sales on costs while using a demand shock as the instrument. The results in Table 3 provide the reduced-form and first-stage estimates that would correspond to such an IV regression. Under Hall’s (1988) assumptions, taking the ratio of the cumulative demand shock-driven increase in sales in year t and $t + 1$ (column 2) to that for costs (column 3) implies an estimated homogeneous markup of 1.15. This already provides suggestive evidence for the fact that U-AEP—a setting in which all wedges are homogeneous and equal to 1—does not hold in our context. We conduct the formal version of this test, and obtain an estimate of the sales-weighted average wedge $\mathbb{E}_{\lambda} [\bar{\mu}]$ that does not rely on the assumption of wedge homogeneity, in Sections 5 and 6.

firms. The first such category accounts for almost all of the effect on total sales.

⁴²To ensure a balanced sample of workers, this analysis focuses on the time window around the first lottery in which a firm participated.

4.2 Treatment Effect Heterogeneity by Firm Characteristics

Testing for C-AEP concerns not just the average level of the value marginal product of any given input across firms, but also the heterogeneity in such marginal products—or equivalently, the dispersion in wedges $\bar{\mu}_{ij,m}$ in equation (3). In advance of the full estimation of the distribution of such wedges in Section 6, we present here a simple exploration of treatment effect heterogeneity by observable firm characteristics.

First, we look at firms with different levels of pre-treatment sales (i.e., from 2008, the year before the start of the lottery scheme).⁴³ Panel (a) of Figure 6 analyzes impacts of lottery-driven demand shocks on sales separately for firms with above- and below-median pre-treatment sales. These two estimates are strikingly similar. Panel (b) further disaggregates firms into quintiles of pre-treatment sales, again with similar results. Responses are also homogeneous across firms that differ along other pre-treatment characteristics: profitability (sales divided by costs), number of employees, number of suppliers, labor intensity (labor costs divided by sales), and third-party reported sales (Appendix Figures A.6–A.10).

Homogeneity in firms’ sales responses to a demand shock does not necessarily imply a low dispersion in firms’ marginal products because firms could differ in the amount of inputs needed to achieve these output responses. However, as Figures 7 and A.11 make clear, we also see very homogeneous responses in labor and total costs across firms of different sizes. Combined, Figures 6, 7, and A.7–A.11 therefore paint the picture of a particularly strong form of marginal product homogeneity, with low heterogeneity across firms in both their output and input responses to a demand shock.⁴⁴

A key limitation of the evidence discussed so far is that it only considers observable characteristics. The methods in the next two sections are designed to test for, and quantify, wedge heterogeneity even when allowing for arbitrary forms of potential heterogeneity across firms.

⁴³This analysis exclude firms with no sales in 2008.

⁴⁴A finding of homogeneous output and input responses is sufficient, but not necessary, for the homogeneity of value marginal products (since marginal products concern the ratio of the two responses, so heterogeneity in output and input responses could offset one another while maintaining marginal product homogeneity).

5 Testing for Allocative Efficiency

5.1 Test Implementation

Proposition 1 proposes a test for AEP in both its unconstrained and constrained forms. To implement this test, we evaluate $\Delta\Pi_i(\boldsymbol{\chi})$ for each firm i at the null-hypothesis value of $\boldsymbol{\chi} = \mathbf{1}$ in the case of U-AEP and at a range of values of $\boldsymbol{\chi} > 0$ in the case of C-AEP. In either case, we then test for the statistical independence of $\Delta\Pi_i(\boldsymbol{\chi})$ from the instrument Z_i (built from lottery-driven demand shocks) via the test for $f(\cdot) = 0$ in equation (11). When statistical independence is rejected, allocative efficiency is rejected too.

We follow Ding et al. (2016), who develop procedures for tests of treatment effect heterogeneity in randomized trials.⁴⁵ This is relevant to our setting since the test for C-AEP is analogous to a test for no heterogeneous responses (of firms' outputs to an exogenously driven increase in their input use) and the U-AEP test is a nested case in which, under the null, responses are not just homogeneous but also equal to one. In particular, we estimate quantile regressions (for nine evenly spaced quantiles) of $\Delta\Pi_i(\boldsymbol{\chi})$ on Z_i . We then use the largest coefficient (in absolute value) as our test statistic TS . If any coefficient differs substantially from zero, and hence TS is large, then the null of $f(\cdot) = 0$ is false and hence the hypothesis of independence should be rejected.

We then take advantage of the fact that the null hypothesis is sharp, and the distribution of Z_i in our setting is known, in order to use randomization inference (Fisher, 1935) to calculate the probability of obtaining a given value of the test statistic under the null. This avoids the need to make asymptotic distributional assumptions about ε_i in equation (9) (such as how technology shocks are correlated across time and firms).

We simulate the null distribution by calculating TS_l in a set of simulations $l = 1 \dots L$, in which the winning firm of each procurement lottery at time 1 is drawn randomly from its known stochastic process. The p-value for a given test is the percentage of simulations in which the value of TS when using the actual lottery winners is smaller than TS_l . Intuitively, when statistical independence of $\Delta\Pi_i(\boldsymbol{\chi})$ and Z_i is violated, we should see a stronger

⁴⁵Further details about our test procedures are described in Appendix D.

correlation between $\Delta\Pi_i(\boldsymbol{\chi})$ and Z_i in the actual data than is likely to have occurred by chance—that is, we should see a larger value of TS for actual lottery realizations than the values TS_l found in many simulations l of alternative lottery realizations. If so, this indicates that the input reallocations caused by the demand shocks generate changes in aggregate output. Via Proposition 1, this implies the economy is not allocatively efficient at time 0.

This procedure can be performed on any cross-section of changes (i.e., from any “time-0” to any “time-1”). In order to ensure that we capture the entire two-calendar year time-path of responses seen above, we use changes in two-year averages; that is, the average of years $t - 1$ and $t - 2$ is “time-0” and that of years t and $t + 1$ is “time-1”. The change $\Delta\Pi_{it}(\boldsymbol{\chi})$ in equation (10) is defined analogously.⁴⁶ Similarly, we define the instrument for firm i in year t as $Z_{it} \equiv D_{i,t} - D_{i,t-2}$. Since our test can be performed on any set of changes, we stack all five such changes (given our data from 2008 to 2015) and perform the joint test that the null hypothesis of U-AEP is true throughout.

5.2 Results

Figure 8 reports the p-value of the test described above when $\Delta\Pi_{it}(\boldsymbol{\chi})$ is evaluated at a range of values of $\boldsymbol{\chi}$ (which confines attention to the case in which all inputs m have the same value of χ_m —denoted by the scalar χ —so that one dimension of the $\boldsymbol{\chi}$ space can be easily illustrated). Beginning with the test for U-AEP ($\boldsymbol{\chi} = \mathbf{1}$), this corresponds to the point $\chi = 1$ on the x-axis. The p-value at this point is extremely low (below 0.001), which indicates that the null of unconstrained allocative efficiency (all wedges $\bar{\mu}_{ij,m}$ are equal to one) is resoundingly rejected. This implies that the treatment effect of demand-driven cost increases on sales is (at standard levels of significance) different from one, a finding that was already anticipated by the discussion in Section 4.1.

We now turn to an evaluation of C-AEP. Testing the null of C-AEP in principle involves conducting our previous test at every value of χ_m , separately for each input type m , and checking whether there is a value of χ_m (potentially a separate one for each input type) at

⁴⁶For example, the component $\bar{w}_m \Delta x_{im}$ in $\Delta\Pi_i(\boldsymbol{\chi})$ is defined as $[\frac{1}{2}(\bar{w}_{m,t-1} + \bar{w}_{m,t-2})][\frac{1}{2}(x_{im,t} + x_{im,t+1}) - \frac{1}{2}(x_{im,t-1} + x_{im,t-2})]$. This requires separate data on quantities and prices, which is not always possible in our data. However, as shown in Section 4.1, the potential violation of Assumption 3 caused by this measurement issue seems likely to be small.

which the test does not reject. Figure 8 shows results from a partial version of this multi-dimensional search, in which χ_m is common (i.e., $\chi_m = \chi$) for all inputs. As it turns out, even this partial search reveals a value at which the test does not reject at standard levels of statistical significance. The p-value of the randomization inference test peaks sharply at the value of $\chi = 1.140$ (where the p-value is 0.35). The null of constrained allocative efficiency is therefore not rejected in this setting.

These findings are consistent with the results in Section 4.2 but go beyond them in three respects. First, they test for cross-firm wedge heterogeneity along arbitrary, unobserved dimensions. Second, they evaluate heterogeneity in value marginal products rather than exploring heterogeneity in output and input responses separately. And third, they allow for sampling variation and deliver a formal p-value.

6 The Cost of Misallocation

Our final analysis goes beyond testing and aims to arrive at point estimates of moments of the wedge distribution as well as an estimate of the welfare cost of misallocation in our context.

6.1 Estimating Moments of the Wedge Distribution

We now apply the result in Proposition 2 to estimate weighted moments of the wedge distribution in our context. In particular, we focus on the two moments— $\mathbb{E}_{\bar{\lambda}}[\bar{\mu}]$ and $\mathbb{V}ar_{\bar{\lambda}}[\bar{\mu}]$ —appearing in the formula for the cost of misallocation in equation (16).

Recall that this expression was derived for a setting in which there is assumed to be no within-firm dispersion of wedges across inputs. This assumption is necessary since our demand-driven source of exogenous variation is unsuited to recover such within-firm dispersion in our context.⁴⁷ It is therefore reassuring that the component of misallocation due to

⁴⁷As Proposition 2 explains, both within-firm and across-firm dispersion can be identified if a separate instrument Z_{im} is available for each type of input, m . It is possible, in principle, to use demand shocks as the source of such input-specific instruments, to the extent that firms' technologies are non-homothetic and the econometrician has access to observable proxies for such features, which would require survey data on firm production.

within-firm wedge dispersion is typically found to be relatively small (e.g., about 15% of the estimated total misallocation cost for China in Hsieh and Klenow (2009, p. 1442)).⁴⁸

In a setting with no within-firm wedge dispersion equation (14) becomes

$$\bar{p}_i \Delta y_i = \bar{\mu}_i \sum_{m \in \mathcal{M}} \bar{w}_m \Delta x_{im} + \varepsilon_i. \quad (19)$$

We begin by applying the procedure described in Section 2.4 to this equation to estimate the sales-weighted first moment of the distribution of coefficients $\bar{\mu}_i$ on the single endogenous regressor $\sum_{m \in \mathcal{M}} \bar{w}_m \Delta x_{im}$, while using a single instrument Z_i constructed from lottery-based demand shocks. Provided that the instrument satisfies Assumption 4, which we discuss below, this delivers a consistent estimate of $\mathbb{E}_{\bar{\lambda}}[\bar{\mu}]$.⁴⁹ Squaring equation (19) yields a second regression specification

$$(\bar{p}_i \Delta y_i)^2 = (\bar{\mu}_i)^2 \left(\sum_{m \in \mathcal{M}} \bar{w}_m \Delta x_{im} \right)^2 + 2\bar{\mu}_i \varepsilon_i \sum_{m \in \mathcal{M}} \bar{w}_m \Delta x_{im} + (\varepsilon_i)^2. \quad (20)$$

Applying our procedure to this equation, the coefficient on the regressor $(\sum_{m \in \mathcal{M}} \bar{w}_m \Delta x_{im})^2$ provides a consistent estimate of $\mathbb{E}_{\bar{\lambda}}[(\bar{\mu})^2]$. We then calculate the centered second moment from $\text{Var}_{\bar{\lambda}}[\bar{\mu}] = \mathbb{E}_{\bar{\lambda}}[(\bar{\mu})^2] - (\mathbb{E}_{\bar{\lambda}}[\bar{\mu}])^2$.⁵⁰

Many details of our implementation are analogous to those in the test procedure in Section 5. In addition, the implementation of the IVCRC procedure requires the choice of a bandwidth and kernel (these are used in a sub-component of the estimation routine that obtains a smoothed estimate of the expected value of the coefficient at each rank of the conditional first-stage distribution). Our baseline analysis follows the defaults in Benson et al. (2022) and uses the rule-of-thumb bandwidth proposed by Fan and Gijbels (1996) and an Epanechnikov kernel. We explore alternatives below.

⁴⁸If wedges are truly dispersed across inputs then the single-instrument version of our procedure will recover moments of the distribution of each firm's (cost share-weighted) average wedge, $\frac{\sum_m \bar{w}_m \bar{x}_{im} \bar{\mu}_{im}}{\sum_m \bar{w}_m \bar{x}_{im}}$, to the extent that the instrument is uncorrelated with changes in relative input prices as seems plausible in our context (and is consistent with the results in Section 4.1).

⁴⁹As discussed in Section 2.4, we obtain the weighted expectation of the distribution of coefficients $\bar{\mu}_i$ by interacting the regressor with indicators for five bins based on firms' 2008 sales shares (the quintiles from Figure 6(b)). Our sensitivity analysis below explores this choice.

⁵⁰We impose that variance must be non-negative by setting $\text{Var}_{\bar{\lambda}}[\bar{\mu}] = \max\{0, \mathbb{E}_{\bar{\lambda}}[(\bar{\mu})^2] - (\mathbb{E}_{\bar{\lambda}}[\bar{\mu}])^2\}$.

Because the IVCRC estimand cannot be represented as a sharp null hypothesis, randomization inference approaches (as applied in Section 5) cannot be used for confidence intervals. We therefore use a block-bootstrap procedure (with 100 bootstrap samples created by drawing firms with replacement) to assess the uncertainty in our estimates. To deal with outliers, to which the quadratic form in equation (20) may be especially sensitive, we trim the sample at the top and bottom 1% of the ratio of (in absolute value) change in revenues to change in costs. We report a range of alternatives below and show this to be a conservative approach.

Finally, we require the instrument to satisfy Assumption 4. The instrument Z_{it} , built from lottery-driven demand shocks, plausibly satisfies independence, part (a) of this assumption. The strong input adjustment responses seen in Section 4.1 provide support for part (b), instrument relevance. And the wide range of lottery sizes means that Z_{it} features multiple points of support and hence satisfies part (d). But it is not *a priori* clear whether it satisfies part (b), first-stage rank-invariance. To explore the accuracy of this assumption, Appendix E presents a simulated version of the economy assumed here, but where firms have a known technology (which, following Hsieh and Klenow (2009), we assume exhibits constant returns-to-scale). We then calibrate this simulation to match our data and calculate the extent of first-stage rank reversals in response to lottery-driven demand shocks, according to the metric proposed by Gollin and Udry (2021). This simulation demonstrates that such reversals are rare because firm heterogeneity in input growth due to non-lottery factors is much greater than that due to any amount of lottery shocks. Violations of Assumption 4(b) therefore seem unlikely to be consequential for our conclusions.⁵¹

Results

The first row of Table 4 (panel a) reports our baseline IVCRC estimates of $\mathbb{E}_{\bar{\chi}}[\bar{\mu}]$ and $\text{Var}_{\bar{\chi}}[\bar{\mu}]$: 1.126 and 0.014, respectively. These values are consistent with the results in the previous section, where we saw that, at the common wedge value of $\chi = 1.140$, the test of C-AEP fails to reject. We should therefore expect a value for an average wedge of approximately that χ value, with little variation in wedges around it. And that is exactly what our point

⁵¹As discussed in Section 2.4, an IV that satisfies first-stage rank-invariance when estimating equation (19) will also satisfy it for equation (20). This is because the first regressor is an endogenous variable but one that is derived from a known function of the second regressor, which is itself the only regressor in (19).

estimates of $\mathbb{E}_{\bar{\lambda}} [\bar{\mu}]$ and $\text{Var}_{\bar{\lambda}} [\bar{\mu}]$ imply.

Table 4 also reports the block-bootstrapped confidence intervals—a two-sided interval for $\mathbb{E}_{\bar{\lambda}} [\bar{\mu}]$, but a one-sided version for $\text{Var}_{\bar{\lambda}} [\bar{\mu}]$ since we know that the true value of this variance cannot be negative. The implied uncertainty about $\mathbb{E}_{\bar{\lambda}} [\bar{\mu}]$ is minimal, and we can hence reject an average wedge equal to one at standard levels of statistical significance. But the confidence interval for $\text{Var}_{\bar{\lambda}} [\bar{\mu}]$ is considerably wider. To explore this, panel (a) of Appendix Figure A.12 plots the 100 bootstrap sample estimates of the variance of wedges that underpin this confidence interval. The vast majority of estimates are very low, but there is also a tail of large estimates in absolute value. This is perhaps unsurprising given the quadratic nature of the terms in the estimating equation (20).

The next five rows of Table 4 describe the robustness of our estimates to alternative specification choices. For example, rather than trimming at the top and bottom 1% level, we can either not trim at all (row 2) or do so at the top and bottom 2.5% level (row 3). Or, rather than using a Epanechnikov kernel for the IVCRC estimator, we can use a Gaussian or uniform kernel (rows 4 and 5). Finally, rather than calculating the weighted moments $\mathbb{E}_{\bar{\lambda}} [\bar{\mu}]$ and $\text{Var}_{\bar{\lambda}} [\bar{\mu}]$ by using five bins of firm sales $\bar{\lambda}$, we can use ten bins. The estimates of $\mathbb{E}_{\bar{\lambda}} [\bar{\mu}]$ are highly insensitive to these variations, with point estimates ranging from approximately 1.11 to 1.13. And the estimates of $\text{Var}_{\bar{\lambda}} [\bar{\mu}]$ are always small in the sense that they are less than or equal to the baseline estimate of 0.014.

6.2 Estimating the Cost of Misallocation

We now use the estimates of $\mathbb{E}_{\bar{\lambda}} [\bar{\mu}]$ and $\text{Var}_{\bar{\lambda}} [\bar{\mu}]$ in Table 4 to calculate the total cost of misallocation in our context, following the formula in equation (16). Doing so requires a value for the two elasticity parameters: that across firms in consumption (θ) and that for aggregate input supply to these firms (η). For the former, we follow Hsieh and Klenow (2009) to set $\theta = 3$; and for the latter we set $\eta = 3$, a value at the upper end of estimates in the literature on labor supply. In addition, equation (16) depends on $\bar{\omega}_C$, the share of the representative household’s shadow income that is spent on consumption of goods. We begin with the midpoint value of $\bar{\omega}_C = 0.5$ but, as we document below, our conclusions are not very sensitive to the value chosen.

The results we obtain when using these parameter values are reported in column (3) of Table 4 (panel a). The baseline point estimate is $\Delta W/W = 0.016$. This implies that the total cost of misallocation is 1.6% of the overall expenditure on products in our model economy (i.e. on the goods and services produced by the firms taking part in procurement lotteries, and the leisure consumed by households). The block-bootstrapped 95% one-sided confidence interval spans from zero to 26.1%. We therefore cannot rule out considerably larger values at standard levels of confidence. However, as with the case of $\text{Var}_{\bar{\lambda}}[\bar{\mu}]$ discussed above, the distribution of bootstrap estimates (reported in panel (b) of Appendix Figure A.12) demonstrates that this result is highly affected by the tails of this distribution; for example, 80% of the bootstrap values of $(\Delta W/W)$ lie below 7%.

The remaining rows of Table 4 (panel a) describe a range of robustness checks on our conclusions about the total cost of misallocation $\Delta W/W$. The first five rows extend the sensitivity checks on $\mathbb{E}_{\bar{\lambda}}[\bar{\mu}]$ and $\text{Var}_{\bar{\lambda}}[\bar{\mu}]$ discussed earlier. Two additional rows consider alternative values of $\bar{\omega}_C$ (of 0.75 and 0.25) centered around our baseline value of 0.5. These seven alternatives suggest that our estimation choices have been conservative. The only exception is $\bar{\omega}_C = 0.75$. But if we were to set this parameter to its maximal value of $\bar{\omega}_C = 1$, this would imply an upper bound on losses from misallocation equal to 2.1%.

As discussed above, the total cost of misallocation $\Delta W/W$ can be decomposed into two components due to: (a) departures from C-AEP while holding constant the availability of aggregate inputs; and (b) misallocation of the aggregate input amounts themselves. Component (b) can be calculated by a transformation of the values in column (1). Doing so, we obtain 0.6% throughout the alternative specifications. Component (a) is a transformation of the values in column (2). It ranges from zero to 1.05% across alternative specifications.

Put together, the estimates in Table 4 (panel a) imply that the allocation of inputs in our context, both to firms and across them, appears to be close to the efficient point. In interpreting this finding, it is important to recall that it refers to the efficiency of the actual allocation, rather than the mechanisms through which this allocation arises. For example, given active government involvement in this sector, our findings do not necessarily imply that a *laissez-faire* policy stance would achieve near-efficiency.

6.3 Comparison to Method Assuming Common Scale Elasticities

The results in Section 6.2 may seem surprising, especially when compared to previous work that typically estimates large costs of misallocation. One possibility is that firms in Ecuador’s construction services sector are simply different from those in other contexts. However, as discussed in the introduction, existing methods typically assume that firms use similar production technologies. This could create a biased impression of misallocation—especially the heterogeneity in wedges inherent to departures from C-AEP—if firms’ actual technologies are more heterogeneous than assumed.

One way of seeing this is to start from the definition of wedges in equation (2) when applied to a single-product firm whose production function is $y_i = \tilde{F}^{(i)}(\mathbf{x}_i)$. Then, multiplying both sides by $(\bar{w}_m \bar{x}_{im})/(\bar{p}_i \bar{y}_i)$, summing across all inputs $m \in \mathcal{M}$, and focusing on the case with no intra-firm wedge dispersion (i.e., $\bar{\mu}_{i,m} = \bar{\mu}_i$ for all m), we obtain

$$\bar{\mu}_i = \left(\frac{\bar{p}_i \bar{y}_i}{\sum_{m \in \mathcal{M}} \bar{w}_m \bar{x}_{im}} \right) \left(\sum_{m \in \mathcal{M}} \frac{\bar{x}_{im}}{\bar{y}_i} \frac{\partial \tilde{F}^{(i)}(\bar{\mathbf{x}}_i)}{\partial x_{i,m}} \right) = \left(\frac{\bar{p}_i \bar{y}_i}{\sum_{m \in \mathcal{M}} \bar{w}_m \bar{x}_{im}} \right) \gamma_i(\bar{\mathbf{x}}_i), \quad (21)$$

where $\gamma_i(\bar{\mathbf{x}}_i) \equiv (\lambda/\bar{y}_i)(\partial \tilde{F}(\lambda \bar{\mathbf{x}}_i)/\partial \lambda)$ is firm i ’s scale elasticity at $\bar{\mathbf{x}}_i$. That is, firm i ’s wedge is equal to the product of its profitability (the ratio of total sales, $\bar{p}_i \bar{y}_i$, to total costs, $\sum_{m \in \mathcal{M}} \bar{w}_m \bar{x}_{im}$) and its scale elasticity.

Intuitively, knowledge of the scale elasticity $\gamma_i(\bar{\mathbf{x}}_i)$ allows an analyst to convert estimates of average products (such as profitability) into estimates of marginal products (which is what wedges depend on). Similarly, when scale elasticities are assumed to be common across firms, then any cross-firm dispersion in average products is mapped one-to-one into conclusions about dispersion in marginal products, and hence about misallocation. A common version of this is the assumption that all firms use technologies that are globally constant returns-to-scale, or $\gamma_i(\bar{\mathbf{x}}_i) = 1$ at all $\bar{\mathbf{x}}_i$.⁵² A prominent example appears in Hsieh and Klenow

⁵²This assumption is stronger than those we invoke to arrive at the estimates in Table 4 (panel a), in two respects. First, our estimates of moments of the wedge distribution require only that firms use locally differentiable technologies, and hence only that local returns-to-scale are finite. Second, the formula for $\Delta W/W$ in equation (16) invokes the assumption that all firms have *locally* constant returns-to-scale technologies, which has no direct relation to globally constant returns-to-scale; for example, the local version allows for technologies with arbitrary overhead costs, whereas the global version rules out overhead costs.

(2009).

Table 4 panel (b) explores the implications of such parametric restrictions in our context. We do so by imposing the assumption that $\gamma_i(\bar{\mathbf{x}}_i) = \gamma$ for all firms i . This continues to allow for firms' technologies to combine inputs in arbitrarily heterogeneous ways, but requires that firms share a common scale elasticity.⁵³ We begin with the constant returns-to-scale case in which $\gamma = 1$. On the basis of this assumption, computing the wedge $\bar{\mu}_i$ for each firm is a straightforward application of equation (21): each firm's wedge $\bar{\mu}_i$ is equal to its profitability. We then calculate the sales-weighted moments $\mathbb{E}_{\bar{\lambda}}[\bar{\mu}]$ and $\text{Var}_{\bar{\lambda}}[\bar{\mu}]$ of the distribution of such wedges across all firms in 2013.⁵⁴ Row 1 reports these estimates, along with their bootstrapped 95% confidence intervals.

A first clear message resulting from these estimates is that the assumption of constant returns-to-scale does not substantially affect the estimate of the first moment $\mathbb{E}_{\bar{\lambda}}[\bar{\mu}]$: this value rises, but only to 1.240 (in panel b), from our preferred value of 1.126 (in panel a). However, a second clear finding is that assuming constant returns does substantially affect the estimate of the second moment, $\text{Var}_{\bar{\lambda}}[\bar{\mu}]$, which rises from 0.014 in panel (a) to 0.611 in of panel (b). This divergence in estimates of wedge heterogeneity is consistent with the findings discussed in Section 4.2. In particular, if all firms truly exhibited constant returns then—as follows from substituting equation (21) into equation (19)—we would see that their responses to demand shocks would exhibit heterogeneity in accordance with their baseline profitability. Yet such heterogeneous responses are not apparent in Appendix Figure A.6.

Putting columns (1) and (2) together, we see in column (3) row 1 that the estimated total cost of misallocation grows considerably (from 1.6% to 47.9%) as a result of the assumption that all firms use constant returns-to-scale technologies. The bootstrapped confidence interval on this estimate ranges from 42.7% to 57.2%, which does not overlap with the corresponding interval of our preferred specification.

One distinction between the two methods compared here concerns the presence of multi-product firms. Our application of the IVCRC method using a lottery-based instrument will

⁵³Formally, this requires that $\tilde{F}^{(i)}(\bar{\mathbf{x}}_i) = G(g^{(i)}(\bar{\mathbf{x}}_i))$ where $G(\cdot)$ is homogeneous of degree γ and $g^{(i)}(\cdot)$ is normalized to be homogeneous of degree one but otherwise arbitrary.

⁵⁴Following Hsieh and Klenow (2009), these calculations remove the smallest and largest 1% of wedges. Our baseline analysis uses 2013 because it is the year with the median estimated cost of misallocation in our 2008-2015 sample; we discuss alternative year estimates below.

estimate the distribution of wedges in the lottery sector and not that of other sales.⁵⁵ By contrast, a multi-product version of equation (21) with $\gamma = 1$ for all products will estimate a (cost-weighted) average of the wedges across all of each firm’s products. To the extent that wedges on procurement lottery products might be different from those on non-lottery products, these methods could therefore arrive at different conclusions because they are estimating different objects. Figure 9 suggests that this is not the case. Here we zoom in on sub-samples of firms with varying shares of their total sales (over 2009-2014) stemming from contracts obtained through the lottery system. We see that the IVCRC misallocation estimates are consistently lower than those implied by equation (21), even when examining the approximately 1,500 firms for which lottery contracts make up more than half of their total sales.

Sensitivity Analysis

Our final analysis examines various alternative versions of the comparisons in Table 4. First, the remaining rows of Table 4 explore how our conclusions change with $\gamma \neq 1$. Row 2 uses a value of decreasing returns ($\gamma = 0.85$) and row 3 one of increasing returns ($\gamma = 1.15$). As is clear from equation (21), given any data on firms’ sales-to-cost ratios, estimated wedges increase with the value of γ . Unsurprisingly, therefore, estimates in row 1 are straddled by those in rows 2 and 3. The total estimated cost of misallocation ranges from 33.2% in row 2 to 67.4% in row 3. Even the lowest of these remains many times larger than our preferred estimate of 1.6%.

Second, we repeat our investigation of the common scale economies assumption for each year separately. Appendix Table A.1 shows these results (for $\gamma = 1$). The estimated cost of misallocation ranges from 38.5% to 86.9%. This variability over time stems almost entirely from the estimates of $\text{Var}_{\bar{\lambda}}[\bar{\mu}]$.

Finally, recall that the derivation of equation (16) was based on a second-order expansion around a point with wedges $\bar{\mu}_i \simeq 1$ for all i . We expect this approximation to be accurate for our preferred wedge estimates in panel (a), which lie close to this value. However, applying this approximation to the case of the more dispersed wedge estimates in panel (b)

⁵⁵As equation (13) demonstrates, this conclusion is not general. But it follows in our context as a result of the fact that (as seen in Figure 5) our instrument has no effect on firms’ sales to non-lottery clients.

may result in bias of unknown sign. Hsieh and Klenow (2009) show that one useful case—where $\gamma = 1$, and firms’ TFPs and wedges are jointly log-normally distributed—results in no approximation error at all, as long as the formula in equation (16) is appropriately adjusted to reflect *unweighted* moments of *log* wedges. Applying this finding, we obtain a cost of misallocation equal to 55.7% (Appendix Table A.2 row 1), which suggests that the use of the approximation in equation (16) is conservative. On the other hand, if we derive the analog of equation (16) for an expansion around the point of $\ln \bar{\mu}_i \simeq 0$ for all i , then the appropriate formula involves *weighted* moments of log wedges. In this case (Appendix Table A.2 row 2) we estimate the cost of misallocation to be far smaller, at 6.3%, though this is still four times larger than our preferred estimate of 1.6%.

Taken together, the findings in Table 4 suggest that the firms in our setting do have heterogeneous degrees of scale economies, and that a measurement approach that restricts this type of technological heterogeneity by assumption would infer more misallocation from the data in this setting than appears to be correct. When the objective is to learn each firm’s wedge individually, assumptions about firm technologies (such as a common scale elasticity) are indispensable. But such assumptions may be overly restrictive when the goal is to estimate moments of the distribution of firms’ wedges, as is typically sufficient for quantifying the cost of misallocation.

7 Conclusion

In this paper we have developed new tools for assessing the allocative efficiency of production, both in the absence and presence of an aggregate input constraint, among any given set of firms. We have developed new procedures that can estimate features of the distribution of wedges—ratios of the value marginal product of an input divided by its price—across all firms, products, and inputs in an economy. By drawing on sources of exogenous variation in firm input changes, these methods proceed without the need to specify production functions, demand functions, or the underlying nature of the distortions that lead to potential inefficiency. Instead, they simply seek to estimate the “treatment effects” of (price-adjusted) inputs on outputs across firms. This then allows for a comparison of such treatment effects

to an idealized efficient allocation, in which they would not differ across input uses.

Our results imply that the firms in the context of our application produce at an allocation that is close to allocative efficiency. This holds both in terms of firms' relative use of given aggregate inputs and in terms of the amount of the aggregate input levels themselves. Our analysis would arrive at a different conclusion if we were to apply commonly-used parametric approaches that assume firms use technologies with similar features (such as a common degree of returns-to-scale). This may be unsurprising given that the essence of our approach has been to allow for technological heterogeneity that could create misleading impressions of misallocation if it were ignored.

Our methods rely on researchers' access to exogenous variation in firm input use, such as that derived from the lottery-based demand shocks at our disposal. The studies highlighted in the Introduction document a wide range of settings where ingenious sources of variation in output demand and input supply—policy changes, income shocks, market integration, bank expansions, exchange rate fluctuations, immigration events, narrow auction winnings, etc.—have been identified by researchers for the purposes of assessing how firms respond to such forces. The procedures developed in this paper provide a toolkit for using such variation to assess misallocation in firm production.

References

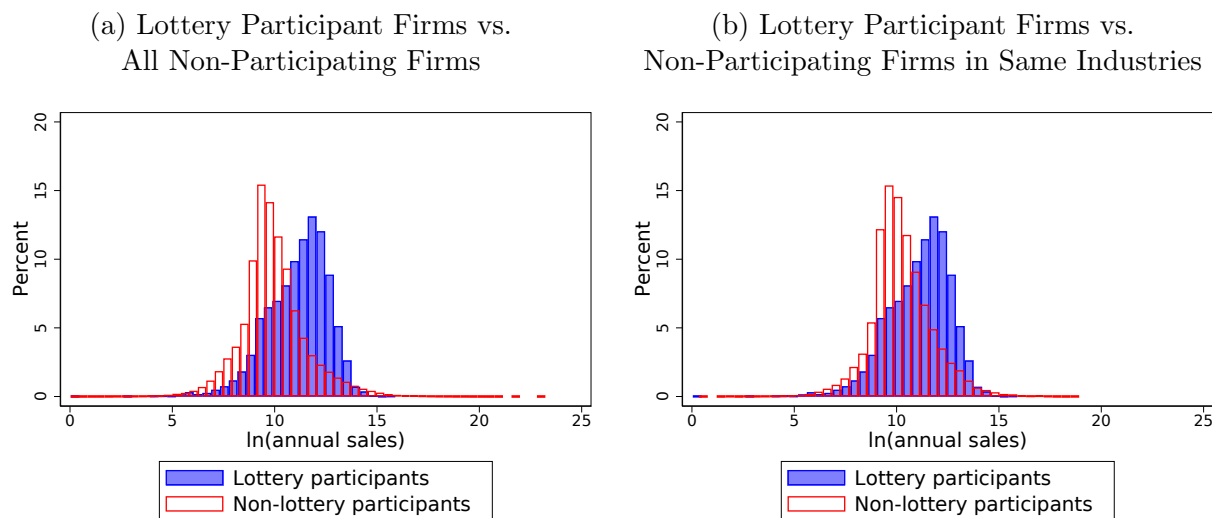
- Adao, Rodrigo, Paul Carrillo, Arnaud Costinot, Dave Donaldson, and Dina Pomeranz**, “Imports, Exports, and Earnings Inequality: Measures of Exposure and Estimates of Incidence,” *The Quarterly Journal of Economics*, 2022, *137* (3), 1553–1614.
- Alfaro-Ureña, Alonso, Isabela Manelici, and Jose P Vasquez**, “The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages,” *The Quarterly Journal of Economics*, 01 2022, *137* (3), 1495–1552.
- Amiti, Mary, Cédric Duprez, Jozef Konings, and John Van Reenen**, “FDI and superstar spillovers: Evidence from firm-to-firm transactions,” Technical Report, National Bureau of Economic Research 2023.
- , **Oleg Itskhoki, and Jozef Konings**, “International Shocks, Variable Markups, and Domestic Prices,” *The Review of Economic Studies*, 02 2019, *86* (6), 2356–2402.
- Atkeson, Andrew and Ariel Burstein**, “Pricing-to-Market, Trade Costs, and International Relative Prices,” *American Economic Review*, December 2008, *98* (5), 1998–2031.
- Atkin, David, Amit K. Khandelwal, and Adam Osman**, “Exporting and Firm Performance: Evidence from a Randomized Experiment,” *The Quarterly Journal of Economics*, 02 2017, *132* (2), 551–615.
- Bai, Jie, Maggie X Chen, Jin Liu, Xiaosheng Mu, and Daniel Yi Xu**, “Stand Out from the Millions: Market Congestion and Information Friction on Global E-Commerce Platforms,” 2022.
- Banerjee, Abhijit V and Esther Duflo**, “Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program,” *Review of Economic Studies*, 2014, *81* (2), 572–607.
- Baqae, David Rezza and Emmanuel Farhi**, “Productivity and Misallocation in General Equilibrium,” *The Quarterly Journal of Economics*, 2020, *135* (1), 105–163.
- Bau, Natalie and Adrien Matray**, “Misallocation and Capital Market Integration: Evidence from India,” *Econometrica*, 2023, *91* (1), 67–106.
- Beerli, Andreas, Jan Ruffner, Michael Siegenthaler, and Giovanni Peri**, “The Abolition of Immigration Restrictions and the Performance of Firms and Workers: Evidence from Switzerland,” *American Economic Review*, 2021, *111* (3), 976–1012.
- Benson, David, Matthew A Masten, and Alexander Torgovitsky**, “ivcrc: An Instrumental-Variables Estimator for the Correlated Random-Coefficients Model,” *The Stata Journal*, 2022, *22* (3), 469–495.
- Bergquist, Lauren Falcao and Michael Dinerstein**, “Competition and Entry in Agricultural Markets: Experimental Evidence from Kenya,” *American Economic Review*, 2020, *110* (12), 3705–3747.
- Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts**, “Does Management Matter? Evidence from India,” *The Quarterly Journal of Economics*, 2013, *128* (1), 1–51.
- Boehm, Johannes and Ezra Oberfield**, “Misallocation in the Market for Inputs: Enforcement and the Organization of Production,” *The Quarterly Journal of Economics*, 06 2020, *135* (4), 2007–2058.
- Borusyak, Kirill and Peter Hull**, “Non-Random Exposure to Exogenous Shocks: Theory and Applications,” *Econometrica*, forthcoming.
- Brugués, Felipe, Javier Brugués, and Samuele Giambra**, “Political Connections and Misallocation of Procurement Contracts: Evidence from Ecuador,” STEG Working Paper, 2022.
- Bun, Maurice JG, Theoplasti Kolaiti, and Tolga Özden**, “Misallocation and Productivity Growth: a Meta-analysis,” Technical Report, De Nederlandsche Bank NV 2023.
- Busso, Matias and Sebastian Galiani**, “The Causal Effect of Competition on Prices and Quality: Evidence from a Field Experiment,” *American Economic Journal: Applied Economics*, 2019, *11* (1), 33–56.

- de Mel, Suresh, David McKenzie, and Christopher Woodruff**, “Returns to Capital in Microenterprises: Evidence from a Field Experiment,” *The Quarterly Journal of Economics*, 2008, *123* (4), 1329–1372.
- , —, and —, “Labor Drops: Experimental Evidence on the Return to Additional Labor in Microenterprises,” *American Economic Journal: Applied Economics*, January 2019, *11* (1), 202–35.
- Demir, Banu, Ana Cecília Fieler, Daniel Xu, and Kelly Kaili Yang**, “O-Ring Production Networks,” *Journal of Political Economy*, forthcoming.
- Ding, Peng, Avi Feller, and Luke Miratrix**, “Randomization Inference for Treatment Effect Variation,” *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, 2016, pp. 655–671.
- Doran, Kirk, Alexander Gelber, and Adam Isen**, “The Effects of High-Skilled Immigration Policy on Firms: Evidence from Visa Lotteries,” *Journal of Political Economy*, 2022, *130* (10), 2501–2533.
- Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker**, “General equilibrium effects of cash transfers: experimental evidence from Kenya,” *Econometrica*, 2022, *90* (6), 2603–2643.
- Fadic, Milenko**, “Letting Luck Decide: Government Procurement and the Growth of Small Firms,” *The Journal of Development Studies*, 2020, *56* (7), 1263–1276.
- Fan, Jianqing and Irene Gijbels**, *Local Polynomial Modelling and its Applications: Monographs on Statistics and Applied Probability*, Vol. 66, CRC Press, 1996.
- Felix, Mayara**, “Trade, Labor Market Concentration, and Wages,” Working Paper, 2021.
- Ferraz, Claudio, Frederico Finan, and Dimitri Sberman**, “Procuring Firm Growth: The Effects of Government Purchases on Firm Dynamics,” NBER Working Paper No. 21219, 2015.
- Fisher, RA**, *The Design of Experiments*, Oliver and Boyd, Edinburgh, 1935.
- Giorcelli, Michela**, “The Long-Term Effects of Management and Technology Transfers,” *American Economic Review*, 2019, *109* (1), 121–52.
- Goldberg, Pinelopi Koujianou, Amit K. Khandelwal, Nina Pavcnik, and Petia Topalova**, “Imported intermediate inputs and domestic product growth: Evidence from India,” *The Quarterly Journal of Economics*, 2010, *125* (4), 1727–1767.
- Gollin, Douglas and Christopher Udry**, “Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture,” *Journal of Political Economy*, 2021, *129* (1), 1–80.
- Goolsbee, Austan and Chad Syverson**, “How do incumbents respond to the threat of entry? Evidence from the major airlines,” *The Quarterly Journal of Economics*, 2008, *123* (4), 1611–1633.
- Hall, Robert E.**, “The Relation between Price and Marginal Cost in U.S. Industry,” *Journal of Political Economy*, 1988, *96* (5), 921–947.
- , “Using Empirical Marginal Cost to Measure Market Power in the US Economy,” NBER Working Paper No. 25251, 2018.
- Haltiwanger, John, Robert Kulick, and Chad Syverson**, “Misallocation Measures: The Distortion that Ate the Residual,” NBER Working Paper No. 24199, 2018.
- Hopenhayn, Hugo A.**, “Firms, Misallocation, and Aggregate Productivity: A Review,” *Annual Review of Economics*, 2014, *6* (1), 735–770.
- Hsieh, Chang-Tai and Peter J Klenow**, “Misallocation and Manufacturing TFP in China and India,” *The Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.
- Hughes, David and Jeremy Majerovitz**, “Measuring Misallocation with Experiments,” Working Paper, 2023.

- Hummels, David, Rasmus Jørgensen, Jakob Munch, and Chong Xiang**, “The wage effects of offshoring: Evidence from Danish matched worker-firm data,” *American Economic Review*, 2014, *104* (6), 1597–1629.
- Hvide, Hans K and Tom G Meling**, “Do Temporary Demand Shocks Have Long-Term Effects for Startups?,” *The Review of Financial Studies*, 2023, *36* (1), 317–350.
- Jensen, Robert and Nolan Miller**, “Market Integration, Demand and the Growth of Firms: Evidence from a Natural Experiment in India,” *American Economic Review*, 2018, *108* (12), 3583–3625.
- Kaboski, Joseph P and Robert M Townsend**, “A Structural Evaluation of a Large-Scale Quasi-Experimental Microfinance Initiative,” *Econometrica*, 2011, *79* (5), 1357–1406.
- Klette, Tor Jakob**, “Market Power, Scale Economies and Productivity: Estimates from a Panel of Establishment Data,” *The Journal of Industrial Economics*, 1999, *47* (4), 451–476.
- Kroft, Kory, Yao Luo, Magne Mogstad, and Bradley Setzler**, “Imperfect Competition and Rents in Labor and Product Markets: The Case of the Construction Industry,” NBER Working Paper No. 27325, 2022.
- Lane, Nathan**, “Manufacturing Revolutions: Industrial Policy and Industrialization in South Korea,” *Quarterly Journal of Economics*, forthcoming.
- Lee, Munseob**, “Government Purchases, Firm Growth and Industry Dynamics: Quasi-Experimental Evidence from Government Auctions,” Working Paper, 2017.
- Loecker, Jan De, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik**, “Prices, Markups, and Trade Reform,” *Econometrica*, 2016, *84* (2), 445–510.
- Masten, Matthew A and Alexander Torgovitsky**, “Identification of Instrumental Variable Correlated Random Coefficients Models,” *Review of Economics and Statistics*, 2016, *98* (5), 1001–1005.
- McCaig, Brian and Nina Pavcnik**, “Export Markets and Labor Allocation In a Low-Income Country,” *American Economic Review*, 2018, *108* (7), 1899–1941.
- Restuccia, Diego and Richard Rogerson**, “Policy Distortions and Aggregate Productivity with Heterogeneous Establishments,” *Review of Economic Dynamics*, 2008, *11* (4), 707–720.
- and — , “The Causes and Costs of Misallocation,” *Journal of Economic Perspectives*, 2017, *31* (3), 151–174.
- Rotemberg, Martin**, “Equilibrium Effects of Firm Subsidies,” *American Economic Review*, 2019, *109* (10), 3475–3513.
- SERCOP**, “Public Procurement Contracts, 2009-2014,” 2014. Last accessed in 4/29/2016.
- Sraer, David and David Thesmar**, “How to Use Natural Experiments to Estimate Misallocation,” *American Economic Review*, 2023, *113* (4), 906–938.
- SRI**, “F101 and F102 tax forms, 2008-2015,” 2015. Last retrieved in 6/12/2018.
- , “Purchases annexes forms, 2008-2015,” 2015. Last retrieved in 12/29/2016.
- , “Employment records, 2007-2017,” 2017. Last retrieved in 2/15/2019.
- Zwick, Eric and James Mahon**, “Tax Policy and Heterogeneous Investment Behavior,” *American Economic Review*, January 2017, *107* (1), 217–48.

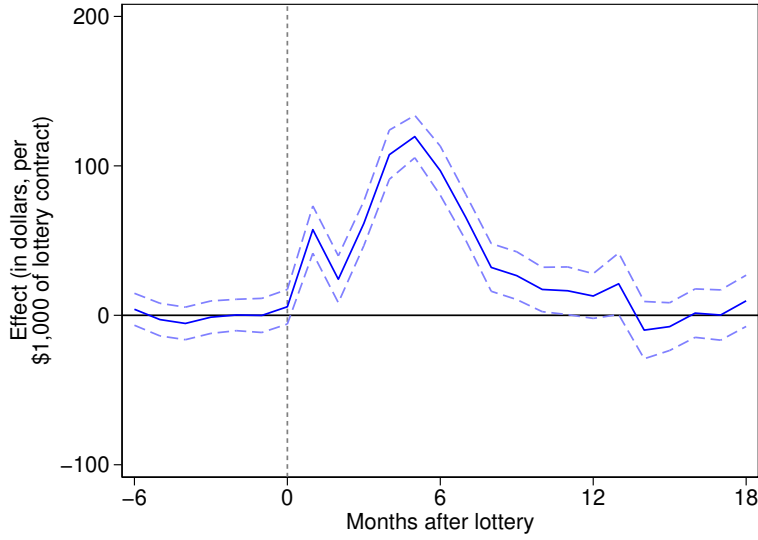
Figures and Tables

Figure 1: Comparison of Firm Size Distributions



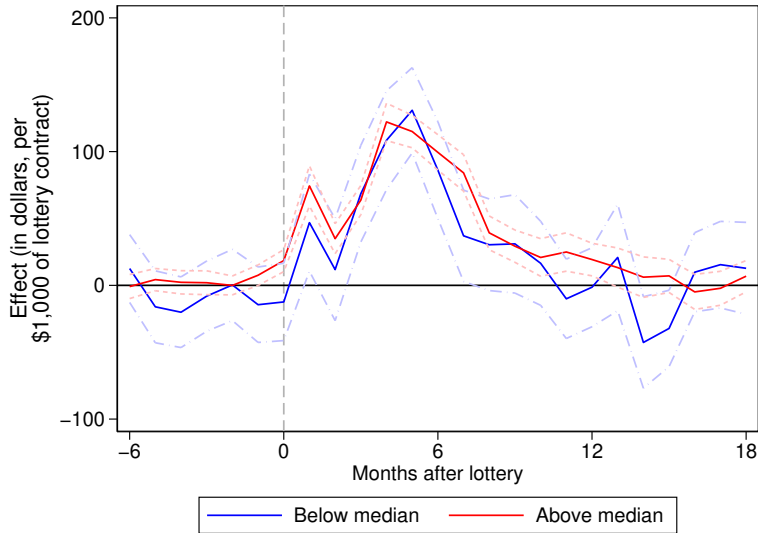
Notes: This figure plots histograms of sales in 2008, the year prior to the start of the procurement lottery system, using firms' annual income tax data (excluding firms with no sales). Lottery participants are firms that participate in at least one lottery during 2009–2014. Non-participant firms are all other firms that were economically active in 2008. Panel (b) restricts the non-participant sample to firms in the construction or engineering industries.

Figure 2: Effects on Total Sales



Notes: This figure plots estimates of the monthly effects of an additional \$1,000 in procurement winnings shocks on total (third-party reported) sales following equation (18). Total sales are based on monthly purchase annexes reported by client entities' VAT filings. Dashed lines indicate 95% confidence intervals that allow for clustering at the firm level.

Figure 3: Effects on Total Sales by Contract Size



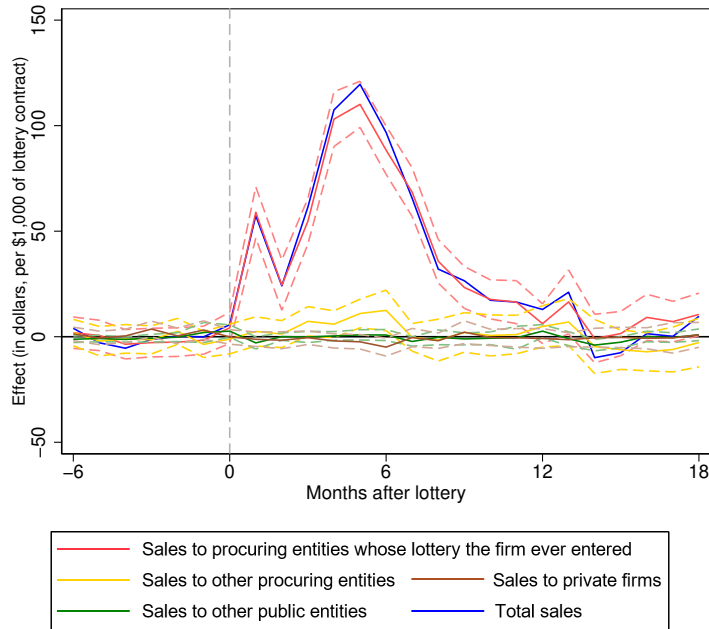
Notes: This figure extends the analysis of Figure 2, estimating monthly effects of an additional \$1,000 in procurement winnings shocks on total sales separately for lotteries with large vs. small contracts (below- vs. above-median contract amount), following equation (18). Total sales are based on monthly purchase annexes reported by clients entities' VAT filings. Dashed lines indicate 95% confidence intervals that allow for clustering at the firm level.

Figure 4: Effects on Employment



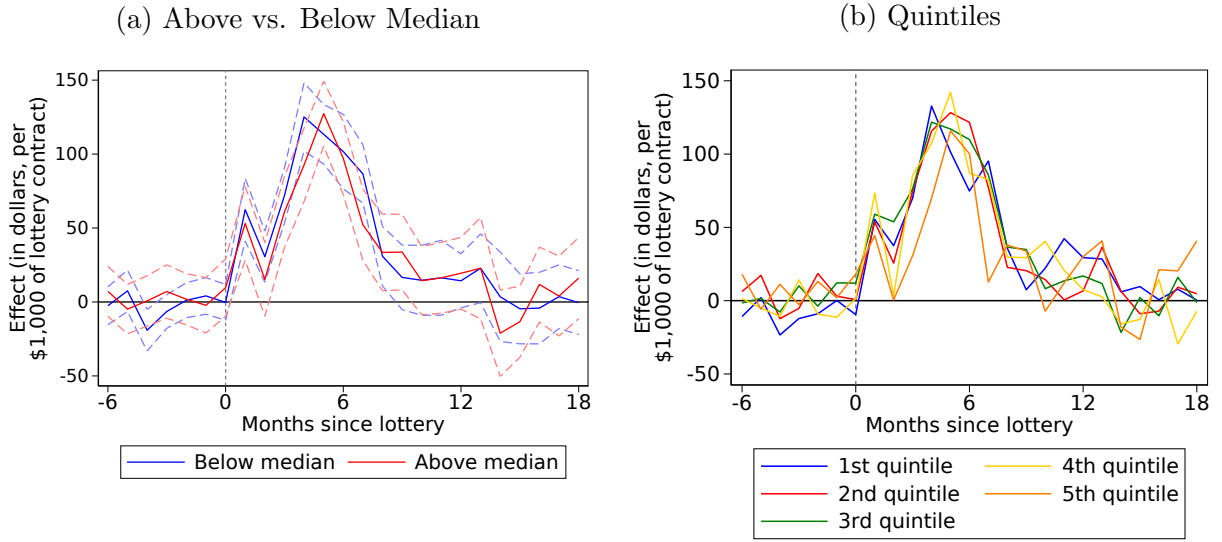
Notes: This figure plots estimates of the monthly effects of an additional \$1,000 in procurement winnings shocks on employment following equation (18), using data from social security records. Panel (a) presents effects on the number of employees, and panel (b) on total wages paid. Dashed lines indicate 95% confidence intervals that allow for clustering at the firm level.

Figure 5: Effects on Sales to Different Types of Clients



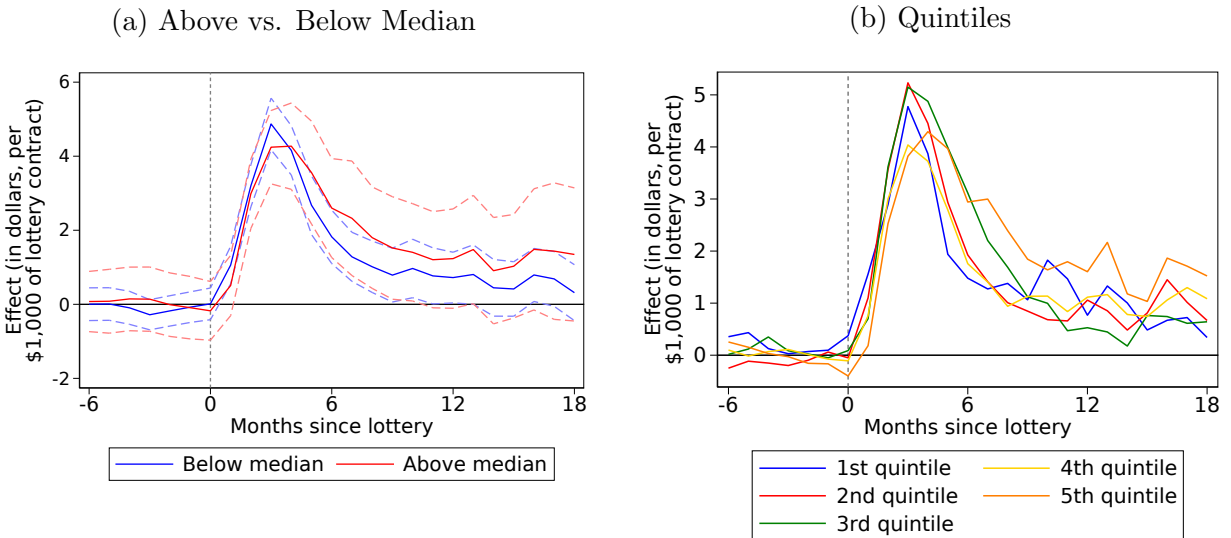
Notes: This figure extends the analysis of Figure 2, estimating the monthly effects of an additional \$1,000 in procurement winnings shocks on sales to mutually-exclusive categories of clients, following equation (18). These clients are: procuring entities with at least one lottery in our study period that the firm participated in (in red); other procuring entities, i.e., other entities that made at least one purchase through the lottery system in our study period (yellow); other public entities that made no purchases through the lottery system (green); and private firms (brown). All sales measures are based on monthly purchase annexes reported by client entities' VAT filings. Dashed lines 95% confidence intervals that allow for clustering at the firm level.

Figure 6: Effects on Total Sales by Firm Size



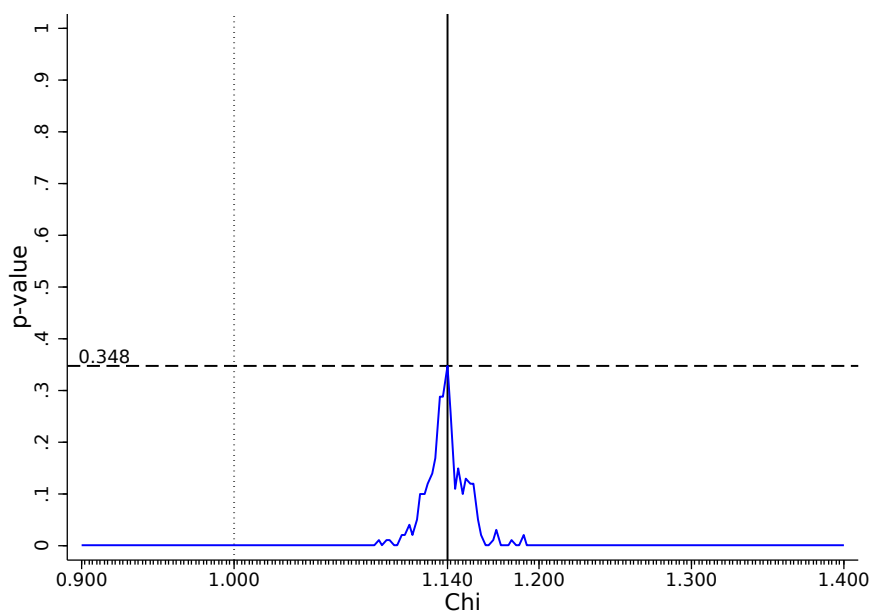
Notes: This figure extends the analysis of Figure 2, estimating the monthly effects of an additional \$1,000 in procurement winnings shocks on total sales by firm size, following equation (18). Total sales are based on monthly purchase annexes reported by client entities' VAT filings. Panel (a) presents estimates for firms with above- and below-median total sales prior to the start of the lottery system (i.e., in 2008), based on firms' annual income tax filings. Panel (b) shows the same but partitioning the sample by quintiles of 2008 sales. Dashed lines in panel (a) indicate 95% confidence intervals that allow for clustering at the firm level.

Figure 7: Effects on Total Wage Payments by Firm Size



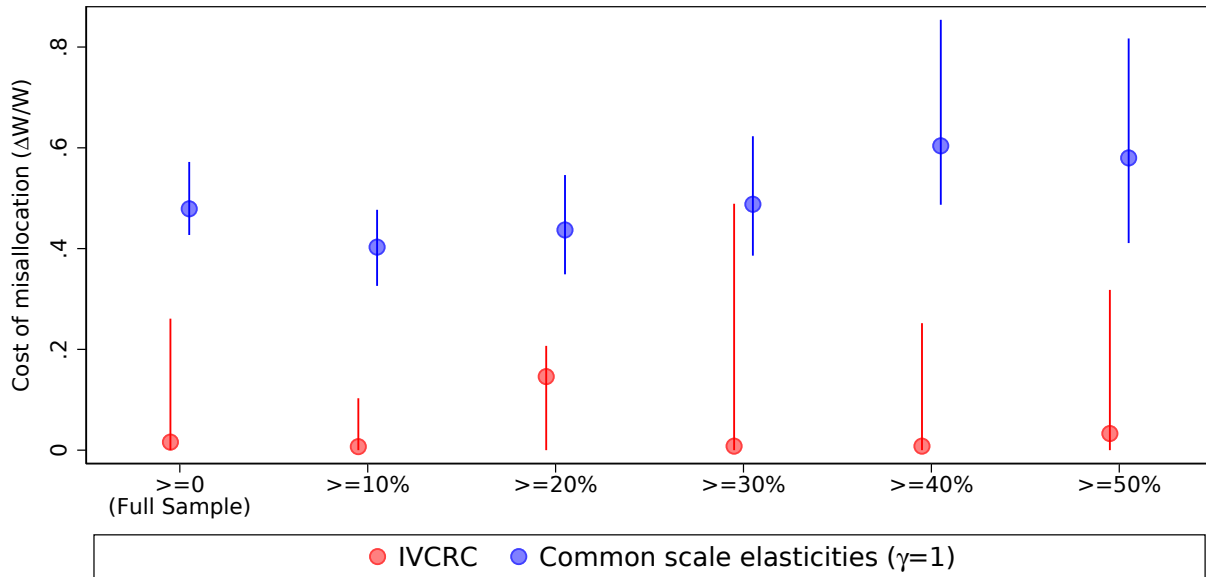
Notes: This figure extends the analysis of Figure 4 Panel (b), estimating the monthly effects of an additional \$1,000 in procurement winnings shocks on total wage payments by firm size, following equation (18), using data from social security records. Panel (a) presents estimates for firms with above- and below-median total sales prior to the start of the lottery system (i.e., in 2008), based on firms' annual income tax filings. Panel (b) shows the same but partitioning the sample by quintiles of 2008 sales. Dashed lines in panel (a) indicate 95% confidence intervals that allow for clustering at the firm level.

Figure 8: Randomization Inference Test for U-AEP and C-AEP



Notes: This figure shows the histogram of results of the randomization inference test for unconstrained allocative efficiency of production (U-AEP) and constrained allocative efficiency of production (C-AEP), as described in Sections 2 and 5. The figure plots the p-values (y-axis) under the null of $\chi = \chi$ (i.e., $\chi_m = \chi$ for all inputs m), for different values of χ (x-axis). The dotted vertical line indicates the null value of $\chi = 1$ used for the test for U-AEP, whereas the solid vertical line indicates the value of χ that corresponds to the highest p-value (0.35) obtained in this range of χ . These results imply that the null of U-AEP is rejected at standard levels, but that of C-AEP is not.

Figure 9: Estimated Cost of Misallocation:
Sub-Samples with Different Shares of Total Sales from Lottery Contracts



Notes: This figure plots the estimated cost of misallocation for sub-samples of firms with different shares of their total sales stemming from procurement lottery contracts. For example, the “> 10%” sample includes those firms for whom lottery contracts make up at least 10% of their total sales. We define sales from a procurement lottery contract as sales to a procuring entity whose lottery the firm won, in the year of the winnings and the year after. We then divide the sum of these sales by total sales in 2009-2014. The red (lower) dots in the figure show estimated cost of misallocation for the different samples with our estimation method using IVCRC, described in Section 6.1 (using the baseline specification as in Table 4 panel (a) row 1). The blue (higher) dots show estimates following the parametric alternative procedure, which assumes that firms use technologies with common scale elasticities, as described in Section 6.3, using constant returns to scale (i.e. $\gamma = 1$) as shown in Table 4 panel (b) row 1. Vertical lines indicate 95% confidence intervals using the block bootstrap procedure.

Table 1: Summary Statistics

	Mean	Median	Std. Dev.	N
Lotteries				
Contract amount (USD)	46,523	31,597	40,989	18,474
Contract anticipated duration (days)	64.5	60	34.5	18,467
Number of participants	10.1	4	15.7	18,474
Firms' Lottery Participation				
Lotteries entered per year	3.5	1	7.2	9,393
Lotteries won per year	0.35	0	0.55	9,393
Firms' Characteristics				
Firm age (years)	11.21	10	11.06	9,393
Is incorporated	0.18	0	0.38	9,393
Number of clients (third-party reported)	4.74	2	17.72	9,393
Sales (third-party reported, USD)	132,707	47,651	271,306	9,393
Sales (self-reported, USD)	141,184	53,360	287,409	9,393
Costs (self-reported, USD)	123,907	42,058	267,508	9,393
Profits (self-reported, USD)	17,278	11,180	36,360	9,393
Employees (social security)	4.40	2	10.10	9,393
Wages (social security, USD)	7,399	2,880	25,926	9,393

Notes: This table presents summary statistics of procurement lotteries (with at least two participating firms) and of firms that participated in any such lotteries. Observations concerning firm characteristics are from each firm's first year of lottery participation. Self-reported tax variables are based on firms' annual income tax filings. Third-party reported variables are based on (annualized) monthly purchase annexes of client entities' VAT filings. 'Social security' indicates data from social security filings. Contract duration is missing for seven observations, as discussed in Appendix C.

Table 2: Balance of Randomization

Coefficient on \$1000 in procurement winnings shocks	
Firm Characteristics	
Firm age (years)	-0.0004 (0.0012)
Is incorporated	0.0001 (0.0000)
Lottery Variables ($t - 1$)	
Lotteries entered	0.0048 (0.0033)
Expected winnings	3.47 (6.98)
Lotteries won	0.0003 (0.0002)
Amount won	-6.02 (8.56)
Social Security Data ($t - 1$)	
Employees	0.0001 (0.0014)
Wages	0.3001 (1.97)
Tax Data ($t - 2$)	
Sales (third-party reported)	-12.59 (32.57)
Sales to procuring entities (third-party reported)	-18.10 (27.65)
Sales to other government entities (third-party reported)	2.39 (2.22)
Sales to private sector (third-party reported)	7.16 (9.79)
Sales (self-reported)	-17.51 (33.92)
Costs (self-reported)	-13.60 (30.80)
Annual income tax (self-reported)	-0.4017 (0.4814)
Number of clients (third-party reported)	0.0009 (0.0018)
P-value of joint F-test:	0.29

Notes: This table presents balance tests of procurement winnings shocks (in \$1,000s) on pre-treatment variables. Each displayed coefficient (and corresponding standard error in parentheses) stems from a separate regression of the outcome variable on procurement winnings shocks in year (t). For time-varying variables, we use data from year ($t-1$). For tax variables, we use information for year ($t-2$) to ensure that they reflect pre-lottery filings. Self-reported tax variables are based on firms' annual income tax filings. Third-party reported variables are based on monthly purchase annexes of client entities' VAT filings. Monthly variables are summed up over the year. Observations cover 2009-2014 for non-lagged variables, 2009-2013 for (lagged) lottery variables, 2008-2013 for (lagged) social security variables, and 2008-2012 for (double lagged) tax data variables. As in all our analysis, the sample only includes observations for years after the firm started to be economically active. For the joint F-test, to include the whole sample, we impute zero for missing values and add a dummy indicator. All monetary variables in USD, winsorized at the top 1%. Standard errors clustered at the firm level in parentheses. *** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$.

Table 3:
Effects on Sales, Costs, and Profits

	(1)	(2)	(3)	(4)	(5)	(6)
	Sales	Sales	Total	Labor	Non-labor	Profits
	(Third-party reported)	(Self- reported)	costs	costs	costs	
Procurement Lottery Shocks:						
Year t	438.71*** (37.31)	430.14*** (39.19)	382.47*** (36.22)	14.39** (4.70)	368.09*** (33.73)	47.67*** (5.80)
Year $t + 1$	269.76*** (48.70)	239.00*** (51.18)	200.57*** (46.49)	11.03 (6.15)	189.54*** (43.05)	38.43*** (7.00)
Number of observations	53,107	53,107	53,107	53,107	53,107	53,107
Number of firms	9,368	9,368	9,368	9,368	9,368	9,368

Notes: This table shows estimates of the effect of procurement winnings shocks (in \$1,000's) on firms' sales, costs and profits in the year of the shock (t) and the next ($t+1$), following equation (18). Third-party reported sales are based on monthly purchase annexes of client entities' VAT filings, summed up over the year. All other variables are based on firms' annual income tax filings. Observations are firm-years. Standard errors clustered at the firm level are reported in parentheses. *** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$.

Table 4: Estimated Cost of Misallocation

	$\mathbb{E}_{\bar{\lambda}}[\bar{\mu}]$	$\text{Var}_{\bar{\lambda}}[\bar{\mu}]$	$\frac{\Delta W}{W}$
	(1)	(2)	(3)
Panel (a): IVCRC estimates			
Baseline	1.126 [1.093, 1.161]	0.014 [0, 0.341]	0.016 [0, 0.261]
No trimming	1.129 [1.098, 1.188]	0 [0, 0.329]	0.006 [0, 0.253]
5% trimming	1.111 [1.078, 1.157]	0 [0, 0.394]	0.005 [0, 0.301]
Gaussian kernel	1.125 [1.095, 1.145]	0 [0, 0.040]	0.006 [0, 0.035]
Uniform kernel	1.126 [1.093, 1.161]	0.014 [0, 0.341]	0.016 [0, 0.261]
10 sales bins	1.115 [1.067, 1.158]	0 [0, 0.617]	0.005 [0, 0.468]
$\bar{\omega}_C = 0.75$	1.126 [1.093, 1.161]	0.014 [0, 0.341]	0.020 [0, 0.387]
$\bar{\omega}_C = 0.25$	1.126 [1.093, 1.161]	0.014 [0, 0.341]	0.010 [0, 0.132]
Panel (b): Alternative procedure assuming common scale elasticities			
Constant returns-to-scale ($\gamma = 1$)	1.240 [1.223, 1.257]	0.611 [0.544, 0.730]	0.479 [0.427, 0.572]
Decreasing returns-to-scale ($\gamma = 0.85$)	1.054 [1.040, 1.068]	0.441 [0.393, 0.528]	0.332 [0.296, 0.397]
Increasing returns-to-scale ($\gamma = 1.15$)	1.426 [1.407, 1.445]	0.807 [0.720, 0.966]	0.674 [0.601, 0.798]

Notes: Columns (1) and (2) report estimates of the 2008 sales-weighted expectation and variance of the wedge distribution that enter the total cost of misallocation formula in equation (16); variance estimates are truncated at zero from below. Column (3) reports the corresponding estimated cost of misallocation $\Delta W/W = \frac{1}{2}\bar{\omega}_C\theta\text{Var}_{\bar{\lambda}}[\bar{\mu}] + \frac{1}{2}\bar{\omega}_C(1 - \bar{\omega}_C)\eta(\mathbb{E}_{\bar{\lambda}}[\bar{\mu}] - 1)^2$ implied by that formula (using parameter values of $\theta = 3$, $\eta = 3$, and $\bar{\omega}_C = 0.5$, unless noted otherwise). Panel (a) shows estimates from the IVCRC method described in Section 6.1 with different specifications (with 2% trimming, Epanechnikov kernel, and 5 sales bins, unless noted otherwise). Panel (b) follows the alternative parametric procedure for estimating wedges, which assumes that firms use technologies with common scale elasticities γ , as described in Section 6.3. The ranges reported in square brackets are two-sided 95% confidence intervals in column (1) of panel (a) and all of panel (b), but one-sided intervals for columns (2) and (3) of panel (a); all such intervals are calculated on the basis of a block bootstrap procedure, with values reported in Appendix Figure A.12.

Misallocation in Firm Production:
A Nonparametric Analysis Using Procurement Lotteries

Paul Carrillo

Dave Donaldson

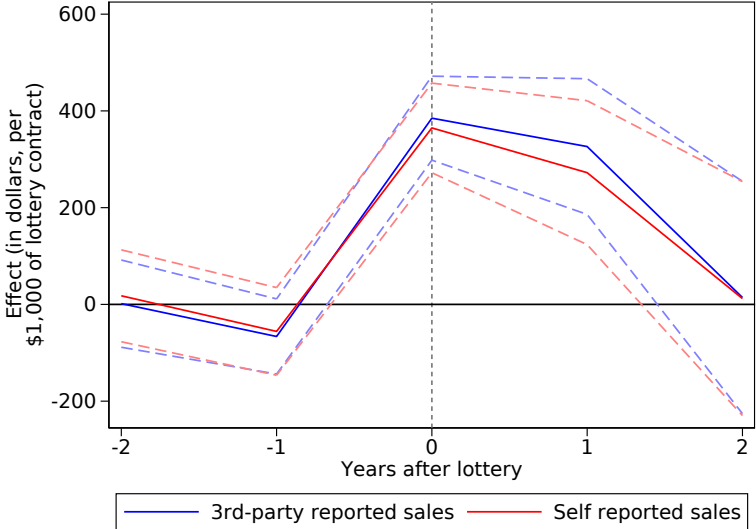
Dina Pomeranz

Monica Singhal

Appendices
(For Online Publication Only)

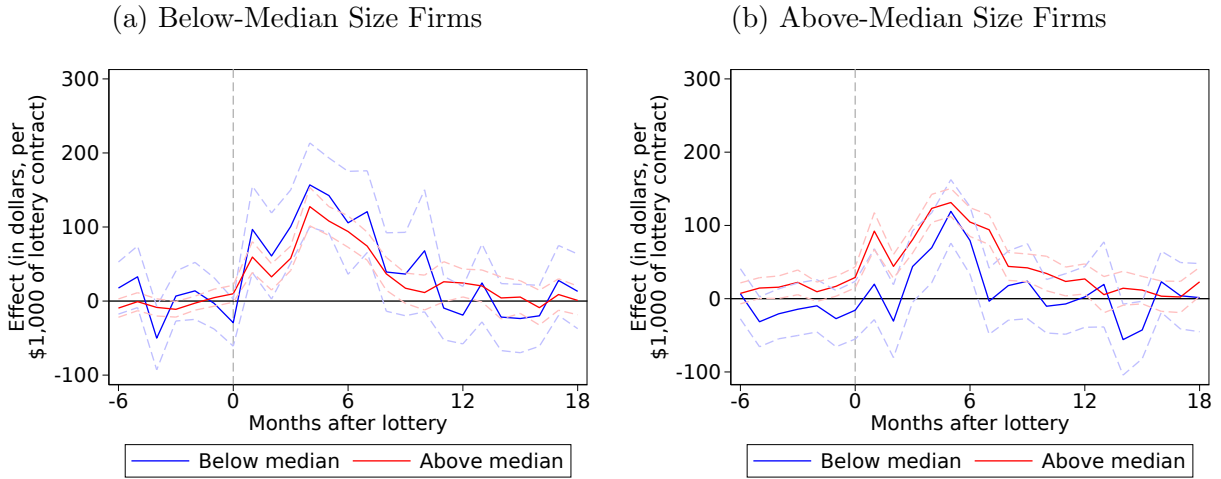
A Appendix Figures and Tables

Figure A.1: Annual Total Sales, Third-Party Reported vs. Self-Reported



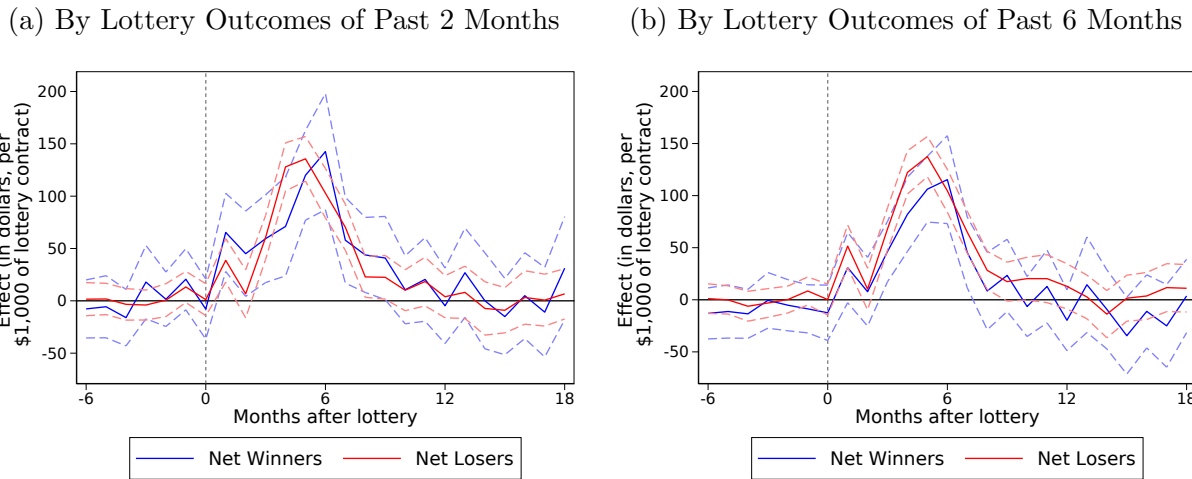
Notes: This figure plots estimates of the annual effect of an additional \$1,000 in procurement winnings shocks on total sales, following equation (18). Third-party reported sales are based on monthly purchase annexes reported by the client entities' VAT filings (annualized over the year) and self-reported sales are based on firms' annual income tax filings. Dashed lines indicate 95% confidence intervals that allow for clustering at the firm level.

Figure A.2: Effects on Total Sales by Contract Size and Firm Size



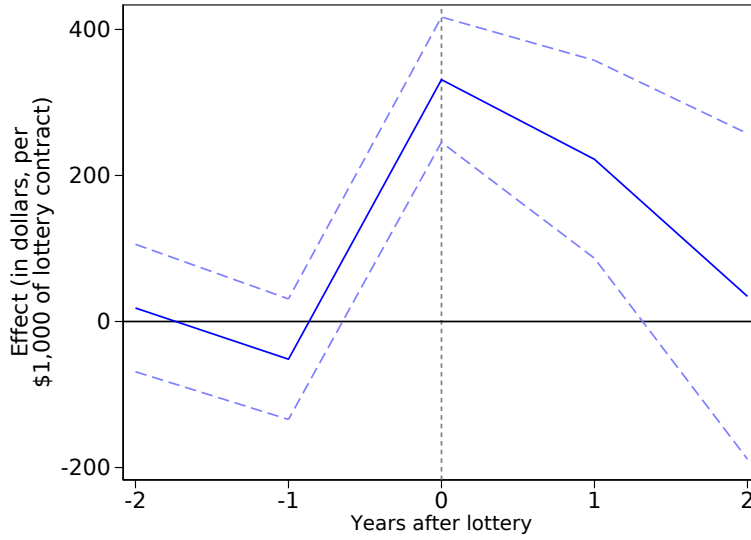
Notes: This figure extends the analysis of Figures 3 and 6, estimating monthly effects of an additional \$1,000 in procurement winnings shocks on total sales by firm size, separately for lotteries with large vs. small contracts (below-, in blue, vs. above-median, in red, contract amounts), following equation (18). Total sales are based on monthly purchase annexes reported by client entities' VAT filings. Panel (a) presents estimates for firms with below-median sales prior to the start of the lottery system (i.e., in 2008), based on firms' annual income tax filings. Panel (b) shows the same but for firms with above-median self-reported sales. Dashed lines indicate 95% confidence intervals that allow for clustering at the firm level.

Figure A.3: Effects on Total Sales by Recent Lottery Outcomes



Notes: This figure extends the analysis of Figure 2, estimating monthly effects of an additional \$1,000 in procurement winnings shocks on total sales, following equation (18), separately by the extent of recent lottery success. Total sales are based on monthly purchase annexes reported by client entities' VAT filings. Net winners (losers) are firms that received a total positive (negative) winnings shock over the previous months. Panel (a) presents estimates summing up a firm's winnings shocks of the past two months and panel (b) does the same for the past six months. Dashed lines indicate 95% confidence intervals that allow for clustering at the firm level.

Figure A.4: Effects on Total Costs



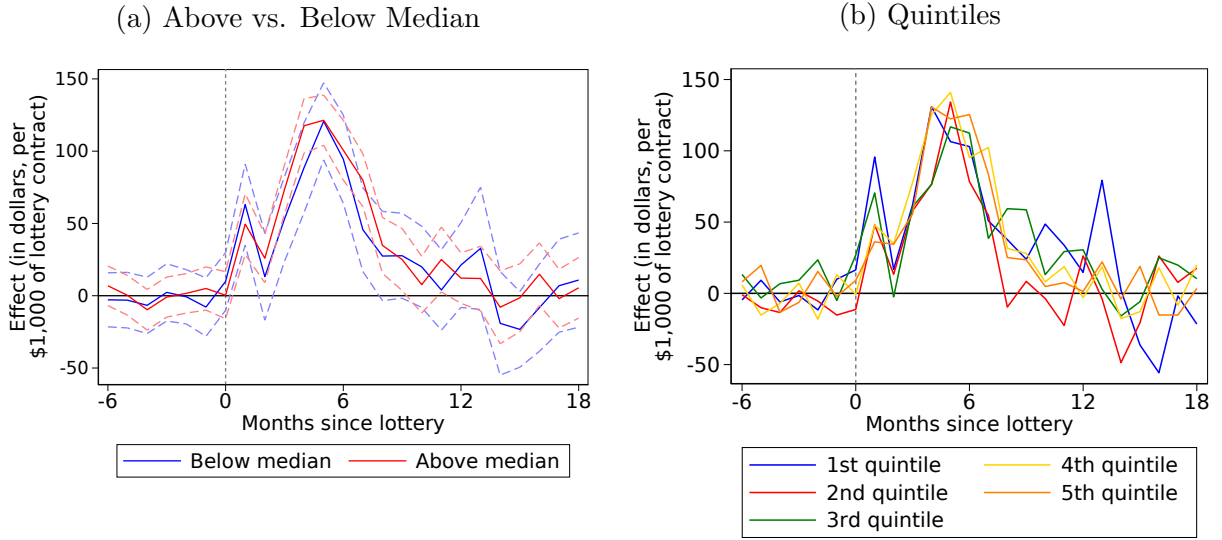
Notes: This figure plots estimates of the annual effects of an additional \$1,000 in procurement winnings shocks on total costs, following equation (18). Total costs include both labor and non-labor costs and are based on firms' annual income tax filings. Dashed lines indicate 95% confidence intervals that allow for clustering at the firm level.

Figure A.5: Effects on Wages



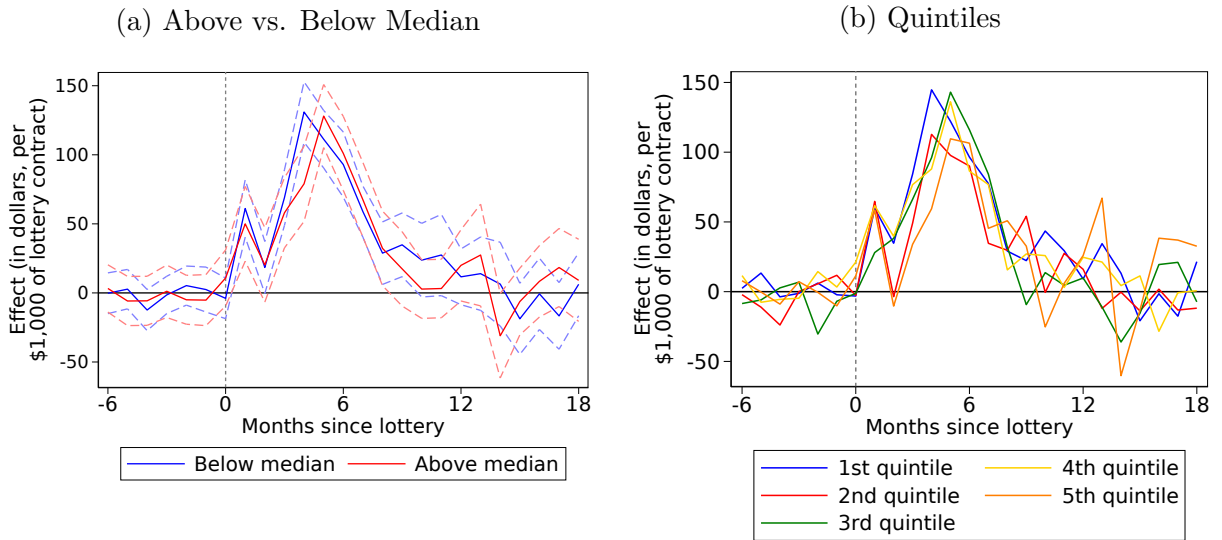
Notes: This figure plots estimates of the monthly effects of an additional \$1,000 in procurement winnings shocks on the average monthly wage using data from social security records. Panel (a) follows equation (18) and presents estimates of the effect on the average monthly wage for all employees. Panel (b) shows the same for continuing workers, to avoid worker composition changes (i.e., new hires or attrition). Specifically, it only includes employees who worked continuously at the firm from 6 months before until 18 months after the first lottery that the firm participated in. Because this analysis only uses the winnings shocks of the first lottery and thus disregards any potential subsequent shocks, we estimate the coefficients in a separate regression for every lead and lag. Dashed lines indicate 95% confidence intervals that allow for clustering at the firm level.

Figure A.6: Effects on Total Sales by Profitability



Notes: This figure extends the analysis of Figure 2, estimating the monthly effects of an additional \$1,000 in procurement winnings shocks on total sales separately by profitability (defined as sales divided by costs), following equation (18). Total sales are based on monthly purchase annexes reported by client entities' VAT filings. Panel (a) presents estimates for firms with above- and below-median profitability prior to the start of the lottery system (i.e., in 2008). Panel (b) shows the same but partitioning the same by quintiles of profitability in 2008. Dashed lines in panel (a) indicate 95% confidence intervals that allow for clustering at the firm level.

Figure A.7: Effects on Total Sales by Number of Employees

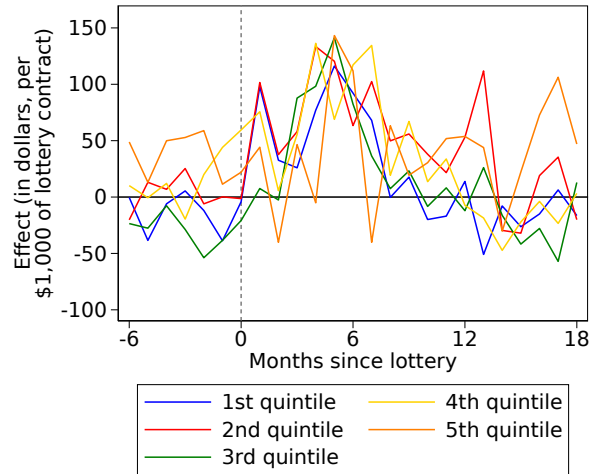
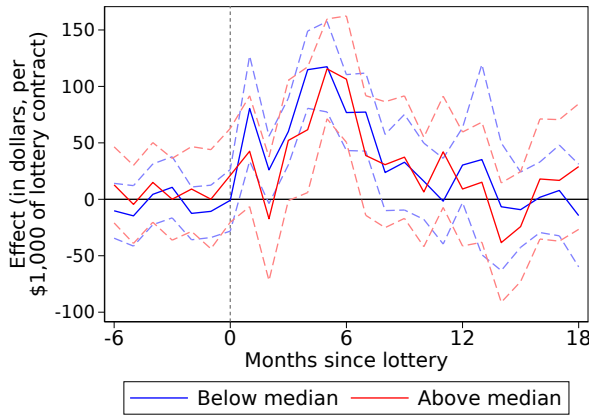


Notes: This figure extends the analysis of Figure 2, estimating the monthly effects of an additional \$1,000 in procurement winnings shocks on total sales by number of employees, following equation (18). Total sales are based on monthly purchase annexes reported by client entities' VAT filings. Panel (a) presents estimates for firms with above- and below-median number of employees prior to the start of the lottery system (i.e., in 2008). Panel (b) shows the same but partitioning the sample by quintiles of the number of employees in 2008. Dashed lines in panel (a) indicate 95% confidence intervals that allow for clustering at the firm level.

Figure A.8: Effects on Total Sales by Number of Suppliers

(a) Above vs. Below Median

(b) Quintiles

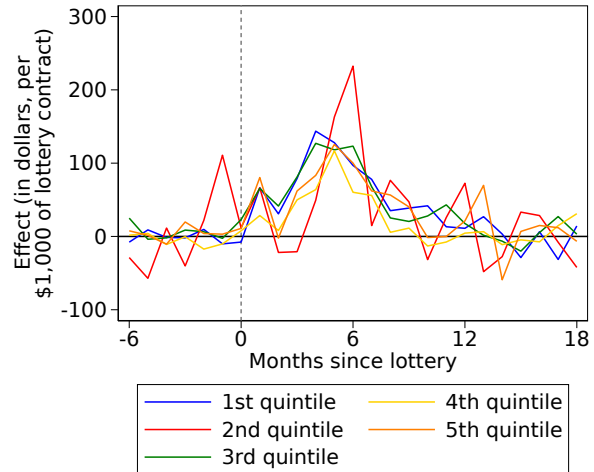
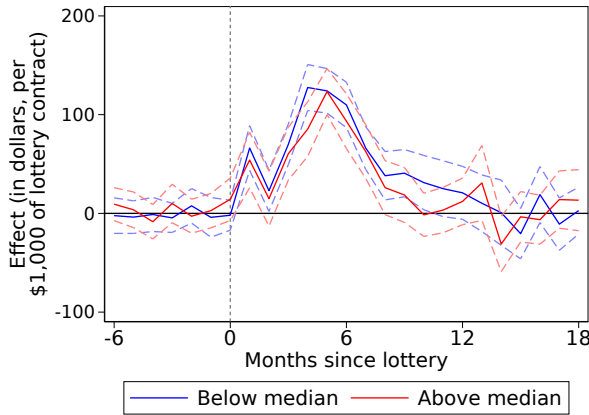


Notes: This figure extends the analysis of Figure 2, estimating the monthly effects of an additional \$1,000 in procurement winnings shocks on total sales for firms with a large vs. small number of suppliers, following equation (18). Total sales are based on monthly purchase annexes reported by client entities' VAT filings. Panel (a) presents estimates for firms with above- and below-median number of suppliers prior to the start of the lottery system (i.e., in 2008), based on firms' annual income tax filings. Panel (b) shows the same but partitioning the sample by quintiles of the number of suppliers in 2008. Dashed lines in panel (a) indicate 95% confidence intervals that allow for clustering at the firm level.

Figure A.9: Effects on Total Sales by Labor Intensity

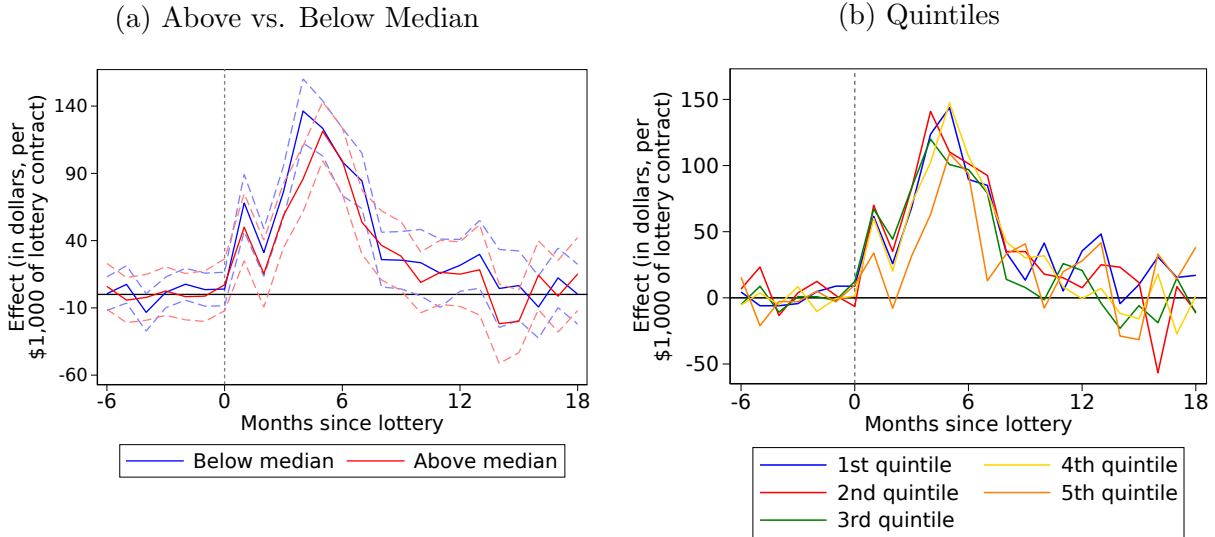
(a) Above vs. Below Median

(b) Quintiles



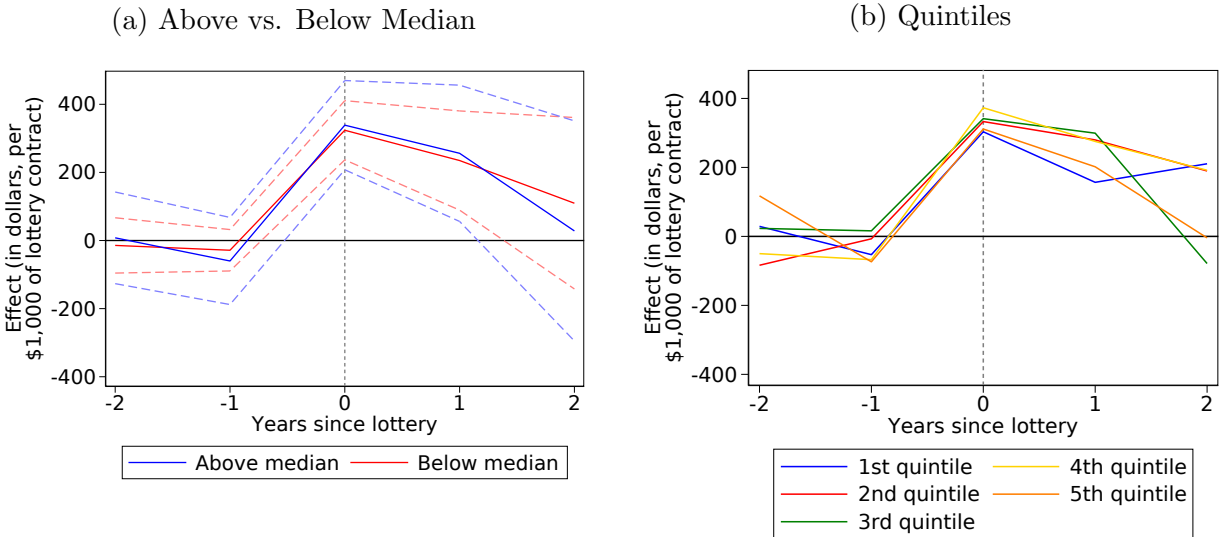
Notes: This figure extends the analysis of Figure 2, estimating the monthly effects of an additional \$1,000 in procurement winnings shocks on total sales by the level of firms' labor intensity in 2008, following equation (18). Total sales are based on monthly purchase annexes reported by client entities' VAT filings. Panel (a) presents estimates for firms with above- and below-median labor intensity prior to the start of the lottery system (i.e., in 2008), based on firms' annual income tax filings. Panel (b) shows the same but partitioning the sample by quintiles of labor intensity in 2008. Labor intensity is defined as the ratio of wage payments over self-reported sales. Dashed lines in panel (a) indicate 95% confidence intervals that allow for clustering at the firm level.

Figure A.10: Effects on Total Sales by Firm Size (Based on Third-Party Reported Sales)



Notes: This figure extends the analysis of Figure 2, estimating the monthly effects of an additional \$1,000 in procurement winnings shocks on total sales by the level of firms' third-party reported sales in 2008, following equation (18). Total sales are based on monthly purchase annexes reported by client entities' VAT filings. Panel (a) presents estimates for firms with above- and below-median third-party reported sales prior to the start of the lottery system (i.e., in 2008), based on firms' annual income tax filings. Panel (b) shows the same but partitioning the sample by quintiles of third-party reported sales in 2008. Dashed lines in panel (a) indicate 95% confidence intervals that allow for clustering at the firm level.

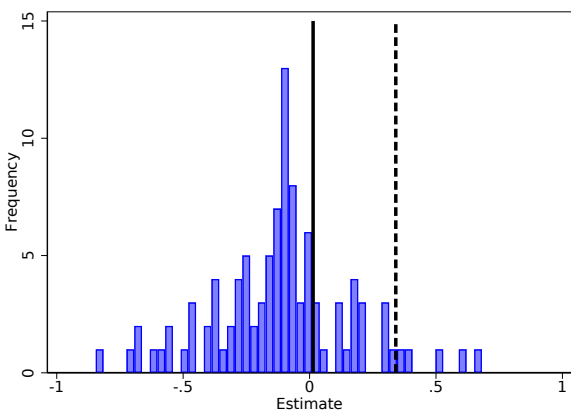
Figure A.11: Effects on Total Costs by Firm Size



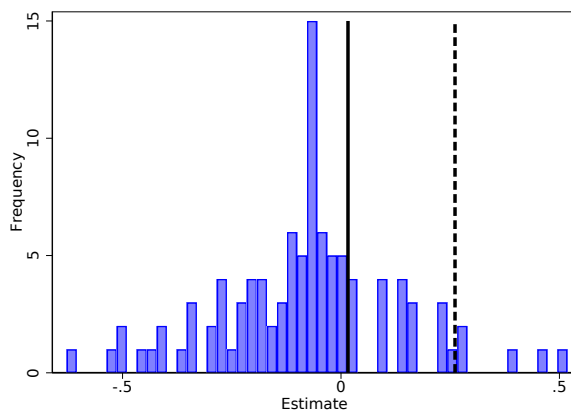
Notes: This figure extends the analysis of Figure A.4, estimating the annual effects of an additional \$1,000 in procurement winnings shocks on annual total costs by firm size, following equation (18). Total costs include both labor and non-labor costs and are based on firms' annual income tax filings. Panel (a) presents estimates for firms with above- and below-median total sales prior to the start of the lottery system (i.e., in 2008), based on firms' annual income tax filings. Panel (b) shows the same but partitioning the sample by quintiles of 2008 sales. Dashed lines in panel (a) indicate 95% confidence intervals that allow for clustering at the firm level.

Figure A.12: Results from a Bootstrapping Procedure of IVCRC Estimates

(a) Sales-Weighted Variance of Wedges



(b) Welfare Cost of Misallocation



Notes: This figure shows histograms of results from a bootstrapping procedure of estimates of the non-truncated sales-weighted variance of wedges, $\mathbb{E}_{\lambda} [(\bar{\mu})^2] - (\mathbb{E}_{\lambda} [\bar{\mu}])^2$, in panel (a), and of the corresponding implied welfare cost of misallocation, $\frac{\Delta W}{W}$, as described in equation (16) (prior to the left-truncation of the variance at zero), in panel (b). The solid vertical lines show the observed point estimates when using the original sample, as opposed to the bootstrapped samples. The dashed vertical lines show the corresponding 95% confidence limit on right-tailed tests for whether $\text{Var}_{\lambda}[\bar{\mu}]$ or $\frac{\Delta W}{W}$ are greater than zero, respectively.

Table A.1: Estimated Cost of Misallocation:
Based on Assumption of Common Scale Elasticities

	$\mathbb{E}_{\bar{\lambda}}[\bar{\mu}]$	$\text{Var}_{\bar{\lambda}}[\bar{\mu}]$	$\frac{\Delta W}{W}$
	(1)	(2)	(3)
2008	1.258 [1.243, 1.295]	0.857 [0.678, 1.835]	0.668 [0.532, 1.410]
2009	1.215 [1.200, 1.230]	0.674 [0.453, 0.992]	0.523 [0.355, 0.767]
2010	1.240 [1.227, 1.250]	0.484 [0.401, 0.588]	0.385 [0.321, 0.464]
2011	1.228 [1.216, 1.240]	0.506 [0.402, 0.665]	0.399 [0.318, 0.520]
2012	1.236 [1.221, 1.250]	0.562 [0.462, 0.674]	0.442 [0.366, 0.529]
2013	1.240 [1.223, 1.257]	0.611 [0.544, 0.730]	0.479 [0.427, 0.572]
2014	1.264 [1.242, 1.281]	0.866 [0.724, 1.042]	0.676 [0.566, 0.811]
2015	1.286 [1.263, 1.314]	1.118 [0.954, 1.310]	0.869 [0.744, 1.018]

Notes: This table shows results from panel (b) of Table 4 row 1 for each year in 2008 to 2015. These are estimates of the cost of misallocation following an alternative procedure that assumes all firms have technologies featuring global constant returns-to-scale ($\gamma = 1$). Columns (1) and (2) report estimates of the sales-weighted expectation and variance of the wedge distribution that enter the total cost of misallocation formula in equation (16). Column (3) shows the corresponding estimated cost of misallocation $\Delta W/W = \frac{1}{2}\bar{\omega}_C\theta\text{Var}_{\bar{\lambda}}[\bar{\mu}] + \frac{1}{2}\bar{\omega}_C(1 - \bar{\omega}_C)\eta(\mathbb{E}_{\bar{\lambda}}[\bar{\mu}] - 1)^2$ implied by that formula (when using parameters values $\theta = 3$, $\eta = 3$, and $\bar{\omega}_C = 0.5$, unless noted otherwise). The ranges reported in square brackets are two-sided 95% confidence intervals calculated using the block bootstrap procedure.

Table A.2: Estimated Cost of Misallocation:
 Alternatives for Method Based on Assumption of Common Scale Elasticities

	$\mathbb{E}_{\bar{\lambda}}[\ln \bar{\mu}]$	$\text{Var}_{\bar{\lambda}}[\ln \bar{\mu}]$	$\frac{\Delta W}{W}$
	(1)	(2)	(3)
Unweighted ($\lambda = \frac{1}{N}$)	0.375 [0.350, 0.404]	0.672 [0.619, 0.740]	0.557 [0.516, 0.602]
Weighted	0.157 [0.148, 0.165]	0.072 [0.065, 0.081]	0.063 [0.058, 0.070]

Notes: This table reports estimates of the cost of misallocation, according to the alternative method based on common scale elasticities, obtained when using alternative approximations. The first row reports the unweighted expectation and variance of log wedges (obtained when using this method) in columns (1) and (2), and the resulting cost of misallocation in column (3). Following Hsieh and Klenow (2009), this is appropriate when $\gamma = 1$ and firms' wedges and TFPs are assumed to be joint log-normally distributed. Similarly, the second row reports values of the sales-weighted expectation and variance of the log wedge distribution, and the resulting cost of misallocation. This is appropriate for a second-order approximation to the welfare cost of misallocation when log wedges are small. In both cases, the calculations in column (3) are given by $\Delta W/W = \frac{1}{2}\bar{\omega}_C\theta\text{Var}_{\bar{\lambda}}[\ln \bar{\mu}] + \frac{1}{2}\bar{\omega}_C(1 - \bar{\omega}_C)\eta(\mathbb{E}_{\bar{\lambda}}[\ln \bar{\mu}])^2$, using parameter values $\theta = 3$, $\eta = 3$, and $\bar{\omega}_C = 0.5$. The ranges reported in square brackets are two-sided 95% confidence intervals calculated using a block bootstrap procedure.

B Further Examples for Quantifying the Cost of Misallocation

Example #1: Multiple input types and a single sector

Consider an economy with single-product, profit-maximizing firms, each producing with a Cobb-Douglas production function that combines capital (denoted x_{iK}) and labor (x_{iL}) via $y_i = A_i(x_{iK})^{\bar{\alpha}_K}(x_{iL})^{1-\bar{\alpha}_K}$. Capital and labor are both in fixed aggregate supply to these firms, so the only source of misallocation will concern the extent to which each input is used across firms, not the overall extent of input use; that is, in this example there is no distinction between U-AEP and C-AEP. The representative consumer has CES preferences for the firms' outputs, with elasticity of substitution θ . Finally, suppose that the time-0 allocation $(\bar{\mathbf{y}}, \bar{\mathbf{x}})$ and prices $(\bar{\mathbf{p}}, \bar{\mathbf{w}})$ imply (via the definition in equation 3) that firm i 's capital and labor wedges are given by $\bar{\mu}_{i,K}$ and $\bar{\mu}_{i,L}$, respectively. This is therefore the economy of Hsieh and Klenow (2009) but with a single sector.

Using the results in Baqaee and Farhi (2020), it is straightforward to show that, up to a second-order approximation, the increase in welfare (relative to the initial level) from eliminating the wedges in this economy is given by

$$\frac{\Delta W}{W} = \frac{1}{2}\bar{\alpha}_K(1 - \bar{\alpha}_K)\text{Var}_{\bar{\lambda}}[\bar{\mu}_K - \bar{\mu}_L] + \frac{1}{2}\theta\text{Var}_{\bar{\lambda}}[\bar{\alpha}_K\bar{\mu}_K + (1 - \bar{\alpha}_K)\bar{\mu}_L], \quad (22)$$

where $\text{Var}_a[b] \equiv \mathbb{E}_a[b^2] - (\mathbb{E}_a[b])^2$ denotes the variance of the vector $\{b_k\}$ weighted by the vector $\{a_k\}$, and $\mathbb{E}_a[b]$ denotes the expectation of the vector b weighted by the vector a . In this case, the variances are weighted by the firms' sales shares, denoted $\bar{\lambda}_i \equiv \frac{\bar{p}_i\bar{y}_i}{\sum_{i'}\bar{p}_{i'}\bar{y}_{i'}}$.

The total cost of misallocation in this example derives from misallocation of inputs across firms even while aggregate inputs are held fixed, but this misallocation can be split into two forms, as is apparent in equation (22). The first term captures the effects of within-firm substitution (which has an elasticity of one in this Cobb-Douglas case) to the potential dispersion in wedges across the two inputs within a firm. And the second term captures the effects of across-firm substitution, on the behalf of consumers (and hence scaling by θ), to the potential differences in cost-weighted average wedges (i.e., $\bar{\alpha}_K\bar{\mu}_{i,K} + (1 - \bar{\alpha}_K)\bar{\mu}_{i,L}$) of

different firms. Because both types of inputs in this economy are in fixed aggregate supply, the relevant features of the wedge distribution that matter for misallocation all concern *dispersion* rather than average levels. However, both of the variance measures that matter here are weighted by $\bar{\lambda}$, since the wedges in larger firms are more costly.

Proposition 2 highlights how all weighted, uncentered moments of the distribution of wedges across firms can be identified. This is directly applicable to the cost of misallocation in equation (22), which can be easily converted from centered second-order weighted moments to ones that are uncentered. Hence, in a setting with instruments that satisfy Assumption 4, and with knowledge of the demand parameter θ , the cost of misallocation (up to a second-order approximation) in this example can be estimated consistently.

Example #2: Multiple sectors and endogenous input supply

We continue with a setting featuring single-product, profit-maximizing firms. But we now allow each firm to have its own arbitrary production function (involving an arbitrary set of inputs) so long as the technology satisfies Assumption 1 and displays constant returns locally to the time-0 allocation.⁵⁶ The representative consumer has nested CES preferences over the products produced by these firms; in particular, the consumer's elasticity of substitution between the firms within sector s (which we denote by $\mathcal{I}(s)$) is given by θ_s and that for substitution across sector-specific bundles is given by ρ .

In contrast to the previous example, we now allow the representative consumer household to also supply the inputs x_{im} to these firms. The household is endowed with a fixed amount of time that it can freely convert into inputs or retain as leisure; it has an elasticity of substitution η between leisure and the bundle of final consumption goods. This endogenous input supply means that the total cost of misallocation will derive from both a component due to misallocation of the aggregate inputs \bar{X}_m that are supplied at time-0, and from a second component due to the misallocation of those total amounts themselves. Finally, we assume that each firm has the same wedge on each of its inputs, which we denote by $\bar{\mu}_{i,m} = \bar{\mu}_i$ for all m . This would be the case if, for example, the underlying cause of potential wedges is firms' market power in their product markets, and/or taxes and subsidies on firms' sales.

⁵⁶A natural technology that fits this form is one with arbitrary overhead costs and constant marginal costs (at fixed input prices and wedges).

Again, the tools in Baqaee and Farhi (2020) make it easy to calculate the total cost of misallocation due to wedges in this economy. As before, we let $\bar{\lambda}_i \equiv \frac{\bar{p}_i \bar{y}_i}{\sum_{i'} \bar{p}_{i'} \bar{y}_{i'}}$ denote the share of firm i 's sales in total goods consumption. We also let $\bar{\psi}_s \equiv \frac{\sum_{i \in \mathcal{I}(s)} \bar{p}_i \bar{y}_i}{\sum_{i'} \bar{p}_{i'} \bar{y}_{i'}}$ denote the share of goods consumption expenditure devoted to sector s and let $\bar{\chi}_{i(s)} \equiv \bar{\lambda}_i / \bar{\psi}_s$ denote the share of firm i within sector s . Finally, we let $\bar{\omega}_C$ denote the share of the household's virtual income spent on consumption goods.⁵⁷ Then we have

$$\frac{\Delta W}{W} = \frac{1}{2} \bar{\omega}_C \sum_s \theta_s \bar{\psi}_s \text{Var}_{\bar{\chi}_{(s)}} [\bar{\mu}] + \frac{1}{2} \bar{\omega}_C \rho \text{Var}_{\bar{\psi}} [\mathbb{E}_{\bar{\chi}_{(s)}} [\bar{\mu}]] + \frac{1}{2} \bar{\omega}_C (1 - \bar{\omega}_C) \eta (\mathbb{E}_{\bar{\lambda}} [\bar{\mu}] - 1)^2. \quad (23)$$

To unpack this expression, we begin by noting that all terms are multiplied by ω_C , as the only wedges in this economy are in consumption (rather than leisure). The first term captures the average effect of within-sector dispersion in wedges across firms. It therefore scales with the size of the within-sector demand elasticity of substitution, θ_s . In particular, $\text{Var}_{\bar{\chi}_{(s)}} (\bar{\mu})$ measures the amount of such dispersion within sector s , weighted by the size of each firm relative to the sector (i.e., by $\bar{\chi}_{(s)}$). The second term captures cross-sector dispersion in the average wedge within each sector (i.e., $\mathbb{E}_{\bar{\chi}_{(s)}} [\bar{\mu}]$). This scales with ρ , the consumer's substitution elasticity across sectors.

The misallocation of the aggregates \bar{X}_m across firms and sectors causes a welfare cost equal to the sum of these first two terms. By contrast, the final term arises due to misallocation of each \bar{X}_m itself. In particular, this component of misallocation exists when the consumption sector-wide (sales-weighted) *average* level of wedges is different from one. Unsurprisingly, this term scales with both the elasticity of input supply (η) and the size of leisure in the economy ($1 - \bar{\omega}_C$) since it is the division of aggregate inputs between firm production and leisure that is potentially misallocated. Finally, a notable feature of expression (23) is that technological features (of each firm's production function) do not enter, since within-firm dispersion is zero. However, this expression can be augmented to include such phenomena by simply adding components such as the first term in equation (22).

⁵⁷If the household's time endowment is \bar{T} and it earns the price \bar{w} for selling inputs then $\bar{\omega}_C \equiv \frac{\sum_i \bar{p}_i \bar{y}_i}{\bar{w} \bar{T}}$.

C Data Appendix

This appendix outlines the process followed to clean the data described in Section 3 and how we arrive at our final sample for analysis.

C.1 Public Procurement Lotteries and SERCOP Contracts

The Ecuadorian public procurement lottery system started in 2009. Through this system, contracts for public construction projects below a certain value were allocated through randomized lotteries among qualified suppliers. The threshold value is 0.00007% of the central government’s annual budget, which corresponds to \$134,176 in 2009 and \$240,100 in 2014. The qualification process for contract lotteries has several steps.

First, any tax-compliant firm can choose to register in the system. Second, government entities initiate a given procurement contract by sending specifications and an expected budget to the national procurement office (SERCOP).⁵⁸ Third, firms that are registered to provide the designated type of service, and are in good standing in regards to past contracts, are invited by SERCOP to submit applications, which include proof of relevant qualifications.⁵⁹ Examples of such services include construction or maintenance of public buildings, small roads and town squares, schools, sewerage, and wells.⁶⁰ Fourth, the procuring entity determines which applicants qualify for the contract. We refer to this set of firms as “lottery participants.” Finally, an automatic and centralized program at SERCOP determines the winner of the contract through a lottery among the participants. As a result, even though participation in any given lottery is the result of deliberate selection on both sides, the allocation of the contract among participants will be randomly determined.

We scrape information from the SERCOP contracting portal between 2009 and 2014 on the application procedure for 18,474 contracts under the procurement process “Contratos de Menor Cuantía de Obras”, which applied to all construction projects below a certain dollar

⁵⁸Roughly 5% of procuring entities are formally private-sector entities with a large share of state ownership.

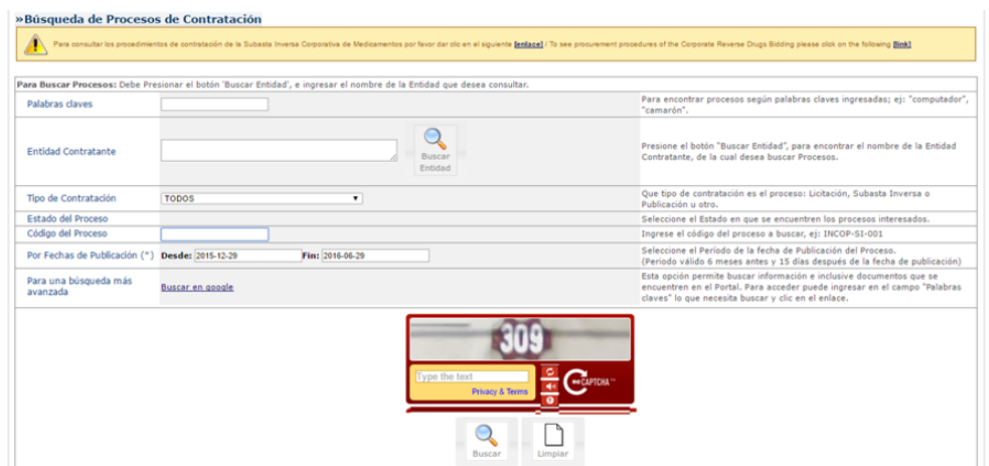
⁵⁹The invitation process is sometimes made over several rounds and often favors SMEs and local firms. In addition, starting in 2013, SERCOP required that the total amount of contracts a firm may enter at any given time is limited to the maximal allowed contract value (i.e., 0.0007% of the government budget).

⁶⁰Included are both physical construction and related services such as those of architects.

threshold.⁶¹ We then clean the collected information to create a dataset in which a single record corresponds to a particular firm’s involvement for a particular lottery contract.

Figures C.1 and C.2 show, respectively, the website from which we scrape the information, and a sample of the data that the portal provides for each contract.

Figure C.1: SERCOP contracting portal interface



Notes: This figure shows the interface of the SERCOP contracting portal. The user inputs the contract number and a range of dates. After entering the captcha, the website displays the contract information.

Source: SERCOP contracting portal website.

We gather the following information for each contract: which firms were invited to submit proposals for it, their IDs, whether they accepted the invitation, whether they submitted a proposal, whether they were deemed eligible for the lottery, and whether they ended up winning the contract. Other variables we retrieve are: the description of the contract, its category, amount, and the delivery deadline.⁶²

Starting from 38,813 scraped contracts, we kept only those with complete information and for which there exists more than one eligible participant and exactly one winner. These cleaning steps remove 20,339 (around 52.4%) of our initial observations, leaving us with 18,474 distinct contracts in total.⁶³ Then, we sent the remaining contracts to a third party

⁶¹This information is publicly available in the following website (working as of March 17, 2023): <https://www.compraspublicas.gob.ec/ProcesoContratacion/compras/SL/view/BusquedaDeProcesos.cpe>. In practice, we use an executable version of `scrape.py` (`scrape.exe`) created using `py2exe`. For more information, see <http://www.py2exe.org/index.cgi/Tutorial>.

⁶²There are seven contracts for which the duration is extremely long and likely wrong. We discard these observations, hence the difference in the number of observations in the first three rows of Table 1.

⁶³Most contracts are removed due to not having more than one eligible participant, while only three

Figure C.2: Relevant information for a sample contract

MARSEPON CIA. LTDA 0890051130001	ESTRUCTURAS [DERROCAMIENTO DE INSTALACIONES DEL CAMBIAMIENTO DE TERMOSMERALDAS]	No cumple	No cumple	No cumple	No cumple	Cumple
MOREIRA CERVERA LUIS MIGUEL 0800429546001	SERVICIOS DE DERRIBO Y DEMOLICION DE EDIFICIOS Y OTRAS ESTRUCTURAS [DERROCAMIENTO DE INSTALACIONES DEL CAMBIAMIENTO DE TERMOSMERALDAS]	Cumple	Cumple	No cumple	No cumple	Cumple
PICO NAVARRO MARIA ELENA 1304932617001	SERVICIOS DE DERRIBO Y DEMOLICION DE EDIFICIOS Y OTRAS ESTRUCTURAS [DERROCAMIENTO DE INSTALACIONES DEL CAMBIAMIENTO DE TERMOSMERALDAS]	Cumple	Cumple	No cumple	No cumple	Cumple

Proveedores habilitados para el sorteo por parte de la Entidad Contratante

Nro.	Proveedor	Descripción	Estado
1	ALVARADO ALMEIDA ERISTRI ELIAS	cumple con todos los requisitos técnicos solicitados en los pliegos	Habilitado
2	MOREIRA CERVERA LUIS MIGUEL	cumple con todos los requisitos técnicos solicitados en los pliegos	Habilitado
3	PICO NAVARRO MARIA ELENA	cumple con todos los requisitos técnicos	Habilitado

Bien/Obra/Servicio Adjudicados

Categoría	Descripción Bien/Obra/Servicio	Proveedor	Cantidad Adjudicada	Precio Unitario	Subtotal	Tiempo de Entrega	Razón Adjudicación	Estado
543100011	SERVICIOS DE DERRIBO Y DEMOLICION DE EDIFICIOS Y OTRAS ESTRUCTURAS [DERROCAMIENTO DE INSTALACIONES DEL CAMBIAMIENTO DE TERMOSMERALDAS]	Pico Navarro Maria Elena 1304932617001	1	USD 167,180.14	USD 167,180.14	30	Proveedor ganador en el Sorteo MC-Obras con estos productos	Adjudicado

Notes: This figure shows information about a sample contract. The lottery winner, the value of the contract and the time to delivery, among other variables, are inside yellow squares.

Source: SERCOP contracting portal website.

to check that all the information was correct. Specifically, this checking involved comparing the scraped values to the actual contractual documentation (available in the same website).⁶⁴ We use the contract values as confirmed through this process.

C.2 Income Tax Forms

The firms' income tax filings provide self-reported information of firms (i.e., their revenue, costs, tax liability, among others).

F101/102 forms

In Ecuador, all incorporated firms are legally obligated to submit an annual detailed corporate income tax form (F101), independent of how large their revenues, costs or assets are. This also applies to all publicly-owned firms. On the other hand, unincorporated firms (which mainly consist of self-employed individuals) are required to file the income tax form F102 if their annual revenue exceeds a standardized deduction amount (which was approximately \$10,000 in our sample period). These forms include self-reported information of the firms' revenue and costs (broken down by certain sub-categories). Small unincorporated contracts are removed due to not having exactly one winner after removing submissions made without documents or by non-eligible firms.

⁶⁴We found errors on around 1% of the contracts and manually fixed them. These errors were due to discrepancies between the contracts and the website itself, not due to our scraping procedure.

firms submit a simplified version of the income tax form (“short form F102”) which corresponds to the personal income tax form. Large unincorporated firms have to file an extended income tax (“long form F102”) consisting of two parts: a part for reporting business income corresponding to the form for incorporated firms and a part for reporting individual income mirroring the short form F102.

In very rare cases (i.e., 0.005% of firm-year observations in the annual income tax data), a firm reports both a F101 and a F102 form. Whenever this is the case and the forms were filed on the same date, we proceed as follows. If the duplicates share the same value in the same tax items, in the first instance we keep the form that the firm was expected to correctly file. In the second instance, we track the first year in which the firm uniquely filed a form and keep the one that matches the type of this first unique filing. If the duplicates have different values in the same tax item, then we keep the observation that has the highest value for each item. On the other hand, if the forms were filed on different dates, then we keep the most recent form.

Self-reported sales and costs measures

Our measures of self-reported sales and costs are constructed based on the annual income tax filings. In line with the theory developed in Section 2, we seek to capture (i) the firm’s income from selling its goods or services and (ii) the costs associated with producing this output.

Accordingly, the total sales measure includes domestic sales, net exports, and other revenues related to selling goods or services. Table C.1 lists the sales items from the income tax form included in our measure of self-reported sales. We deliberately exclude revenue from dividends, financial rents, donations and contributions, and capital gains on sales of fixed assets. The total costs measure comprises labor costs and non-labor costs as shown in Table C.2. We deliberately exclude reported costs due to losses from the sales of assets, and we also disregard tax liabilities deducted from the income tax (e.g., property taxes, excise tax, non-creditable VAT).⁶⁵

⁶⁵Specifically, we compute our sales and costs measure by subtracting the excluded line items from the reported total sales and costs measures, respectively. This has the advantage that a few firms only report information on their total costs or revenues but not on sub-categories. In the very few instances in which the adjusted total sales or costs measure is negative due to reporting errors, we replace it with a zero.

Table C.1: Sales Line Items in Forms 101 and 102

Sales item
Domestic sales subject to 12% tax rate
Domestic sales subject to 0% tax rate
Exports
Other income from abroad
Other taxable income
Other exempted income
Change in inventory of products produced by the firm
Sales related to registered business activities*
Received fees*
Received agricultural income*
Received income from abroad*
Received wages*

This table shows the line items included in the definition of total sales for the annual income tax form. * marks line items that are only present for F102 and related to the “short form” section of the form.

Table C.2: Cost Line Items in Forms 101 and 102

Costs items	Category
Wages, salaries and other taxable remunerations	Labor costs
Social benefits and other non-taxable compensation	Labor costs
Contribution to social security (including reserve fund)	Labor costs
Obligatory profit share passed on to employees	Labor costs
Costs related to wages*	Labor costs
Net purchases of domestic goods not produced by the firm	Non-labor costs
Imports not produced by the firm	Non-labor costs
Net domestic purchases of raw material	Non-labor costs
Imports of raw material	Non-labor costs
Professional fees and expenses	Non-labor costs
Fees to foreigners for one-time expenses	Non-labor costs
Real estate rent	Non-labor costs
Maintenance and repairs	Non-labor costs
Fuel	Non-labor costs
Marketing	Non-labor costs
Supplies and materials	Non-labor costs
Transportation	Non-labor costs
Commercial leasing	Non-labor costs
Commissions	Non-labor costs
Insurance and reinsurance intermediaries	Non-labor costs
Administrative costs	Non-labor costs
Travel expenses	Non-labor costs
Public services	Non-labor costs
Payment for other goods and services	Non-labor costs
Bank interest	Non-labor costs
Interest paid to third parties	Non-labor costs
Amortization	Non-labor costs
Change in inventory of goods not produced by the firm	Non-labor costs
Change in inventory of raw material	Non-labor costs
Depreciation of fixed assets	Non-labor costs
Provisions	Non-labor costs
Other losses	Non-labor costs
Indirect costs incurred from abroad by related parties	Non-labor costs
Costs related to registered business activities*	Non-labor costs
Costs related to professional activity or liberal occupation*	Non-labor costs
Costs related to real estate rents*	Non-labor costs
Costs related to other assets*	Non-labor costs
Costs related to agricultural income*	Non-labor costs
Costs related to other income*	Non-labor costs

This table shows the line items included in the definition of total cost for the annual income tax form, and which are included in the definition of implied labor and non-labor costs. * marks line items that are only present for F102 and related to the “short form” section of the form.

We construct these sales and costs measures to represent prices gross of sales taxes (e.g., VAT and excise tax). This attempts to consistently follow our theory which represents sales at the actual expense of the buyer (i.e., including any statutory sales tax). On the F101/102, firms report revenues and costs net-of-VAT, but gross of other sales taxes.⁶⁶ Hence, we adjust our self-reported sales measure as follows.

1. Firms report sales subject to 12% VAT in a dedicated line item.
2. We multiply the amount reported by 0.12 and add it to our measure of total self-reported sales.

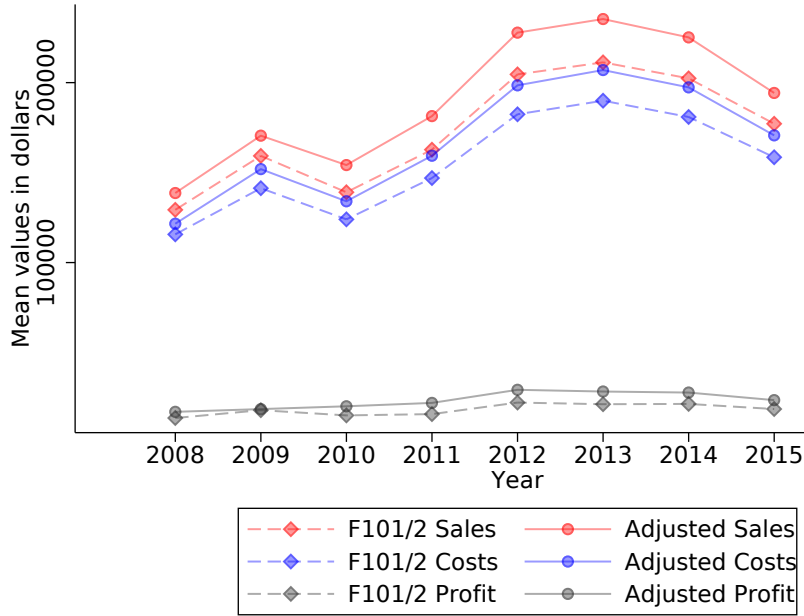
Since costs are not reported separately by their VAT rate, we proceed as follows to compute the gross-of-VAT costs based on the self-reported net-of-VAT amounts.

1. We identify the line items that are subject to VAT and calculate their total amount (denoted V_i)
2. If a firm files purchase annexes, we calculate the (observed) share of purchases subject to 12% VAT (denoted s_i). For firms that do not file purchase annexes, we predict this share by their level of sales and incorporation status.
3. We calculate the VAT amount as $V_i \times s_i \times 0.12$ and add it to our measure of total self-reported costs.

All in all, our adjusted measures of sales and costs (and profits) are consistent with the total revenue and costs items (and profits) from the annual tax filling. They are also consistent across years. This consistency is shown in Figure C.3, which plots the evolution over time of these variables in their raw and adjusted versions.

⁶⁶In Ecuador, sales are subject to either a 12% or 0% VAT rate, or are exempt from the VAT.

Figure C.3: Sales and Cost Measures



Notes: This figure shows the evolution across years of the sales, costs and profit measures for their raw (F101/2) and adjusted versions.

C.3 Purchase Annexes

The monthly purchase annexes provide both self-reported firms' purchases information (i.e., their own purchases) and, implicitly, third-party reported information about other firms' sales.

All incorporated firms (F101 filers) are required to submit purchase annexes every month. F102-filing firms must keep accounting records and file purchase annexes on a monthly basis if their annual revenues surpass a certain threshold (e.g., \$100,000 in 2012). They are also mandated to do this if their annual costs and expenses exceed a given value (\$80,000 in 2012), or if they begin economic activities with capital above a certain threshold (\$60,000 in 2012).⁶⁷

To quantify transactions between firms using the purchase annex, we follow a process similar to that described in Adao et al. (2022). For each transaction, we observe the anonymized

⁶⁷A considerable number of firms that are not required to do so end up voluntarily filing purchase annexes. However, for these smaller firms (who represent a smaller share of the country's economic activity, naturally) records may not be complete. See Adao et al. (2022) for further details.

tax ID of both the buyer and the seller, the value of the purchase, the VAT paid, what part of the value was subject to either a tax rate of 12% or 0%, and the transaction’s date. The data feature some inconsistencies. For instance, we drop transactions with a negative VAT value. When there is a positive value for VAT but the transaction value is inconsistent with it, we rectify the transaction value. In cases where there is a positive transaction value but VAT is missing, we calculate the correct VAT based on the transaction value. We also do not consider purchases whose value exceeds a value greater than two times the maximum between the buyer’s annual cost and the seller’s annual revenue, as such transactions are likely the result of data entry errors. Finally, we also drop transactions where the buyer and the seller are the same firm. The cleaned purchased annexes are used to compute monthly, half-yearly and annual transactions between firms.⁶⁸

Since the transactions are reported purchases by the buyer, we compute a measure of *third-party* reported sales based on the purchase annex data.⁶⁹

C.4 Social Security and Firm Characteristics

First, we combine two sources of data for employment. One covers the period from 2007 to 2017 and is extracted from the social security record that keeps track of all the employees whose employer filed the social security for them; the other one spans from 2009 to 2016, and has data from either the social security record or the F107 form (a tax declaration form filed by firms with information about their workers). We restrict the sample to the universe of employees that ever work for lottery firms. Then, we identify the month in which an employee is hired by a firm whenever there are no payments from the employer in the previous month. Similarly, we consider the employee to be a new employee if they were hired in the last 12 months, and we define documented employees as those that have worked for any firm in this same period.

Second, we have basic information of every firm registered with the tax authority up to

⁶⁸We use the register date when available. Otherwise, we use the purchase date. We drop transactions where neither of these two is available.

⁶⁹Similar to the self-reported sales measure, we construct third-party reported sales gross of sales taxes. To obtain this measure, we sum up the reported net-of-VAT transaction amount and the VAT reported for that transaction to create a measure of third-party reported sales gross-of-VAT.

2012. This dataset provides a snapshot of the registered firms’ characteristics in this year for variables such as: the province where it is located, the year in which it became active and the industry code (ISIC 3.1).

C.5 Combining the Datasets

We can combine the different datasets based on consistently anonymized firm identifiers across all datasets.

To identify sales from lottery participants to different types of entities, we join the lottery data with the transaction-level purchase annexes, information from the procurement contracts, and information on firm characteristics. This allows us to identify whether the reporting firm (i.e., the buyer in the purchase annex) is a procuring entity, another public entity, or a private firm.⁷⁰ We then aggregate the reported purchases on the month-seller level to obtain measures of third-party reported sales to different types of entities as well as total third-party reported sales.

Finally, we define a firm to be economically active from the first time it self-reports positive sales or costs in its income tax forms, or appears in any of the above datasets either as a lottery participant, an employer (in the social security data), or a supplier (in another firm’s purchase annex). We then exclude firm-year observations from the period before a firm is economically inactive and impute zeroes for any missing values after a firm has become economically active.

⁷⁰We define “public entities” as firms that are either government agencies or utilize the lottery procurement system. Similarly, “private firms” are defined as firms that are neither a government agency nor utilize the lottery procurement system.

D Test of statistical independence

This appendix provides further details about the test of statistical independence used in Section 5 to test for U-AEP and C-AEP.

D.1 Testing for U-AEP

As stated in Proposition 1, the null of U-AEP can be carried out by testing for whether $f(\cdot) = 0$ in equation (11) or not. Following Ding et al. (2016), one way to do this is via a quantile regression of $\Delta\Pi_i(\mathbf{1})$ on Z_i . This corresponds to the quantile model

$$Q_{\Delta\Pi(\mathbf{1})}(\tau | Z) = \alpha(\tau) + \beta(\tau)Z, \quad (24)$$

for any quantile τ , where the null hypothesis of U-AEP requires that $\beta(\tau) = 0$ for all τ . We therefore use the test statistic

$$TS \equiv \max_{\tau \in \mathcal{T}} |\beta(\tau)|, \quad (25)$$

where \mathcal{T} denotes some finite set of quantiles to be evaluated. The null should be rejected when TS is large.

We implement a randomization inference version of this test. Following Fisher (1935), the basic idea is to compare TS based on the actual data to the distribution of statistics TS_l that one obtains in a series of simulations $l = 1 \dots L$ in which the winners of each lottery are randomly re-drawn from the set of actual entrants into each lottery and the sharp null value is used to generate simulated outcome data for each observation. Since the sharp null in question is one of zero treatment effect (of Z_i on $\Delta\Pi_i(\mathbf{1})$), this procedure amounts to performing a quantile regression of the actual data $\Delta\Pi_i(\mathbf{1})$ on the values of Z_i generated by simulated counterfactual lottery winnings and repeating this across many simulations. Specifically, we proceed as follows:

1. Estimate TS on the actual data and denote this value by \widehat{TS} . That is, estimate a quantile regression of $\Delta\Pi_{i,t}(\mathbf{1})$ on $Z_{i,t}$, obtain the quantile coefficient estimates $\widehat{\beta}(\tau)$,

and compute the corresponding \widehat{TS} based on (25).⁷¹ We set $\mathcal{T} = \{0.1, \dots, 0.9\}$.

2. Re-randomize the identity of the winning firm in each lottery. Refer to the values of Z_i that are obtained under this re-randomization as Z_i^l .
3. Estimate a quantile regression of $\Delta\Pi_{i,t}(\mathbf{1})$ on Z_i^l . Calculate the corresponding value of TS_l based on (25), again with $\mathcal{T} = \{0.1, \dots, 0.9\}$.
4. Repeat steps #2 and #3 an additional $L - 1$ times. In practice, we set $L = 100$.
5. The resulting L values of TS_l provide an estimate of the exact null distribution of the test statistic TS . Therefore, calculate the exact p-value for the null hypothesis from the share of simulations l in which $TS_l > \widehat{TS}$. If $p < \alpha$, we reject the null at the α level.

Figure 8, at the value displayed as $\chi = 1$ on the x-axis, reports the p-value from performing this test in our context.

D.2 Testing for C-AEP

Recall from Proposition 1 that the null of C-AEP is the same for that of U-AEP apart from the fact that rather than testing on the basis of $\Delta\Pi_i(\mathbf{1})$, we must repeat the test on the basis of $\Delta\Pi_i(\boldsymbol{\chi})$ for all values of $\boldsymbol{\chi} > \mathbf{0}$. As discussed in Section 5.2, we confine attention to values of $\boldsymbol{\chi}$ in which all elements of this vector are equal to each other and equal to the scalar χ (i.e., $\chi_m = \chi$ for all m). Our test for C-AEP therefore simply amounts to repeating the above test for U-AEP (which was effectively done for $\chi = 1$) at a wider range of values of χ . In practice, we do this for 201 uniformly spaced values in the interval $[0.9, 1.4]$. Figure 8 reports the p-values for each of these tests. The test for C-AEP fails to reject (at the 5% level) if any of the p-values exceeds 0.05.

⁷¹As described in Section 5, the lack of available data on prices of outputs and inputs in our context requires that we proxy for $\bar{p}_i \Delta y_i$ with $\Delta(p_i y_i)$, etc. We therefore have $\Delta\Pi_i(\mathbf{1}) = \Delta(p_i y_i) - \sum_m \Delta(w_m x_{im})$. Further, when computing time-differences “ Δ ” we use differences in two-year averages, such as $\Delta(p_{i,t} y_{i,t}) \equiv \frac{1}{2}(p_{i,t} y_{i,t} + p_{i,t+1} y_{i,t+1}) - \frac{1}{2}(p_{i,t-1} y_{i,t-1} + p_{i,t-2} y_{i,t-2})$. We then stack the cross-sections from all such available changes over 2008–2015, but only use the firm-year observations corresponding to firms that are active in 2008 so that our testing sample aligns with those firms that enter our sample for estimating sales-weighted wedge distributions in Section 6. Finally, the instrument is defined as $Z_{i,t} = D_{i,t} - D_{i,t-2}$.

E Assessing First-Stage Rank-Invariance

Section 2.4 describes the use of the IVCRC estimator developed in Masten and Torgovitsky (2016) to estimate the sales-weighted first and second moment of the distribution of wedges. As outlined in Assumption 4, one condition for consistency of IVCRC is what Masten and Torgovitsky (2016) refer to as “first-stage rank-invariance”: that the first-stage relationship can be characterized by $\bar{w}_m \Delta x_{im} = h_m(Z_{im}, V_{im})$ for some unknown $h_m(\cdot)$ and scalar V_{im} , with $\frac{\partial h_m}{\partial V_{im}} > 0$ for all $m \in \bar{\mathcal{M}}(i)$ and all Z_{im} . This condition implies that (for any input type m), if we set Z_{im} to any value z , then the ranking of firms in terms of their input changes (i.e., $\bar{w}_m \Delta x_{im}$) would not depend on the value of z chosen. Put differently, firms can respond to Z_{im} heterogeneously but only to the extent that their rank in the input change distribution is unchanged by such response heterogeneity.

In order to gauge the plausibility of this assumption in our context, in this Appendix we conduct a simulation of a model inspired by features of Hsieh and Klenow (2009), calibrated to our regression sample, that suggests that first-stage rank-invariance may hold approximately in our context.

E.1 Setup

We consider a set of N firms indexed by i . Each produces a differentiated product y_i using a single input x_i according to the technology

$$y_i = a_i x_i, \tag{26}$$

where a_i denotes firm i 's productivity. The firm faces two sources of demand: (a) private-sector buyers whose demand is given (in expenditure terms) by $E_i = p_i^{1-\sigma}$, where σ is a constant price elasticity of demand; and (b) procurement lottery-based demand denoted (again in expenditure terms) by W_i . The firm's total sales ($S_i \equiv p_i y_i$) are therefore given by $S_i = E_i + W_i$.

The definition of the wedge in equation (2) implies, in this model, that $\mu_i = p_i a_i / w$, where w denotes the price of the input. Starting from the “time-0” allocation (at which all

variables are denoted with a bar), we consider a small change in the input Δx_i (between time-0 and a subsequent time period) that results from an arbitrary change in a_i and W_i , but where, for the sake of simplicity in this simulation, μ_i and w are held constant. This will imply that the first-stage of our IVCRC estimation equation is given by

$$\bar{w}\Delta x_i = [(\sigma - 1)\bar{w}^{1-\sigma}(\bar{\mu}_i)^{-\sigma}\bar{a}_i^{\sigma-1}] \hat{a}_i + (\bar{\mu}_i)^{-1}\Delta W_i, \quad (27)$$

where $\hat{a}_i \equiv \Delta a_i/\bar{a}_i$.

As discussed in Section 3.3, procurement lottery-based demand W_i satisfies $W_i = D_i + \mathbb{E}[W_i \mid \{A_k, N_k\}_{k \in \mathcal{K}_i}]$, where A_k denotes the contract value of lottery k , N_k denotes the number of participants in lottery k , \mathcal{K}_i denotes the set of lotteries that firm i participates in during the given time period, and D_i denotes the deviation between actual lottery demand W_i and expected lottery demand given lottery participation. Given the definition of our instrument, $Z_i \equiv \Delta D_i$, we therefore have

$$\bar{w}\Delta x_i = b_i + (\bar{\mu}_i)^{-1}Z_i, \quad (28)$$

where we have defined

$$b_i \equiv [(\sigma - 1)\bar{w}^{1-\sigma}(\bar{\mu}_i)^{-\sigma}\bar{a}_i^{\sigma-1}] \hat{a}_i + (\bar{\mu}_i)^{-1}\Delta \mathbb{E}[W_i \mid \{A_k, N_k\}_{k \in \mathcal{K}_i}]. \quad (29)$$

Input changes ($\bar{w}\Delta x_i$) therefore derive from two components. First, the component b_i arises due to factors unconnected from the instrument (the technology and lottery participation changes in equation 29). And second, the component $(\bar{\mu}_i)^{-1}Z_i$, which does depend on the instrument. In this model, first-stage rank-invariance would be satisfied if there were no heterogeneity in $\bar{\mu}_i$ because in such a case equation (28) could trivially be written in the form $\Delta x_i = h(Z_i, V_i)$ with a scalar V_i and with $\frac{\partial h}{\partial V_i} > 0$ at any value of Z_i . More generally, if there is heterogeneity in $\bar{\mu}_i$ and also in b_i , and $Corr(\bar{\mu}_i, b_i) < 0$, then first-stage rank-invariance is not guaranteed. In what follows, we use the structure of the model in this Appendix in order to calibrate the values of b_i and $\bar{\mu}_i$ implied by our data, and thereby calculate how often rank-invariance is violated.

E.2 Quantifying First-Stage Rank-Reversals

In principle, one could assess the prevalence of first-stage rank-invariance in this model by: (a) measuring b_i and $\bar{\mu}_i$ for every firm i ; (b) computing each firm’s rank in the distribution of $\bar{w}\Delta x_i$ implied by equation (28) at the lowest possible value of the instrument Z ; and then (c) repeating this at every other possible value of Z in order to keep track of the number of times that firms’ ranks (in the distribution of $\bar{w}\Delta x_i$) reverse across the full support of Z values. In practice, however, this is both computationally costly (given the number of potential calculations involved) and also conceptually unclear, given that Z is in principle a continuous variable. We therefore follow Gollin and Udry (2021) in drawing a random sample of comparisons (both pairs of firm observations whose rank is to be assessed and sets of values of Z at which we assess whether rank-reversal has occurred) from the set of all possible comparisons. To the extent that this random sample is large, it should provide a reliable impression.

We calibrate the model as follows. First, we set $\sigma = 3$ as in Section 6.3 (and, for example, in Hsieh and Klenow (2009)). Second, we let \bar{w} equal the average annual wage for a worker in our dataset. Third, we pursue two versions for the calibration of $\bar{\mu}_i$ in parallel: (a) a version in which these come from the sales bin-specific average wedge estimate obtained from the IVCRC estimation procedure in Section 6.1; and (b) a version that assumes all firms have constant returns-to-scale technologies, as in Section 6.3 and, for example, Hsieh and Klenow (2009).⁷² Fourth, given data on S_i and W_i , manipulations of the above supply and demand equations can be used to solve for \bar{a}_i and \hat{a}_i .⁷³ Finally, $\mathbb{E}[W_i \mid \{A_k, N_k\}_{k \in \mathcal{K}_i}]$ is directly observable. Applying these steps we therefore know the values of b_i and $\bar{\mu}_i$ (one for each separate method for calibrating $\bar{\mu}_i$) corresponding to every firm-year observation in our dataset.

Using these values of b_i and $\bar{\mu}_i$, we quantify rank-reversals using the following procedure (done twice for each separate method for calibrating $\bar{\mu}_i$):

1. Randomly partition all firm-year observations into pairs r . Let r_1 and r_2 indicate the

⁷²As discussed in Section 6.3, under constant returns-to-scale we have $\bar{\mu}_i = \bar{S}_i / (\sum_m \bar{w}_m \bar{x}_{im})$.

⁷³That is, combining the equations $S_i = E_i + W_i$, $E_i = p_i^{1-\sigma}$, and $\mu_i = p_i a_i / w$, we have $a_i = w \mu_i (S_i - W_i)^{1/\sigma}$.

two observations in pair r . Let there be P pairs in total.

2. Begin with pair $r = 1$. Randomly draw a value of the instrument Z from the distribution of the deviations from expected winnings and denote that value by Z_a . Evaluate the first-stage equation (28) at $Z_i = Z_a$ for observation r_1 and denote the result by $F(r_1, Z_a)$. Do the same for observation r_2 and denote the result by $F(r_2, Z_a)$.
3. Repeat step #2 for a second randomly drawn value of the instrument, denoted Z_b . Hence calculate $R(r) \equiv [F(r_1, Z_a) - F(r_2, Z_a)][F(r_1, Z_b) - F(r_2, Z_b)]$. This provides one possible comparison (between two observations and two values of the instrument).
4. Repeat steps #2 and #3 for nine additional sets of randomly drawn instrument value pairs. This yields ten values for $R(r)$ in total for the first pair, $r = 1$.
5. Repeat steps #2-#4 for all remaining pairs $r = 2 \dots P$.
6. Rank-reversals have occurred whenever $R(r) < 0$. Calculate the share of comparisons (out of $10P$ total comparisons) in which this has happened.

Table E.1 reports the share of rank reversals in the randomly chosen comparisons illuminated by the above simulation. Unsurprisingly, the substantially smaller wedge dispersion used in the IVCRC-based version of this simulation (columns 1-3) shows far less propensity for rank-reversals than does the constant returns-based version (columns 4-6). But even in the latter case, rank reversals are quite rare, happening in less than 6.2% of pairwise comparisons.

Table E.1: Frequency of Rank-Reversals in Simulated Economy

IVCRC wedges			CRTS wedges		
Rank reversals (1)	No. of comparisons (2)	Share (3)	Rank reversals (4)	No. of comparisons (5)	Share (6)
1,256	141,820	0.0089	7,565	123,070	0.0615

Notes: This table shows results on the frequency of first-stage rank-reversals across randomly drawn pairwise comparisons of observations and instrument values. “IVCRC wedges” assigns firms’ wedges based on the sales-bin specific point estimates of average wedges obtained when using the IVCRC procedure described in Section 6.1. “CRTS wedges” instead uses the formula $\bar{\mu}_i = (\bar{p}_i \bar{y}_i) / (\sum_m \bar{w}_m \bar{x}_{im})$, which follows from assuming that all firms have constant returns-to-scale technologies as in Section 6.3, in order to estimate wedges. The number of comparisons differs across these alternatives due to the fact that the latter version cannot be computed for firms with zero costs.