Grouped Variation in Factor Shares: An Application to Misallocation

by

Jose Asturias
U.S. Census Bureau

Jack Rossbach
Georgetown University Qatar

CES 22-33 August 2022

The research program of the Center for Economic Studies (CES) produces a wide range of economic analyses to improve the statistical programs of the U.S. Census Bureau. Many of these analyses take the form of CES research papers. The papers have not undergone the review accorded Census Bureau publications and no endorsement should be inferred. Any opinions and conclusions expressed herein are those of the author(s) and do not represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Republication in whole or part must be cleared with the authors.

To obtain information about the series, see www.census.gov/ces or contact Christopher Goetz, Editor, Discussion Papers, U.S. Census Bureau, Center for Economic Studies, 4600 Silver Hill Road, Washington, DC 20233, CES.Working.Papers@census.gov. To subscribe to the series, please click here.
Abstract

A striking feature of micro-level plant data is the presence of significant variation in factor cost shares across plants within an industry. We develop a methodology to decompose cost shares into idiosyncratic and group-specific components. In particular, we carry out a cluster analysis to recover the number and membership of groups using breaks in the dispersion of factor cost shares across plants. We apply our methodology to Chilean plant-level data and find that group-specific variation accounts for approximately one-third of the variation in factor shares across firms. We also study the implications of these groups in cost shares on the gains from eliminating misallocation. We place bounds on their importance and find that ignoring them can overstate the gains from eliminating misallocation by up to one-third.

**Keyword:** Cluster Analysis, Misallocation

**JEL Classification:** D24, O11, O47, O14

---

* This paper was previously circulated with the title “Misallocation in the Presence of Multiple Production Technologies.” We thank Manuel Garcia-Santana, Timothy Kehoe, Pau Pujolas, James Schmitz, Kirk White, and Chunying Xie for their feedback on this paper. Any opinions and conclusions expressed herein are those of the author(s) and do not represent the views of the U.S. Census Bureau. Asturias: jose.asturias@census.gov; Rossbach: Jack.Rossbach@georgetown.edu.
1 Introduction

Research using micro-level firm data has found that factor cost shares (expenditures on each factor input divided by revenues) vary widely across firms, even within narrowly defined products and industries. This variation is often interpreted as arising from unobserved distortions, which leads to the misallocation of factor inputs within and across firms.

In this paper, we show, by applying a novel methodology, that there is a clustering pattern in these cost shares. Specifically, we develop a methodology based on cluster analysis to decompose cost shares into a firm-specific idiosyncratic component and a group-specific component that is common across all firms in a cluster (hereafter, we refer to this component as the cluster-specific component).

Cluster analysis works by identifying groups of observations that share similar characteristics — in our case, cost shares — and is common in applications in which pattern recognition is important. For example, cluster analysis is commonly used in image processing and population inference based on genotype data.\(^1\) Importantly, our methodology does not require us to specify the number of clusters present within an industry or firm membership among these clusters a priori, but instead recovers these objects as part of our analysis by detecting gaps in the dispersion of factor cost shares across firms.

We next integrate our methodology with a standard model of firm heterogeneity. In our model, we allow for firm idiosyncratic and group distortions as well as different technologies within an industry. We use the model to show how the cluster analysis should be applied to the data, how to interpret the results, and under what conditions our methodology will successfully identify the number and assignment of clusters across firms.

We apply our methodology to Chilean plant-level manufacturing data and find strong evidence for the existence of clusters in cost shares for the majority of industries.\(^2\) We find that 14 of the 23 industries we study have more than one cluster: 7 industries have 2 clusters, 5 industries have 3 clusters, and 2 industries have 4 clusters. We conduct a variance decomposition and find that cluster-specific variation, the variation accounted for by factors that are common across firms in a cluster, accounts for a substantial fraction of the overall variation: 23 percent of the variation is cluster-specific in industries with 2 clusters and 47 percent in industries with 4 clusters on average. Note that this cluster-specific variation is often referred to as between-cluster variation.\(^3\) We

---

1. See Jain (2010) for a survey of the use of clustering in pattern recognition. For examples in the field of image processing, see Park, Yun, and Lee (1998) and Ray and Turi (1999); for examples in the field of population inference, see Jakobsson and Rosenberg (2007) and Pritchard, Stephens, and Donnelly (2000).
2. The methodology that we develop is intended primarily for intra-industry detection of clusters. It may not be appropriate to apply the methodology across broadly defined industries because the main identification of the methodology comes from only the cost shares of firms and it is possible for distinct industries to have similar cost shares.
3. The cluster-specific variation is related to the between-cluster variation in cost shares, whereas the idiosyncratic
also find that the importance of cluster-specific variation across factor inputs varies substantially. For industries with 2 or more clusters, cluster-specific variation explains only 7.2 percent of the variation in intermediate input cost shares; however, it explains 37.5 percent of the variation in unskilled labor cost shares and 54.7 percent of the variation in capital cost shares.

Our framework indicates that the magnitude of the gains from eliminating misallocation depends on the underlying cause of clustering in cost shares. If the clustering in cost shares is the result of cluster-specific distortions (e.g., a capital subsidy to all firms in that cluster), not from different technologies, all of the variation in cost shares would still be interpreted by the model as being a result of distortions. On the other hand, if clustering in cost shares are driven by differing technologies, then the model would interpret some of the variation in cost shares to these differing technologies, thereby reducing the gains from eliminating misallocation. Therefore, the gains from eliminating misallocation would be overstated if the variation in cost shares is, at least in part, caused by a difference in technologies.

We conduct two exercises which suggest that differences in clusters are related to the technology used by plants. First, we find consistent patterns between the size of plants and the clusters to which they are assigned in Chile (e.g., larger plants use capital more intensively relative to labor) that are similar to those found using U.S. manufacturing data. This comparison to the United States is useful as this economy is typically considered to be undistorted in the literature and thus these differences in cost shares and plant size would be technological in nature.

Second, we conduct a case study of the Indian steel manufacturing industry, which has available data that allow us to identify different production technologies used within the industry using a methodology similar to Collard-Wexler and De Loecker (2015). Indian steel plants fall into two general categories: vertically integrated plants and those that use electric furnaces. We find that our methodology recovers clusters that correlate with those different production technologies. This result, along with the comparison with U.S. manufacturing, indicates that clusters identified by our methodology are correlated with production technologies.

To quantify the potential bias that can arise in estimating gains from eliminating misallocation because of the difference in technologies, we conduct two exercises to place bounds on the importance of differing technologies across clusters in accounting for the gains from eliminating misallocation. In the first exercise, we suppose that the differences in clusters are driven solely by cluster-specific distortions. We find that in this case, the gains from eliminating misallocation are similar to the gains suggested by a standard framework in which there is only one cluster. In the second exercise, we suppose that the differences in clusters are entirely a result of different technologies (and not cluster-specific distortions). We find that in this case, the gains from eliminating misallocation decline by approximately 30 percent relative to the standard framework.4 This result

4In this case, the differing production technologies across clusters reduces the correlation between the marginal revenue products for skilled and unskilled labor and firm size, which Restuccia and Rogerson (2008) and Hopenhayn...
suggests that the gains from eliminating misallocation would be substantially overstated if clusters arise because of differences in technologies.

We finally perform a decomposition in which we assume that differences between clusters that can be attributed to differences in firm sizes are a result of cluster-specific distortions rather than different technologies. This assumption allows us to back out an estimate for both technologies and cluster-specific distortions. In this case, we still find that technology accounts for two-thirds of the differences in cost shares across clusters. In addition, when we consider the gains from eliminating misallocation, our results are virtually unchanged by allowing for both cluster-specific distortions and cluster-specific technologies compared to when we allow for only differing technologies.

**Literature review** Economists have made significant advances over the past decade in understanding how misallocation of inputs across firms can be a major factor in explaining the gaps in output and total factor productivity (TFP) between rich and poor countries, as first pioneered by Hsieh and Klenow (2009).  

Our findings have implications for the burgeoning literature that asks what fraction of measured misallocation reflects true misallocation as opposed to other sources of variation—for example, model misspecification and data mismeasurement. Hopenhayn (2014) discusses possibilities such as measurement error, the degree of decreasing returns to scale, and capital adjustment costs. Asker, Collard-Wexler, and De Loecker (2014) find that capital adjustment costs can potentially account for a large portion of the wedges on capital and therefore the observed dispersion in the marginal revenue product (MRP) of capital. Other papers that explore sources of mismeasurement include those by Li (2015), Schelkle (2016), Rotemberg and White (2020), and Bils, Klenow, and Ruane (2021).

Our findings are complementary to alternative explanations for measured misallocation in the literature, such as adjustment costs and measurement error. For example, adjustment costs lead to dispersion in cost shares around a single point instead of multiple centers around which there is dispersion in cost shares. Therefore, we are targeting a different type of dispersion compared to these other papers, although we would not be surprised if adjustment costs and measurement error explain a significant amount of the misallocation that remains after controlling for multiple clusters. Likewise, multiple clusters can explain some features of the data that are not easily explained by these alternatives. For example, the misallocation of unskilled and skilled labor, where capital adjustment costs are not relevant, is a significant source of misallocation in our data.

(2014) show amplifies the misallocation that results from the dispersion and leads to larger losses in overall output and TFP.

5 Applications of this methodology are found in a wide variety of settings, including the study of misallocation during times of crisis (Oberfield (2013), Sandleris and Wright (2014)), misallocation in southern Europe (Dias, Marques, and Richmond (2016), Garcia-Santana, Moral-Benito, Pijoan-Mas, and Ramos (2019), Gopinath, Kalemli-Ozcan, Karabarbounis, and Villegas-Sanchez (2015)), and misallocation in the United States in the 1800s (Ziebarth (2013)) and across Latin American countries (Busso, Madrigal, and Pages (2013)).
Furthermore, the distortions for those inputs are correlated with firm size. If clusters are driven by differences in technologies, then our finding that larger firms use less labor-intensive technologies provides a natural explanation for this correlation, whereas alternatives such as measurement error may struggle in this regard.

Our paper is also complementary to Song and Wu (2015), which focuses on measuring capital misallocation in China. They allow for multiple production technologies in the sense that they assume the capital intensity in the Cobb-Douglas production function to be normally distributed across firms. Normally distributed differences in capital intensity would likely explain some fraction of the idiosyncratic variation in cost shares that remains after we separate out the cluster-specific variation in cost shares. In general, even if production technologies vary continuously, as Peter and Ruane (2018) suggest, there can still be cluster-specific differences, which our methodology would pick up. Again, the main difference is that when we perform our counterfactual experiments, allowing for idiosyncratic differences in production technologies would further reduce the potential gains from eliminating misallocation. In Section A.8, we perform a simple exercise in which we embed the concept of idiosyncratic technologies from Hsieh and Klenow (2009), Song and Wu (2015), and David and Venkateswaran (2019) in our framework and show quantitatively that our concept of cluster-specific technologies is complementary to, rather than subsumed by, their concept of idiosyncratic technologies.

Our work is also connected to previous papers in economics that use cluster analysis. An early and influential application of cluster analysis in economics was recovering descriptive groups among qualitative marketing and business data. For example, Gartner (1990) surveys researchers and business owners on their personal definition of entrepreneurship and uses cluster analysis to classify responses into two distinct views of what entrepreneurship entails. Kotey and Meredith (1997) survey small furniture manufacturers about the relationship between owners and managers and use cluster analysis to group respondents into three relationship types. Leuz, Nanda, and Wysocki (2003) conduct a descriptive cluster analysis in which they group countries using data on legal and institutional characteristics. More recent applications of cluster analysis include recovering trend patterns in quantitative data. Crone (2005) uses cluster analysis to group states into economic regions based on similarity in business cycle characteristics, Levy-Yeyati and Sturzenegger (2005) classify countries into exchange rate regimes using data on exchange rates and international reserves, and Humphries (2017) classifies individuals according to their life-cycle employment profiles. Advances have also been made on the theoretical side; for example, Bonhomme and Manresa (2015) have extended the standard K-means clustering algorithm to apply to panel data. Our paper and previous applications of cluster analysis have two major differences. First, we develop an economic model that tells us which variables should be included in the cluster analysis, how the cluster analysis should be applied, and how the output can be tied to specific structural parameters in our model. Second, we develop a new methodology for statistically determining the
number of groups in our cluster analysis, whereas previous applications have relied on heuristic methods for determining the number of groups.

2 Model

In this section, we present the theoretical framework we use to motivate our methodology for identifying cluster-specific variation in cost shares. We first lay out our framework for an industry featuring both cluster-specific and idiosyncratic variation in cost shares in Section 2.1. The cluster-specific component can be either due to differences in technologies or due to cluster-specific distortions. At this point, we do not specify which features of the economy are observable in the data. In Section 2.2, we then motivate which parameters of the economy we treat as observable in the data and describe the intuition behind our strategy for identifying parameters that are not directly observable.

2.1 An Industry with Multiple Clusters

The framework we use is a standard model of monopolistic competition with heterogeneous firms. Our methodology for identifying clusters within an industry will only require data for that industry; therefore, we focus this section on firms within a single industry. We consider a multi-industry extension of the model when we perform counterfactual exercises in Section 5.5.

The industry contains $M$ firms, where each firm produces a differentiated product according to a constant returns to scale Cobb-Douglas technology. Our framework is similar to that of Hsieh and Klenow (2009), except that we allow for an arbitrary number of inputs, $I$, as opposed to only capital and labor, and we have multiple clusters of firms within an industry, where both production technologies and distortions are allowed to vary across clusters.

Within an industry, there are $H$ total clusters, and the production function for firm $m$ utilizing the production technology for cluster $h$ is

$$y_m = z_m \prod_{i=1}^{I} \left( x_{im}^{i} \right)^{\theta_{ih}}$$

where $y_m$ is firm $m$’s output, $z_m$ is its productivity, $x_{im}^{i}$ is its quantity of input $i$, and $\theta_{ih}$ is the factor intensity for input $i$ by firms in cluster $h$.\(^6\) We follow the standard assumption that production

\(^6\)In Section A.2.1, we show that our methodology is consistent with a constant elasticity of substitution (CES) production function and is not restricted to a Cobb-Douglas production function. In particular, we show that the CES production function implies a continuum of factor intensities across firms, and, while the production function is more flexible, it cannot account for the relationships between the marginal product of inputs and plant size that we find in Sections 4 and 5. In addition, recent work in industrial organization has proposed very flexible functional forms for the production function (e.g., Gandhi, Navarro, and Rivers, 2020). These flexible functional forms, similarly to CES, could create a dispersion in cost shares even if there is only one technology present. As long as these functional forms do not generate discrete jumps in the cost shares of firms, then our methodology would likely determine that it was generated by only one cluster.
technologies exhibit constant returns to scale; therefore, \( \sum_{i=1}^{I} \theta_i^h = 1 \) for all clusters \( h \). We assume that a firm’s cluster is exogenously given rather than an endogenous choice.

The industry has a perfectly competitive bundler that combines the differentiated outputs into an industry output according to a CES aggregator,

\[
Y = \left( \sum_{m=1}^{M} (y_m)^\rho \right)^{1/\rho},
\]

where \( 1/(1 - \rho) \) is the elasticity of substitution across differentiated outputs. This implies that firm \( m \) faces demand for its output given by

\[
y_m = \frac{E}{(p_m)^{1-\rho} (P)^{\frac{-\rho}{1-\rho}}},
\]

where \( E \) is total expenditures on the industry output, which are constant, \( p_m \) is the price of firm \( m \)’s output, and \( P \) is the industry price index given by

\[
P = \left( \sum_{m=1}^{M} (p_m)^{\frac{\rho}{1-\rho}} \right)^{\frac{1-\rho}{\rho}}.
\]

Firms face distortions that lead to variation in the input prices they face.\(^7\) In particular, firm \( m \) faces an effective input cost of \( \tau_{i,h,m}^i p^i \) for input \( i \), where \( p^i \) is the price of factor input \( i \), and \( \tau_{i,h,m}^i \) is the distortion that firm \( m \) in cluster \( h \) faces for input \( i \). We assume that distortions can be separated into three multiplicative components. In particular we assume that

\[
\tau_{i,h,m}^i = \tau_h \tau_{i,m}^i \tau_{i,h}^i
\]

where \( \tau_h \) is a cluster-specific distortion that is common to all firms in cluster \( h \) and does not depend on the specific input, \( \tau_{i,h}^i \) is a cluster-input-specific distortion for input \( i \) and is common to all firms in cluster \( h \), and \( \tau_{i,m}^i \) is an idiosyncratic distortion that is specific to input \( i \) for firm \( m \).

The firm solves the following profit maximization problem, taking as given the demand for the firm’s output and effective input prices:

\[
\max \quad p_m y_m - \sum_{i=1}^{I} \tau_{i,h,m}^i p^i x_m^i.
\]

\(^7\)When working with capital and labor as the only two inputs, a common assumption is that firms face only an output distortion and a capital distortion. In Section A.2.2, we consider the implications of including output distortions. We find that output distortions would imply that the MRP of inputs would increase uniformly for all inputs. We see in Section 5, however, that this is not the case (see Figure 3). Hsieh and Olken (2014) make a similar argument when discussing that markups, which would map to an output distortion in our framework, do not appear to drive the relationship between the marginal revenue products and size of plants in India, Indonesia, and Mexico.
This problem has the solution that firms charge a constant markup, equal to \(1/\rho\), over their marginal cost,

\[ p_m = \frac{1}{\rho} \frac{1}{z_m} \prod_{i=1}^{I} \left( \frac{\tau_{h,m}^i p^i}{\theta_{h}^i} \right)^{\theta_{h}^i}, \tag{7} \]

where the marginal cost for firm \(m\) depends on the firm’s productivity, the firm’s distortions, and the firm’s production technology. This expression also highlights that the impact that distortions have on a firm’s marginal cost depends on the firm’s production technology.

The presence of these distortions affects the choices of firms when choosing inputs and leads to dispersion in the MRP of inputs across firms. In this case, the MRP of input \(i\) for firm \(m\) in cluster \(h\) is given by

\[ MRP_{h,m}^i = \tau_{h,m}^i p^i. \tag{8} \]

Thus, if a firm faces a large distortion, \(\tau_{h,m}^i\), on input \(i\), then the firm will use less of that input, and that input will have a correspondingly high MRP. When we perform counterfactual experiments in Section 4, we interpret these distortions as arising due to taxes or subsidies which vary across firms, inputs, and clusters, so as to cause firms to use inefficient amounts of each input, leading to misallocation.

Cost minimization implies that the share of a firm’s total expenditures, including distortions, spent on a given input will be equal to the firm’s factor intensity for that input:

\[ \theta_{h}^i = \frac{\tau_{h,m}^i p^i x_{m}^i}{\sum_{j=1}^{I} \tau_{h,m}^j p^j x_{m}^j}. \tag{9} \]

Finally, we can show that the following condition holds:

\[ \frac{1}{\rho} = \frac{p_m y_m}{\sum_{j=1}^{I} \tau_{h,m}^j p^j x_{m}^j}, \tag{10} \]

which relates markups to revenues and factor expenditures with distortions. This condition will be useful for identifying model parameters in the following section.

### 2.2 Connecting the Model to Data

Up to this section, we have not discussed which features of the model are typically available in the data and which are not. The motivation for this paper, however, is that many of the parameters in Section 2.1 are not available in standard data, which limits the ability of researchers to identify clusters of firms that vary systematically in their production technologies or distortions. For example, if factor intensities were directly observable, then we would instantly observe whether different firms were employing different production technologies. Likewise, if we were able to observe each firm’s input expenditures with the distortions included, then we would be able to use equation 9 to directly recover the factor intensities and therefore the production techniques.
We are interested in cases in which clusters of firms are not directly observable in the data. Furthermore, whereas some datasets break down values into prices and quantities, many datasets contain only expenditures. Therefore, we assume that we are only able to observe expenditures, excluding distortions, \( p^i x_m^i \), for each input \( i \) and revenues, \( p_m y_m \), for each firm \( m \) in the industry. Note that our observed expenditures on inputs do not include the unobserved idiosyncratic distortions or the technology-specific distortions. In addition, we assume that we are not able to observe markups or factor intensities, nor are we able to observe the production technology used by each firm.

To derive expressions that can be used to recover the unobservable parameters, we first define the cost share for input \( i \) as

\[
\tilde{\theta}_m^i \equiv \frac{p^i x_m^i}{p_m y_m}.
\] (11)

These cost shares do not include distortions and are therefore directly observable in the data. We can then combine equation 9, which relates input expenditures to distortions and factor intensities, with equation 10, which relates input expenditures to the markup over marginal cost, to derive the following expression for the cost share for firm \( m \) for input \( i \):

\[
\tilde{\theta}_m^i = \frac{\rho \theta^i_h}{\tau^i_h \tau^i_m}.
\] (12)

This expression will be the main equation we use throughout the paper to identify the unobservable parameters and to identify the production technology used by each firm. The key benefit of this equation is that the cost share for a given input does not depend on the factor intensities or distortions for other inputs. In contrast to the case in which we divide by revenue, if we were to instead use observed input expenditures as a fraction of total observed expenditures, \( p^i x_m^i / \sum_{j=1}^I p^j x_m^j \), then the cost share for each input would depend on all of the idiosyncratic distortions simultaneously.\(^8\)

3 Recovering Clusters

In this section, we lay out our methodology for identifying the clusters each firm belongs to within an industry, where clusters differ in terms of their production technologies, cluster-specific input distortions, or both. Our methodology is based on cluster analysis, which is a set of processes used to group observations so that they are similar within groups. Our application of cluster analysis will be based on the expression for cost shares derived in the previous section and on the intuition

---

\(^8\)In our framework, markups are common across all firms in an industry. To the extent that there are cluster-specific differences in markups, these differences would be absorbed into the cluster-specific distortions, \( \tau_h \). If there are idiosyncratic differences in markups, these would be absorbed into a common term that scales \( \tau^i_m \) for all of the factor inputs for a given firm (i.e., firms with high markups would have a higher \( \tau^i_m \) for all inputs).
that, in an industry with multiple clusters, we should expect gaps in the dispersion of cost shares
to arise between sets of firms belonging to different clusters.

We present our methodology in several steps. First, in a preliminary step, we transform the
cost shares to prepare for the cluster analysis. Second, we partition the set of firms in each industry
using the K-means clustering algorithm for an arbitrary number of groups. Third, after applying
the clustering algorithm to a range of possible groups, we use those results to recover the number
of clusters within each industry. Fourth, we show how to use our clustering results to identify the
underlying parameters of the model: markups, factor intensities, and distortions.

In Section 3.6, we describe simulation exercises in which we generate synthetic data to verify
that our methodology successfully recovers the number of clusters in each industry, the assignment
of firms to these clusters, and the parameters of the model.

3.1 Transforming Data

Our methodology relies on exploiting the structure of equation 12. Before applying cluster anal-
ysis, we take the log of the equation to get

\[
\log \hat{\theta}_m = \log \rho + \log \theta^h_i - \log \tau^i - \log \tau^h_i - \log \tau^i_m. \tag{13}
\]

This transformation is useful because it linearizes the relationship between cost shares and the key
unobserved parameters of our model. In addition, a large literature has shown that the K-means
clustering algorithm works particularly well for clustering Gaussian-distributed data. When we
apply our methodology to Chilean plant-level data in Section 4, we find that the idiosyncratic
distortions are well approximated by a lognormal distribution. Therefore, transforming the data
by taking logs helps to ensure the accuracy of our results.

The crux of our strategy for recovering the unobserved parameters relies on using equation 13
and exploiting variation in cost shares. Note that \( \log \rho + \log \theta^h_i - \log \tau^i - \log \tau^h_i \) is common for all
firms within a cluster.\textsuperscript{11} Thus, all variation in cost shares in cluster \( h \) arises from \( \log \tau^i_m \).

\textsuperscript{9}Note that equation 12 is similar to equation 18 in Hsieh and Klenow (2009), which they use to identify the
distortions in their model. Papers in the misallocation literature tend to rely on identifying distortions based on
variation in factor cost shares. This means that our framework is relatively standard and our methodology for
identifying clusters of firms within an industry can easily be incorporated into frameworks in the misallocation
literature with alternative specifications.

\textsuperscript{10}Jain (2010) has an overview on the use and development of the K-means algorithm and points out that the
algorithm does well in finding spherical clusters in the data, which is consistent with a Gaussian distribution.

\textsuperscript{11}Note that if firms within an industry sell products that use a production function with different factor inten-
sities, \( \theta^h_i \), or face different cluster-input-specific distortions, \( \tau^i_h \), then clusters may reflect differences in the products
that firms produce.
3.2 Conditions on the Distribution of Distortions

Equation 5 shows that the input distortion $\tau_{i,h,m}$ can be decomposed into three multiplicative components — cluster-specific distortions ($\tau_h$ and $\tau_{i,h}$) and idiosyncratic distortions ($\tau_{i,m}$). We now describe conditions that will allow us to determine the components and level of $\tau_{i,h,m}$, technology parameters, and demand parameters ($\rho$).

First, we use the following condition to pin down the level of $\tau_h$:

$$\frac{1}{H} \sum_{h=1}^{H} \tau_h^{-1} = 1. \quad (14)$$

This equation implies that the harmonic mean of $\tau_h$ is equal to 1. Second, we have the following condition on the level of the cluster-input-specific distortions, $\tau_{i,h}$:

$$\sum_{i=1}^{I} \frac{1}{\tau_{i,h}} = 1, \quad \forall h \in \{1, ..., H\}. \quad (15)$$

This equation implies that the weighted harmonic mean of cluster-input-specific distortions in cluster $h$ is equal to 1.\(^{12}\) Note that this condition uses the factor intensity of the production function, $\theta_{i,h}$, as weights when calculating the weighted harmonic average. Third, we have the following condition on the level of idiosyncratic distortions:

$$\frac{\sum_{m=1}^{M} \mathbb{I}_{k(m)=h} (\tau_{i,m})^{-1}}{M_h} = 1, \quad \forall h \in \{1, ..., H\}, \ \forall i \in \{1, ..., I\}, \quad (16)$$

where $M_h \equiv \sum_{m=1}^{M} \mathbb{I}_{k(m)=h}$ is the number of firms that belong to cluster $h$. This equation implies that the harmonic mean of the idiosyncratic distortions within each cluster is equal to 1. We will show in Section 3.5 that the use of harmonic means in equations 14-16 yields intuitive expressions for model parameters that can easily be taken to the data.

In Section A.5, we show that raising any of the harmonic means characterized in equations 14-16 would increase the input distortions $\tau_{i,h,m}$ proportionally for all firms. Thus, it will not affect the dispersion in the MRP of inputs within an industry since it affects all firms in the same multiplicative manner. It would, however, affect our welfare analysis in two ways. First, it would raise the overall level of distortions. Second, it would also affect our estimate of $\rho$.\(^{13}\) Thus, in

---

\(^{12}\)Equation 15 is consistent with constant returns to scale. For example, in the case that $\tau_{i,h} = 1$ for all $i$ and $h$, then the condition becomes $\sum_{i=1}^{I} \theta_{i,h} = 1$. Note that our framework would remain largely unchanged if we relax this assumption such that $\sum_{i=1}^{I} \theta_{i,h} = R$ where $R \neq 1$. In particular, equation 13 would remain the same and thus not affect the clustering methodology. The main changes in Section 3.1 would be that the right-hand side of equation 15 would be $R$. Section A.5.2 discusses the implications of $R \neq 1$ on the recovered parameters in Section 3.5. The main difference is that it would affect the value of $\rho$.

\(^{13}\)Because raising the harmonic mean in equations 14, 15, and 16 affects the estimates of the input distortions, $\tau_{i,h,m}$, and $\rho$ in the same manner, two of these three conditions are normalizations. See Section A.5 for full details.
Appendix A.2.3, we conduct robustness exercises to examine how our results change as we vary $\rho$ and vary the levels of total misallocation in the economy.\textsuperscript{14}

We now describe two additional assumptions that are useful for the clustering exercise. First, we assume that, within each cluster, the idiosyncratic distortions follow a multivariate lognormal distribution. This condition also implies that the cost share of firms, $\bar{\hat{\theta}}_{m}$, will be lognormally distributed within each cluster. The reason is that, as discussed in Section 3.1, all variation in cost shares in cluster $h$ arises from the idiosyncratic distortions, $\tau_{m}$. It is also useful to point out that our methodology is flexible in that it allows both for distortions to be correlated across factor inputs — capturing that firms with a high capital cost distortion may be more likely to have a high labor cost distortion — and for the distribution of distortions to differ across clusters within an industry. Additionally, the assumption of lognormality appears to be well supported by our data, which we show using kernel density plots in Section 4.\textsuperscript{15}

Second, we suppose that each firm’s cost share is closest to the average cost share for its cluster. Although there is no way to verify this assumption in the data without having ex ante knowledge about which group each firm belongs to, in Section A.3, we use simulations to explore the robustness of our methodology when this condition is not satisfied (e.g., if there is some overlap in clusters). We find that the methodology still works well, although we tend to understate the true number of clusters.

3.3 Partitioning Firms by Cluster

In order to recover the number of clusters, we follow a two-step strategy. First, in this subsection, we show how to optimally partition the set of firms in order to determine which cluster each firm is assigned to under the assumption that there are $H$ clusters in the industry. Second, in the next subsection, we show how to use the sum of squared errors (SSE) from the optimal partition for $H = 1, \ldots, H_{\text{max}}$, where $H_{\text{max}}$ is the maximum number of clusters we consider, to identify the actual number of clusters, $H$, in an industry.

Under the assumption that there are $H$ clusters in an industry, our goal is to partition the set of firms into $H$ groups where each firm is assigned to a single group, $h \in \{1, \ldots, H\}$, and each group has at least one firm assigned to it. For each grouping, our optimal partition minimizes the SSE between the logged cost shares and the mean logged cost share for each firm’s group. In particular, the optimal partition achieves

\textsuperscript{14}In our quantitative analysis, we find that the implied mean and median $\rho$ from the condition in equations 14-16 is 0.80 (an elasticity of substitution of 5), which is in line with estimates of elasticities from the existing literature and thus appears to imply reasonable elasticities.

\textsuperscript{15}When we perform the Kolmogorov-Smirnov test on the logged cost shares for each input after clustering the data, we fail to reject normality 92.4 percent of the time (out of 184 input-cluster-industry tuples) at a 5 percent significance level.
\[ SSE_H \equiv \min_{\{h(m)\}} \sum_{h=1}^{H} \sum_{m=1}^{M} \sum_{i=1}^{I} \mathbb{I}_{h(m)=h} \left( \log \tilde{\theta}_m^i - \xi_h^i \right)^2, \]  

where \( h(m) \) is the mapping that returns the group that firm \( m \) is assigned to, \( \mathbb{I}_{h(m)=h} \) is an indicator function that takes on the value of one if firm \( m \) is assigned to group \( h \) and zero otherwise, and \( \xi_h^i \) is the mean value of the logged cost shares for input \( i \) for firms assigned to group \( h \).

In general, it is impossible to find the globally optimal partition that minimizes the SSE to solve the problem in equation 17 because of the high dimensionality of the problem, which belongs to the class of NP-hard problems. Even with only two clusters and 50 firms, there are \( 2^{49} \) possible groupings, which highlights the infeasibility of the brute force method of evaluating all possible combinations to find the partition that minimizes the SSE. To tackle this problem, we apply the heuristic K-means++ clustering algorithm developed by Arthur and Vassilvitskii (2007) to find a grouping that achieves a local minimum of SSE. The K-means++ algorithm is initialized by selecting \( H \) centers randomly from the set of observations. The first center is chosen uniformly at random, and subsequent centers are chosen with probability inversely proportional to their Euclidean distance from the closest previously chosen center (in the standard K-means algorithm, all \( H \) centers are chosen uniformly at random). This initialization strategy ensures that the starting points tend to be far away from each other and improves the speed and accuracy of the algorithm compared to selecting all centers with uniform probability. Following initialization, the algorithm iterates between the following two steps until convergence:

- **Assignment step:** Assign each observation to the closest center:

\[ \hat{h}(m) = \arg \min_{h \in \{1,...,H\}} \sum_{i=1}^{I} \left( \log \tilde{\theta}_m^i - \xi_h^i \right)^2. \]  

- **Recentering step:** After all observations have been assigned, recompute centers as the component-wise mean values for observations in each group:

\[ \xi_h^i = \frac{1}{\sum_{m=1}^{M} \mathbb{I}_{h(m)=h}} \sum_{m=1}^{M} \mathbb{I}_{h(m)=h} \log \tilde{\theta}_m^i. \]  

After convergence is achieved, meaning that the centers remain the same and no observations switch group assignment, the algorithm will yield a partition where each firm is assigned to one of \( H \) groups. If \( H \) is equal to the number of clusters in the industry — determining that this is the focus of the next step of our methodology — and the partition is the globally optimal one, then this will correctly separate firms by their clusters. To avoid selecting a locally optimal partition far from the globally optimal partition, we reinitialize the algorithm 1,000 times\(^{16}\) with different randomly drawn starting points and then select the partition with the lowest SSE.

\(^{16}\)Only one industry had a firm switch clusters compared to when we used 10 initializations, and increasing the number of initializations to 10,000 times yielded no changes in our results.
3.4 Determining the Number of Clusters

Our goal is to recover the total number of clusters, \( H \), within the industry. After determining the number of clusters, we can then partition the set of firms following Section 3.3 to recover the cluster each individual firm belongs to. Our strategy for recovering \( H \) will again be based on the intuition that there should be gaps in the distributions of cost shares between firms with different clusters, which result from different production technologies, different cluster-specific distortions, or both. Effectively, our clustering algorithm in Section 3.3 ensures that firms are similar within groups, while our methodology in this section ensures that clusters are sufficiently distinct so as to avoid incorrectly partitioning a single cluster into multiple clusters. This indicates that our methodology will be conservative in the number of clusters recovered.

To determine the number of clusters in an industry, we cluster our data using the algorithm in Section 3.3 with \( H = 1, \ldots, H_{\text{max}} \) clusters, where again \( H_{\text{max}} \) is the maximum number of clusters we consider. Our strategy relies on computing the SSE of the optimal partition, \( \text{SSE}_H \), for each number of clusters, \( H \), and then comparing the reduction in SSE of going from \( H - 1 \) to \( H \) clusters.

We want to be conservative in our estimate of the number of clusters in order to avoid claiming the existence of multiple clusters when there is only a single one. Therefore, we evaluate the presence of additional clusters in a sequential matter. This means we first test whether there is evidence of two clusters in the industry versus the null hypothesis that there is only one. Increasing the number of clusters in the clustering algorithm always leads to a decrease in the SSE (up to the number of observations, at which point the SSE = 0). Therefore, we need to test whether the reduction in the SSE from clustering the data with two clusters instead of one is greater than the reduction we would expect if we knew there was only a single cluster, that is, if all firms use the same production technology and there are no cluster-specific distortions. Conditional on finding a sufficient reduction in the SSE to reject the null hypothesis that there is only a single cluster, we continue the process. Specifically, we examine whether the reduction in the SSE from clustering the data with three groups instead of two is greater than the reduction we would expect if there were only two clusters, and so on.

In order to compute the expected reduction in the SSE conditional on having a given number of clusters, we require an assumption on the data-generating process.\(^{17}\) In particular, we assume that, within each cluster, the distortions on factor inputs, \( \tau_{im}^i \), follow a multivariate lognormal distribution as mentioned in Section 3.2. We compute the expected reduction in the SSE from clustering with two groups conditional on having only one cluster in three steps. First, we fit a single multivariate lognormal distribution to the data. Second, we generate 1,000 synthetic data sets.

\(^{17}\)The widely used C-H index from Calinski and Harabasz (1974) is a widely used heuristic measure used to determine the number of clusters. A drawback of the C-H index is that it cannot be used to evaluate whether there is only a single cluster. When applied to industries we identify as having more than one technology, however, the C-H index yields results identical to those from our methodology when applied to Chilean plant-level data in Section 4.
datasets — the cutoffs are similar with 500 and 2,000 synthetic datasets — with the same number of observations as the original dataset, where each observation is drawn independently from the fitted multivariate lognormal distribution. Third, we apply our clustering algorithm from Section 3.3 individually to each synthetic dataset and calculate

$$R_{2\mid 1}^{\text{syn}} = \frac{\text{SSE}_{2\mid \text{syn}}}{\text{SSE}_{1\mid \text{syn}}},$$

(20)

where $\text{SSE}_{2\mid \text{syn}}$ and $\text{SSE}_{1\mid \text{syn}}$ are calculated using equation 17. The ratio $R_{2\mid 1}^{\text{syn}}$ is the ratio of the SSE from running our clustering algorithm with two groups versus one group for the synthetic dataset, where, by construction, the synthetic dataset has only one true cluster. A lower $R_{2\mid 1}^{\text{syn}}$ therefore indicates a greater reduction in the SSE from clustering with two groups versus one group.

After computing $R_{2\mid 1}^{\text{syn}}$ for each synthetic dataset, we define $\hat{R}_{2\mid 1}^{\text{syn}}$ to be the 20th percentile of $R_{2\mid 1}^{\text{syn}}$ values, so that 80 percent of the synthetic datasets have a reduction in the SSE that is smaller than $R_{2\mid 1}^{\text{syn}}$. We chose the 20th percentile as it leads to robust results when applied to our data, which we discuss in more detail in Section A.10.1. We use $\hat{R}_{2\mid 1}^{\text{syn}}$ as our cutoff test statistic for determining whether the actual data contain a single cluster or multiple clusters. In particular, we compute $R_{2\mid 1} = \text{SSE}_{2}/\text{SSE}_{1}$ for our original data and then run the following comparison:

$$\text{Conclude two or more clusters } \iff \ R_{2\mid 1} < \hat{R}_{2\mid 1}^{\text{syn}}.$$  (21)

This is similar to performing a hypothesis test where the null hypothesis is that there is a single cluster, and we reject the null hypothesis if $R_{2\mid 1} < \hat{R}_{2\mid 1}^{\text{syn}}$. This test is based on the intuition that we are unlikely to get a greater reduction in the SSE in the data without having detectable gaps in the distribution of cost shares.

Conditional on concluding there are two or more clusters in the industry, we repeat our exercise to evaluate whether there is evidence for three or more clusters, and we continue this process sequentially. We describe this process in more detail in Section A.10.2. We set $R_{\mathcal{H}\mid \mathcal{H}-1}^{\text{syn}}$ as the 20th percentile of $R_{\mathcal{H}\mid \mathcal{H}-1}^{\text{syn}}$ datasets, and we conclude there is evidence for $\mathcal{H}$ or more clusters if the reduction in the SSE in the data is greater than this cutoff. We do this using a condition analogous to equation 21, and we continue the above process sequentially for $\mathcal{H} = 1, \ldots, \mathcal{H}_{\max}$ until we fail to conclude that there are $\mathcal{H}$ or more clusters. For example, we would conclude that there are three clusters if we find evidence both in favor of there being three or more clusters and against there being four or more clusters.

Formally, we conclude that there are $\mathcal{H} \in \{1, \ldots, \mathcal{H}_{\max} - 1\}$ clusters if $\mathcal{H}$ is the minimum value that satisfies the following criteria:

$$R_{\mathcal{H}+1\mid \mathcal{H}} \geq R_{\mathcal{H}+1\mid \mathcal{H}}^{\text{syn}},$$

(22)

If there are no $\mathcal{H} \in \{1, \ldots, \mathcal{H}_{\max} - 1\}$ that satisfy the condition in equation 22 (i.e., there is evidence that there are at least $\mathcal{H}_{\max}$ clusters), then we set $\mathcal{H} = \mathcal{H}_{\max}$. 

14
3.5 Identifying Cluster-Specific Variation in Cost Shares

After identifying the total number of clusters following Section 3.4, we are able to use our clustering algorithm from Section 3.3 to determine the specific cluster each firm belongs to. We now assume that we have correctly determined the number of clusters and assigned firms to the correct cluster, topics which we will further discuss in Section 3.6, and use this information to identify the remaining unobservable parameters in our framework: markups, and distortions. We discuss our methodology for recovering each of the remaining parameters in the following subsections.

3.5.1 Markups

Markups are determined by the elasticity of substitution in equation 10, and all firms in an industry charge the same constant markup over marginal cost. To derive an expression for markups, we sum equation 12 for a given input \( i \) across all firms belonging to cluster \( h \):

\[
\sum_{m=1}^{M} \tilde{\theta}_m^i = \rho \left( \tau_h \tilde{\theta}_h^i \right) - 1 \theta_h^i. \tag{23}
\]

In the above equation, we removed idiosyncratic distortions by plugging in our assumption from equation 16. We next sum equation 23 across all inputs for firms that belong to cluster \( h \) to get

\[
\sum_{i=1}^{I} \sum_{m=1}^{M} \tilde{\theta}_m^i = \rho \left( \tau_h \right) - 1 \sum_{i=1}^{I} \tau_h^i \theta^i_h. \tag{24}
\]

We can plug equation 15 into the above equation and then sum over all clusters to get

\[
\rho \sum_{h=1}^{H} \left( \tau_h \right) - 1 = \sum_{h=1}^{H} \sum_{i=1}^{I} \sum_{m=1}^{M} \tilde{\theta}_m^i / M_h. \tag{25}
\]

The above steps reveal our need for an assumption on the average cluster-specific distortion in order to compute markups. In particular, we use the condition from equation 14 to arrive at

\[
\rho = \frac{1}{H} \sum_{h=1}^{H} \sum_{i=1}^{I} \sum_{m=1}^{M} \tilde{\theta}_m^i / M_h, \tag{26}
\]

which allows us to determine the markup for firms in an industry.\(^{18}\) This is because \( \tilde{\theta}_m^i \) is directly observable, and we know \( \hat{h}(m) \) and \( M_h \) from our clustering methodology in Sections 3.3 and 3.4. For an example of how to interpret the above equation, suppose that observed expenditures on inputs are a small fraction of revenue on average within an industry. Equation 26 implies that \( \rho \) will be low, which is consistent with high markups in the industry.

\(^{18}\)In Section A.2.3, we show that our results are robust to alternative estimates for the elasticity of substitution.
3.5.2 Cluster-Specific Variation: Cluster-Specific Distortions and Factor Intensities

Without additional data, we cannot fully disentangle factor intensities from cluster-specific distortions. Nevertheless, we can still decompose variation in cost shares across firms into a cluster-specific component and an idiosyncratic component in order to evaluate their relative importance. To recover the cluster-specific common distortion, \( \tau_h \), we exploit that it does not vary across factor inputs or across firms that share a cluster. Therefore, we can rearrange equation 24, where we have averaged across firms within a given cluster and across inputs, to derive the following relationship:

\[
\tau_h = \rho \left( \sum_{i=1}^{I} \sum_{m=1}^{M} \frac{I_{(m)=h} \tilde{\theta}_{im}}{M_h} \right)^{-1}.
\]  

(27)

The intuition of the expression is that if observed expenditures are a smaller fraction of revenues for firms in one cluster, then \( \tau_h \) will be high, which means that firms in that cluster face higher input costs.

As mentioned previously, we cannot separate \( \theta_{ih} \) from \( \tau_{ih} \) without additional data; however, we can still characterize them jointly. To characterize the cluster-input-specific distortions and factor intensities, we average cost shares across firms within each cluster and then adjust for the parameters that affect the average cost share. We rearrange equation 23 to get the following expression:

\[
\frac{1}{\tau_{ih}} \theta_{ih} = \frac{\tau_h \sum_{m=1}^{M} I_{(m)=h} \tilde{\theta}_{im}}{\rho M_h}.
\]  

(28)

Given estimates of \( \tau_h \) and \( \rho \), we can then calculate the combined \( \frac{1}{\tau_{ih}} \theta_{ih} \) for input \( i \) and cluster \( h \). The intuition behind this expression is that a higher cluster-specific common distortion, \( \tau_h \), will deflate the mean cost share within a cluster, whereas because our strategy relies on using cost shares (factor expenditures divided by total revenue) rather than simply factor expenditures divided by total expenditures, higher markups deflate the ratio of expenditures to total revenues on any given factor. If the right-hand side of equation 28 is high, then the model rationalizes it with a high \( \theta_{ih} \), which is technological, or a low \( \tau_{ih} \), which is related to policy.

3.5.3 Idiosyncratic Distortions

The final unobserved variables we have to recover are the idiosyncratic firm-level distortions, \( \tau_{im} \), for each firm \( m \) and input \( i \). We find these firm-level distortions by rearranging equation 12 as follows:

\[
\tau_{im} = \frac{\rho \frac{1}{\tau_{ih}} \theta_{im}}{\tau_{ih} \tilde{\theta}_{im}}.
\]  

(29)

This equation sets \( \tau_{im} \) so that the model matches the cost shares in the data. Based on equation 13, we know that variation in cost shares across firms within a cluster is entirely a result of the idiosyncratic distortions (\( \tau_{im} \)) of those firms. Thus, the correlation in input distortions (\( \tau_{ih,m} \)) across firms within a cluster is the same as the correlation in cost shares across those firms in the data.
3.6 Applying Clustering Methodology to Synthetic Data

In Section A.3, we use simulated data to show that we can use the described methodology to recover the number of clusters in each industry and the assignment of firms to these clusters using simulations. We find that the clustering methodology works well in recovering the correct number of clusters, assigning firms to these clusters, and estimating the remaining model parameters. We also explore the robustness of the methodology under alternative scenarios such as correlated idiosyncratic distortions and overlapping clusters. The methodology correctly identifies the number of clusters when idiosyncratic distortions are correlated. In the case of overlapping clusters, we find that the methodology generally works well but tends to understate the true number of clusters.

4 Application to Chilean Plant-Level Data

To examine the importance of cluster-specific variation versus idiosyncratic variation in cost shares, we apply our methodology from Section 3 to Chilean plant-level data.

4.1 Data Description and Preparation

We use the 2005 vintage of the Encuesta Nacional de Industria Anual (ENIA) dataset collected by the Instituto Nacional de Estadisticas (INE), which covers all active manufacturing plants in Chile with more than 10 employees. The ENIA plant-level data have been thoroughly examined by previous studies, including studies on misallocation and on estimating production functions.\textsuperscript{19} Therefore, the reliability of the Chilean data is well understood and established from these previous studies.

The ENIA dataset classifies each plant according to its four-digit ISIC Revision 3 industry code, which is what we use as our definition for an industry. Along with each plant’s industry, we require information on revenues as well as expenditures on capital, unskilled labor, skilled labor, and intermediate inputs.\textsuperscript{20} Previous studies on misallocation have primarily used only capital and

\textsuperscript{19}Papers that have used Chilean plant-level data include Alvarez and Lopez (2005), Asturias, Hur, Kehoe, and Ruhl (Forthcoming), Bergoeing and Repetto (2006), Levinsohn (1999), Levinsohn and Petrin (2003), Oberfield (2013), Pavcnik (2002), and Petrin and Levinsohn (2012).

\textsuperscript{20}We follow Liu and Tybout (1996) and Tybout (1996) in the preparation of the data. Gross output is the sum of total income (sales of goods produced; goods shipped to other establishments; resales of products; and work, repairs, and installations for third parties), electricity sold, buildings produced for own use, machinery produced for own use, vehicles produced for own use, goods produced that go to inventory (final inventory of goods in process plus final inventory of goods produced minus initial inventory of goods in process minus initial inventory of goods produced). For intermediate inputs, we include the purchases of intermediates (materials, fuels, goods purchased for resale, cost of work done by third parties, water, greases, and oil), electricity, and the materials used from inventories (initial inventories minus final inventories). For the labor input, we include white-collar production workers, executives, administrative workers, and salespeople as white-collar workers. We include blue collar production and non-production workers as blue-collar workers.
labor; however, our inclusion of intermediate inputs and unskilled labor separately from skilled labor is useful for distinguishing when firms belong to different clusters by giving us more margins along which to apply our clustering algorithm. Future studies may benefit from breaking down intermediate input usage even further — for example, by separating electricity usage, fuel usage, and water usage — to help identify the various clusters within an industry. We refrain from including these inputs separately, however, because of the prevalence of missing observations for these inputs in many industries.\footnote{Care should be taken to avoid including inputs arbitrarily. For instance, if we include inputs that would not be expected to display constant factor cost shares, this would violate our assumption of Cobb-Douglas production functions and may lead us to incorrectly estimate the number of clusters.}

For revenues we use the nominal gross output reported by plants. For expenditures on skilled and unskilled labor, we use nominal labor remuneration for each type of labor. For intermediate inputs, we use the nominal value reported by plants. Unlike expenditures on intermediate inputs and labor, we do not have a direct measure of the user cost of capital. We utilize a strategy similar to that in Young (1995) to recover the user cost of capital by using the following no-arbitrage condition for each type of capital that we consider:

\[
R_{tj} = 1 + r_t - (1 - \delta_j) \frac{P_t}{P_{t+1}} \frac{P^K_{t+1}}{P^K_t},
\]

where \(R_{tj}\) is the user cost of capital type \(j\), \(\delta_j\) is the depreciation rate of capital type \(j\), \(P_t\) is the price level of the aggregate economy, \(P^K_t\) is the price of a unit of capital, and \(r_t\) is the real interest rate. We use the economy-wide real interest rate for \(r_t\), the GDP deflator to determine \(P_t/P_{t+1}\), and the investment deflator to determine \(P^K_{t+1}/P^K_t\), all of which we obtain from the World Development Indicators database. We assume depreciation rates of 20 percent for vehicles, 10 percent for machinery, and 5 percent for buildings, which is consistent with that in Levinsohn and Petrin (2003). We multiply the user cost of capital for each type of capital, found using equation 30, by the nominal value of that capital. To determine the nominal value of each type of capital, we use the plant’s reported book value of capital for vehicles, machinery, and buildings.

\subsection*{4.2 Clustering Results}

We use the methodology described in Section 3 on the Chilean plant-level data described in Section 4.1. For the analysis, we apply our methodology separately for each industry. We require a sufficient number of plants per cluster to be able to detect clusters. Therefore, we exclude industries with fewer than 50 plants, and we cap the potential number of clusters per industry by setting \(H_{\text{max}} = 4\). After we exclude industries with insufficient observations, we are left with 23 industries containing a total of 2,754 plants. For each industry, our methodology yields the number of clusters, which cluster each plant belongs to, the industry markup over marginal cost, factor intensities for each cluster’s production technology, and idiosyncratic distortions for each plant. In order to place
our results in context, we also conduct an analysis in which we back out each of these objects under the assumption that there is a single cluster in each industry. We do this by repeating the steps from Section 3.5 under the assumption that $H = 1$ and $\hat{h}(m) = 1$ for each plant $m$.

Table 1 reports the industries included in our analysis, the number of plants operating in each industry, and the number of clusters identified by our methodology. We find multiple clusters in a majority of the industries in our sample: 14 industries have multiple clusters, compared to 9 industries for which we cannot identify more than a single cluster. There is no relationship between the number of plants in an industry and the number of clusters we recover.

To understand the importance of cluster-specific factors ($\theta_i^h$, $\tau_h$, and $\tau_i^h$) in explaining the variation in cost shares, we decompose the variance of observed cost shares into cluster-specific and idiosyncratic factors. Taking the variance of equation 13 and rearranging yields our decomposition

\[
\frac{\text{Var}(\log \theta_i^h - \log \tau_h - \log \tau_i^h)}{\text{Var}(\log \theta_i^h)} + \frac{\text{Var}(\log \tau_h)}{\text{Var}(\log \theta_i^h)} + 2\frac{\text{Cov}(\log \theta_i^h - \log \tau_h - \log \tau_i^h, \log \tau_i^h)}{\text{Var}(\log \theta_i^h)} = 1, \tag{31}
\]

where the first term captures the contribution of cluster-specific factors, the second term captures the contribution of idiosyncratic factors, and the third term captures the importance of correlations between cluster-specific and idiosyncratic factors. The first term is related to the between-cluster variation in cost shares, whereas the second term is related to the within-cluster variation. Note that this decomposition applies to each input. From equation 28, we can see that this decomposition does not depend on whether clusters arise from differences in production technologies or from cluster-specific distortions.

Table 2 reports the results of our variance decomposition by the number of clusters in the industry, where each share is reported as a percentage. The results are averaged across industries and across each of the four inputs. We find that cluster-specific variation can account for significant portions of the overall variation: cluster-specific factors account for an average of 23.0 percent for industries with two clusters and increase to an average of 46.9 percent for industries with four clusters. We also find that the covariance share remains small regardless of the number of clusters, indicating that the cluster-specific factors and idiosyncratic factors are largely independent. Next, we aim to study whether the fraction of variation explained by the cluster-specific component varies across inputs. Table 3 reports the decomposition separately for each input, averaging across all industries with more than one cluster. We find that there is wide variation across inputs. Cluster-specific factors appear relatively unimportant for intermediate inputs. On the other hand, cluster-specific factors explain a significant share of the variation in cost shares of capital, accounting for 54.7 percent of the overall variation on average. The fraction of variation explained by the cluster-specific components of skilled and unskilled labor is 26.8 and 37.5 percent, respectively.

To better understand our result, consider the industry of manufacturing of non-ferrous and

\[\text{22} \]Our estimate of $\rho$ is not necessary in calculating this variance decomposition, as can be seen in equation 13.
precious metals (ISIC 2720), which has multiple clusters. In Chile this industry is primarily composed of plants involved in copper processing (henceforth, we will refer to ISIC 2720 as the copper manufacturing industry). It is the largest industry by gross output in our dataset, and exports of copper products accounted for over 50 percent of Chile’s exports in 2005 according to data from the United Nation’s Comtrade database. Our methodology finds that the copper manufacturing industry has two clusters. We find that 19 plants operate in cluster 1 and 43 plants operate in cluster 2, reflecting that a large number of plants are operating in each cluster. We find that there are significant differences in plant size across clusters. Figure 1 shows the kernel density plots of gross output by cluster. We see that cluster 1 has significantly larger plants. In fact, the median plant assigned to cluster 1 has a gross output that is 14.9 times that of the median plant in cluster 2.

**Figure 1**

**Kernel Density Plot of Gross Output**

**Copper Manufacturing Industry**

![Kernel Density Plot](image)

Figure 1 shows the kernel density plot of the gross output of plants in the copper manufacturing industry by the recovered technology of plants.

We now investigate the relationship between plant size and the cluster-input-specific term $\frac{\theta^i_h}{\tau^i_h}$ that we recovered in equation 28, which will allow us to better understand the characteristics of clusters as they relate to plant size. To do so, we regress the log of $\frac{\theta^i_h}{\tau^i_h}$ on the log of gross output of plants, where $h$ is the cluster to which the plant was assigned. We include industry fixed effects in order to compare large and small plants within the same industry. We estimate this regression separately for each of the four inputs and report the results of these regressions in Table 4. We find that larger plants tend to have higher $\frac{\theta^i_h}{\tau^i_h}$ for intermediate inputs. These larger plants also tend to have lower $\frac{\theta^i_h}{\tau^i_h}$ for skilled and unskilled labor. Conversely, we do
not find a significant difference in $\theta_{ih}/\tau_{ih}$ for capital across large and small plants. These findings imply that larger plants are assigned to clusters with a higher capital expenditure relative to their labor expenditure. Furthermore, larger plants are assigned to clusters with higher skilled labor expenditures relative to unskilled labor expenditures.

It is useful to compare our results with empirical relationships found in data from U.S. manufacturing. The reason is that the U.S. economy is typically considered to be an undistorted economy in the literature. Thus, if we observe similar patterns in the United States and Chile, then it would suggest that the differences in clusters are correlated with technology ($\theta_{ih}$) and not cluster-input-specific distortions ($\tau_{ih}$). Table 4 shows that in Chile: 1) larger plants use capital more intensively relative to labor, and 2) larger plants use skilled labor more intensively relative to unskilled labor. If these relationships were driven solely by systematic distortions, then we would not expect those relationships to hold in the U.S. data. However, Raval (2019) finds a positive relationship between the size of U.S. manufacturing plants (measured in value-added terms) and the plant’s capital cost share (also measured in value-added terms) within industries. This finding is consistent with the first fact mentioned before in Table 4. Similarly, Holmes and Mitchell (2008) and Oi and Idson (1999) have emphasized that larger U.S. firms are both more capital intensive and more skilled worker intensive. Furthermore, many of the stylized facts that have emerged in the international trade literature are also consistent with these previous facts. For example, Bernard, Eaton, Jensen, and Kortum (2003) and Bernard, Jensen, Redding, and Schott (2007) have found that manufacturing exporters in the United States tend to be larger and at the same time use both capital and skilled workers more intensively.

5 Implications of Clustering on Gains from Eliminating Misallocation

The gains from eliminating misallocation vary depending on the extent to which clusters are the result of different production technologies ($\theta_{ih}$) or cluster-input-specific distortions ($\tau_{ih}$). For example, suppose that we interpret clusters as being primarily driven by differing technologies. Then the model would interpret some of the variation in cost shares as arising from these different technologies and thus reducing the gains from eliminating misallocation. Conversely, if these differences in cost shares are driven primarily by cluster-input-specific distortions ($\tau_{ih}$), then the variation in cost shares would still be interpreted by the model as being from distortions.

In this section, we use our framework to estimate the gains from eliminating misallocation in each case and thus providing bounds on the gains from eliminating misallocation. However, as discussed above, our methodology does not separately identify differences in factor intensities ($\theta_{ih}$) from cluster-input-specific distortions ($\tau_{ih}$). Therefore, it would be useful to discuss, in the first instance, whether clusters identified by our methodology are correlated with production technologies. In Section 5.1, we conduct a case study of the Indian steel manufacturing industry, which has
available data for identifying different production technologies used within the industry.\textsuperscript{23} We find that our methodology recovers clusters that correlate with those different production technologies, therefore indicating that clusters identified by our methodology are correlated with production technologies.\textsuperscript{24} Sections 5.2 to 5.5 describe the exercise in which we place bounds on the gains from eliminating misallocation. In Section 5.6, we conduct another exercise in which we allow for both cluster-specific technologies and cluster-input-specific distortions and show how to decompose them with an additional assumption on the nature of the distortions.

### 5.1 Are Clusters Correlated with Technologies or Distortions? A Case Study of Indian Steel Plants

As previously discussed, we cannot determine whether differences in clusters are driven by technology, $\theta^i_h$, or cluster-input-specific distortions, $\tau^i_h$. Thus, a useful exercise is to study an industry that has available data for determining the type of production technology used by each plant. With this information, we can examine whether the cluster assignments of plants are consistent with the technology that they use (e.g., are plants that use a production technology that is more capital intensive assigned to the cluster that uses capital more intensively?). In Section A.4, we describe an exercise in which we study Indian steel producers using data from the Annual Survey of Industries (ASI) for the fiscal year 2005-2006. The data from India have the advantage of providing detailed product-level information on inputs and outputs, which is useful for identifying the technologies of plants. We first apply the clustering methodology using data from the industry of the manufacture of basic iron and steel (NIC 2710). We find that the industry has two clusters, and one of these two clusters has a higher capital cost share relative to the other cluster.

As a second step, we apply a methodology similar to Collard-Wexler and De Loecker (2015) in order to determine the technology used by each plant. Like the U.S. steel plants that are studied in Collard-Wexler and De Loecker (2015), Indian steel plants fall into two general categories: vertically integrated plants and those that use electric furnaces. After applying the methodology used by Collard-Wexler and De Loecker (2015), we find that vertically integrated plants tend to be larger and more capital intensive than plants that use electric furnaces, which is consistent

\textsuperscript{23}Section A.9 describes in more detail the copper processing industry in Chile. We find that copper manufacturers use two main technologies used by to produce copper cathodes: the traditional method and the newer SX/EW method. Unfortunately, without additional information on intermediate input usage or product-level outputs, we are unable to conclusively map our recovered technologies to the technologies present in the copper industry.

\textsuperscript{24}Many potential explanations exist for the relationship between the size of plants and the clusters in which they are assigned. In Section A.7, we study whether unobserved labor costs, the presence of SEOs, overhead labor costs, and labor-contracting frictions can account for the result that larger plants are assigned to clusters with lower labor cost shares. We find that they tend to either be inconsistent with our data or would not lead to multiple clusters when our clustering methodology is applied. We also document size-dependent policies in Chile that are related to labor and capital (e.g., capital taxation based on size of plants). We similarly find that these policies cannot account for the documented patterns.
with the structure of the steel industry in India as well as in the United States.\textsuperscript{25} This finding is consistent with the results from our clustering algorithm, in which 95 percent of the vertically integrated plants are assigned to the cluster that has a higher capital cost share. Furthermore, 69 percent of the plants using electric furnaces are assigned to the cluster with the lower capital cost share. Thus, the cluster assignments of these steel plants tend to be related to the technologies that they utilize.

### 5.2 Conditions on Distortions

We now describe additional conditions on the distortions needed to study the two cases that we consider; that is, differences in clusters are a result of technologies ($\theta^i_h$) or cluster-input-specific distortions ($\tau^i_h$).

In the first case, we have that $\tau^i_h = 1$ for all $i$ and $h$, meaning that cluster-input-specific distortions play no role in the determination of clusters. We can use this condition along with equation 28 to determine $\theta^i_h$ for each cluster $h$. In the second case, we use the following condition to pin down the level of cluster-input-specific distortions:

$$\frac{1}{H} \sum_{h=1}^{H} (\tau^i_h)^{-1} = 1, \forall i \in \{1, ..., I\}.$$  \hspace{1cm} (32)

This equation implies that the harmonic mean of cluster-input-specific distortions for a given input across clusters is equal to 1. In Section A.5, we show that raising the harmonic mean in equation 32 would increase the input distortions $\tau^i_{h,m}$ proportionally for all firms. Thus, it would affect the overall level of distortions; however, it will not affect the dispersion in the MRP of inputs within an industry since it affects all firms in the same multiplicative manner.

We can sum equation 28 over all clusters and then substitute equation 32. This allows us to identify the factor intensities as follows:

$$\theta^i = \frac{1}{H} \sum_{h=1}^{H} \frac{\tau^i_h \sum_{m=1}^{M} \mathbb{I}_{k(m)=h} \theta^i_m}{M_h}. \hspace{1cm} (33)$$

We can substitute $\theta^i$ into 28 to recover the cluster-input-specific distortions, $\tau^i_h$, for each cluster.

\textsuperscript{25}For example, Collard-Wexler and De Loecker (2015) find that plants that use electric arc furnaces (minimills) are less capital intensive relative to vertically integrated plants (see pages 137 and 148). These findings are consistent with other studies of the steel industry in industrialized economies, including Stubbles (2009) (page 137) and Fenton (2005) (page 9), and with statements made by industry participants, such as those in Genet (2012) and Genet (2013). In the Indian context, steel producers use two types of electric furnaces: electric arc furnaces and induction furnaces. Dasgupta (2017) describes capital requirements for each type of plant in India and finds that induction furnaces have significantly lower capital requirements than electric arc furnaces.
5.3 Recovered Factor Intensities

We apply the conditions in Section 5.2 in order to recover the factor intensities under the two cases described. We additionally consider the standard case in the literature in which each industry has one cluster and one technology. As an example of our results, Table 5 describes the recovered factor intensities for the copper manufacturing industry under these three scenarios. The first row reports the factor intensities supposing that there is only one cluster. Rows 2 and 3 report the factor intensities for the two recovered technologies if we allow for cluster-specific technologies. We find that the implied factor intensities are effectively a weighted average of the factor intensities when we assume that there is only one cluster. For example, we find that technology 1 has a factor intensity of 0.87 for intermediate inputs and technology 2 has a factor intensity of 0.78. If we only allow for one cluster, then the inferred factor intensity is 0.80. Row 4 reports the factor intensities when we allow for cluster-input-specific distortions but not cluster-specific technologies. Note that assuming a single production technology with no cluster-specific distortions does not yield the same factor intensities as when we allow for cluster-input-specific distortions. This is because the number of firms in each cluster may differ, which leads to bias in the recovered factor intensities when cluster-input-specific distortions are ignored.

5.4 Dispersion in MRPs

We now examine the implications for the dispersion in the MRP of each input. In our framework, dispersion in the MRPs necessarily implies that there is misallocation, since it would be efficient to reallocate inputs from plants with low to high MRP until MRPs are equalized across firms. As mentioned in Section 2.1, the MRP of input $i$ for firm $m$ is

$$MRP_i = q_h^i q_m^i p^i.$$ \hspace{1cm} (34)

Following the misallocation literature, we focus on the dispersion of logged MRPs. We can ignore input prices in our analysis without affecting our chosen measure of dispersion, the standard deviation of logged MRPs, because $p^i$ is constant for all firms within an industry.

Table 6 lists the percentage reductions in the standard deviations of logged MRPs for each of the four inputs under the interpretation of clusters being due to differences in production technologies.\textsuperscript{26} We report the reduction for each of the industries we identified as having more than one production technology in Table 1. We also report the aggregate reduction in the dispersion of MRPs for each input, which is the reduction in the standard deviation when we jointly consider all plants in the economy across all industries regardless of the number of production technologies.

\textsuperscript{26}If instead we interpret clusters as arising due to differences in cluster-specific distortions, we would have only minor differences in MRPs compared with the standard framework. As we show in Section 5.5, however, our clustering methodology is still important in this case since it allows us to decompose misallocation caused by variation in MRPs into cluster-specific and idiosyncratic components.
The aggregate measures require assuming that input prices are constant across industries. We find significant variation in the magnitude of the reduction across industries and inputs. Overall, we find the largest reductions for capital, with an aggregate decline of 15.0 percent. We also see reductions in the dispersion of the MRP for unskilled labor of 12.1 percent and for skilled labor of 6.8 percent. There are minimal reductions in dispersion for intermediate input usage, in which we saw an aggregate decline of only 1.3 percent and small increases in some industries.

**Figure 2**

**Kernel Density Plots of Marginal Revenue Product of Inputs One vs. Multiple Technologies**

Panel A of Figure 2 shows the kernel density plot of the MRP of capital under one technology and multiple technologies; Panel B shows the same plot for intermediate inputs; Panel C shows the same plot for skilled labor; and Panel D shows the same plot for unskilled labor. We use a Epanechnikov kernel function for the kernel density plot.

Figure 2 displays the kernel density plots of the MRPs for each input for all plants in our sample. The kernel density plots serve to flexibly characterize the distributions of the MRPs by non-parametrically estimating the probability density functions for each of the four inputs. From these figures we can observe that, if we interpret clusters as being technologies, the dispersion in the MRPs of capital exhibits the largest decline, which is consistent with the results of Table 6. In all three cases, there is relatively little dispersion in the MRP of intermediate inputs under both specifications. This finding is consistent with the findings of Petrin and Sivadasan (2013), who find
low levels of misallocation in intermediate inputs among Chilean plants using ENIA plant-level data covering the period 1982-1994.

Restuccia and Rogerson (2008) point out that the correlations between idiosyncratic distortions and the size of firms can play an important role in determining the impact these distortions have on overall misallocation. In particular, if the MRP of an input is increasing with firm size, this indicates large potential gains from eliminating distortions. Eliminating distortions would then lead to the reallocation of the input from smaller firms to large firms, which tend to be more productive. In our framework, this same insight applies. Thus, it is informative to investigate how the relationship between the MRP of inputs and size changes after we account for cluster-specific factors. In Section A.1, we show that similar patterns hold when we use productivity directly instead of gross output.

In Figure 3 we plot the relationship between the MRPs of each input and the gross output of plants. These figures are constructed by fitting the data with a lowess locally weighted regression, which allows us to non-parametrically and flexibly characterize the relationship between MRPs and size. To get a sense of the scale of plants, 100 million pesos is equivalent to approximately 180,000 U.S. dollars using the average 2005 exchange rate provided by the World Development Indicators database.

We find a strong, positive relationship between size and the MRPs of both skilled and unskilled labor. This finding suggests that large plants use too little skilled and unskilled labor, and therefore it would be efficient to reallocate both types of labor from small to large plants. The correlation between size and MRP is positive under all three specifications; however, the magnitude of the relationship decreases significantly if we allow for multiple clusters and interpret clusters as arising from different production technologies. For example, under the interpretation of multiple production technologies, the predicted MRP of skilled labor for plants in the 95th percentile of size is 2.44 times that of the plant that is in the 5th percentile of size. Using a single production technology overstates this ratio by over 30 percent, yielding a ratio of 3.20. For unskilled labor, the overstatement is significantly larger. The ratio is 4.50 with multiple production technologies and 8.24 for a single production technology, meaning that the ratio is overstated by over 80 percent if we only use a single production technology. Our findings come from the fact that large firms face a higher marginal cost for hiring unskilled labor, but also that large firms employ production technologies that require less unskilled labor. If large firms use systematically different production technologies than small firms, then assuming one technology leads to overstating the correlation between size and MRPs.

For capital and intermediate inputs, we find much weaker relationships between size and MRPs, with ratios of 1.27 and 1.23, respectively, for the MRPs of capital and intermediate inputs for firms.

Hopenhayn (2014) derives a sufficient statistic that allows for the comparison of the TFP losses from two different sets of distortions. Using this condition, he shows that distortions that are correlated to productivity reduce TFP by a greater amount than non-correlated distortions (that is, distortions chosen at random).
Panel A of Figure 3 shows the lowess fitted curve for the MRP of capital by gross output under the assumption of a single cluster and allowing for multiple clusters. For multiple clusters, we present two curves, one under the assumption that clusters are differentiated solely by production technologies and the other under the assumption that clusters are differentiated solely by cluster-specific distortions. Panel B shows the same plot for intermediate inputs; Panel C shows the same plot for skilled labor; and Panel D shows the same plot for unskilled labor. We use least square smoothing and a bandwidth of 0.8 for the lowess locally weighted regression.
inputs would have a much larger impact than the same distortion applied to unskilled labor, so this information does not imply that misallocation in intermediate inputs is unimportant. In our framework, misallocation in intermediate inputs ends up being a non-negligible part of overall misallocation despite the low dispersion in MRPs arising from the importance of intermediate inputs in the production functions we estimate for firms.

5.5 Gains from Eliminating Distortions

We now want to understand the importance of cluster-specific factors for the estimated gains in manufacturing output and TFP from eliminating misallocation through removing distortions in the economy.\(^{28}\) We again present our results by contrasting two extreme cases, one in which we do not allow for cluster-input-specific distortions, and another in which we do not allow for cluster-specific technologies. Our results from the previous section highlight two main reasons why accounting for cluster-specific production technologies would be expected to reduce the predicted gains from eliminating misallocation. The first reason is that accounting for cluster-specific production technologies reduces the estimated dispersion in MRPs of inputs across firms, particularly for capital, suggesting smaller distortions and therefore smaller gains from eliminating them. The second reason is that accounting for different technology usage across large and small firms can explain a large part of the relationship between MRPs and the size of firms for skilled and unskilled labor. By reducing the magnitude of this association, we can be expected to reduce the predicted gains from eliminating distortions. If we instead interpret clusters as arising from cluster-input-specific distortions, we no longer necessarily expect the gains from eliminating misallocation to decrease. We can, however, decomposed the gains into those due to eliminating cluster-specific distortions and those due to eliminating idiosyncratic distortions to understand the relative importance of the two factors.

We want to allow for reallocation across industries; therefore, we impose additional structure on the model presented in Section 2. In particular, we consider a multi-industry extension where the output of each industry (let the subscript \(s\) denote the industry for each variable from Section 2) is used to create a single final consumption good, \(Y\), by a perfectly competitive bundler with the following Cobb-Douglas production function:

\[
Y = \prod_{s=1}^{S} (Y_s)^{\alpha_s},
\]

where \(Y_s\) is the output from industry \(s\), which we get from equation 2, \(\alpha_s\) is industry \(s\)’s share of

\(^{28}\)We follow the typical definition of misallocation in the literature and define it as losses that result from the presence of distortions. In our framework, eliminating markup heterogeneity across industries would generate an increase in aggregate output; however, we do not classify this as misallocation or remove markup heterogeneity in our counterfactual experiments. The papers by Peters (2020) and Asturias, Garcia-Santana, and Ramos (2019) are recent examples of studies that examine the losses in output that result from markup heterogeneity.
total expenditures, and \( S \) is the total number of industries in the economy.

We assume that factors are fully mobile across industries so that the income received for supplying one unit of each input will be the same regardless of which industry or firm uses that input. The budget constraint of the representative consumer implies that total expenditures will be equal to income from factor inputs, \( p^i X^i \) for input \( i \), profits from firms, \( \Pi \), and transfers from firms due to distortions, \( T \). Total expenditures across all industries in the economy are therefore given by

\[
PY = \sum_{i=1}^{I} p^i X^i + \Pi + T, \tag{36}
\]

where \( X^i \) denotes the exogenously determined and inelastically supplied total amount of factor \( i \) in the economy and \( P \) is the aggregate price index:

\[
P = \prod_{s=1}^{S} \left( \frac{P_s}{\alpha_s} \right)^{\alpha_s}, \tag{37}
\]

where \( P_s \) is defined in equation 4. Profits are the result of markups over marginal cost and are given by

\[
\Pi = \sum_{s=1}^{S} \left( \frac{1 - \rho_s}{\rho_s} \right) \left( \sum_{m=1}^{M_s} p_{m,s} y_{m,s} \right). \tag{38}
\]

Note that markups are allowed to vary across industries. Transfers are defined as

\[
T = \sum_{s=1}^{S} \sum_{m=1}^{M_s} \sum_{i=1}^{I} \left( \tau_{h(m),s} \tau_{i(m),s} \tau_{i,m,s} - 1 \right) p^i x_{m,s}^i. \tag{39}
\]

The clearing condition for each input \( i \) is given by

\[
X^i = \sum_{s=1}^{S} \sum_{m=1}^{M_s} x^i_{m,s}, \tag{40}
\]

where the quantity of input demanded by each firm is given by

\[
x^i_{m,s} = \theta^i_{h(m),s} \frac{1}{\rho_s} \frac{p_{m,s} y_{m,s}}{\tau_{h(m),s} \tau_{i(m),s} \tau_{i,m,s} P^i}. \tag{41}
\]

The previous equation shows that firms that face a high distortion for a given input will demand less of that input in equilibrium. In the absence of distortions, we would expect constant factor costshares across firms that share a common production technology.

In order to calculate the gains from eliminating distortions, we need to calibrate the productivity of each plant. We are interested in estimating the percentage change in output given the removal of distortions, which means that we only need to know the relative productivities of firms within each industry. To calculate relative productivities, we first normalize the productivity of one plant, \( \tilde{m} \), in each industry to 1. Let \( h \) represent the technology of firm \( m \) and \( \tilde{h} \) represent the technology
of firm $\tilde{m}$, both in industry $s$. We then combine equations 3 and 7 to find the productivity of plant $m$ relative to that of $\tilde{m}$. We have that the productivity of plant $m$ relative to plant $\tilde{m}$ is

$$z_{m,s} = \frac{z_{m,s}}{z_{\tilde{m},s}} = \left( \frac{p_{m,s}y_{m,s}}{p_{\tilde{m},s}y_{\tilde{m},s}} \right)^{1-\rho_s} \prod_{i=1}^{I} \left( \frac{\tau_{i,m,s}^{\tilde{m},s} \theta_{i,m,s}^{\tilde{m},s} \theta_{i,m,s}^{\tilde{m},s} \tau_{i,m,s}^{\tilde{m},s}}{\tau_{i,n,s}^{\tilde{m},s} \theta_{i,n,s}^{\tilde{m},s} \theta_{i,n,s}^{\tilde{m},s} \tau_{i,n,s}^{\tilde{m},s}} \right)^{-1},$$

(42)

where we normalize $z_{\tilde{m},s} = 1$ to find $z_{m,s}$. Equation 42 ensures that the size distribution of firms is the same in the model and data. Note that if firms share a production technology, their relative productivity is a function of their relative revenues and a weighted geometric mean of their relative distortions, where the weights of the geometric means are the factor intensities.

When we calibrate our model, revenues and input expenditures come from the data; however, our data do not break down values into prices and quantities. We are only interested in changes in output; therefore, we can normalize the total supply of each factor equal to one, that is, set $X^i = 1$, for all $i$. This normalization does not alter our results because it gets absorbed into prices and therefore into how we calibrate our productivities. If instead we were able to identify the productivity of each firm independently of calibrating it using equation 42, then changing the total supply of each factor could lead to different results in terms of the gains from reducing misallocation. We can show, for example, that the relative amount of inputs across firms does not change with the total endowment of that input in our calibration. Equation 41 shows that the demand for an input is a function of the revenues of a firm. We take firm revenues directly from the data in our calibration; therefore, the relative amount of inputs across firms remains the same even if we change the total endowment. Observed factor expenditures, $p^i x^i_{m,s}$, are likewise taken from the data, which highlights that an increase in the endowment of an input must be absorbed into an equal and inverse decrease in the price of that input.

After normalizing the total supply of each factor, we can calculate the equilibrium numerically by solving equations 35–42 for input prices, productivities, and output of each firm and therefore output for each industry, $Y_s$, and aggregate output, $Y$. We repeat the same exercise after setting $\tau_{h(m),s} = 1$, $\tau_{\tilde{h}(m),s} = 1$, and $\tau_{m,s} = 1$ for all firms, industries, and inputs. This allows us to predict aggregate output in the absence of distortions $\hat{Y}$. We calculate the percentage gain in aggregate output from eliminating misallocation as

$$\Delta Y \equiv 100 \times \frac{\hat{Y} - Y}{Y}. \quad (43)$$

We similarly calculate the percentage gain in output, $\Delta Y_s$, for each industry, $s$, by substituting industry-level output, $Y_s$ and $\hat{Y}_s$, for aggregate output in the above equation.

Table 7 reports the counterfactual gains in output from eliminating misallocation by removing distortions. First we calculate the gains under the standard assumption of a single cluster in each industry by setting $H = 1$ for all industries and following the rest of our methodology above and in
Section 3.5. We find that the counterfactual gains in aggregate output for eliminating distortions in the standard specification are 38.7 percent. Next, we examine the gains from eliminating distortions when we allow for multiple clusters in each industry according to Table 1. As discussed in Section 2.2, clusters are distinguished by having different values of $\tau_{ih}$ across inputs, which is a combination of cluster-input-specific distortions and differences in factor intensities (production technologies). Without additional information or assumptions, we cannot disentangle these two components. ($\theta_{i}^h = \theta^i$).

When we attribute all differences between clusters to differences in production technologies, the predicted gains in aggregate output from eliminating misallocation drop to 26.8 percent — a reduction of nearly 1/3 relative to the standard framework. Note, that while we are not allowing for cluster-input-specific distortions, $\tau_{ih}$, we do still allow for cluster-specific distortions, $\tau_h$, that are common across inputs. While the fraction of misallocation explained by these cluster-specific distortions is low, they do amplify the gains when eliminated jointly with the idiosyncratic distortions, $\tau_{im}$. To see where these gains are coming from, Table 8 reports the percentage changes in output at the industry level. Whereas most industries experience a growth in output, we also find a decline in output after the removal of distortions for some industries, which highlights the importance of the reallocation of inputs across sectors.

In contrast to our large decrease in aggregate gains in the multiple production technology case, if we attribute all differences between clusters to cluster-specific distortions, the predicted gains remain substantially closer to the standard framework at 35.3 percent — a reduction of approximately 10 percent. Despite this relatively modest change, cluster-specific distortions play an important role in determining the aggregate gains from eliminating misallocation. To see this, Table 7 shows the counterfactual gains when we remove cluster-specific and idiosyncratic distortions separately. We find that eliminating the cluster-specific distortion raises output by 6.7 percent. The elimination of the firm-only distortions raises output by 23.1 percent. Thus, we find that approximately $1/3$ ($= 1 - 23.1/35.3$) of the gains from eliminating misallocation can be attributed to cluster-specific factors, which is both due to directly eliminating cluster-specific distortions (6.7 percent) and due to the additional gains of eliminating cluster-specific distortions after idiosyncratic distortions have already been removed (an additional 5.5 percent).

To understand where the largest gains in productivity arise at the industry level, we conduct a similar exercise for changes in TFP. Table 7 reports the changes in aggregate TFP, and Table 9 reports the changes in TFP at the industry level. To calculate TFP, we first construct a representative production function for each technology (firms that differ only in cluster-specific distortions are considered as using the same technology when we compute TFP). Output for technology $h$ in industry $s$ is

$$Y_{h,s} \equiv \left( \sum_{m=1}^{M_s} \mathbb{I}_{\delta(m)=h}(y_{m,s})^{\rho_s} \right)^{1/\rho_s}.$$ (44)
Thus, the TFP for technology \( h \) is

\[
A_{h,s} = \frac{Y_{h,s}}{\Pi_{i=1}^{l} \left( X_{h,s}^i \right)^{\theta_i}},
\]

(45)

where \( X_{h}^i \) is the total amount of input \( i \) — capital, skilled labor, unskilled labor, and intermediate inputs — used by firms utilizing technology \( h \), which is characterized by

\[
X_{h,s}^i \equiv \sum_{m=1}^{M_s} I_{h(m)=h} x_{m,s}.
\]

(46)

We find the percentage change in TFP for each technology after the removal of distortions, \( A'_{h,s} \). To calculate changes in industry-level TFP, we weight these percentage changes with the share of gross output accounted for by firms that use that technology in the calibrated economy:

\[
\Delta A_s \equiv \sum_{h} \left( \frac{A'_{h,s}}{A_{h,s}} P_{h,s} Y_{h,s} \right).
\]

(47)

Further, it reports the change in aggregate TFP. We compute changes in aggregate TFP as follows:

\[
\Delta A \equiv \sum_{s} (\alpha_s \Delta A_s),
\]

where the change in aggregate TFP, \( \Delta A \), is defined as the weighted average of changes in industry TFPs, \( \Delta A_s \), where the weight for each industry is its Cobb-Douglas expenditure share, \( \alpha_s \). We find that the aggregate gains in TFP are 30.9 percent in the case that we only allow for one cluster. The gains in TFP are 24.3 percent when we interpret clusters as being driven by differences in technology, and 33.7 percent when we interpret clusters as being due to cluster-input-specific distortions. Unlike output, which decreases in some industries when distortions are removed, Table 9 shows that TFP increases in all industries after the removal of distortions. As before, we find a significant reduction – nearly 1/4 – in the predicted gain when we interpret differences in clusters as being technological. Similarly, while the aggregate TFP gains are slightly higher for the cluster-input-specific distortions case compared to the single cluster case, we find that around 1/3 (\( =1 \cdot 23.1/33.7 \)) of the gains in TFP can be attributed to cluster-specific and cluster-input-specific distortions.

Several other papers have investigated the potential role of technology in explaining measured misallocation. For example, Hsieh and Klenow (2009) conduct an exercise to measure the role of technological differences across plants in explaining misallocation. In Section A.8, we contrast our approach with the methodologies used by these papers. We also discuss how the results from each approach inform us of the nature of misallocation.

5.6 Decomposing Technology and Cluster-Specific Distortions

In this section, we show how additional assumptions can be used to decompose differences in average cost shares across clusters into differences arising from technology (factor intensities) and
differences arising from cluster-input-specific distortions. We do so by exploiting within-cluster differences in cost shares across firms of different sizes. In particular, we conduct an exercise in which we attribute all size-related differences in cost shares across clusters to cluster-specific distortions. We consider this exercise to be conservative in terms of the role it leaves for cluster-specific technologies because technology is also likely to be related to size. For example, if operating a technology involves fixed costs, then larger firms may find it worthwhile to adopt a technology that is different from that of smaller firms.

We pin down our estimates of size-dependent differences across clusters by using estimates of size-related variation within clusters and then extrapolating these estimates based on the size differences of firms in different clusters. In particular, we regress logged cost shares on logged gross output with industry-cluster \((FE_{s,h}^i)\) fixed effects:

\[
\log \tilde{\theta}^i_m = \beta^i_{GO} \log p_m y_m + FE_{s,h}^i + \epsilon^i_m. \tag{48}
\]

The regression estimate of \(\beta^i_{GO}\) for each input is reported in Table 10. In the above regression, \(\beta^i_{GO} \log p_m y_m\) captures differences in cost shares that can be attributed to size differences across firms. The results show that, controlling for industry and cluster, larger firms tend to have higher intermediate input cost shares and lower skilled and unskilled labor cost shares.

We use these estimates to generate hypothetical input cost shares, \(\hat{\theta}^i_m\), that we predict for each firm using only information on the firm’s size and industry. In particular, we generate the hypothetical input cost shares according to the equation

\[
\log \hat{\theta}^i_m = \beta^i_{GO} \log p_m y_m, \tag{49}
\]

where the values for \(\beta^i_{GO}\) is taken directly from our estimates of the regression in equation 48.

Using the hypothetical cost shares from equation 49, we rerun our analysis to compute the cluster-input-specific distortions, \(\tau^i_h\). To do so, we use the hypothetical cost shares in equations 28 and 33 (substitute \(\hat{\theta}^i_m\) for \(\tilde{\theta}^i_m\)) with the same cluster assignments as in our main analysis. In particular, when we combine these two expressions to arrive at the following expression for \(\tau^i_h\)

\[
\tau^i_h = \frac{1}{H} \sum_{H'=1}^H \frac{1}{M'} \sum_{m=1}^M \sum_{m'=1}^M \frac{I_{i'(m)=h} \hat{\theta}^i_m}{M'_m} \cdot \sum_{m=1}^M \sum_{m'=1}^M \frac{I_{i'(m)=h} \hat{\theta}^i_m}{M'_m} . \tag{50}
\]

When using the hypothetical cost shares, we assume that all cluster-specific variation can be attributed to cluster-input-specific distortions, which allows us to estimate \(\tau^i_h\). After computing cluster-input-specific distortions, \(\tau^i_h\), we are able to input them directly into equation 28 with the actual observed cost shares, \(\tilde{\theta}^i_m\), where we attribute the remaining variation in observed cost shares to differences in factor intensities, \(\theta^i_h\).\(^{29}\)

\(^{29}\)Note, that the inclusion of \(FE_{s,h}^i\) in equation 49 will not affect our estimates for \(\tau^i_h\) and \(\theta^i_h\). To see why this
We can now compare the relative importance of the factor intensities, $\theta_i$, and cluster-input-specific distortions, $\tau_i$, in explaining the variation of $\theta_i/\tau_i$ that we recovered in the main analysis. Table 4 shows the results from regressing the log of $\theta_i/\tau_i$ of plants on the log of gross output and industry dummies. Table 11 reports the results of the same regression, except that the dependent variable is the log of the updated factor intensities, $\theta_i$; Table 12 reports the same results except, that the dependent variable is the log of the cluster-input-specific distortions, $\tau_i$. Notice that, as expected, if we subtract the estimates of Table 12 from the estimates of Table 11, we recover the same estimates of Table 4. We see that the sign and statistical significance of all the coefficients remain unchanged, indicating that larger firms are still more likely to use technologies with higher factor intensities for intermediate inputs and lower factor intensities for skilled and unskilled labor. The magnitudes are lower, however, by approximately one-third, owing to cluster-specific distortions, whereas two-thirds of the effect remains because of technological differences across clusters.

Table 13 reports the predicted gains in output and TFP by eliminating misallocation due to distortions following Section 5.5. Comparing this table to the results from Table 7, we see that the gains from eliminating misallocation when we allow for both cluster-specific distortions and cluster-input-specific technologies are nearly identical to when we allow for cluster-specific technologies only. For example, the gain in output from eliminating all misallocation is 26.5 percent in the framework where we allow for both cluster-specific technologies and cluster-input-specific distortions versus 26.8 percent when we only allow for cluster-specific technologies and 35.3 percent when we allow for only cluster-input-specific distortions. These results are not surprising given that even when we attribute all cluster-specific differences to cluster-specific distortions, as we do in Table 7, eliminating cluster-specific distortions does not have a large impact on misallocation beyond amplifying the impact of eliminating idiosyncratic distortions. These results further suggest that our analysis in which we allow for only cluster-specific distortions and cluster-specific technologies may be a reasonable approximation even if cluster-input-specific distortions exist and are non-negligible.

6 Conclusion

In this paper, we developed a methodology to identify variation in cost shares due to cluster-specific and idiosyncratic components. We find evidence for multiple clusters in the majority of industries in the Chilean manufacturing sector and find that cluster-specific variation can account for a large fraction of the overall variation in cost shares—ranging from 23 percent on average for industries with two clusters up to 47 percent of industries with four clusters.

is the case, we exponentiate both sides equation 49 to find $\hat{\theta}_m = (p_m y_m)^{\beta_{GO} e^{F E^{i}_{s,h}}}$, which implies that $F E^{i}_{s,h}$ has a proportional effect on $\hat{\theta}_m$. Equation 50 shows that including $F E^{i}_{s,h}$ will have the same proportional effect on the numerator and denominator of the expression and will thus have no effect on the estimated $\tau_i$. 

34
Given the findings of this paper, a natural next step would be to try to identify the fraction of variation in cost shares that is attributable to differences in technology as opposed to cluster-specific distortions. With that goal in mind, it would be useful to apply the clustering methodology in this paper to U.S. manufacturing data. Because the United States is typically considered to be an undistorted economy, the clusters that emerge from this exercise could be interpreted as coming from different technologies. We would thus be able to compare the clustering patterns in the United States and in those in developing countries such as Chile, which would indicate whether our results in Chile are driven by technological factors. Furthermore, under the assumption that firms across countries use the same technology, one could use the technologies identified in the U.S. data and apply them to developing countries in the spirit of Hsieh and Klenow (2009) to quantify the gains from eliminating misallocation.\footnote{It would also be useful to implement a clustering procedure that allows for hidden groups when directly estimating the production function parameters. For example, the use of a mixture model would allow us to use maximum likelihood estimation and thus to jointly estimate parameters related to the clusters and the production function parameters for each cluster. The main advantage would be that it would allow for more flexibility in the production function since it would not require us to derive an equation like 13 to apply the clustering methodology.}

In this paper, we utilized the K-means++ clustering algorithm, which assigns each firm to a single production technology. This choice makes it straightforward to conduct counterfactual experiments, as we do not need to keep track of a probabilistic set of potential distortions for each firm when we consider the effect of eliminating misallocation. For other applications, in which researchers are less interested in counterfactual experiments and more interested in characterizing the features of each cluster, researchers may be interested in exploring alternative clustering and machine learning algorithms. For example, mixed-membership models are used by search engines to categorize objects such as news articles, which may fall into multiple categories and could be useful for analyzing multi-product plants that employ cluster-specific technologies simultaneously. One of the primary contributions of this paper is the use of an economic model to justify which variables the clustering algorithm should be applied to, in what cases the methodology will be successful, and how to use the results of the algorithm to recover specific unobservable parameters in the original model. Therefore, although we expect these algorithms to be useful for analyzing the increasingly disaggregated plant-level data available to researchers, we also want to emphasize the importance of connecting the algorithms employed to specific economic models, as we do in this paper.
### Table 1

**NUMBER OF PRODUCTION TECHNOLOGIES BY INDUSTRY**

<table>
<thead>
<tr>
<th>ISIC</th>
<th>Industry Description</th>
<th>Firms</th>
<th>Clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>1511</td>
<td>Processing and preserving of meat products</td>
<td>113</td>
<td>2</td>
</tr>
<tr>
<td>1512</td>
<td>Processing and preserving of fish products</td>
<td>152</td>
<td>3</td>
</tr>
<tr>
<td>1513</td>
<td>Processing and preserving of fruits and vegetables</td>
<td>80</td>
<td>1</td>
</tr>
<tr>
<td>1520</td>
<td>Manufacture of dairy products</td>
<td>56</td>
<td>1</td>
</tr>
<tr>
<td>1531</td>
<td>Manufacture of grain mill products</td>
<td>88</td>
<td>1</td>
</tr>
<tr>
<td>1541</td>
<td>Manufacture of bakery products</td>
<td>499</td>
<td>1</td>
</tr>
<tr>
<td>1552</td>
<td>Manufacture of wines</td>
<td>110</td>
<td>4</td>
</tr>
<tr>
<td>1711</td>
<td>Preparation and weaving of textiles</td>
<td>53</td>
<td>3</td>
</tr>
<tr>
<td>1810</td>
<td>Manufacture of wearing apparel</td>
<td>188</td>
<td>1</td>
</tr>
<tr>
<td>1920</td>
<td>Manufacture of footwear</td>
<td>60</td>
<td>1</td>
</tr>
<tr>
<td>2010</td>
<td>Sawmilling and planing of wood</td>
<td>196</td>
<td>2</td>
</tr>
<tr>
<td>2102</td>
<td>Manufacture of corrugated paper and paperboard</td>
<td>63</td>
<td>3</td>
</tr>
<tr>
<td>2211</td>
<td>Publishing of books, brochures, and other publications</td>
<td>52</td>
<td>1</td>
</tr>
<tr>
<td>2221</td>
<td>Printing</td>
<td>64</td>
<td>4</td>
</tr>
<tr>
<td>2411</td>
<td>Manufacture of basic chemicals</td>
<td>66</td>
<td>3</td>
</tr>
<tr>
<td>2424</td>
<td>Manufacture of soap, cleaning preparations, and perfumes</td>
<td>58</td>
<td>2</td>
</tr>
<tr>
<td>2429</td>
<td>Manufacture of other chemical products n.e.c.</td>
<td>61</td>
<td>2</td>
</tr>
<tr>
<td>2520</td>
<td>Manufacture of plastics products</td>
<td>242</td>
<td>2</td>
</tr>
<tr>
<td>2695</td>
<td>Manufacture of articles of concrete, cement and plaster</td>
<td>103</td>
<td>2</td>
</tr>
<tr>
<td>2720</td>
<td>Manufacture of basic precious and non-ferrous metals</td>
<td>62</td>
<td>2</td>
</tr>
<tr>
<td>2811</td>
<td>Manufacture of structural metal products</td>
<td>120</td>
<td>3</td>
</tr>
<tr>
<td>2899</td>
<td>Manufacture of other fabricated metal products n.e.c.</td>
<td>125</td>
<td>1</td>
</tr>
<tr>
<td>3610</td>
<td>Manufacture of furniture</td>
<td>143</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1 shows the number of technologies determined by the clustering methodology at the industry level. *ISIC* is the four-digit ISIC Rev. 3 code for each industry we evaluate with our methodology (>50 firms). *Industry Description* lists a shortened description of the economic activities that make up each industry. *Firms* lists the number of firms (plants) in each industry. *Prod. Technologies* shows the number of production technologies operating in each industry as recovered by our methodology.
Table 2

Variance Decomposition by Number of Clusters

<table>
<thead>
<tr>
<th>Number of Clusters</th>
<th>Idiosyncratic Share</th>
<th>Cluster-Specific Share</th>
<th>Covariance Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>77.2</td>
<td>23.0</td>
<td>-0.2</td>
</tr>
<tr>
<td>3</td>
<td>60.1</td>
<td>37.4</td>
<td>2.5</td>
</tr>
<tr>
<td>4</td>
<td>56.1</td>
<td>46.9</td>
<td>-3.0</td>
</tr>
</tbody>
</table>

Table 2 shows the variance decomposition for factor ratios across firms within an industry. Shares are constructed separately for each input and industry and then averaged across all inputs across industries with the same number of clusters and reported as a percentage. *Idiosyncratic Share* is defined as $\frac{\text{Var}(\log \tau^i_m)}{\text{Var}(\log \tilde{\theta}^i_m)}$ and captures within-cluster variation. *Cluster-Specific Share* is defined as $\frac{\text{Var}(\log \theta^i_h - \log \tau^i_h \tau^i_m)}{\text{Var}(\log \tilde{\theta}^i_m)}$ and captures between-cluster variation. *Covariance Share* is the residual and calculated as 1 minus the other two shares.

Table 3

Variance Decomposition by Input

<table>
<thead>
<tr>
<th>Input</th>
<th>Idiosyncratic Share</th>
<th>Cluster-Specific Share</th>
<th>Covariance Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capital</td>
<td>47.9</td>
<td>54.7</td>
<td>-2.7</td>
</tr>
<tr>
<td>Skilled Labor</td>
<td>71.2</td>
<td>26.8</td>
<td>2.0</td>
</tr>
<tr>
<td>Unskilled Labor</td>
<td>59.2</td>
<td>37.5</td>
<td>3.3</td>
</tr>
<tr>
<td>Intermediate Inputs</td>
<td>94.0</td>
<td>7.2</td>
<td>-1.1</td>
</tr>
</tbody>
</table>

Table 3 shows the variance decomposition for factor ratios across firms within an industry. Shares are constructed separately for each input and industry and then averaged across all industries with at least two clusters and reported as a percentage. *Idiosyncratic Share* is defined as $\frac{\text{Var}(\log \tau^i_m)}{\text{Var}(\log \tilde{\theta}^i_m)}$ and captures within-cluster variation. *Cluster-Specific Share* is defined as $\frac{\text{Var}(\log \theta^i_h - \log \tau^i_h \tau^i_m)}{\text{Var}(\log \tilde{\theta}^i_m)}$ and captures between-cluster variation. *Covariance Share* is the residual and calculated as 1 minus the other two shares.
Table 4

Cluster-Specific Intensities ($θ_i^h/τ_i^h$) vs. Gross Output

<table>
<thead>
<tr>
<th>Logged Cluster-Specific Intensity</th>
<th>Capital</th>
<th>Skilled Labor</th>
<th>Unskilled Labor</th>
<th>Intermediate Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logged Gross Output</td>
<td>-0.0106</td>
<td>-0.0382***</td>
<td>-0.0846***</td>
<td>0.0119***</td>
</tr>
<tr>
<td></td>
<td>(0.0130)</td>
<td>(0.00590)</td>
<td>(0.00892)</td>
<td>(0.00206)</td>
</tr>
<tr>
<td>Industry FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>1463</td>
<td>1463</td>
<td>1463</td>
<td>1463</td>
</tr>
</tbody>
</table>

Standard errors in parentheses; * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Table 4 reports the results for the regression of the log of gross output on the log of each cluster-specific intensity ($θ_i^h/τ_i^h$). Each column reports the result of the regression for the specified input, where industry fixed effects are included to adjust for differences across industries. A value greater than zero indicates that larger firms are more likely to operate technologies with a greater cluster-specific intensity for the specified input.

Table 5

Recovered Production Technologies
Manufacturing of Non-Ferrous and Precious Metals (ISIC 2720)

<table>
<thead>
<tr>
<th>Case</th>
<th>Technology</th>
<th>Count</th>
<th>$θ_{sl}$</th>
<th>$θ_{ul}$</th>
<th>$θ_k$</th>
<th>$θ_x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Technology</td>
<td>62</td>
<td>0.07</td>
<td>0.05</td>
<td>0.08</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>Multiple Technologies</td>
<td>1</td>
<td>19</td>
<td>0.03</td>
<td>0.01</td>
<td>0.10</td>
<td>0.87</td>
</tr>
<tr>
<td>Multiple Technologies</td>
<td>2</td>
<td>43</td>
<td>0.08</td>
<td>0.07</td>
<td>0.07</td>
<td>0.78</td>
</tr>
<tr>
<td>Cluster-Specific Distortions</td>
<td>62</td>
<td>0.05</td>
<td>0.04</td>
<td>0.09</td>
<td>0.82</td>
<td></td>
</tr>
</tbody>
</table>

Table 5 shows the recovered factor intensities for the manufacturing of non-ferrous and precious metals (ISIC 2720). Case indicates whether we report the case in which we impose a single technology or allow for multiple technologies. Technology assigns each technology an identifying number, which is useful if there is more than one technology in the industry. These identifiers are used in Figure 1. Count lists the number of plants that use each technology. Columns 4-7 report the factor intensities for skilled labor, unskilled labor, capital, and intermediate inputs, respectively.
Table 6 reports the percentage reduction in the standard deviation of logged MRPs of each input after accounting for the presence of multiple production technologies. Only industries with more than one production technology are included in the table. The row labeled Mean reports the average reduction taken across the industries listed in the table. The row labeled Aggregate reports the reduction when we compute the standard deviation in logged MRPs across all plants in the economy while ignoring industries. Columns 2-5 show the percentage decline in the standard deviation of the logged MRP for capital, skilled labor, unskilled labor, and intermediate inputs, respectively, in each industry. The percentage decline of the standard deviation for input \( i \) is calculated as \( 100 \times (\sigma_{i,\text{multi}}/\sigma_{i,\text{1tech}} - 1) \), where \( \sigma_{i,\text{multi}} \) is the standard deviation of the logged MRP for input \( i \) allowing for multiple technologies and \( \sigma_{i,\text{1tech}} \) is the same statistic for the case in which we only allow for a single production technology in each industry.

<table>
<thead>
<tr>
<th>ISIC</th>
<th>Capital</th>
<th>Skilled Labor</th>
<th>Unskilled Labor</th>
<th>Intermediate Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1511</td>
<td>34.7</td>
<td>3.0</td>
<td>10.9</td>
<td>3.9</td>
</tr>
<tr>
<td>1512</td>
<td>40.2</td>
<td>20.9</td>
<td>9.3</td>
<td>-6.4</td>
</tr>
<tr>
<td>1552</td>
<td>37.4</td>
<td>16.0</td>
<td>34.9</td>
<td>-0.5</td>
</tr>
<tr>
<td>1711</td>
<td>41.9</td>
<td>15.2</td>
<td>35.4</td>
<td>8.5</td>
</tr>
<tr>
<td>2010</td>
<td>26.4</td>
<td>-0.2</td>
<td>18.0</td>
<td>2.4</td>
</tr>
<tr>
<td>2102</td>
<td>31.4</td>
<td>17.3</td>
<td>33.6</td>
<td>15.4</td>
</tr>
<tr>
<td>2221</td>
<td>30.7</td>
<td>39.9</td>
<td>47.3</td>
<td>0.6</td>
</tr>
<tr>
<td>2411</td>
<td>40.4</td>
<td>25.9</td>
<td>29.1</td>
<td>1.3</td>
</tr>
<tr>
<td>2424</td>
<td>28.0</td>
<td>17.5</td>
<td>7.9</td>
<td>3.3</td>
</tr>
<tr>
<td>2429</td>
<td>26.9</td>
<td>10.5</td>
<td>6.7</td>
<td>4.3</td>
</tr>
<tr>
<td>2520</td>
<td>31.0</td>
<td>0.8</td>
<td>8.8</td>
<td>-1.5</td>
</tr>
<tr>
<td>2695</td>
<td>30.7</td>
<td>11.9</td>
<td>7.4</td>
<td>3.0</td>
</tr>
<tr>
<td>2720</td>
<td>-2.4</td>
<td>21.5</td>
<td>37.0</td>
<td>-2.1</td>
</tr>
<tr>
<td>2811</td>
<td>36.6</td>
<td>11.2</td>
<td>44.4</td>
<td>0.8</td>
</tr>
<tr>
<td>Mean</td>
<td>31.0</td>
<td>15.1</td>
<td>23.6</td>
<td>2.4</td>
</tr>
<tr>
<td>Aggregate</td>
<td>15.0</td>
<td>6.8</td>
<td>12.1</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Table 6 reports the percentage reduction in the standard deviation of logged MRPs of each input after accounting for the presence of multiple production technologies. Only industries with more than one production technology are included in the table. The row labeled Mean reports the average reduction taken across the industries listed in the table. The row labeled Aggregate reports the reduction when we compute the standard deviation in logged MRPs across all plants in the economy while ignoring industries. Columns 2-5 show the percentage decline in the standard deviation of the logged MRP for capital, skilled labor, unskilled labor, and intermediate inputs, respectively, in each industry. The percentage decline of the standard deviation for input \( i \) is calculated as \( 100 \times (\sigma_{i,\text{multi}}/\sigma_{i,\text{1tech}} - 1) \), where \( \sigma_{i,\text{multi}} \) is the standard deviation of the logged MRP for input \( i \) allowing for multiple technologies and \( \sigma_{i,\text{1tech}} \) is the same statistic for the case in which we only allow for a single production technology in each industry.
Table 7
**Aggregate Gains from Eliminating Distortions**

<table>
<thead>
<tr>
<th>Case</th>
<th>Cluster-Specific Technologies</th>
<th>Cluster-Input-Specific Distortions</th>
<th>Single Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Cluster Only</td>
<td>Firm Only</td>
</tr>
<tr>
<td>Output</td>
<td>26.8</td>
<td>0.6</td>
<td>23.9</td>
</tr>
<tr>
<td>TFP</td>
<td>24.3</td>
<td>1.5</td>
<td>20.4</td>
</tr>
</tbody>
</table>

Table 7 shows the aggregate gains in output and TFP from removing distortions. *Cluster-Specific Technologies* lists the calculated increase in output when we allow for cluster-specific technologies and cluster-specific distortions but not cluster-input-specific distortions. *Cluster-Input-Specific Distortions* shows the same calculation when we allow for cluster-input-specific distortions but not cluster-specific technologies. *Single Cluster* reports the same calculation when we do not allow for cluster-specific technologies or cluster-specific distortions. *All* indicates that both cluster-specific distortions ($\tau_h$, $\tau_i^h$) and idiosyncratic distortions ($\tau_{im}$) were removed. *Cluster Only* indicates that only the cluster-specific distortions were removed and that idiosyncratic distortions were left unchanged. *Firm Only* indicates that only the idiosyncratic distortions were removed and that cluster-specific distortions were left unchanged.
### Table 8

**Percentage Increase in Output from the Elimination of Misallocation**

<table>
<thead>
<tr>
<th>ISIC</th>
<th>Cluster-Specific Technologies</th>
<th>Cluster-Input-Specific Distortions</th>
<th>Single Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>1511</td>
<td>15.3</td>
<td>25.0</td>
<td>16.8</td>
</tr>
<tr>
<td>1512</td>
<td>47.3</td>
<td>29.3</td>
<td>32.5</td>
</tr>
<tr>
<td>1513</td>
<td>25.0</td>
<td>31.6</td>
<td>21.9</td>
</tr>
<tr>
<td>1520</td>
<td>17.4</td>
<td>22.6</td>
<td>16.3</td>
</tr>
<tr>
<td>1531</td>
<td>18.9</td>
<td>24.2</td>
<td>20.4</td>
</tr>
<tr>
<td>1541</td>
<td>36.7</td>
<td>41.4</td>
<td>31.3</td>
</tr>
<tr>
<td>1711</td>
<td>-6.9</td>
<td>-4.0</td>
<td>-3.1</td>
</tr>
<tr>
<td>1810</td>
<td>5.5</td>
<td>9.4</td>
<td>-2.5</td>
</tr>
<tr>
<td>1920</td>
<td>9.3</td>
<td>13.8</td>
<td>3.5</td>
</tr>
<tr>
<td>2010</td>
<td>2.8</td>
<td>7.1</td>
<td>6.6</td>
</tr>
<tr>
<td>2102</td>
<td>-8.1</td>
<td>-2.5</td>
<td>-7.7</td>
</tr>
<tr>
<td>2211</td>
<td>-6.4</td>
<td>-1.1</td>
<td>-14.3</td>
</tr>
<tr>
<td>2221</td>
<td>-4.6</td>
<td>2.0</td>
<td>6.5</td>
</tr>
<tr>
<td>2411</td>
<td>27.3</td>
<td>50.1</td>
<td>55.8</td>
</tr>
<tr>
<td>2424</td>
<td>21.1</td>
<td>17.5</td>
<td>10.4</td>
</tr>
<tr>
<td>2429</td>
<td>16.5</td>
<td>18.9</td>
<td>18.3</td>
</tr>
<tr>
<td>2520</td>
<td>1.5</td>
<td>2.6</td>
<td>-0.0</td>
</tr>
<tr>
<td>2695</td>
<td>43.3</td>
<td>58.6</td>
<td>45.8</td>
</tr>
<tr>
<td>2720</td>
<td>40.9</td>
<td>43.2</td>
<td>58.2</td>
</tr>
<tr>
<td>2811</td>
<td>1.1</td>
<td>7.5</td>
<td>2.9</td>
</tr>
<tr>
<td>2899</td>
<td>6.2</td>
<td>10.8</td>
<td>-1.6</td>
</tr>
<tr>
<td>3610</td>
<td>9.4</td>
<td>14.2</td>
<td>0.7</td>
</tr>
<tr>
<td>Aggregate</td>
<td>26.8</td>
<td>35.3</td>
<td>38.7</td>
</tr>
</tbody>
</table>

Table 8 reports the percentage increase in industry-level output from removing distortions to eliminate misallocation both within and across industries. Column *ISIC* reports the industry code for each of the industries included in the exercise. ISIC 1552 (Manufacture of wines) is excluded because of a negative implied markup when only a single production technology is used. *Cluster-Specific Technologies* lists the calculated increase in output when we allow for cluster-specific technologies and cluster-specific distortions but not cluster-input-specific distortions. *Cluster-Input-Specific Distortions* shows the same calculation when we allow for cluster-input-specific distortions but not cluster-specific technologies. *Single Cluster* reports the same calculation when we do not allow for cluster-specific technologies or cluster-specific distortions. The row *Aggregate* reports the overall gain in output, where overall output is calculated according to equation 35.
Table 9
TFP GAINS FROM REALLOCATION
ONE VS. MULTIPLE TECHNOLOGIES

<table>
<thead>
<tr>
<th>ISIC</th>
<th>Cluster-Specific Technologies</th>
<th>Cluster-Input-Specific Distortions</th>
<th>Single Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>1511</td>
<td>27.0</td>
<td>29.3</td>
<td>32.5</td>
</tr>
<tr>
<td>1512</td>
<td>34.4</td>
<td>57.1</td>
<td>38.2</td>
</tr>
<tr>
<td>1513</td>
<td>33.1</td>
<td>33.1</td>
<td>33.1</td>
</tr>
<tr>
<td>1520</td>
<td>21.2</td>
<td>21.2</td>
<td>21.2</td>
</tr>
<tr>
<td>1531</td>
<td>29.7</td>
<td>29.7</td>
<td>29.7</td>
</tr>
<tr>
<td>1541</td>
<td>28.0</td>
<td>28.0</td>
<td>27.9</td>
</tr>
<tr>
<td>1711</td>
<td>9.4</td>
<td>9.5</td>
<td>17.7</td>
</tr>
<tr>
<td>1810</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
</tr>
<tr>
<td>1920</td>
<td>14.1</td>
<td>14.1</td>
<td>14.1</td>
</tr>
<tr>
<td>2010</td>
<td>16.5</td>
<td>18.8</td>
<td>24.1</td>
</tr>
<tr>
<td>2102</td>
<td>7.8</td>
<td>8.6</td>
<td>13.9</td>
</tr>
<tr>
<td>2211</td>
<td>14.6</td>
<td>14.6</td>
<td>15.3</td>
</tr>
<tr>
<td>2221</td>
<td>6.8</td>
<td>18.6</td>
<td>36.2</td>
</tr>
<tr>
<td>2411</td>
<td>10.3</td>
<td>32.1</td>
<td>17.4</td>
</tr>
<tr>
<td>2424</td>
<td>10.7</td>
<td>14.4</td>
<td>14.0</td>
</tr>
<tr>
<td>2429</td>
<td>22.7</td>
<td>24.6</td>
<td>30.9</td>
</tr>
<tr>
<td>2520</td>
<td>15.3</td>
<td>15.4</td>
<td>18.3</td>
</tr>
<tr>
<td>2695</td>
<td>30.7</td>
<td>38.3</td>
<td>38.1</td>
</tr>
<tr>
<td>2720</td>
<td>42.3</td>
<td>43.3</td>
<td>51.9</td>
</tr>
<tr>
<td>2811</td>
<td>19.6</td>
<td>22.8</td>
<td>26.5</td>
</tr>
<tr>
<td>2899</td>
<td>27.2</td>
<td>27.2</td>
<td>27.2</td>
</tr>
<tr>
<td>3610</td>
<td>23.0</td>
<td>23.0</td>
<td>23.0</td>
</tr>
<tr>
<td>Aggregate</td>
<td>24.3</td>
<td>33.7</td>
<td>30.9</td>
</tr>
</tbody>
</table>

Table 9 reports the percentage increase in industry-level TFP from removing distortions to eliminate misallocation both within and across industries. Column ISIC reports the industry code for each of the industries included in the exercise. ISIC 1552 (Manufacture of Wines) is excluded because of a negative implied markup when only a single production technology is used. Cluster-Specific Technologies lists the calculated increase in output when we allow for cluster-specific technologies and cluster-specific distortions but not cluster-input-specific distortions. Cluster-Input-Specific Distortions shows the same calculation when we allow for cluster-input-specific distortions but not cluster-specific technologies. Single Cluster reports the same calculation when we do not allow for cluster-specific technologies or cluster-specific distortions. The row Aggregate reports the overall gain in output, where overall output is calculated according to equation 35.
Table 10
Cost Shares vs. Gross Output with Cluster Fixed Effects

<table>
<thead>
<tr>
<th>Logged Cost Share</th>
<th>Capital</th>
<th>Skilled Labor</th>
<th>Unskilled Labor</th>
<th>Intermediate Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logged Gross Output</td>
<td>0.0249</td>
<td>-0.120***</td>
<td>-0.202***</td>
<td>0.0397***</td>
</tr>
<tr>
<td></td>
<td>(0.0134)</td>
<td>(0.0108)</td>
<td>(0.0112)</td>
<td>(0.00524)</td>
</tr>
<tr>
<td>Cluster FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>2754</td>
<td>2754</td>
<td>2754</td>
<td>2754</td>
</tr>
</tbody>
</table>

Standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001

Table 10 reports the results for the regression of the log of gross output on the log of each input cost share. Each column reports the result of the regression for the specified input, where industry and cluster fixed effects are included to adjust for differences in average cost shares across industries. A value greater than zero indicates that larger firms tend to have a higher cost share for a given input compared to smaller firms that are assigned to the same cluster.

Table 11
Factor Intensities vs. Gross Output

<table>
<thead>
<tr>
<th>Logged Factor Intensity</th>
<th>Capital</th>
<th>Skilled Labor</th>
<th>Unskilled Labor</th>
<th>Intermediate Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logged Gross Output</td>
<td>-0.0126</td>
<td>-0.0227***</td>
<td>-0.0589***</td>
<td>0.00808***</td>
</tr>
<tr>
<td></td>
<td>(0.0131)</td>
<td>(0.00546)</td>
<td>(0.00778)</td>
<td>(0.00192)</td>
</tr>
<tr>
<td>Industry FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>1463</td>
<td>1463</td>
<td>1463</td>
<td>1463</td>
</tr>
</tbody>
</table>

Standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001

Table 11 reports the results for the regression of the log of gross output on the log of each factor intensity when we allow for both cluster-specific technologies and cluster-input-specific distortions. Each column reports the results of the regression for the specified factor intensity, where industry fixed effects are included to adjust for differences in factor intensities across industries. A value greater than zero indicates that larger firms are more likely to be assigned to a cluster with a greater factor intensity for the specified input.
### Table 12

**Cluster-Specific Distortions vs. Gross Output**

<table>
<thead>
<tr>
<th>Logged Cluster-Input-Specific Distortion</th>
<th>Capital</th>
<th>Skilled Labor</th>
<th>Unskilled Labor</th>
<th>Intermediate Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logged Gross Output</td>
<td>-0.00207***</td>
<td>0.0155***</td>
<td>0.0258***</td>
<td>-0.00385***</td>
</tr>
<tr>
<td></td>
<td>(0.000221)</td>
<td>(0.00112)</td>
<td>(0.00185)</td>
<td>(0.000320)</td>
</tr>
<tr>
<td>Industry FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>1463</td>
<td>1463</td>
<td>1463</td>
<td>1463</td>
</tr>
</tbody>
</table>

Standard errors in parentheses; ∗ p < 0.05, ∗∗ p < 0.01, ∗∗∗ p < 0.001

Table 12 reports the results for the regression of the log of gross output on the log of each cluster-input-specific distortion when we allow for both cluster-specific technologies and cluster-input-specific distortions. Each column reports the results of the regression for the specified input, where industry fixed effects are included to adjust for differences across industries. A value greater than zero indicates that larger firms are more likely to be assigned to a cluster with a greater cluster-input-specific distortion for the specified input.

### Table 13

**Aggregate Gains from Eliminating Distortions**

<table>
<thead>
<tr>
<th>Case</th>
<th>All</th>
<th>Cluster Only</th>
<th>Firm Only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output</td>
<td>26.5</td>
<td>1.1</td>
<td>22.5</td>
</tr>
<tr>
<td>TFP</td>
<td>24.1</td>
<td>1.7</td>
<td>19.6</td>
</tr>
</tbody>
</table>

Table 13 shows the aggregate gains in output and TFP from removing distortions when we allow for both cluster-specific technologies and cluster-input-specific distortions, which we decompose according to Section 5.6. All indicates that cluster-specific distortions ($\tau_h$), cluster-input-specific distortions ($\tau_{ih}$), and idiosyncratic distortions ($\tau_{im}$) were all removed. Cluster Only indicates that only the cluster-specific and cluster-input-specific distortions were removed and that idiosyncratic distortions were left unchanged. Firm Only indicates that only the idiosyncratic distortions were removed and that cluster-specific and cluster-input-specific distortions were left unchanged.
References


A Appendix

A.1 MRPs versus Productivity

In Figure 3 we plot the MRP for each input against gross output. We find that under the assumption of a single cluster, the largest firms in the economy appear to have an MRP of skilled labor that is approximately 3.2 times higher than the smallest firms in the economy and roughly 8 times higher for unskilled labor. If we allow for multiple clusters and interpret these clusters as arising due to differences in production technologies employed by firms, the magnitude of the relationship between size and the MRPs of skilled and unskilled labor decreases significantly. As Hopenhayn (2014) discusses, however, the more important relationship for capturing the impact of distortions on misallocation is the relationship between MRPs and the productivity of firms. Size is often used as a proxy for productivity; however, if distortions are large enough, then highly productive firms may be small in terms of gross output. Similarly, distortions may cause low productivity firms to become large.

To address this concern, we construct similar graphs highlighting the relationship between the MRP for each input against productivity. The productivity of each firm is backed out using our model, during which we normalize the productivity of the smallest firm in each industry to 1. This normalization means that productivity, unlike gross output, is not directly comparable across different industries. To capture the additional importance of firms in larger industries (in our model, the size of each industry is independent from the distortions that firms within the industry face), we weight each firm’s productivity by its industry’s share of gross output in the economy. Finally, we normalize the smallest weighted productivity in the economy to 1. Figure A.1 displays the loess locally weighted regressions between the MRP of each input and weighted productivity. The relationship between MRPs and productivity is qualitatively the same as the relationship between MRPs and gross output. More productive and larger firms face higher MRPs of unskilled and skilled labor, and to a much greater extent if we assume there is only a single cluster in each industry or if clusters are generated from cluster-specific distortions as opposed to cluster-specific production technologies.

A.2 Comparison with Alternative Approaches

In this section, we consider two alternative approaches. First, we consider a CES production function that allows for an elasticity of substitution across inputs that is different from one. We also consider output distortions within our setup. Our main findings do not appear to be the result of production function misspecification or the lack of considering output distortions.
Figure A.1
Marginal Revenue Product of Inputs by Productivity
Single versus Multiple Clusters

Panel A of Figure A.1 shows the loess fitted curve for the MRP of capital by productivity under the assumption of a single cluster and allowing for multiple clusters. To account for differences in industry size, we weight each firm’s productivity by its industry’s and cluster’s share of total gross output in the economy, and then we normalize the smallest weighted productivity in the economy equal to 1. For multiple clusters, we present two curves, one under the assumption that clusters are differentiated solely by production technologies and the other under the assumption that clusters are differentiated solely by cluster-specific distortions. Panel B shows the same plot for intermediate inputs, Panel C shows the same plot for skilled labor, and Panel D shows the same plot for unskilled labor. We use least square smoothing and a bandwidth of 0.8 for the loess locally weighted regression.

A.2.1 Relaxing the Cobb-Douglas Assumption

In our baseline model, we assumed a Cobb-Douglas production function, which has the feature that all variation in cost shares across firms is due to either differences in factor intensities or differences in distortions. In this section, we explore how clustering would work if production functions were CES. We also discuss whether our results may be driven by the use of a Cobb-Douglas production function.

We will again focus on a single industry with $M$ firms. Demand across firms is CES and the same as in Section 2.1. Each firm has a CES production function, in particular, firm $m$ in cluster $h$ has the production function
where $y_m$ is firm $m$’s output, $z_{h,m}$ is its productivity, $x^i_m$ is its quantity of input $i$, $\sigma$ is the elasticity of substitution across inputs, and $\theta^i_h$ is the CES share parameter which varies across clusters and indicates the importance of input $i$ for firms in cluster $h$. Assume that $z_{h,m} = z_h z_m$ so there is both a cluster-specific component to productivity and an idiosyncratic component. Firm $m$ in cluster $h$ faces distortion $\tau^i_{h,m} = \tau_h \tau^i_h \tau^i_m$ for input $i$. Firms maximize profits subject to their production function, as in equation 6, which leads to firms charging a constant markup, $1/\rho$, over their marginal cost. Relative expenditures by firm $m$ in cluster $h$ on inputs $i$ and $j$ are given by

$$
\frac{p^i \tau^i_{h,m} x^i_m}{p^j \tau^j_{h,m} x^j_m} = \left( \frac{\theta^i_h}{\theta^j_h} \right)^{\sigma} \left( \frac{p^i \tau^i_{h,m}}{p^j \tau^j_{h,m}} \right)^{\sigma - 1}.
$$

(52)

This highlights that under a CES production function, relative input expenditures depend on relative CES intensities as well as the relative prices of each good and the elasticity of substitution. This contrasts with our Cobb-Douglas production function specification, where relative input expenditures depend only on relative factor intensities. The unit cost function for firm $m$ in cluster $h$ is given by

$$
1/z_{h,m} \left( \sum_{i=1}^I (\theta^i_h)^{\sigma} (\tau^i_{h,m} p^i)^{1-\sigma} \right)^{1/\sigma},
$$

which allows us to solve for firm $m$’s observed cost share, $\tilde{\theta}^i_m$, for input $i$,

$$
\tilde{\theta}^i_m = \frac{p^i x^i_m}{p_m y_m} = \rho \left( \frac{1}{\tau^i_{h,m}} \right)^{\sigma} \left( \frac{\theta^i_h}{\theta^j_h} \right)^{\sigma} \left( \sum_{i=1}^I (\theta^i_h)^{\sigma} (\tau^i_{h,m} p^i)^{1-\sigma} \right)^{1-\sigma}.
$$

(53)

This expression can be contrasted with that of equation 12. Whereas under a Cobb-Douglas specification, cost shares will vary only due to differences in distortions or factor intensities for a given input, equation 53 shows that distortions for all inputs factor into variation in cost shares in a CES specification. Additionally, the average cost share in an industry will depend on the price of the input and the elasticity of substitution across inputs. We can follow the same process as we do in Section 3 to decompose variation in factor cost shares into a systematic cluster-specific component and an idiosyncratic component. In particular, taking the log of equation 53 and rearranging, we get

$$
\log \tilde{\theta}^i_m = \log \rho - \sigma \left( \log \tau^i_m - \log \tau_h + \log \frac{1}{\tau^i_h} \right) - \log \left( \sum_{i=1}^I (\theta^i_h)^{\sigma} (\tau^i_{h,m} p^i)^{1-\sigma} \right).
$$

(54)

Equation 54 helps to address concerns that misspecification might lead us to incorrectly separate between-cluster variation from idiosyncratic variation in factor cost shares. In particular, Tables 1–3 should be unaffected if our production function is misspecified from not using CES. It is also important to note, however, that the key relationships we uncovered between the marginal products of inputs and firm size cannot be explained by this misspecification. Recent literature, such as
Oberfield and Raval (2014), emphasizes that intermediate inputs are complementary to capital and labor in a CES production function. To better understand whether this potential misspecification is driving our results, suppose that we have a production function in which capital and labor have an elasticity of substitution of one and are nested within a CES production function with intermediate inputs. This would imply that the elasticity of substitutions between intermediate inputs and both capital and labor are less than one. Suppose that we increase the distortions on labor input. In the CES case, this makes it more expensive to use labor, which decreases the labor cost share. Capital and labor are substitutable, and as a result the cost share of capital would increase. Finally, the cost share of intermediate inputs, which is complementary to labor, would decline. These relationships, however, are not consistent with the main patterns that we find in Section 4. The results in Table 4 show that larger plants are assigned to clusters that have high labor cost shares, which matches our example by design. We find, however, that larger plants are assigned to clusters that do not have higher capital cost shares but instead higher intermediate input shares — the opposite of what a CES production function would imply in the absence of cluster-specific differences in cost shares.

A.2.2 Output Distortions

Hsieh and Klenow (2014) work with a Cobb-Douglas production function that has capital and labor as the only inputs. With only data on firm-level factor input expenditures and gross output, it is not possible to separately identify output distortions if there are distortions on each of the factor inputs. Therefore, the literature generally assumes that there are unobserved distortions on only output and capital, leaving labor undistorted. In contrast, we leave output undistorted and place a distortion on each of the inputs. While this choice does not affect our clustering results, it can matter for our counterfactual experiments and when searching for policies that might contribute to measured misallocation.

Suppose we were to incorporate an unobserved output distortion, \( \tau_{ym}^\bullet \), which is interpreted as a tax or subsidy on output. We would rewrite the firm’s problem in equation 6 as

\[
\max \tau_{ym}^\bullet p_m y_m - \sum_{i=1}^{l} \tau_{ih,m}^i p_i^j x_{im}.
\]

Equation 55 shows that the factor input distortion that we previously recovered includes both \( \tau_{im}^i \) and \( \tau_{ym}^\bullet \). Notice, however, that \( \tau_{ym}^\bullet \) affects the recovered distortions across all factor inputs. For example, if \( \tau_{ym}^\bullet < 1 \), meaning that a firm is taxed on its output, then the firm would have a higher MRP across each of its factor inputs. Suppose that there are size-dependent output distortions.
that target large firms. In that case, we would expect the MRP of all inputs to increase with firm size. Figure 3, however, shows that larger plants do not have uniformly high MRPs. Even if output distortions were not correlated with firm size, if output distortions were large, we would expect to see a high positive correlation across the recovered input distortions. We find, however, that the recovered input distortions are weakly correlated within an industry, with correlations that can be negative as well as positive. Table A.1 reports the correlation of the logged MRP of inputs when there is one cluster in each industry.\textsuperscript{31} Table A.2 reports the same statistics if we allow multiple clusters and the clusters are driven by different technologies (i.e., $\tau^i_h = 1$ for all $h$ and $i$). Table A.3 reports the variance of the logged MRP of inputs for these two cases. Overall, the patterns we recover suggest that output distortions are unlikely to be an important source of misallocation or a driver of the patterns we recover in the Chilean data.

Our suggestion that output distortions are unlikely to be important may initially seem at odds with the literature. For example, Hsieh and Klenow (2009) argue for large roles for output distortions. To this end, we point out that their concept of output distortions maps to distortions on factor inputs in our framework. This is because, as mentioned above, it is not possible to identify output distortions separately from a distortion on each input. In Section A.8, we perform an exercise that is equivalent to having output distortions with no distortions on intermediate inputs and we find numbers that are in line with the literature. This finding makes it clear that our framework mechanically finds the same results as other papers; however, a decision needs to be made whether to attribute these results to output distortions or to distortions on factor inputs. We are not aware of any paper that considers both specifications and concludes that output distortions are more consistent with their data compared to having a distortion on each factor input. There are also good theoretical reasons to expect misallocation to be specific to intermediate inputs. For example, there is a large literature following Monahan (1984) on optimal quantity discount pricing for suppliers. Boehm and Oberfield (2020) find misallocation in intermediate input usage due to contract enforcement strength, and Aggarwal, Giera, Jeong, Robinson, and Spearot (2018) find that location affects market access and transportation costs—both of which affect intermediate input cost and usage.

### A.2.3 Robustness to Changes in the Elasticity of Substitution

The elasticity of substitution across output from different firms, $1/(1 - \rho)$, is a key parameter for measuring the gains from eliminating misallocation. In this paper, we estimated the elasticity of substitution separately for each industry by imposing restrictions on the mean of the cluster-specific distortions in equations 14 and 15, and backing out $\rho$ using equation 26, effectively estimating $\rho$.

\textsuperscript{31}If we combine skilled and unskilled labor in the case in which there is one cluster in each industry, then the correlation of the logged MRP of capital and labor increases to 0.22. Furthermore, if we use a value-added production function in which the two inputs are labor and capital, then the correlation of the logged MRP of these inputs increases to 0.32.
Table A.1
CORRELATION OF LOGGED MRP OF INPUTS WITH ONE TECHNOLOGY IN EACH INDUSTRY

<table>
<thead>
<tr>
<th></th>
<th>Unskilled labor</th>
<th>Skilled labor</th>
<th>Capital</th>
<th>Intermediate inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unskilled labor</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skilled labor</td>
<td>0.06</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capital</td>
<td>0.16</td>
<td>0.14</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Intermediate inputs</td>
<td>-0.19</td>
<td>-0.15</td>
<td>-0.08</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table A.1 shows the correlation of the logged MRP of inputs when we assume that there is one cluster in each industry. To construct this table, we use the logged input distortions $\tau_{i,h,m}$ that were constructed as described in Section 3 under the assumption that there is only one cluster. We then subtract the industry mean of these logged input distortions to find the dispersion within the industry. Finally, we calculate the pairwise correlations using all of the plants in the data.

Table A.2
CORRELATION OF LOGGED MRP OF INPUTS WITH MULTIPLE TECHNOLOGIES IN EACH INDUSTRY (i.e., $\tau_h^i = 1$ for all $h$ and $i$)

<table>
<thead>
<tr>
<th></th>
<th>Unskilled labor</th>
<th>Skilled labor</th>
<th>Capital</th>
<th>Intermediate inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unskilled labor</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skilled labor</td>
<td>-0.03</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capital</td>
<td>0.05</td>
<td>0.07</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Intermediate inputs</td>
<td>-0.16</td>
<td>-0.12</td>
<td>-0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table A.2 shows the correlation of the logged MRP of inputs when there are multiple technologies in each industry (i.e., $\tau_h^i = 1$ for all $h$ and $i$). To construct this table, we use the logged input distortions $\tau_{i,h,m}$ that were constructed as described in Section 3 and set $\tau_h^i = 1$. We then subtract the industry mean of these logged input distortions to find the dispersion within the industry. Finally, we calculate the pairwise correlations using all of the plants in the data.

Table A.3
STANDARD DEVIATION OF LOGGED MRP OF INPUTS

<table>
<thead>
<tr>
<th></th>
<th>Multiple technologies and $\tau_h^i = 1$ for all $h$ and $i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unskilled labor</td>
<td>1.04</td>
</tr>
<tr>
<td>Skilled labor</td>
<td>0.91</td>
</tr>
<tr>
<td>Capital</td>
<td>1.22</td>
</tr>
<tr>
<td>Intermediate inputs</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Table A.3 shows the standard deviation of the logged MRP of inputs. See notes for Table A.1 regarding the construction of the logged MRP of inputs with one technology. See notes for Table A.2 regarding the construction of the logged MRP of inputs with multiple technologies.
Table A.4
AGGREGATE GAINS FROM ELIMINATING DISTORTIONS (PERCENTAGES)

<table>
<thead>
<tr>
<th>Case</th>
<th>Default ρ’s</th>
<th>ρ = 0.8</th>
<th>ρ = 0.67</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Multiple Clusters</td>
<td>Single Cluster</td>
<td>Multiple Clusters</td>
</tr>
<tr>
<td>Output</td>
<td>26.8</td>
<td>38.7</td>
<td>30.6</td>
</tr>
<tr>
<td>TFP</td>
<td>24.3</td>
<td>30.9</td>
<td>27.6</td>
</tr>
</tbody>
</table>

Table A.4 shows the aggregate gains in output and TFP from removing distortions with different estimates of ρ, which affect the elasticity of substitution, 1/(1 − ρ). Multiple Clusters reports the counterfactual gains in a framework that allows for cluster-specific production technologies and distortions and Single Cluster under the assumption that there are no cluster-specific differences in production technologies or distortions.

Using the average markup for firms in the industry. To investigate the sensitivity of our results to our estimates of ρ we show how our estimated gains from eliminating misallocation change with different elasticities.

Although there is variation across industries in our estimates of ρ, both the median and mean value of ρ that we estimate are equal to 0.80. To eliminate the impact that variation in ρ across industries has on our results, we set ρ = 0.8 for all industries (an elasticity of substitution equal to 5). This approach allows us to drop our restriction on cluster-specific distortions in equation 14. We back out τₜ for each cluster using equation 27. This leaves τₜ/ρ, and therefore the rest of our estimated parameters, unchanged as we change ρ. We perform an additional robustness check with ρ = 0.67 (an elasticity of substitution equal to 3), which is the value used in Hsieh and Klenow (2009), but smaller than the estimated ρ in each of our industries.

Table A.4 reports the estimated gains from eliminating misallocation using our baseline ρ’s (also reported in Table 7) along with the results when ρ = 0.8 and ρ = 0.67 for all industries. As expected, the gains from eliminating misallocation are smaller when ρ = 0.67 relative to ρ = 0.8 because the former implies a lower elasticity of substitution. The fraction of misallocation explained by allowing for cluster-specific technologies, however, is larger with lower elasticities of substitution. Allowing for cluster-specific technologies reduces the estimated output gains from eliminating misallocation by 36.4 percent (=(20.6−32.4)/32.4) when ρ = 0.67, by 26.6 percent when ρ = 0.8, and by 30.7 percent for our default elasticities. Therefore, while a small elasticity leads to conservative estimates for the gains of eliminating misallocation, it leads to large estimates for the importance of accounting for cluster-specific technologies. Therefore, we view our estimates of ρ in our main analysis as striking a reasonable balance between these two forces.
A.3 Clustering Methodology Details

This section presents a simple example of our clustering methodology and evaluates the performance of our clustering methodology using simulations. Our clustering methodology is composed of two distinct parts. The k-means++ algorithm, which is a standard heuristic for partitioning a set of observations into $k$ distinct groups, and our simulation based approach for determining the number of clusters. Since the k-means++ algorithm is relatively standard, this section focuses on the details for our methodology for determining the number of clusters or groups in a dataset.

A.3.1 A Simple Example of the Clustering Methodology

Here we present a simple example of our clustering methodology applied to an artificial dataset where the only inputs are capital and labor and two technologies are present in the data: a capital-intensive technology and a labor-intensive technology. This simple example is not based on actual data and is meant for illustrative purposes to give a better understanding of the clustering algorithm. Figure A.2 shows the synthetic data used in our example, which we generated using a multivariate lognormal distribution. Note that the cost shares are lognormally distributed within each cluster as discussed in Section 3.2, which is consistent with spherical shape of the data for each cluster. Furthermore, note that the axes of the figures are logged cost shares. Panel A presents the actual data, and Panels B, C, and D present the data when clustered with one, two, and three clusters, respectively. Panel C shows that when we apply the k-means++ algorithm with two clusters, we successfully recover each firm’s true cluster in our example. When we move from one cluster to two clusters, $R_{21}^2 = SSE_2/SSE_1 = 0.063$, and when we move from two clusters to three clusters, $R_{32}^2 = SSE_3/SSE_2 = 0.774$.

Figure A.3 shows the simulated datasets used by our clustering methodology to determine the number of clusters. Panel E shows one of the simulated datasets, which is generated by fitting a single lognormal distribution to Panel B of Figure A.2. Panel F shows the partitioning of the data when we apply the k-means++ algorithm with two clusters to this single lognormal distribution. When we cluster this simulated data with two clusters, we have $R_{21}^{syn} = SSE_{2}^{syn}/SSE_{1}^{syn} = 0.439$. This is a smaller reduction in SSE than in the original data; therefore, according to equation 21 from our clustering methodology section, we conclude that there are two or more technologies in the original data. The original data have a much larger reduction than the simulated data because two clusters are in the original data, and the gap between the two clusters leads to a

Note that when we generate the simulated data, we generate a large number of simulated datasets. This is due only to sample size error, since when there are a small number of observations, the reduction in SSE in the simulated datasets will vary slightly across simulations. As we discuss in Section 3.4, we use a cutoff equal to the top 20th percentile for the reduction in SSE. As the number of observations increases, the reduction in SSE across simulations converges until the cutoff plays no role. The simulated datasets we present in Figure A.3 are the cutoff simulations used in determining the number of clusters.
Panel A of Figure A.2 shows the original data used in our example. These data were generated from two technologies: a capital-intensive technology and a labor-intensive technology. Panel B shows the data when only a single cluster is applied; as we can see, all points are treated the same. Panel C shows the data when the k-means++ algorithm is applied with two clusters. Panel D shows the data when the k-means++ algorithm is applied with three clusters. A large reduction in SSE when moving from one to two clusters. This is what we mean when we say that our clustering methodology effectively works by detecting gaps in the distribution of cost shares. Panel G shows one of the simulated datasets, which is generated by fitting two lognormal distributions to the original data, one for each cluster from Panel C of Figure A.2. Panel H shows the partitioning of the data when we apply the k-means++ algorithm with three clusters to the simulated data generated by two lognormal distributions. When we cluster this simulated data...
with three clusters, we have $R_{3|2}^{syn} \equiv \frac{SSE_{3}^{syn}}{SSE_{2}^{syn}} = 0.659$. Since the reduction in the simulated data is greater than the reduction in the original data, $R_{3|2} > R_{3|2}^{syn}$, we conclude that there are not three or more clusters and therefore there are exactly two clusters in the original data — which is correct.

A.3.2 Simulation Strategy

This section reports the process we use to generate the simulated data used to evaluate our methodology. For each simulation, we generate a single industry which we assign to have one, two, or three production technologies with equal probability. Each production technology is defined as a Cobb-Douglas production function with four inputs and varying factor intensities. We focus on the case in which all systematic variation in cost shares is due to differences in multiple production technologies (i.e., $\tau_{ih} = 1$ for all inputs $i$ and clusters $h$) and try to determine whether our methodology correctly identifies the number of technologies and which firms use each production technology. Although in this exercise we focus on the case in which all systematic distortion is due to different technologies for purposes of illustration, our results would be similar if we had instead simulated data in which differences in cost shares are due to cluster-input-specific distortions ($\tau_{ih}^l$). We find that our methodology performs well in recovering unobserved production technologies when data are generated according to the model in Section 2 and there is no overlap in the clusters. In order to explore the robustness of our methodology we also evaluate our methodology under a variety of sample sizes and methodological choices including those in which the assumption that each firm’s cost share is closest to the center of its cluster as described in Section 3.2 is violated.

In the data, frequently one or two factor intensities are much larger than the others; therefore, we choose the factor intensities in a two-step process. For each production technology, we randomly choose the number of inputs that the technology will be more intensive in, where the technology can be more intensive in one, two, or three inputs, chosen with equal probability. Conditional on a given production technology being more intensive in a specific number of inputs, we randomly assign which factor inputs are the more intensive inputs with equal probability, and we assign the remaining factor inputs as less intensive. In order to ensure that the production technologies are reasonably distinct and thereby satisfy the condition that each firm’s cost share is closest to the center of its cluster, we require that no two production technologies in the industry are identical in terms of what their more-intensive and less-intensive factor inputs are. For factors classified as more intensive, we assign them a preliminary value drawn from a uniform distribution with support of [0.20, 1.00], whereas factors classified as less intensive have a preliminary value drawn from a uniform distribution with support of [0.01,0.10]. This leads to a median ratio between the largest and smallest factor intensities of 22, compared to 23 in the data. We then calculate the factor intensities, $\theta_{ih}$, for each production technology, $h$, by normalizing the preliminary values to sum to 1.
Panel E of Figure A.3 shows one of the simulated datasets where we fit and generate a single multivariate lognormal distribution to the data from Figure A.3. Panel F shows the partitioning of the simulated data when we apply the k-means++ algorithm with two clusters; note that clustering the simulated data with two clusters does not reduce SSE as much as in the original data. This suggests there are two or more clusters in the original data. Panel G shows one of the simulated datasets where we fit and generate a two multivariate lognormal distribution to the data from Figure A.3. Panel H shows the partitioning of the simulated data when we apply the k-means++ algorithm with three clusters. Clustering with three clusters reduces SSE around the same and slightly more in the simulated data than in the original data. This suggests there are not three or more clusters in the original data; leading us to conclude that there are exactly two clusters, which is correct in this example.

For each production technology, we assign a random number of firms to use the technology, where we use a uniform distribution to generate between 30 and 70 firms for each technology. This
### Table A.5

**Selection of Parameters for Monte Carlo Exercise**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Generating Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of production technologies ($H$)</td>
<td>Uniform: 1, 2, and 3</td>
</tr>
<tr>
<td>Number of inputs ($I$)</td>
<td>4</td>
</tr>
<tr>
<td>Factor intensity for input $i$ for production technology $h$ ($\theta^h_i$)</td>
<td>Uniform distribution over $[0.01,0.1]$ and $[0.2,1]$ with equal probability (rescaled to sum to 1)</td>
</tr>
<tr>
<td>Number of firms assigned to production technology $h$</td>
<td>Uniform: [30,70]</td>
</tr>
<tr>
<td>Standard deviation of lognormal distribution of input $i$ distortion</td>
<td>Uniform: [0.1,0.3]</td>
</tr>
<tr>
<td>Technology-specific distortion ($\tau^h$)</td>
<td>Uniform $[0.5,1.5]$ (rescaled to mean of 1)</td>
</tr>
<tr>
<td>Parameter that governs elasticity of substitution ($\rho$)</td>
<td>Markup $(1/\rho)$ is uniform: [1.01,1.50]</td>
</tr>
</tbody>
</table>

Table A.5 summarizes the data-generating process for the parameters of the Monte Carlo exercise. Column 1 shows the parameter of interest. Column 2 describes the generating process for that parameter.

The process ensures that we have enough firms to apply our methodology for each simulation. For each firm, $m$, we generate the distortions, $\tau^m_i$, for each input, $i$, drawn from a multivariate lognormal distribution. We set the parameters so the mean of the logged distribution is zero for all inputs, and the standard deviations are chosen from a uniform distribution with support [0.1, 0.3]. Using a multivariate distribution allows for the possibility of correlation between the distortions. In the base scenario, we assume the distortions are independently distributed; however, we evaluate correlated distortions in Table A.9. The average correlation in the data is close to zero as reported in Section A.2.3; however, the standard deviations tend to be larger than the standard deviations we use in our simulations. We chose the standard deviations to be smaller to ensure that each firm’s cost share is closest to the center of its cluster; however, we later explore the robustness of our results to generating distortions from distributions with larger standard deviations.

We also require the elasticity of substitution and the technology-specific distortions. The median markup in the data is 1.25, with a range between 1.0 and 1.5. Therefore, we select the elasticity of substitutions so that markups are uniformly distributed across simulations, ranging between a 1 percent markup and a 50 percent markup. The technology-specific distortion is chosen by drawing a preliminary value from a standard uniform distribution for each production technology and then normalizing the preliminary values so that the mean value is 1. Because our methodology only requires cost shares, and not revenues and expenditures separately, other parameters such as the firm-level productivity, $z_m$, of each firm are irrelevant for our evaluation.

Table A.5 summarizes the parameters used for the Monte Carlo exercise.
Table A.6

**TRUE VERSUS RECOVERED NUMBER OF PRODUCTION TECHNOLOGIES**

<table>
<thead>
<tr>
<th>True Number of Technologies</th>
<th>Number of Technologies Recovered</th>
<th>Percentage Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
<td>100.0</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table A.6 shows the true number of technologies versus the number of production technologies recovered in the Monte Carlo exercise. Column 1 describes the true number of production technologies in the simulated data. Columns 2-4 describe the number of simulations that recovered 1, 2, and 3 technologies, respectively. Column 5 shows the percentage of simulations in which we recovered the correct number of technologies.

Table A.7

**PERCENTAGE OF FIRMS CORRECTLY CLASSIFIED BY TECHNOLOGY**

<table>
<thead>
<tr>
<th>True Number of Technologies</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Percentage</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Table A.7 shows the percentage of firms that were correctly classified by their technology use in the Monte Carlo exercise. The columns labeled 1, 2, and 3, indicate the number of production technologies in the simulated data, while the column Overall indicates the average across all simulations.

A.3.3 Simulation Results

After generating the synthetic data, we run all of the steps of the clustering methodology described in Section 3. Table A.6 compares the true number of technologies (in the row) with the number of technologies we recover (in the column) in each simulation. We successfully recover the true number of technologies 100 percent of the time regardless of whether there are 1, 2, or 3 technologies. Table A.7 shows the percentage of firms that we correctly classify by technology, which is also 100 percent.

Lastly, we evaluate how our methodology performs at recovering the other underlying parameters. We again focus on the cases in which we correctly recover the true number of technologies. We construct the average absolute percentage error as follows:

\[
\text{Average Absolute Percentage Error} \equiv \left| \frac{\Omega_{\text{true}} \Omega_{\text{recovered}}}{\Omega_{\text{true}}} \right|.
\]

The results for the factor intensities, distortions, and markups are reported in Table A.8. For all variables, the average absolute percentage error is less than 10 percent, indicating a high degree of accuracy in parameter recovery. For example, the value for factor intensities is around 3 percent, indicating that if the true factor intensity is 0.35, our results suggest that the recovered intensity
Table A.8
AVERAGE ABSOLUTE PERCENT ERROR BETWEEN TRUE AND RECOVERED PARAMETERS

<table>
<thead>
<tr>
<th>True Number of Technologies</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor Intensities ($\theta_{ih}$)</td>
<td>2.7</td>
<td>2.8</td>
<td>2.7</td>
<td>2.7</td>
</tr>
<tr>
<td>Idiosyncratic Distortions ($\tau_{im}$)</td>
<td>4.6</td>
<td>4.6</td>
<td>4.4</td>
<td>4.5</td>
</tr>
<tr>
<td>Tech-Specific Distortions ($\tau_{ih}$)</td>
<td>0.0</td>
<td>4.9</td>
<td>6.5</td>
<td>3.8</td>
</tr>
<tr>
<td>markup</td>
<td>4.2</td>
<td>8.0</td>
<td>9.5</td>
<td>7.2</td>
</tr>
</tbody>
</table>

Table A.8 shows the average absolute percentage error (as described in equation 57) for factor intensity, distortions, and markups in the Monte Carlo exercise. Column 1 reports the variable of interest. Columns 2-4 indicate the true number of production technologies in the simulated data.

will likely be between 0.34 and 0.36, centered around 0.35. If we allow positive errors to offset negative errors and only compute the average percentage error without taking absolute values, then the errors are effectively zero, meaning that our recovered parameters are centered around the true parameters.

In order to explore the robustness of our methodology we report the success rate of our methodology under two additional scenarios. Table A.9 reports the percentage of simulations in which we correctly recover the true number of production technologies for each scenario. *Base* refers to our base case scenario that we describe in Section A.3.2. The other columns of the table report deviations from the base scenario as follows. *Correlated Distortions* refers to the case in which we allow the distortions to be correlated across factor inputs, where the correlations are drawn randomly from a uniform distribution with support $[-0.2, 0.2]$. *Overlapping Clusters* refers to the case in which we increase the standard deviation of the distortions — we generate them from a uniform distribution with support $[0.3, 0.7]$ compared to the base scenario support of $[0.1, 0.3]$ — which leads to the assumption that each firm’s cost share is closest to the center of its cluster being violated for some observations. Under both scenarios, we find that our methodology performs well overall, finding the true number of technologies in 99.3 percent of simulations with correlated distortions and 93.7 percent of simulations with overlapping clusters. When the clusters overlap, however, we find we are more likely to understate the true number of clusters compared to the base scenario. This result indicates the conservative nature of our methodology and the inherent difficulty of disentangling closely related technologies.
### Table A.9

**Robustness of Methodology in Alternative Scenarios**

<table>
<thead>
<tr>
<th>Number of Technologies</th>
<th>True Number of Technologies</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base Scenario</td>
<td>100.0 100.0 100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Correlated Distortions</td>
<td>99.0 100.0 100.0</td>
<td>99.7</td>
</tr>
<tr>
<td>Overlapping Clusters</td>
<td>100.0 100.0 81.0</td>
<td>93.7</td>
</tr>
</tbody>
</table>

Table A.9 reports the success rate of our methodology — the percentage of simulations in which we successfully uncover the true number of production technologies — under three scenarios. The first column reports the scenario we are using to generate our simulations. Columns 2-4 report the success rates for simulations for each number of production technologies. The column *Overall* indicates the success rate across all simulations.
A.4 Distinguishing Vertically Integrated Steel Plants from Plants Using Electric Furnaces in India

We apply our methodology to data from Indian steel producers. To assess whether the clusters identified by our methodology correlate with technology, we can use a methodology similar to Collard-Wexler and De Loecker (2015) to distinguish between two types of steel plants—vertically integrated steel plants and those that use electric furnaces. Being able to identify these two different types of steel plants in the data allows us to study whether the clustering methodology can sort plants by the type of technology utilized to produce steel.

India is a major producer of steel—the country ranked as the third largest producer of steel in the world in 2016, behind only China and Japan. Two main types of steel plants operate in India.\textsuperscript{33} First, are the integrated steel producers. An integrated steel producer manufactures steel starting from the raw materials such as iron ore. The plant starts by combining iron ore and other inputs to make molten iron and then uses basic oxygen furnaces to convert the molten iron into molten steel. Finally, they use continuous casting and rolling plants to produce solid steel products. The second type of plant is composed of steel producers that use electric furnaces (instead of the basic oxygen furnace), which use a combination of scrap steel and direct reduction iron (instead of iron ore) as their main inputs.\textsuperscript{34}

For our analysis, we use plant-level data for the steel industry from the Annual Survey of Industries (ASI) for the years 2005-2006. The ASI covers plants that have more than 10 workers if they have electricity and 20 if they do not. The information is similar to that found in plant-level datasets and includes information such as sales, employment, wage bill, capital stock, and intermediate goods usage. The dataset also contains product-level data on intermediate input usage, which we use in our analysis to distinguish between vertically integrated plants and those that use electric furnaces. We construct the same variables as for the Chilean manufacturing data, with the exception of labor; the ASI data are not detailed enough for us to separate skilled labor from unskilled labor.

\textsuperscript{33} For an overview of steel production in India and government policy, see Dasgupta (2017), Ministry of Steel (2005), and Ministry of Steel (2017). See Rogers (2009) for more information on the structure of the U.S. steel industry.

\textsuperscript{34} Steelmakers in India use one of two types of electric furnaces—electric arc furnaces that are used by minimills and the induction furnace, which is used almost exclusively in India in the mass production of steel. The main difference between the two furnaces is that they rely on different mixes of scrap steel and direct reduction iron, a distinction that does not play an important role for the purposes of our analysis. In addition to allowing for a greater reliance on direct reduction iron, induction furnaces allow for a much smaller scale of operation relative to electric arc furnaces. For example, estimates from the Ministry of Steel indicate that the average plant using an electric arc furnace has an annual capacity of 230,000 tons of steel, whereas the average annual capacity of plants that use an induction furnace is only 17,000 tons of steel in 2005-2006 (786 plants use an induction furnace with a total capacity of 13.22 million tons per year, and 38 plants use electric arc furnaces with a total capacity of 8.73 million tons per year). See page 28 of Ministry of Steel (2007) for more information.
We apply criteria similar to Collard-Wexler and De Loecker (2015) to distinguish between vertically integrated steel plants and those that use electric furnaces. First, we categorize a plant as using an electric furnace if it reports scrap steel as being more than 20 percent of the total cost of materials or if a plant reports using direct reduction iron.\textsuperscript{35} Second, we categorize a plant as vertically integrated if it reports using iron ore as an input.\textsuperscript{36} Third, to be considered either a vertically integrated plant or a plant that uses an electric furnace, the plant needs to produce either finished or semi-finished steel.\textsuperscript{37} If a plant is flagged as being both vertically integrated and one that uses an electric furnace, we follow Collard-Wexler and De Loecker (2015) and designate it as being vertically integrated. For example, vertically integrated plants are engaged in many activities, including in some cases as having both an electric arc furnace and a basic oxygen furnace.

The characteristics of the plants that are classified as vertically integrated or using electric furnaces are consistent with the characteristics of their production technology. For example, we find that vertically integrated plants have a gross output premium of 3.88, meaning that when we regress the logged gross output on a dummy that indicates whether a plant is vertically integrated, the coefficient is 3.88. In addition, vertically integrated plants have a capital cost share premium of 1.41, where capital cost is the ratio of the user cost of capital and gross output. Thus, vertically integrated plants are larger and operate a capital-intensive technology, which is consistent with the structure of the steel industry, as discussed in footnote 25. We also find that vertically integrated plants have an intermediate input cost share premium of -0.25, which is also consistent with the structure of the steel industry.\textsuperscript{38} As a final check on the robustness of the results, we find that the electricity cost as a fraction of the total intermediate input usage premium for vertically integrated plants is -0.74. Thus, plants utilizing electric furnaces use electricity more intensively than vertically integrated plants.

We apply the clustering methodology on the industry that manufactures basic iron and steel, which is industry code 2710 in the NIC 04 classification system. Note that this industry includes

\textsuperscript{35}The criteria for identifying plants that use electric furnaces are identical to those used by Collard-Wexler and De Loecker (2015). It is important to note that Indian steel producers, unlike those in the United States, use two types of electric furnaces: electric arc furnaces (minimills) and induction furnaces. The same criteria can still be applied, however, since both types of plants use scrap steel and direct reduction iron.

\textsuperscript{36}The criteria used by Collard-Wexler and De Loecker (2015) could not be implemented because the product categories used in the U.S. data are different from those used by the ASI. For example, Collard-Wexler and De Loecker (2015) identify a plant as being vertically integrated if it reports producing “Coke Oven or Blast Furnace Products,” which is not a product category in the ASI data.

\textsuperscript{37}These plants are classified under the following industry subcategories of NIC 2720: NIC industry codes 2714 manufacture of semi-finished products (ingots/ billets/ blooms/ slabs); 2715 manufacture of non-flat steel products (bars, rods, structural, rails); 2716 manufacture of flat steel products not coated; and 2717 manufacture of flat products, coated with zinc, tin, chromium, or other materials.

\textsuperscript{38}Cost models by economists from Steelonthenet.com, a company that specializes in providing industry intelligence to participants in the steel industry, find that plants using basic oxygen furnaces, such as vertically integrated plants, tend to have a lower intermediate input cost share relative to plants using electric arc furnaces.
Table A.10 reports $\theta^i_{i,h}/\tau^i_{h}$ for inputs using data from plants in the manufactures of basic iron and steel (NIC 2720) in the ASI Indian plant-level data described in Section A.4. To arrive at these figures, we apply the methodology described in Section 3; $\theta^i_{i,h}/\tau^i_{h}$ is characterized as in equation 28.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Labor</th>
<th>Capital</th>
<th>Intermediate inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.02</td>
<td>0.01</td>
<td>0.97</td>
</tr>
<tr>
<td>2</td>
<td>0.07</td>
<td>0.13</td>
<td>0.80</td>
</tr>
</tbody>
</table>

The two types of steel plants previously identified and additionally has other types of plants (e.g., rolling mills that convert steel into final pieces such as wires or tubes). We are motivated to apply the clustering algorithm on the whole industry because the NIC 04 classification system used in the Indian data is similar to the ISIC Rev 3 classification system used in the Chilean data. Thus, our goal is to see whether the clustering algorithm sorts these two types of plants into clusters that we would consider to be reasonable given what we know about these two technologies.

The clustering methodology recovers two clusters; the recovered $\theta^i_{i,h}/\tau^i_{h}$ for each cluster are reported in Table A.10. Cluster 2 uses capital more intensively relative to cluster 1. Furthermore, cluster 1 uses intermediate inputs more intensively. The algorithm performs reasonably well in sorting the two different types of steel plants into two separate clusters—95 percent of vertically integrated plants are assigned to cluster 2, and 69 percent of plants using electric furnaces are assigned to cluster 1. It is useful to note that the NIC 2720 industry has producers of non-steel products as well, which are included in the clustering analysis (e.g., NIC 2720 has producers of intermediate inputs used by the industry, such as direct reduction iron). Thus, vertically integrated plants and plants that use electric furnaces are assigned to the cluster that most closely resembles their cost shares. Furthermore, given the heterogeneity of products within the industry, it is likely that there are other clusters that we cannot identify based only on variation in cost shares.
A.5 Alternative Assumptions

In Section A.5.1, we study in more detail the implications of the conditions on distortions described in equations 14-16 and equation 32. We show how the inferred parameters would change if we relaxed these conditions on the distortions.

A.5.1 Relaxing Conditions on the Distribution of Distortions

Equations 26-29 In Theorem 1, we characterize equations 26-29 without imposing the conditions in equations 14-16. We will use these characterizations in the subsequent derivations.

Theorem 1. The equations that characterize parameters $\rho$, $\tau_h$, $\theta_i^h$, and $\tau_i^m$ in equations 26-29 without the use of conditions in equations 14-16 are

$$
\rho = \frac{\sum_{h=1}^{H} \sum_{i=1}^{I} \sum_{m=1}^{M} \mathbb{1}_{h(m)=h} \theta_i^m}{\sum_{h=1}^{H} \left( (\tau_h)^{-1} \sum_{i=1}^{I} \left[ \frac{1}{\tau_h} \sum_{m=1}^{M} \mathbb{1}_{h(m)=h} \left( \frac{(\tau_i^m)^{-1}}{M_h} \right) \right] \right)}
$$

(58)

$$
\tau_h = \rho \left[ \frac{\sum_{i=1}^{I} \sum_{m=1}^{M} \mathbb{1}_{h(m)=h} \theta_i^m}{\sum_{i=1}^{I} \sum_{m=1}^{M} \mathbb{1}_{h(m)=h} \theta_i^m} \right] M_h
$$

(59)

$$
\frac{\theta_i^h}{\tau_h} = \frac{\tau_h}{\rho} \left( \sum_{m=1}^{M} \mathbb{1}_{h(m)=h} \tilde{\theta}_m \right) \left( \sum_{m=1}^{M} \mathbb{1}_{h(m)=h} \left( \frac{\tau_i^m}{M_h} \right)^{-1} \right)^{-1}
$$

(60)

$$
\frac{\tau_i^m}{\tau_h} = \frac{\rho \theta_i^h}{\tau_h \theta_i^m}
$$

(61)

Proof. See Appendix A.6 for a detailed proof.

We next apply these equations to the two cases that we study in Section 5 to see how it would affect the parameters that we infer.

Case in Which Differences in Clusters Are Entirely a Result of Differences in Technology We now study the case in which differences in clusters are driven by technologies (i.e., $\tau_i^h = 1$ for all $h$ and $i$), which we study in Section 5. In Theorem 2, we characterize the parameters in Theorem 1 under this case. In Theorem 3, we study the implications of relaxing the conditions laid out in equations 14-16.

Theorem 2. Suppose that $\tau_i^h = 1$ for all $h$ and $i$. Then the expressions for $\rho$, $\tau_h$, $\theta_i^h$, and $\tau_i^m$ from Theorem 1 would become

$$
\rho = \frac{\sum_{h=1}^{H} \sum_{i=1}^{I} \sum_{m=1}^{M} \mathbb{1}_{h(m)=h} \theta_i^m}{\sum_{h=1}^{H} \left( (\tau_h)^{-1} \sum_{i=1}^{I} \left[ \frac{1}{\tau_h} \sum_{m=1}^{M} \mathbb{1}_{h(m)=h} \left( \frac{(\tau_i^m)^{-1}}{M_h} \right) \right] \right)}
$$

(62)
\[
\tau_h = \rho \frac{\sum_{i=1}^I \left[ \theta_h \frac{\sum_{m=1}^M \left( I_{h(m)=h} \left( \tau_{i,m}^i \right)^{-1} \right)}{M_h} \right]}{\sum_{i=1}^I \frac{\sum_{m=1}^M I_{h(m)=h} \theta_{i,m}^i}{M_h}} \tag{63}
\]

\[
\theta_h^i = \frac{\tau_h}{\rho} \left( \sum_{m=1}^M I_{h(m)=h} \hat{\theta}_{i,m}^i \right) \left( \frac{\sum_{m=1}^M I_{h(m)=h} \left( \tau_{i,m}^i \right)^{-1}}{M_h} \right)^{-1} \tag{64}
\]

\[
\tau_{i,m}^i = \frac{\rho \theta_{i,m}^i}{\tau_h \hat{\theta}_{i,m}^i} \tag{65}
\]

**Proof.** We substitute \( \tau_h^i = 1 \) for all \( h \) and \( i \) into equations 58-61.

Given the conditions in Theorem 2, we can now study how these parameters would change if we relax the assumptions in equations 14 and 16. Note that when clusters are entirely driven by different technologies, since \( \tau_h^i = 1 \), equation 15 implies that \( \sum_{i=1}^I \theta_h^i = 1 \), which is consistent with constant returns to scale production technology.

**Theorem 3.** Suppose that we altered equations 14 and 16 to be

\[
\frac{\sum_{m=1}^M I_{h(m)=h} \left( \tau_{i,m}^i \right)^{-1}}{M_h} = \gamma, \forall h \in \{1, \ldots, H\}, \forall i \in \{1, \ldots, I\}, \tag{66}
\]

\[
\frac{1}{H} \sum_{h=1}^H (\hat{\tau}_h)^{-1} = \eta, \forall i \in \{1, \ldots, I\}, \tag{67}
\]

where \( \hat{\tau}_h \) and \( \hat{\tau}_h^i \) are the parameters that we would recover under these new conditions. Note that equations 66 and 67 imply harmonic means of \( 1/\gamma \) and \( 1/\eta \), respectively.

Then the new parameters \( \hat{\rho}, \tau_{h,m}^i, \) and \( \hat{\theta}_h^i \) would be

\[
\hat{\rho} = \frac{1}{\eta \gamma} \rho, \tag{68}
\]

\[
\hat{\tau}_{h,m}^i = \frac{1}{\eta \gamma} \tau_{h,m}^i, \tag{69}
\]

\[
\hat{\theta}_h^i = \theta_h^i, \tag{70}
\]

where \( \rho, \theta_h^i, \) and \( \tau_{h,m}^i \) are the parameters from the baseline specification. Furthermore, the input distortion \( \tau_{h,m}^i \) consists of the following two components:

\[
\hat{\tau}_h = \frac{1}{\eta} \tau_h, \tag{71}
\]

\[
\hat{\tau}_m^i = \frac{1}{\gamma} \tau_m^i, \tag{72}
\]

where \( \tau_h \) and \( \tau_m^i \) are the parameters from the baseline specification.

**Proof.** See Section A.6 for a detailed proof. 

69
Theorem 3 shows that raising the harmonic mean in equation 67 will result in a proportional increase in \( \tau_{i,h,m} \) and also imply a lower markup (higher \( \rho \)). Given the proportional increase in \( \tau_{i,h,m} \) for all firms, changes in \( \eta \) will not affect the dispersion in the MRP of inputs, as described in equation 34. A second important point is that raising the harmonic mean in equation 66 affects the industry in the same manner as the harmonic mean in equation 67. Thus, having an industry in which the harmonic mean of the idiosyncratic distortions is higher is equivalent to an industry in which the harmonic mean of \( \tau_h \) is higher. In that sense, one of these two conditions is a normalization. Third, note that \( \theta^i_h \) does not depend on either \( \eta \) or \( \gamma \).

Case in Which Differences in Clusters Are a Result of Cluster-Input-Specific Distortions (\( \tau^i_h \)) We now study the case in which differences in clusters are driven by cluster-input-specific distortions, \( \tau^i_h \), which we study in Section 5. In Theorem 4, we characterize the parameters in Theorem 1 under this case. In Theorem 5, we study the implications of relaxing the conditions laid out in equations 14-16.

**Theorem 4.** Suppose that all differences in clusters are driven by cluster-input-specific distortions, \( \tau^i_h \), which implies that there is only one technology in each industry, or \( \theta^i = \theta^i_h \) for all \( h \). Then the expressions for \( \rho \), \( \tau_h \), \( \theta^i \), \( \tau^i_h \), and \( \tau^i_m \) from Theorem 1 would become

\[
\rho = \frac{\sum_{h=1}^{H} \sum_{i=1}^{I} \sum_{m=1}^{M} \frac{1}{\tau^{i,m}} \tilde{\theta}_m}{\sum_{h=1}^{H} \left[ \left( \tau_h \right)^{-1} \sum_{i=1}^{I} \frac{1}{\tau^i_h} \sum_{m=1}^{M} \frac{\tilde{I}_{h(m)=h}^i (\tau^i_m)^{-1}}{M_h} \right]}
\]

(73)

\[
\tau_h = \rho \left[ \sum_{i=1}^{I} \frac{\sum_{m=1}^{M} \tilde{I}_{h(m)=h}^i (\tau^i_m)^{-1}}{M_h} \right]^{-1}
\]

(74)

\[
\theta^i = \left( \sum_{h=1}^{H} \frac{1}{\tau^i_h} \right)^{-1} \sum_{h=1}^{H} \left[ \frac{\sum_{m=1}^{M} \tilde{I}_{h(m)=h}^i (\tau^i_m)^{-1}}{\rho \sum_{m=1}^{M} \tilde{I}_{h(m)=h}^i \theta^i_m} \right]
\]

(75)

\[
\tau^i_h = \frac{\rho \tilde{\tau^i_h}}{\tau^i_h}
\]

(76)

\[
\tau^i_m = \frac{\rho \tilde{\tau^i_m}}{\tau^i_h}
\]

(77)

**Proof.** Equations 73, 74, and 77 are the same as those in equations 58, 59, and 61, respectively. To derive equation 75, we set \( \theta^i_h = \theta^i \) in equation 60. We then sum this condition over \( h \) and solve for \( \theta^i \). To solve for the expression in equation 76, we rearrange the expression in equation 60 to solve for \( \tau^i_h \).

**Theorem 5.** Suppose that we altered equations 14-16 and 32 to be

\[
\sum_{m=1}^{M} \tilde{I}_{h(m)=h} (\tau^{i}_m)^{-1} = \gamma, \forall h \in \{1, \ldots, H\}, \forall i \in \{1, \ldots, I\},
\]

(78)
\[ \frac{1}{H} \sum_{h=1}^{H} (\hat{\tau}_h)^{-1} = \eta, \forall i \in \{1, \ldots, I\} \] (79)

\[ \sum_{i=1}^{I} \frac{1}{\hat{\tau}_i} = \kappa, \] (80)

\[ \frac{1}{H} \sum_{h=1}^{H} \frac{1}{\hat{\tau}_h} = \kappa, \] (81)

respectively, where \( \hat{\tau}_m, \hat{\tau}_h, \hat{\tau}_i, \) and \( \hat{\theta}^i \) are the parameters that we would recover under these new conditions. Note that the conditions in equations 78 and 79 imply harmonic means of \( 1/\gamma \) and \( 1/\eta \), respectively; likewise, the conditions in equations 80 and 81 both imply a harmonic mean of \( 1/\kappa \).

Then the new parameters \( \hat{\rho}, \hat{\tau}_{h,m}, \) and \( \hat{\theta}^i \) would be

\[ \hat{\rho} = \frac{1}{\gamma \kappa \eta} \rho, \] (82)

\[ \hat{\tau}_{h,m} = \frac{1}{\eta \kappa} \tau_{h,m}, \] (83)

\[ \hat{\theta}^i = \theta^i, \] (84)

where \( \rho, \theta^i, \) and \( \tau_{h,m} \) are the parameters from the baseline specification. Furthermore, the input distortion \( \hat{\tau}_{h,m} \) is broken down into the following two components:

\[ \hat{\tau}_m = \frac{1}{\gamma} \tau_{m}, \] (85)

\[ \hat{\tau}_h = \frac{1}{\eta} \tau_{h}, \] (86)

\[ \hat{\tau}_i = \frac{1}{\kappa} \tau_{i}, \] (87)

where \( \tau_{m}, \tau_{h}, \tau_{i} \) are the parameters from the baseline specification.

**Proof.** See Appendix A.6 for a detailed proof. \( \square \)

The results in Theorem 5 are similar to those in Theorem 3. One difference is that now the weighted harmonic mean of \( \tau_i \) in cluster \( h \) is equal to \( 1/\kappa \). Raising \( 1/\kappa \) results in a proportionate increase in \( \tau_{h,m} \) for all firms, implies a higher \( \rho \), and leaves \( \theta^i \) unchanged.\(^{39}\)

\(^{39}\)Note that constant returns to scale implies that the right-hand sides of equations 80 and 81 are the same. To see why this is the case, sum both sides of equation 80 over \( h \) to find that \( \sum_{h=1}^{H} \sum_{i=1}^{I} \frac{1}{\hat{\tau}_i} = \sum_{h=1}^{H} \kappa \), which implies that \( \sum_{i=1}^{I} \theta^i \sum_{h=1}^{H} \frac{1}{\hat{\tau}_h} = H \kappa \). This condition will hold only if the right-hand side of equation 81 is \( \kappa \) and constant returns to scale holds.
A.5.2 Constant Returns to Scale

We now analyze the case in which we relax the assumption of constant returns to scale (i.e., we allow for $\sum_{i=1}^{I} \theta_{h}^{i} = R$). It is useful to note that equation 13 would remain the same as we can rederive this equation under the new condition. Furthermore, we now show that the differences in the returns to scale will be reflected in a different estimate for $\rho$.

**Theorem 6.** Consider the case in which equations 58-61 characterize the parameters $\rho$, $\tau_{h}$, $\theta_{h}^{i}/\tau_{h}^{i}$, and $\tau_{m}^{i}$. Suppose that we altered equations 14-16 to be

$$\sum_{m=1}^{M} \mathbb{I}_{h(m)=h} \left( \hat{\tau}_{m}^{i} \right)^{-1} M_{h} = 1, \forall h \in \{1, \ldots, H\}, \forall i \in \{1, \ldots, I\},$$

$$\frac{1}{H} \sum_{h=1}^{H} (\hat{\tau}_{h})^{-1} = 1, \forall i \in \{1, \ldots, I\},$$

$$\sum_{i=1}^{I} \frac{1}{\hat{\tau}_{h}^{i}} = R,$$

respectively, where $\hat{\tau}_{m}^{i}$, $\hat{\tau}_{h}$, $\hat{\tau}_{h}^{i}$, and $\hat{\theta}_{h}^{i}$ are the parameters that we would recover under these new conditions. Furthermore, suppose that we modify the constant returns to scale condition to be

$$\sum_{i=1}^{I} \hat{\theta}_{h}^{i} = R.$$

Then the new $\hat{\rho}$ would be

$$\hat{\rho} = \frac{1}{R} \rho,$$

where $\rho$ is the parameter from the baseline specification. Furthermore, the new distortions ($\hat{\tau}_{h}$, $\hat{\tau}_{m}^{i}$, and $\hat{\tau}_{h}^{i}$) and technology parameters ($\hat{\theta}_{h}^{i}$) are as follows:

$$\hat{\tau}_{h} = \tau_{h},$$

$$\frac{\hat{\theta}_{h}^{i}}{\tau_{h}^{i}} = \frac{\theta_{h}^{i}}{\tau_{h}^{i}},$$

$$\hat{\tau}_{m}^{i} = \tau_{m}^{i},$$

where $\tau_{h}$, $\tau_{m}^{i}$, $\tau_{h}^{i}$, and $\theta_{h}^{i}$ are the parameters from the baseline specification.

**Proof.** See Appendix A.6 for a detailed proof. \qed
A.6 Proofs

A.6.1 Proof of Theorem 1

Proof. The proof laid out here closely resembles the steps taken in Section 3.5 except that we carry out the derivations without using the conditions in equations 14-16.

As a first step, we derive the general form of equation 24. We sum equation 12 for a given input $i$ across all firms belonging to cluster $h$ to find the condition that corresponds to equation 23:

$$\sum_{m=1}^{M} \mathbb{I}_{h(m)=h} \tilde{\theta}_m^i = \rho \left( \tau_h \right)^{-1} \theta_h^i \sum_{m=1}^{M} \mathbb{I}_{h(m)=h} \left( \tau_m^i \right)^{-1}. \quad (96)$$

We next sum equation 96 across all inputs for firms that belong to cluster $h$ to find the condition

$$\sum_{i=1}^{I} \sum_{m=1}^{M} \mathbb{I}_{h(m)=h} \tilde{\theta}_m^i = \rho \left( \tau_h \right)^{-1} \sum_{i=1}^{I} \left[ \frac{1}{\tau_h} \sum_{m=1}^{M} \mathbb{I}_{h(m)=h} \left( \tau_m^i \right)^{-1} \right]. \quad (97)$$

This condition corresponds to equation 24.

To find the expression for $\rho$ in equation 58, we sum equation 97 over all clusters and solve for $\rho$. To solve for $\tau_h$ in equation 59, we use equation 97 and rearrange it to solve for $\tau_h$. To find $\theta_h^i/\tau_h^i$ in equation 60, we rearrange equation 96 to solve for $\theta_h^i/\tau_h^i$. To find $\tau_m^i$ in equation 61, we rearrange equation 12.

\[\blacksquare\]

A.6.2 Proof of Theorem 3

Proof. To find the condition for $\hat{\rho}$, we plug in equations 66 and 67 into the analogous condition of equation 62 to find

$$\hat{\rho} = \frac{1}{\eta \gamma} \frac{1}{H} \sum_{h=1}^{H} \sum_{i=1}^{I} \sum_{m=1}^{M} \mathbb{I}_{h(m)=h} \hat{\theta}_m^i \frac{M_h}{M_h}. \quad (98)$$

Note that this condition is the same as $\hat{\rho} = 1 / (\eta \gamma) \rho$ where $\rho$ is the value from the baseline specification found in equation 26.

To find the expression for $\hat{\tau}_h$, we plug in equations 66, 67, and 68 into the analogous condition of equation 63 to find

$$\hat{\tau}_h = \frac{1}{\eta} \rho \left( \sum_{i=1}^{I} \sum_{m=1}^{M} \mathbb{I}_{h(m)=h} \hat{\theta}_m^i \frac{M_h}{M_h} \right)^{-1}. \quad (99)$$

Note that this condition is the same as $\hat{\tau}_h = (1/\eta) \tau_h$ where $\tau_h$ is the value from the baseline specification found in equation 27.

To find the expression for $\hat{\theta}_h^i$, we plug in equations 66, 68, and 71 into the analogous condition of equation 64 to find

$$\hat{\theta}_h^i = \frac{1}{H} \sum_{h=1}^{H} \frac{\tau_h \sum_{m=1}^{M} \mathbb{I}_{h(m)=h} \hat{\theta}_m^i}{\rho \frac{M_h}{M_h}}. \quad (100)$$

Note that this condition is the same as $\hat{\theta}_h^i = \theta_h^i$ where $\theta_h^i$ is the value from the baseline specification found in equation 33.
To find the expression for $\hat{\tau}_m$, we plug in equations 68, 70, and 71 into the analogous condition of equation 65 to find

$$\hat{\tau}_m = \frac{1}{\gamma} \tau_m^i,$$

(101)

where $\tau_m^i$ is the value from the baseline specification.

A.6.3 Proof of Theorem 5

Proof. To find the condition for $\hat{\rho}$, we plug in equations 78, 79, and 80 into the analogous condition of equation 73 to find

$$\hat{\rho} = \frac{1}{\eta \gamma \kappa} \frac{1}{H} \sum_{h=1}^{H} \sum_{i=1}^{I} \frac{\sum_{m=1}^{M} \mathbb{I}_{\theta_m = \hat{\theta}_m}}{M_h}.$$

(102)

Note that this condition is the same as $\hat{\rho} = 1/ (\eta \gamma \kappa) \rho$ where $\rho$ is the value from the baseline specification found in equation 26.

To find the expression for $\hat{\tau}_h$, we plug in equations 78, 80, and 82 into the analogous condition of equation 74 to find

$$\hat{\tau}_h = \frac{1}{\eta} \rho \left( \frac{\sum_{i=1}^{I} \sum_{m=1}^{M} \mathbb{I}_{\theta_m = \hat{\theta}_m}}{M_h} \right)^{-1}.$$

(103)

Note that this condition is the same as $\hat{\tau}_h = (1/\eta) \tau_h$ where $\tau_h$ is the value from the baseline specification found in equation 27.

To find the expression for $\hat{\theta}_h^i$, we plug in equations 79, 78, 82, and 86 into the analogous condition of equation 75:

$$\hat{\theta}_h^i = \frac{1}{H} \sum_{h=1}^{H} \frac{\tau_h \sum_{m=1}^{M} \mathbb{I}_{\theta_m = \hat{\theta}_m}}{M_h}.$$

(104)

Note that this condition is the same as $\hat{\theta}_h^i = \theta_h^i$ where $\theta_h^i$ is the value from the baseline specification found in equation 33.

To find the expression for $\hat{\tau}_i^h$, we plug in equations 78, 82, 84, and 86 into the analogous condition of equation 76 to find

$$\hat{\tau}_i^h = \frac{1}{\kappa} \theta^i \rho \left( \frac{\sum_{m=1}^{M} \mathbb{I}_{\theta_m = \hat{\theta}_m}}{M_h} \right)^{-1}.$$

(105)

Note that this condition is the same as $\hat{\tau}_i^h = (1/\kappa) \tau_i^h$ where $\tau_i^h$ is the value from the baseline specification.

To find the expression for $\hat{\tau}_m$, we plug in equations 82, 84, 86, and 87 into the analogous condition of equation 77 to find

$$\hat{\tau}_m = \frac{1}{\gamma} \tau_m^i,$$

(106)

where $\tau_m^i$ is the value from the baseline specification. □
A.6.4 Proof of Theorem 6

Proof. To find the condition for $\hat{\rho}$, we plug in equations 88, 89, 90, and 91 into the analogous condition of equation 58 to find

$$\hat{\rho} = \frac{1}{R} \frac{1}{H} \sum_{h=1}^{H} \sum_{i=1}^{I} \sum_{m=1}^{M} \frac{I_{h(m)=h} \hat{\theta}_{m}^{i}}{M_{h}}. \quad (107)$$

Note that this condition is the same as $\hat{\rho} = (1/R) \rho$ where $\rho$ is the value from the baseline specification found in equation 26.

To find the expression for $\hat{\tau}_{h}$, we plug in equations 92, 88, and 90 into the analogous condition of equation 59 to find

$$\hat{\tau}_{h} = \rho \left( \sum_{i=1}^{I} \sum_{m=1}^{M} \frac{I_{h(m)=h} \hat{\theta}_{m}^{i}}{M_{h}} \right)^{-1}. \quad (108)$$

Note that this condition is the same as $\tau_{h} = \tau_{h}$ where $\tau_{h}$ is the value from the baseline specification found in equation 27.

To find the expression for $\hat{\theta}_{h}^{i}/\hat{\tau}_{h}^{i}$, we plug in equations 92, 93, and 88 into the analogous condition of equation 60 to find

$$\frac{\hat{\theta}_{h}^{i}}{\hat{\tau}_{h}^{i}} = \frac{R \tau_{h} \sum_{m=1}^{M} I_{h(m)=h} \hat{\theta}_{m}^{i}}{\rho \sum_{m=1}^{M} I_{h(m)=h} \hat{\theta}_{m}^{i}}. \quad (109)$$

Note that this condition is the same as $\hat{\theta}_{h}^{i}/\hat{\tau}_{h}^{i} = R \theta_{h}^{i}/\tau_{h}^{i}$ where $\theta_{h}^{i}/\tau_{h}^{i}$ is the value from the baseline specification found in equation 28.

To find the expression for $\hat{\tau}_{m}^{i}$, we plug in equations 92, 93, and 94 into the analogous condition of equation 61 to find

$$\hat{\tau}_{m}^{i} = \tau_{m}^{i}, \quad (110)$$

where $\tau_{m}^{i}$ is the value from the baseline specification.
A.7 Alternative Explanations

In this section, we consider several alternative explanations for our finding in Table 4 that larger plants are assigned to clusters with lower labor cost shares. We argue that these alternative explanations tend to either be inconsistent with our data or would not lead to multiple clusters when our clustering methodology is applied.

A.7.1 Unobserved Labor Costs and State Owned Enterprises

Dirección del Trabajo (2005) shows that unionization rates are higher among larger firms. The report also shows that unionized firms as well as larger firms are more likely to raise employee salaries.\(^{40}\) We may be concerned that these increases in compensation, which are correlated to size, may result in larger plants having lower labor cost shares. However, the measure of employee compensation we use includes a broad range of compensation types including regular pay (including salary, commissions, and overtime pay) as well as additional benefits given by the plant (including bonuses, snacks, education, transportation, and uniforms), and only unobserved differences in wages would generate dispersion in our recovered labor cost shares. To see this, suppose that there is an observed distortion, \(\hat{\tau}_m\), for each firm so that each firm is observed as paying a different input cost. This observed distortion is taken into account when the firm optimizes and therefore will show up in the numerator of the true cost share in equation 9 along with the unobserved distortions. However, since this distortion is observed, it also shows up in the numerator of our observed cost share in equation 11 (i.e., we observe \(\tilde{\theta}_m = \hat{\tau}_m + \tilde{x}_{i}^{p} p_{m} x_{i}^{p}\)). This means that this observed distortion cancels out in equation 12, leaving our methodology for recovering technologies and unobserved distortions unchanged. The intuition for why observed distortions do not affect our results is that our methodology recovers and decomposes differences between the observed cost share and the true cost share (which in our Cobb-Douglas framework is equal to the factor costshare in the production function). If a distortion is observed, then it does not create a wedge between the true and observed cost shares and ultimately does not affect any of our results.

Labor costs may still differ systematically across firms in ways that are not captured in our measure of labor compensation. To address this concern, we exploit data on the ownership status of firms in our dataset in case unobserved labor costs differ across ownership types. Of the 2,741 firms in our dataset, 104 of the firms are either fully or partially state owned (henceforth, SOEs). SOEs are present in 20 of 23 industries, and although they are more concentrated than privately owned firms in terms of which cluster they are assigned to, no clusters are composed solely of SOEs. To compare SOEs with privately owned firms, we regress log observed cost shares for each input on SOE status while controlling for industry and cluster. The results of these regressions are shown in Table A.11 and show that while SOEs appear to have lower skilled and unskilled

\(^{40}\)Refer to tables “Reajuste salarial en el último año” on page 38 and “Reajuste de salarios, según existencia de sindicato” on page 39 in Dirección del Trabajo (2005).
labor cost shares compared to privately owned firms in the same industry and technology, this relationship disappears after controlling for firm size as measured by logged gross output. As we discuss in Section 5.6, there is evidence that larger firms have lower labor cost shares and higher intermediate cost shares; however, these differences are small compared to the average difference across clusters. The fact that SOEs do not differ from privately owned firms of the same size suggests that unobserved labor costs that potentially differ across SOEs and private firms are unlikely to be driving our results.

Table A.11
logged cost shares versus state ownership

<table>
<thead>
<tr>
<th>Capital</th>
<th>Skilled Labor</th>
<th>Unskilled Labor</th>
<th>Intermediate Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>State Ownership</td>
<td>0.101</td>
<td>-0.194*</td>
<td>-0.286**</td>
</tr>
<tr>
<td></td>
<td>(0.105)</td>
<td>(0.0860)</td>
<td>(0.0848)</td>
</tr>
<tr>
<td>Logged Gross Output</td>
<td>0.0237</td>
<td>-0.119***</td>
<td>-0.200***</td>
</tr>
<tr>
<td></td>
<td>(0.0135)</td>
<td>(0.0109)</td>
<td>(0.0113)</td>
</tr>
<tr>
<td>Cluster FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>N</td>
<td>2754</td>
<td>2754</td>
<td>2754</td>
</tr>
</tbody>
</table>

Standard errors in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001

Table A.11 reports the results for the regression of the state ownership on the log of input cost share. Each column reports the result of the regression for the specified factor input cost share, where industry and cluster fixed effects are included to adjust for differences in factor intensities across industries and clusters. We run the regressions both with and without controlling for size with logged gross output. A value greater than zero indicates that firms with state ownership have a higher cost share than firms with only private ownership that are assigned to the same cluster.

A.7.2 Overhead Labor Costs

Suppose that plants incur labor overhead (fixed labor costs) in their production function. Since these costs are fixed, they would not show up in the firm’s first-order conditions or in expression 9. These costs would, however, show up in the observed cost shares in both expressions 11 and 12 and would cause dispersion in observed labor cost shares. Readers may therefore be concerned that overhead labor costs could lead larger firms to be assigned to clusters with lower labor cost shares. Our clustering methodology, however, determines the number of clusters by identifying gaps in the dispersion of observed cost shares. If the distribution of firm sizes is continuous without clear gaps in the distribution of firms, overhead labor costs would result in a negative correlation between size and labor cost shares, but would not yield multiple clusters in our methodology. The distribution of firm sizes in our data indeed appears to be continuous and roughly lognormal both overall and within each industry, lending support to the idea that overhead labor costs would not
be able to generate multiple clusters.

A.7.3 Labor Contracting Frictions

We may also be concerned that there are labor contracting frictions, such as in Grobovsek (2020), whereby poor contract enforcement discourages firm owners from delegating responsibility to mid-level managers. We want to understand whether these contracting frictions may explain why larger plants are assigned to clusters with lower labor cost shares. If this were the case, we would expect to find large distortions for non-production workers relative to production workers as plant size increases. Through the lens of our framework, we would expect that these unobserved distortions would be more pronounced for skilled labor relative to unskilled labor as plant size increases. Thus, the regression of the log of $\theta^i_h/\tau^i_h$ on the log of gross output of plants, where $h$ is the cluster to which the plant is assigned, would give a more negative coefficient on skilled labor. Table 4, however, shows that the coefficient on skilled labor is -0.04 compared to -0.08 for unskilled labor. We conclude that these kinds of contractual frictions cannot account for the differences in cost shares that we observe.

A.7.4 Size-dependent Labor and Capital Policies in Chile

We now attempt to document size-dependent policies related to labor and capital in Chile to see if they can account for the relationship between cluster assignments and plant size, as document in Section 4.2.

First, we are not aware of any significantly distortionary size-dependent labor policies in Chilean labor law. The primary size-dependent regulations for Chilean firms are as follows: firms with more than 25 employees are required to establish a joint health and safety committee, are required to establish a bipartite training committee, and must ensure that at least 85 percent of their workers are Chilean nationals. More significant policies, such as the right to unionize, are constitutional in nature and not size dependent. This is in contrast to countries such as France, which Gourio and Roys (2014) document as having substantial size-dependent policies which apply to firms with more than 50 employees and lead to a noticeable break in the number of firms with more than 50 employees. In the Chilean data, there are no significant breaks in the distribution of firm size by employment. Second, Dinerstein and Patiño Peña (2018) show that while there is some variation in corporate tax rates across firms in Chile, 75 percent of firms have corporate tax rates under the statutory rate of 17 percent. Furthermore, corporate tax rates do not vary significantly when firms are grouped by employment, sales, region, business entity type, or industry. For these reasons,

---
41 We reclassify skilled and unskilled labor compensation as production labor and non-production labor. We find that the correlation of logged cost shares between non-production labor and skilled labor at the plant level is 0.80; we find that the same correlation for production labor and unskilled labor is 0.74. These findings suggest that non-production labor tends to rely more heavily on skilled labor and production labor on unskilled labor.
most of the cluster-specific distortions that apply in other economies do not seem able to explain
our Chilean findings.
A.8 Technology’s Share of Misallocation in Related Literature

Hsieh and Klenow (2009) conduct an exercise to measure the role of technological differences across plants in explaining misallocation. In this exercise, they consider the variation in the capital-labor share for each firm to be due to differences in technology; the remaining variation in TFPR across firms continues to be misallocation. They compare their estimated gains from eliminating misallocation in China and India under this scenario to their baseline case where they consider variation in the capital-labor share to be misallocation. They find that the estimated gains in TFP from equalizing misallocation remain large, with estimated gains between 77–90 percent of the baseline results for China and 65–80 percent of the gains for India. This result suggests that technological differences across plants might reasonably be expected to explain between 10 and 35 percent of the estimated gains from eliminating misallocation. In Table 7 we show that accounting for cluster-specific technologies reduces the estimated gains of eliminating misallocation by a little over 30 percent, which places us on the high end of their estimates, to the extent that our results might be viewed as comparable (we discuss this comparison in the next paragraph). David and Venkateswaran (2019) perform a similar exercise with an equivalent definition of technology. They estimate an upper bound on the importance of technology by using the covariance between the average revenue products of labor and capital. The intuition is that a higher capital intensity in the production function will, all else equal, lead to a higher average revenue product of labor and a lower average revenue product of capital; the implication is that these two things should be negatively correlated in the data if there is significant variation in capital intensities across firms. They find an upper bound for the role of technology of 17 percent for China and 38 percent for the United States, which again would place our results on the upper end of, but still within, their range.

The exercises in our papers have several important differences. Most importantly, when these papers refer to technology, they refer to each firm having a unique production technology and capital intensity, whereas in this paper we consider only a small number of production technologies in each industry, which we identify using our clustering methodology. This means that much of what these papers refer to as technology continues to be considered as misallocation in this paper. Additionally, we allow for the possibility of misallocation in intermediate inputs, which neither of the previously mentioned papers allow for. Therefore, there is no room in these papers for technology to reduce this source of misallocation. Taken together, these differences mean that our role for technology could be larger or smaller than theirs and serves primarily as a complementary concept.

To develop more comparable estimates to these papers using our framework, we perform a simple exercise in which we determine what fraction of our measured misallocation can be explained by idiosyncratic distortions that lead to variation in the capital-labor ratio. In particular, we rewrite $\tau_{sl}^{m} = \tau_{m}^{\text{common}} \left( \tau_{m}^{sl} / \tau_{m}^{\text{common}} \right)$. We similarly rewrite $\tau_{ul}^{m} = \tau_{m}^{\text{common}} \left( \tau_{m}^{ul} / \tau_{m}^{\text{common}} \right)$ and $\tau_{k}^{m} = \tau_{m}^{\text{common}} \left( \tau_{m}^{k} / \tau_{m}^{\text{common}} \right)$.
\(\tau_m^{\text{common}}(\tau_k^m/\tau_m^{\text{common}})\). Notice that \(\tau_m^{\text{common}}\) is a distortion that is common to skilled labor, unskilled labor, and capital.

We then estimate the aggregate gains from eliminating idiosyncratic misallocation related to variation in capital-labor ratios by setting \(\tau_{sl}^m = \tau_{ul}^m = \tau_k^m = \tau_m^{\text{common}}\). We can then compare the fraction of total misallocation that is explained by this exercise to the numbers reported by Hsieh and Klenow (2009) and David and Venkateswaran (2019). As mentioned in Section A.2.2, these papers have output distortions in their frameworks, which affect all factor inputs including intermediate inputs. They do not, however, feature separate distortions on intermediate inputs; therefore, we leave \(\tau_{x}^m\) unchanged. This implies, however, that \(\tau_m^{\text{common}}\) is not identified in the data. For that reason, we consider four possible ways of estimating \(\tau_m^{\text{common}}\) that vary in their conservativeness. Table A.12 reports the results of each of these four methods. For the case Intermediates, we set \(\tau_m^{\text{common}} = \tau_{x}^m\). This case is the one most similar to the exercises in the other papers, since \(\tau_m^{\text{common}}\) can then be interpreted as an output distortion, \(\tau_y^m\), and there would be no distortion for intermediate inputs. In the standard single-cluster framework, we find aggregate gains of 9.8 percent after eliminating the distortions related to the capital-labor ratio variation, which is 25 percent of the total gains of 38.7 percent for eliminating all misallocation and comfortably in the middle of the estimates reported in the literature. Surprisingly, however, misallocation increases by 1.9 percent if we first apply our clustering methodology. Part of this result is because our methodology allows for cluster-specific output distortions, which would be disguised as capital-labor distortions if one were to group firms with different technologies and incorrectly specify their production functions. The more important factor, however, is highlighted in Table 3 and Table 6: although intermediate input distortions are smaller than for the other factors, they are also less affected by accounting for multiple clusters compared to misallocation for the other factors. Therefore, when we set \(\tau_m^{\text{common}} = \tau_{x}^m\), we effectively increase the distortions on some firms.\(^{42}\) Alternative methods for determining \(\tau_m^{\text{common}}\) avoid this issue. For the case Average, we set \(\tau_m^{\text{common}}\) to a factor-intensity weighted average of each of the \(\tau\)'s: \(\tau_m^{\text{common}} = \sum_i \tau_m^i \theta_h^i / (1 - \theta_x^h)\). For the case Smallest, we set \(\tau_m^{\text{common}}\) to the least distorted of the distortions on skilled labor, unskilled labor, and capital: \(\tau_m^{\text{common}} = \arg\min_{i \in \{ul,sl,k\}} |\tau_m^i - 1|\). For the case None, we eliminate idiosyncratic distortions on skilled labor, unskilled labor, and capital entirely by setting \(\tau_m^{\text{common}} = 1\). For each of these cases, we find that output and TFP increase regardless of whether we allow for multiple-clusters or restrict ourselves to the single cluster framework. For the cases Smallest and None, the gains in the single-cluster framework are considerably greater than with cluster-specific technologies, which implies that idiosyncratic technologies may explain some of the misallocation explained by our cluster-specific technologies.

\(^{42}\)Note that if we have output distortions instead of intermediate input distortions, and interpret our clusters as being from cluster-specific distortions only, this implies that there would effectively be room for variation in factor cost shares arising from technology. As pointed out in Section 5.6, if we allow for both cluster-specific distortions and cluster-specific technologies, then our results are similar to when we allow for cluster-specific technologies only.
Table A.12
AGGREGATE GAINS FROM ELIMINATING CAPITAL-LABOR RATIO DISTORTIONS
(PERCENTAGES)

<table>
<thead>
<tr>
<th>Case</th>
<th>Output</th>
<th>TFP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Multiple</td>
<td>Single</td>
</tr>
<tr>
<td></td>
<td>Clusters</td>
<td>Cluster</td>
</tr>
<tr>
<td>Intermediates</td>
<td>-1.9</td>
<td>9.8</td>
</tr>
<tr>
<td>Average</td>
<td>2.3</td>
<td>1.7</td>
</tr>
<tr>
<td>Smallest</td>
<td>1.5</td>
<td>7.4</td>
</tr>
<tr>
<td>None</td>
<td>4.2</td>
<td>16.9</td>
</tr>
<tr>
<td>Total</td>
<td>26.8</td>
<td>38.7</td>
</tr>
</tbody>
</table>

Table A.12 shows the aggregate gains in output and TFP from removing distortions that lead to dispersion in capital-labor ratios (i.e., the gains from equalizing the distortions on skilled labor, unskilled labor, and capital by setting $\tau_m^l = \tau_m^u = \tau_m^k = \tau_m^{common}$). Multiple Clusters reports the counterfactual gains under the framework in which we allow for cluster-specific technologies and distortions. Single Cluster reports the counterfactual gains under the assumption that there are no cluster-specific differences in production technologies or distortions. Case indicates the method we choose for estimating $\tau_m^{common}$, each of which is discussed in the text. The row Total indicates the total aggregate gains from eliminating all misallocation for each framework.

Taken together, the results of these experiments highlight two key findings. The first is that our standard single-cluster framework produces comparable estimates to those from the literature. The second finding is that the concept of cluster-specific technologies that we capture with our clustering methodology is considerably different from, and likely complementary to, the concept of idiosyncratic technologies considered in the literature.
A.9 Copper Processing in Chile

Anecdotally, we can provide circumstantial evidence on the existence of multiple production technologies in Chile’s economy. In the copper industry, for example, two main technologies are used by copper manufacturers to produce copper cathodes, which are 99.99 percent copper. The traditional process involves extracting copper ore (1 percent purity) from open air mines, from which copper concentrates (25-40 percent purity) are produced using mineral flotation (this process is considered to belong to mining rather than manufacturing, and therefore these plants are not included in our sample). The copper concentrates are then sent to smelters that produce copper anodes (97-99 percent purity) and then to refineries that produce copper cathodes (99.99 percent purity). A newer process, solvent extraction electrowinning (SX/EW), is a multi-stage process through which copper cathodes are produced directly from copper ore. SX/EW is significantly less energy intensive due to occurring at ambient temperatures, and works through the combination of a leeching solution, which dissolves the copper, a solvent solution, which binds with the copper, and then electrowinning, which separates the copper from the solution and deposits it onto cathodes. In practice, both the traditional process and the SX/EW process are used internationally and within Chile; however, SX/EW plants tend to be less energy intensive as well as smaller on average. In addition, the copper industry contains downstream manufacturers, referred to as semi-manufacturers, which produce semi-finished products such as copper wires, cables, pipes, and alloys from the copper cathodes. Unfortunately, without additional information on intermediate input usage or product-level outputs, we are unable to conclusively map our recovered technologies to the technologies present in the copper industry, or to rule out that our recovered technologies are due to scale effects that allow larger plants to be less labor intensive in general.
A.10 Details of Clustering Methodology in Section 3.4 in which we Determine the Number of Clusters

This section gives more details about the methodology described in Section 3.4 in which we determine the number of clusters.

A.10.1 20th Percentile Cutoff

We choose the 20th percentile because it leads to robust results when applied to our data in Section 4. In particular, our results in Section 4 are not affected if we switch to using the 15th or 25th percentile. This percentile is the significance level of our test in the simulations, which, in general, is larger than the probability that we conclude there are two or more technologies when there is only one cluster in our true data. The reason for this difference is that the true data-generating process may differ significantly from the fitted distribution we use in our simulations. In Section A.3, we evaluate our methodology using simulated data, and we find that we recover the correct number of clusters 100 percent of the time when we use the 20th percentile of $R^2_1$ for our cutoff and the data are generated according to our model in Section 2. In general, the cost of using lower percentiles, such as the 5th percentile, is that our cutoff becomes too strict and we underpredict the true number of clusters without much gain in accuracy when there is only one cluster. On the other hand, if we use a cutoff much higher than the 20th percentile, our cutoff becomes too lenient, which goes against our goal of being conservative in our estimates for the number of clusters in each industry.

A.10.2 Selecting the Number of Clusters

Our general methodology for concluding there are $H$ or more clusters — conditional on already concluding there are $H-1$ or more clusters — requires an estimate of the expected reduction in the SSE from clustering with $H$ groups if there are only $H-1$ clusters. We calculate this in four steps. First, we cluster the data according to Section 3.3 with $H-1$ groups. Second, we fit a multivariate lognormal distribution independently to each of the $H-1$ groups in our clustered data. Third, we generate a large number of synthetic datasets, where each synthetic dataset is constructed by taking independent draws from each of the $H-1$ multivariate lognormal distributions, such that the number of observations drawn from each distribution is equal to the number of observations in the group to which we fitted the distribution in the original data. For example, if we use $H-1 = 2$ groups in our original data, and after clustering, group 1 contains 15 observations, whereas group 2 contains 20 observations, then our synthetic datasets will contain 15 observations drawn from the multivariate lognormal distribution corresponding to group 1 and 20 observations drawn from the multivariate lognormal distribution corresponding to group 2. Fourth, we apply our clustering algorithm from Section 3.3 to each synthetic dataset individually and calculate $R^\text{Syn}_{H|H-1}$, which we
do using an equation analogous to 20.