# Measuring Factor Misallocation across Industries: General Methods and Evidence on the Great Recession*

Thomas Schelkle†

October 13, 2017

**Abstract**

The prior literature measuring factor misallocation depends on the restrictive assumption that output elasticities are constant across production units. Such approaches cannot provide credible evidence on misallocation across industries. This paper develops more robust methods to measure misallocation and applies them to investigate the capital and labour allocation across 473 U.S. manufacturing industries. This reveals substantial factor misallocation across industries with potential output gains from an efficient reallocation of $22 - 64\%$. The degree of misallocation varies over time and increased during the Great Recession, which contributed about $10 - 60\%$ to the observed decline in manufacturing output.

*JEL codes: E23, E32, O11, D61*
*Keywords: Misallocation, Industries, Great Recession*

*I am grateful to Francesco Caselli, Peter Funk, Dominik Sachs, Andreas Schabert and participants at several seminars and conferences for helpful comments and conversations.

†University of Cologne; Center for Macroeconomic Research; Albertus-Magnus-Platz; 50923 Köln; Germany; Email: schelkle@wiso.uni-koeln.de

# 1 Introduction

In recent years there is a strong interest in factor misallocation following Restuccia and Rogerson (2008) and Hsieh and Klenow (2009) in order to explain income and productivity differences across time and space. Factor misallocation refers to an "inefficient" allocation where marginal value products are not equalised across production units such that the total value of output at current prices is not maximised. However determining whether an empirically observed factor allocation is efficient is extremely challenging. The main reason is that marginal products are not directly observable. So far the literature draws conclusions from differences in observed average products to differences in unobserved marginal products between production units. However differences in average products may in general simply reflect differences in output elasticities without any differences in marginal products. Thus these inferences rely either on the strong assumption that production functions are very similar across production units or on being able to perfectly control for differences in output elasticities by observed differences in factor income shares. This important literature may therefore incorrectly reject an efficient factor allocation simply because these assumptions are not satisfied.

As a consequence prior work focusses on misallocation between different firms within an industry and neglects misallocation across industries. The main reason is that the key identifying assumption is most likely to hold in the former context, but is relatively implausible in the latter. Thus these methods cannot provide credible evidence on misallocation across industries. This is an important limitation because differences in taxes, subsidies and regulation across industries among other factors suggest that misallocation across industries could be substantial. There are also several economic reasons why misallocation across industries may vary over time and contribute to major economic fluctuations like the Great Recession. For instance different industries may face different productivity shocks or consumers may differently adjust their demand for the products of different industries during such a period. But in the presence of adjustment costs or other frictions factors of production cannot efficiently be reallocated in response to these developments such that misallocation becomes more severe, cf. Mian and Sufi (2015) for such arguments. The financial crisis probably had a heterogeneous impact on industries due to their differing need for external finance (Rajan and Zingales 1998), and his may have also worsened the capital allocation across industries. However the quantitative importance of misallocation across industries and how much changes to misallocation contributed to the Great Recession are unclear so far.

This paper makes two contributions to the literature measuring factor misallocation. The first is to provide a novel methodological framework for these measurements which is more robust and can be applied in a larger number of contexts. The second contribution is to provide a quantitative analysis of the capital and labour allocation across 473 manufacturing industries at the six-digit level in the United States before and during the Great Recession. This investigation reveals a significant misallocation of resources across industries. It also provides evidence that misallocation became more severe during the Great Recession and contributed substantially to the observed decline in manufacturing output.

The methodological framework consists of two parts. First, I derive a novel testable implication of equalised marginal products assuming only that production functions are homogeneous of a known degree. This setting is considerably more general than the assumptions on specific functional forms or common values of output elasticities across production units in the literature. It is shown that when marginal products of labour and capital are equalised across production units a high average product of labour of a production unit relative to another unit must be accompanied by a sufficiently low average product of capital and therefore a sufficiently high capital intensity. This theoretical prediction can be compared to data to test the efficiency hypothesis. If an observed factor allocation does not exhibit this property then it cannot possibly be characterised by equalised marginal products for any combination of homogeneous production functions with the assumed degrees of homogeneity. Due to the relatively general assumptions such an approach is more robust to misspecification and has a lower risk of incorrectly rejecting an efficient allocation. This implies that in case the data is inconsistent with this theoretical implication one can be very confident that the efficiency hypothesis is indeed false. In addition one can gain insights into the possible magnitude of marginal product differentials underlying an observed allocation.

The second part of the methodological framework investigates how much total output could potentially be increased when distortions are eliminated and resources are reallocated efficiently. This provides a measure of the economic significance of misallocation. Determining the hypothetical efficient allocation requires additional functional form assumptions on production functions and preferences as in previous work. However the framework does again not require exact knowledge of the magnitude of output elasticities. Instead an implication of homogeneous production functions is used to compute upper and lower bounds on the potential output gain. These bounds provide more robust though of course less precise evidence on the magnitude of such potential gains. An extension allows to flexibly strengthen the assumptions on output elasticities, which translates into a more demanding efficiency test and tighter bounds on potential output gains.

The developed methods can be used in a large number of contexts. A main benefit is that these methods extend the range of applications in which one can credibly measure misallocation for example between industries, sectors, regions and countries. But they can also be applied to misallocation between firms within an industry and for checking the robustness of the prior literature on this topic.

The second contribution of the paper is to provide an analysis of the capital and labour allocation across 473 manufacturing industries at the six-digit level in the United States before and during the Great Recession. As discussed above there are many economic reasons that make one suspect misallocation across industries and an increase of misallocation in the Great Recession. However there is so far no evidence on its quantitative magnitude. Another motivation is that regardless of these reasons many observers still consider the U.S. to be a relatively undistorted economy. If such a belief is correct then it should be straightforward to rationalise this data as an efficient factor allocation when using the general methods developed here. But if one can refute efficiency in this application then it seems likely that misallocation may also be found in many other countries and contexts.

Contrary to the conventional wisdom, the analysis provides strong evidence for economically significant misallocation across industries. In 2005 almost half of all industry pairs do not satisfy the requirements of an efficient allocation based on assuming constant returns to scale in all industries. In other words there is no combination of constant returns to scale production functions that could rationalise this data as an efficient allocation. This result extends to situations where returns to scale differ between industries unless these differences are extremely large. The lower bound of potential output gains from an efficient reallocation of factors takes a value of 22% of actual output. In contrast the upper bound takes a very high value for the most general assumptions. But if one restricts output elasticities to fall within a still fairly large range centred around observed factor income shares then the upper bound on potential output gains is 64% of actual output. In this application a traditional approach that sets output elasticities exactly equal to factor income shares yields a point estimate for potential output gains of 28%, which is surprisingly close to the lower bound here.

An analysis of the dynamics of misallocation during 2005-2009 provides several pieces of evidence for an increase in misallocation during the Great Recession. In 2009 about 60% of all industry pairs are inconsistent with an efficient allocation compared to only about 50% in 2005. The range in which the potential output gains of an efficient reallocation need to fall also shifts upwards over time from $22-64\%$ in 2005 to $28-72\%$ in 2009. Due to the overlap of these ranges, one cannot directly conclude from this that the potential output gain increased. But if one assumes output elasticities to remain constant across years and also narrows the range within which they may fall then one obtains an increase in potential output gains between 2 and 11 percentage points. The increase in misallocation then contributes between 10 and 60% to the observed about 18 percentage point negative deviation of manufacturing output from trend in 2009. If one sets output elasticities exactly equal to observed factor income shares then one obtains an increase in potential output gains of 6.4 percentage points representing a 35% contribution to the fall in manufacturing output. Overall these results suggest that mitigating the underlying frictions has not only the potential to permanently increase total output, but also to dampen economic fluctuations.

The paper relates to a very active recent literature on factor misallocation surveyed by Restuccia and Rogerson (2013, 2017) and Hopenhayn (2014). In particular the paper fits into what Restuccia and Rogerson (2013, 2017) call the "indirect approach" which attempts to measure the overall level of misallocation resulting from the cumulative effect of all distortionary policies, institutions and market imperfections. Prior work such as Hsieh and Klenow (2009), Bartelsman, Haltiwanger, and Scarpetta (2013) and Vollrath (2009) relies on output elasticities being constant across the production units under study, or across countries, or both. This allows them to draw conclusions from average product differences to marginal product differences between production units and from differences in average product dispersion to differences in the degree of misallocation across countries. However as discussed above a failure of this key assumption potentially invalidates such conclusions. There is also only a limited range of applications where this assumption seems plausible. The present paper develops more robust methods for measuring misallocation in order to quantitatively investigate factor misallocation across industries, where this assumption is unlikely to hold.

Another strand of the literature using the "indirect approach" for identifying misallocation relies on factor prices being equal to marginal products. Papers like Banerjee and Duflo (2005) then take factor price differentials as a direct indication of misallocation. Other work combines information on factor income shares and average products to calculate marginal products as in Caselli and Feyrer (2007). A potential limitation here is the possible departure of factor prices from marginal value products as discussed particularly for developing countries in the survey by Rosenzweig (1988). Another problem is that it is very difficult to reliably measure factor income shares for example because labour compensation of the self-employed is often treated as capital income as argued by Gollin (2002). Thus, it is an important advantage that the methods of this paper are independent of any direct assumptions and data requirements concerning factor prices.

One paper that also addresses concerns on heterogeneity of output elasticities among other factors for the identification of misallocation is Song and Wu (2015). These authors pursue a similar aim, but their approach is very different. They impose a lot of structure on the data including distributional assumptions and restrictions across time periods such that they also require panel data to identify misallocation. In contrast this paper shows what inferences can be drawn using only very general assumptions and minimal economic structure such that one needs only cross-sectional data. Thus, these two approaches are complementary.

The paper is also related to the misallocation literature that employs the "direct approach" and studies the effects of specific imperfections and distortionary policies, cf. the surveys of Restuccia and Rogerson (2013, 2017) and Hopenhayn (2014). Examples are financial frictions (Buera, Kabowski, and Shin 2011; Caselli and Gennaioli 2013; Midrigan and Xu 2014; Moll 2014), frictional labour markets (Lagos 2006), size-dependent policies and regulation (Guner, Ventura, and Xu 2008; García-Santana and Pijoan-Mas 2014), imperfect information (David, Hopenhayn, and Venkateswaran 2016; Senga 2016), adjustment costs (Asker, Collard-Wexler, and Loecker 2014) or imperfect output markets (Peters 2013), among others. This literature provides many possible explanations which may be causing the overall level of misallocation measured by this paper and the rest of the literature using an "indirect approach".

The empirical application of the paper provides novel evidence on economically significant misallocation across industries in the United States, which has been neglected by the literature so far. Thus it complements prior work following Hsieh and Klenow (2009) with a focus on misallocation within industries. The analysis also shows that an increase in misallocation contributed substantially to the decline in manufacturing output during the Great Recession. This complements prior evidence on increases in misallocation over time particularly during the U.S. Great Depression (Ziebarth 2015), the Chilean crisis of 1982 (Oberfield 2013), the Argentine crisis of 2001 (Sandleris and Wright 2014), and in Southern Europe around the introduction of the Euro (Dias et al. 2016; García-Santana et al. 2016; Gopinath et al. 2015), and for a decrease in misallocation in Eastern Europe during a period of capital account liberalization (Larrain and Stumpner 2015) and in Chile after the 1982 crisis (Chen and Irarrazabal 2015). All these papers employ the methodology of Hsieh and Klenow (2009) or a close variant of it. In contrast this paper uses more robust methods. This reveals that at least for the Great Recession the conclusions on an increase in misallocation can only be drawn under sufficiently strong assumptions on output elasticities, though these are still more general than used by this literature. Nevertheless in light of this finding it seems important to explore the sensitivity of comparisons of misallocation across time and space in more detail in future work, for which the methods developed here are also useful.

The paper is structured as follows. The testable implication of an efficient factor allocation is developed in section 2 and the bounds on potential output gains in section 3. Section 4 presents the data on U.S. manufacturing industries. The results on the factor allocation across industries in 2005 are provided in section 5 and on the dynamics of misallocation during the Great Recession in section 6. Section 7 presents robustness checks and section 8 concludes.

# 2 A Testable Implication of Efficient Allocations

This section derives observable restrictions on factor allocations for given marginal product differentials, which are valid for all well-behaved and homogenous production functions. These theoretical predictions can be compared to the data which allows to test the hypothesis that an observed factor allocation exhibits equalised marginal products. One can also employ them to characterise the set of possible unobserved marginal product differentials underlying an observed allocation.

## 2.1 Basic Assumptions

There are $N \geq 2$ production units indexed by $i = 1, \ldots, N$, which depending on the context could for example be different firms, industries, sectors or countries. I also frequently refer to a pair of these production units consisting of units $a$ and $b$. The set of all the $N(N-1)/2$ possible pairs $(a, b)$ is denoted by $P = \{(1,2), (1,3,), \ldots, (N-1, N)\}$.

The output of goods of each production unit is denoted by $Y_i$ with associated given output price $p_i$. All production units use labour $L_i$ and capital $K_i$ as common production factors. The total amount of factors that can be allocated between all production units is exogenous and for labour denoted by $\overline{L}$ and for capital by $\overline{K}$. The production units may differ in their production functions. But it is assumed that all production functions are "well-behaved" such that they satisfy standard regularity conditions like continuity, differentiability and are strictly increasing and concave in $K_i$ and $L_i$. Furthermore the production functions are assumed to be homogenous in $K_i$ and $L_i$ of degree $0 < \lambda_i \leq 1$ where the degree of homogeneity $\lambda_i$ may differ between the production units.[1]

In the following I allow for the presence of distortions that drive a wedge between the marginal value products of the production units. The labour wedge for a pair of production units $(a, b)$ is denoted by $d_{ab}^L$ and the capital wedge by $d_{ab}^K$. These wedges are exogenous and capture the cumulative effect of market imperfections, institutions and distortionary policies. For an interior solution the factor allocation is determined by modified marginal value product equations that for each pair of production units $(a, b) \in P$ read as

$$d_{ab}^L \, p_a \frac{\partial Y_a}{\partial L_a} = p_b \frac{\partial Y_b}{\partial L_b} \tag{1}$$

$$d_{ab}^K \, p_a \frac{\partial Y_a}{\partial K_a} = p_b \frac{\partial Y_b}{\partial K_b} \tag{2}$$

and the resource constraints $\sum_{i=1}^N L_i = \overline{L}$ and $\sum_{i=1}^N K_i = \overline{K}$. One may simply view these equations as definitions of the marginal product differentials $d_{ab}^L$ and $d_{ab}^K$. Hence given known differentials $d_{ab}^L$ and $d_{ab}^K$ for all pairs one can also determine the factor allocation by these equations independently of how the factor allocation is determined in reality. Note that by definition the marginal product of labour differentials satisfy $d_{ba}^L = 1/d_{ab}^L$ and $d_{ac}^L = d_{ab}^L \times d_{bc}^L$ where $c$ is a third production unit, and of course the same applies to those of capital. This implies for instance that the $N-1$ differentials $d_{ab}^L$ between one fixed unit $a$ and all other units $b = 1, \ldots, N$ with $b \neq a$ are sufficient to determine the marginal product differentials for any other pair of production units.

A value of the labour (capital) wedge $d_{ab}^L$ ($d_{ab}^K$) above one indicates that the marginal value product of labour (capital) is higher in production unit $b$ than in $a$, and vice versa. If the wedges $d_{ab}^L$ and $d_{ab}^K$ are not equal to one for all pairs of production units $(a, b) \in P$ then marginal value products are not equalised across all production units and accordingly total income is not maximised at this allocation. I refer to such a situation as "factor misallocation" and the allocation

[1] The theoretical properties of allocations with equalised marginal value products derived below are in principle also valid for production functions with degrees of homogeneity larger than one, i.e. for increasing returns to scale. But then an allocation with equalised marginal value products is not necessarily a situation where the value of total output is maximised. Accordingly equalization of marginal value products may not be desirable and testing this empirically would not be so interesting. This is the reason for restricting attention to production functions which are at a maximum linearly homogeneous.

being "inefficient". In contrast an "efficient" allocation is one where marginal value products are equalised ($d^L_{ab} = d^K_{ab} = 1$ for all pairs $(a, b) \in P$) such that the total value of output is maximised. The main aim of this paper is to assess whether an observed factor allocation could or could not be an "efficient" allocation in this total output maximizing sense and how strong the deviation from efficiency is.[2]

## 2.2 Implications of Marginal Product Differentials

The main problem in identifying factor misallocation is that marginal products are unobservable. Thus, I show how marginal product differentials are related to average product differentials and other observable variables. Divide equation (1) by (2) to obtain an equation of marginal rates of technical substitution reading as

$$\frac{d^L_{ab}}{d^K_{ab}} \frac{\frac{\partial Y_a}{\partial L_a}}{\frac{\partial Y_a}{\partial K_a}} = \frac{\frac{\partial Y_b}{\partial L_b}}{\frac{\partial Y_b}{\partial K_b}}. \tag{3}$$

It is more convenient to work with equations (1) and (3). The key step is then to apply two simple "multiply and divide" tricks to equations (1) and (3) given by

$$d^L_{ab}\ p_a \frac{Y_a}{L_a} \frac{\partial Y_a}{\partial L_a} \frac{L_a}{Y_a} = p_b \frac{Y_b}{L_b} \frac{\partial Y_b}{\partial L_b} \frac{L_b}{Y_b} \tag{4}$$

$$\frac{d^L_{ab}}{d^K_{ab}} \frac{K_a \frac{\partial Y_a}{\partial L_a} \frac{L_a}{Y_a}}{L_a \frac{\partial Y_a}{\partial K_a} \frac{K_a}{Y_a}} = \frac{K_b \frac{\partial Y_b}{\partial L_b} \frac{L_b}{Y_b}}{L_b \frac{\partial Y_b}{\partial K_b} \frac{K_b}{Y_b}}. \tag{5}$$

Rearranging and denoting the average value product of labour by $y_i = p_i \frac{Y_i}{L_i}$, the capital intensity by $k_i = \frac{K_i}{L_i}$, the output elasticity of labour by $\varepsilon_{Li} = \frac{\partial Y_i}{\partial L_i} \frac{L_i}{Y_i}$ and the output elasticity of capital by $\varepsilon_{Ki} = \frac{\partial Y_i}{\partial K_i} \frac{K_i}{Y_i}$ for each production unit $i$ yields

$$\frac{y_b}{y_a} = \frac{\varepsilon_{La}}{\varepsilon_{Lb}} d^L_{ab} \tag{6}$$

$$\frac{k_b}{k_a} = \frac{\varepsilon_{La}}{\varepsilon_{Lb}} \frac{\varepsilon_{Kb}}{\varepsilon_{Ka}} \frac{d^L_{ab}}{d^K_{ab}} \tag{7}$$

[2]This way of defining "efficiency" is motivated by a macroeconomic perspective focussed on income comparisons and explaining income differences. However a maximization of the total value of output at current prices is also related to traditional theoretical concepts like productive efficiency and pareto-optimality. Under the standard assumptions on production functions maintained here, the allocation being total output maximizing at current prices is sufficient for productive efficiency and the allocation being on the production possibility frontier. But it is not necessary because the allocation could be on the production possibility frontier and only be total output maximizing for a different set of prices. If one also makes standard assumptions on households, which imply that their marginal rates of substitution are equal to the current price ratio between two goods, then being total output maximizing at current prices is necessary and sufficient for pareto-optimality of the allocation.

where $\frac{y_b}{y_a}$ is the ratio of average value products of labour between two production units $a$ and $b$, which is sometimes also called "relative labour productivity", and $\frac{k_b}{k_a}$ is the ratio of capital intensities. Instead of working with equation (7) one could also work with the equivalent of equation (6) for capital, which can be written as $\frac{y_b/k_b}{y_a/k_a} = \frac{\varepsilon_{Ka}}{\varepsilon_{Kb}} d_{ab}^K$. Here $\frac{y_b/k_b}{y_a/k_a} = \frac{p_b Y_b/K_b}{p_a Y_a/K_a}$ is the ratio of the average value product of capital between the two production units.

Equations (6) and (7) provide a number of key insights. First, there is indeed a meaningful relationship between the average product of labour ratio $\frac{y_b}{y_a}$ and the marginal product of labour ratio $d_{ab}^L$. However one can only draw direct conclusions from one to the other if one also knows the ratio of output elasticities of labour. In general the output elasticities of a production unit fully depend on the specific production function and the amount of factors used and are not constant. Thus, using Cobb-Douglas functions with common parameters can be restrictive in this context because output elasticities are then constant. For example if one makes such an assumption in an analysis of misallocation between different firms within an industry as in Hsieh and Klenow (2009) then one attributes all differences in average products between firms to differences in marginal products. But in general a difference in average products may just as well be driven by a difference in output elasticities without any difference in marginal products. Another example is the question of factor misallocation between the agricultural and non-agricultural sector in different countries as analysed by Vollrath (2009). If one assumes Cobb-Douglas production functions for these two sectors with common parameters in all countries then one will automatically attribute all of the huge differences in $\frac{y_b}{y_a}$ across countries to differences in $d_{ab}^L$. However part of the $\frac{y_b}{y_a}$ variation may simply be caused by variation in output elasticities across countries. Such a variation in output elasticities may result from different production functions or from different amounts of used factors. Another insight is that in principle suitable values of $d_{ab}^L$ and $d_{ab}^K$ can rationalise any observed $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combination.

So far the derivation is without much loss in generality. Now I impose the assumption that the production functions of all production units are homogenous in $K_i$ and $L_i$ of degree $0 < \lambda_i \leq 1$. By Euler's theorem this implies that for each production unit $i$ the sum of output elasticities equals the degree of homogeneity[3]

[3] I refer to the key assumption as being one of homogeneity of production functions with known $\lambda_i$ due to the importance and wide use of this assumption in both theoretical and applied work. However strictly speaking the derived restrictions require only a weaker assumption which is that equation (8) holds for known $\lambda_i$ *at the current allocation.* In other words the value of the sum of output elasticities must be known locally at the current allocation. But production functions need not be homogeneous, which is a global property of production functions.

reading as

$$\varepsilon_{Li} + \varepsilon_{Ki} = \lambda_i. \tag{8}$$

Furthermore the standard regularity conditions of production functions mentioned above require marginal products of $K_i$ and $L_i$ to be positive. Hence output elasticities need to be positive as well such that $\varepsilon_{Li} > 0$ and $\varepsilon_{Ki} > 0$ and together with equation (8) this implies bounds on the output elasticities given by $\varepsilon_{Li} \in (0, \lambda_i)$ and $\varepsilon_{Ki} \in (0, \lambda_i)$. Equations (6), (7) and (8) together with these bounds on output elasticities and given values of $d^L_{ab}$ and $d^K_{ab}$ imply restrictions on the observable quantities $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ for a pair of production units. These restrictions are stated by the following proposition (all proofs are relegated to appendix A).

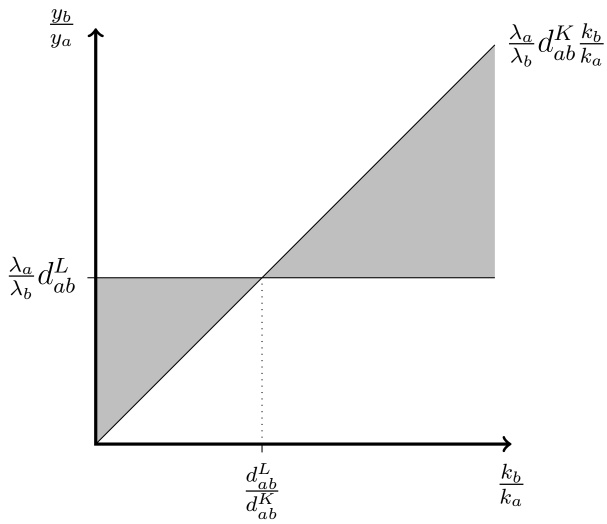
**Proposition 1.** *If two production units $(a, b)$ operate with well-behaved and homogenous production functions of degree $\lambda_a$ and $\lambda_b$ and the factor allocation exhibits marginal product differentials of labour $d^L_{ab}$ and capital $d^K_{ab}$ between these production units then the average product of labour ratio $\frac{y_b}{y_a}$ and capital intensity ratio $\frac{k_b}{k_a}$ at this allocation satisfy either*

$$\begin{array}{lll} \dfrac{k_b}{k_a} > \dfrac{d^L_{ab}}{d^K_{ab}} & \text{and} & \dfrac{\lambda_a}{\lambda_b} d^L_{ab} < \dfrac{y_b}{y_a} < \dfrac{\lambda_a}{\lambda_b}\dfrac{k_b}{k_a} d^K_{ab}, \quad \text{or} \\ \dfrac{k_b}{k_a} < \dfrac{d^L_{ab}}{d^K_{ab}} & \text{and} & \dfrac{\lambda_a}{\lambda_b}\dfrac{k_b}{k_a} d^K_{ab} < \dfrac{y_b}{y_a} < \dfrac{\lambda_a}{\lambda_b} d^L_{ab}, \quad \text{or} \\ \dfrac{k_b}{k_a} = \dfrac{d^L_{ab}}{d^K_{ab}} & \text{and} & \dfrac{y_b}{y_a} = \dfrac{\lambda_a}{\lambda_b} d^L_{ab}. \end{array}$$

Proposition 1 shows that for any pair of production units $(a, b) \in P$ the maintained assumptions on production functions and given values of $\lambda_a$, $\lambda_b$, $d^L_{ab}$ and $d^K_{ab}$ imply that the observable quantities $\frac{y_b}{y_a}$ and $\frac{k_b}{k_a}$ fall within a certain set. This set of possible combinations of $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ is illustrated as the shaded area in figure 1. Though this set is in principle large, the key result here is that "not anything goes". There are combinations of $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ which can never occur for given values of $d^L_{ab}$ and $d^K_{ab}$. Thus hypotheses about specific values of $d^L_{ab}$ and $d^K_{ab}$ can be refuted because one can point to observations of $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combinations which are inconsistent with such hypotheses.

In a nutshell the intuition for the upper right part of the set is as follows. In this region production unit $b$ has a relatively high average product of labour compared to unit $a$ ($\frac{y_b}{y_a} > \frac{\lambda_a}{\lambda_b} d^L_{ab}$). For a given marginal product of labour differential $d^L_{ab}$ this indicates a relatively low output elasticity of labour of unit $b$ ($\frac{\varepsilon_{La}}{\varepsilon_{Lb}} > \frac{\lambda_a}{\lambda_b}$). But for homogeneous production functions a relatively low output

Figure 1: Illustration of Proposition 1




Notes: The shaded area represents the set of $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combinations which are consistent with the basic assumptions and given specific values of $\lambda_a$, $\lambda_b$, $d^L_{ab}$, $d^K_{ab}$ as described in proposition 1.

elasticity of labour is accompanied by a relatively high output elasticity of capital conditional on potential differences in the degree of homogeneity between units $(\frac{\varepsilon_{La}}{\varepsilon_{Lb}} > \frac{\lambda_a}{\lambda_b} \iff \frac{\varepsilon_{Ka}}{\varepsilon_{Kb}} < \frac{\lambda_a}{\lambda_b})$. Thus this unit must have a relatively low average product of capital $(\frac{y_b/k_b}{y_a/k_a} < \frac{\lambda_a}{\lambda_b} d^K_{ab})$. Such a pattern requires the capital intensity of this unit to sufficiently exceed the one of the other unit $(\frac{k_b}{k_a} > \frac{\frac{y_b}{y_a}}{\frac{\lambda_a}{\lambda_b} d^K_{ab}})$. The reverse argument explains the lower left part of the set. Points outside this admissible set would require the output elasticities of labour and capital of a production unit to be simultaneously relatively high or low for the given marginal product differentials. But such a pattern is ruled out by homogeneity of production functions. Thus observations outside this set can only occur for different marginal product differentials.

Note that a certain value of $\frac{y_b}{y_a}$ is consistent with many different values of $\frac{k_b}{k_a}$ because for a given marginal product of labour differential $d^L_{ab}$ the value of $\frac{y_b}{y_a}$ only requires a certain ratio of output elasticities of labour $\frac{\varepsilon_{La}}{\varepsilon_{Lb}}$. There are many possible combinations of $\varepsilon_{La}$ and $\varepsilon_{Lb}$ within their respective admissible bounds which may underly such a value of $\frac{\varepsilon_{La}}{\varepsilon_{Lb}}$. But these possible combinations yield very different values for the ratio of output elasticities of capital $\frac{\varepsilon_{Ka}}{\varepsilon_{Kb}}$ and this in turn determines different values of $\frac{k_b}{k_a}$ depending on the exact underlying elasticity combination.

## 2.3 Confronting the Efficiency Hypothesis with Data

The previous theoretical results can be used to test the null hypothesis that an observed factor allocation is efficient, i.e. that marginal value products are equalised such that $d_{ab}^L = 1$ and $d_{ab}^K = 1$ for all pairs $(a, b) \in P$. The alternative hypothesis is that at least one marginal product differential deviates from 1. The test relies on the assumptions on production functions of the previous section. First one needs to pick specific values for the degree of homogeneity $\lambda_i$ of the production function of each production unit. For example in many applications one may assume constant returns to scale and accordingly choose a value of 1 based on the replication argument. In other applications the researcher may want to use a value below 1 because a fixed factor other than capital or labour like land or managerial skills is also key for production and there are only constant returns to scale to all factors.

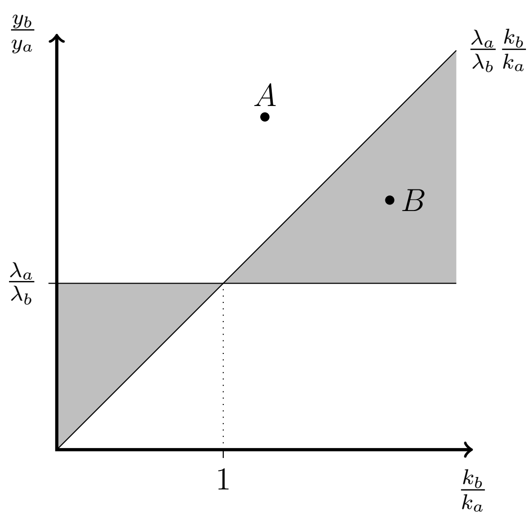
Conducting the test requires observations on the average product of labour ratio $\frac{y_b}{y_a}$ and capital intensity ratio $\frac{k_b}{k_a}$ for all pairs $(a, b)$ at the current allocation. Given the assumed values of $\lambda_i$ the test of the null hypothesis then simply consists in checking whether for each pair $(a, b) \in P$ the observed combination $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ satisfies the conditions of proposition 1 for $d_{ab}^L = d_{ab}^K = 1$, which are either

$$\begin{aligned}
&\frac{k_b}{k_a} > 1 \qquad \text{and} \qquad \frac{\lambda_a}{\lambda_b} < \frac{y_b}{y_a} < \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a}, \qquad \text{or} \\
&\frac{k_b}{k_a} < 1 \qquad \text{and} \qquad \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a} < \frac{y_b}{y_a} < \frac{\lambda_a}{\lambda_b}, \qquad \text{or} \\
&\frac{k_b}{k_a} = 1 \qquad \text{and} \qquad \frac{y_b}{y_a} = \frac{\lambda_a}{\lambda_b}.
\end{aligned}$$

In other words one checks whether for each pair of production units the observed $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combination is an element of the shaded non-rejection region of figure 2, which is the equivalent to figure 1 for the specific values $d_{ab}^L = d_{ab}^K = 1$. The figure also contains two examples $A$ and $B$ of possible observations. If at least one of the observed $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combinations does not satisfy the test conditions like observation $A$, then one rejects the null hypothesis. Under the maintained assumptions such a factor allocation cannot possibly be efficient and necessarily involves a misallocation of resources between production units $a$ and $b$. The reason is that under the null hypothesis observation $A$ can only arise if unit $b$ has simultaneously a relatively low output elasticity of labour and a relatively low output elasticity of capital compared to unit $a$, i.e. $\frac{\varepsilon_{La}}{\varepsilon_{Lb}} > \frac{\lambda_a}{\lambda_b}$ and $\frac{\varepsilon_{Ka}}{\varepsilon_{Kb}} > \frac{\lambda_a}{\lambda_b}$. But such a pattern of elasticities is ruled out by the assumption of homogeneous production functions. Thus marginal products cannot be equalised and the null hypothesis cannot be correct for this observation. In contrast, if the above conditions are satisfied for the observed allocation like for observation $B$, then one cannot reject the null hypothesis. In this case the factor allocation may be efficient because equalised marginal products and the maintained basic assumptions are fully consistent. As in all tests a failure to reject the null hypothesis does not imply that the null is correct. Here this means that even if the factor allocation satisfies the above conditions, it may still be inefficient.

Note that this test for the efficiency of the factor allocation simply consists of a collection of tests for the efficiency of pairwise factor allocations. In the application I then also report for what fraction of the total $N(N-1)/2$ pairs the test rejects pairwise efficiency. This provides an insight into how frequently pairs of production units necessarily deviate from equalised marginal products and in this sense provides a measure of how strong the rejection of overall efficiency is.

Figure 2: The Efficiency Test




Notes: The shaded area represents the non-rejection region and the rest the rejection region. If an observed $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combination is an element of the rejection region then one rejects the null hypothesis of an efficient factor allocation ($d^L_{ab} = 1$, $d^K_{ab} = 1$). If it falls into the non-rejection region, one does not reject the null hypothesis. Points $A$ and $B$ refer to hypothetical examples that one may observe.

The developed test is not a statistical test. Instead it is a simple comparison between a theoretical prediction which should hold under a certain hypothesis and the data. Such a non-statistical approach is also taken by the prior misallocation literature when it draws conclusions from average product to marginal product differences. In this sense the present paper follows the literature.

Nevertheless, it is helpful to explain the pros and cons of the proposed test by using analogies to statistical hypothesis testing. For this purpose the "randomness" underlying the test can be thought of as the uncertainty resulting from our lack of knowledge about the true production functions and output elasticities of the different production units. The main advantage of the test is that it relies on relatively weak assumptions on production functions. Accordingly the probability of making a type 1 error (rejecting a correct null) due to a misspecification of the underlying production model is small. This means that being able to reject the null hypothesis with this test is very informative and should induce a high confidence that the null hypothesis is indeed false. The flip side of this advantage is that there is a higher probability of making a type 2 error (failing to reject a false null) because of the weak assumptions on production functions. Accordingly, the power of the test (the probability of not committing a type 2 error) may be small. In other words one needs to be aware that failing to reject the null is not necessarily very informative on the presence or absence of an efficient allocation. The following subsection provides details on the possible magnitude of marginal product differentials depending on whether the null hypothesis is rejected or not. An extension in section 2.5 allows to tighten the basic assumptions and to trade off the probability of making these two errors.

Whether a low type 1 or type 2 error is more desirable does in general depend on the context. However here a low type 1 error seems to be a key advantage. First note that the standard conventions of statistical hypothesis testing imply that economists seem to strongly value low type 1 errors. Second the hypothesis of efficient factor markets has such an importance and long intellectual history in Economics that it seems reasonable to only consider this hypothesis as refuted when we have very strong evidence against it. Third a refutation of this hypothesis may justify political interventions like for example investments in the transportation infrastructure of regions where misallocation seems to be present. But if economic resources for such interventions are scarce then one clearly wants to be sure that they are not wasted and only invested in regions that really have inefficient factor markets. Thus a low type 1 error is desirable.

It is helpful to contrast the procedure developed here with the previous literature that assumes Cobb-Douglas production functions with some given values for the elasticity parameters. The non-rejection region implicit in such an approach consists of one single point in figure 2. This shows that the approach of the prior literature basically always rejects an efficient factor allocation and these rejections may be incorrect unless the used parameter values are exactly equal to the unknown true elasticities.

Finally, note that instead of testing for an efficient allocation ($d^L_{ab} = 1$ and $d^K_{ab} = 1$ for all pairs) one can of course also test other hypotheses on arbitrary values of $d^L_{ab}$ and $d^K_{ab}$ different from one. In this case one checks whether the observed $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combinations satisfy the conditions of proposition 1 given the specific values of $d^L_{ab}$ and $d^K_{ab}$ that one wishes to test. Thus the developed procedures cannot only shed light on whether the factor allocation deviates from equalised marginal products, but also allow investigations of the magnitude of such deviations.

## 2.4 Set of Possible Marginal Product Differentials

This section characterises the set of unobserved marginal product differentials $d^L_{ab}$ and $d^K_{ab}$ between a pair of production units $(a, b)$ that are consistent with a specific observed $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combination. This provides information on the extent of misallocation in the sense of the possible magnitude of the marginal product differentials underlying an observed allocation. Characterizing this set requires no further assumptions and its form is described by the following corollary which follows directly from proposition 1.

**Corollary 1.** *If two production units $(a, b)$ operate with well-behaved and homogenous production functions of degree $\lambda_a$ and $\lambda_b$ and the factor allocation involves an average product of labour ratio $\frac{y_b}{y_a}$ and a capital intensity ratio $\frac{k_b}{k_a}$ between the production units then the marginal product differentials of labour $d^L_{ab}$ and capital $d^K_{ab}$ between the production units satisfy either*

$$\begin{array}{lcll} d^L_{ab} > \widetilde{d}^L_{ab} & \textit{and} & d^K_{ab} < \widetilde{d}^K_{ab}, & \textit{or} \\ d^L_{ab} < \widetilde{d}^L_{ab} & \textit{and} & d^K_{ab} > \widetilde{d}^K_{ab}, & \textit{or} \\ d^L_{ab} = \widetilde{d}^L_{ab} & \textit{and} & d^K_{ab} = \widetilde{d}^K_{ab}, & \end{array}$$

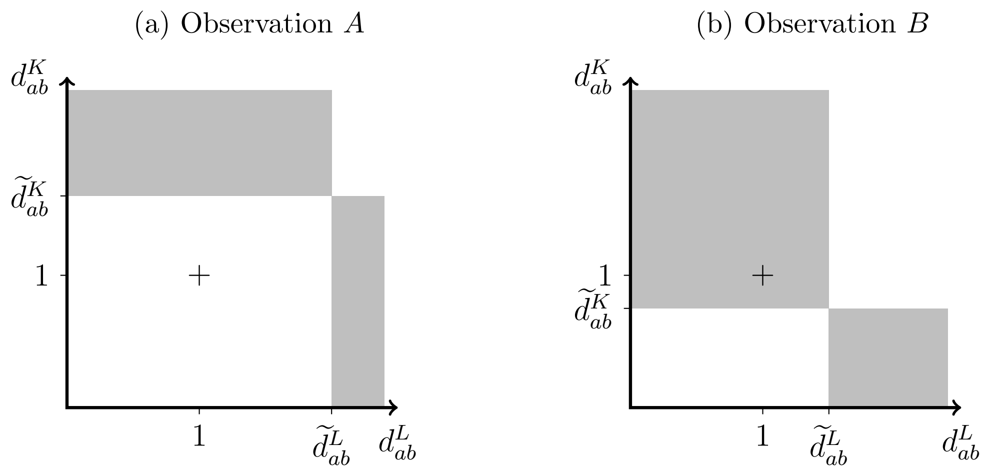
*where*

$$\widetilde{d}^L_{ab} \equiv \frac{\frac{y_b}{y_a}}{\frac{\lambda_a}{\lambda_b}} \quad \textit{and} \quad \widetilde{d}^K_{ab} \equiv \frac{\frac{y_b}{y_a}}{\frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a}}.$$

The corollary shows that each $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combination can in principle be generated by many different combinations of marginal product differentials $(d^L_{ab}, d^K_{ab})$. The set of possible $(d^L_{ab}, d^K_{ab})$ is summarised by the two boundary terms $\widetilde{d}^L_{ab}$ and $\widetilde{d}^K_{ab}$ which in turn depend on the observed $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combination and the assumed degrees of homogeneity $\lambda_a$ and $\lambda_b$. The corollary implies that either $d^L_{ab}$ is larger than $\widetilde{d}^L_{ab}$ or $d^K_{ab}$ is larger than $\widetilde{d}^K_{ab}$. In the former case $d^K_{ab}$ needs to be smaller than

$\widetilde{d}_{ab}^{K}$ and in the latter $d_{ab}^{L}$ needs to be smaller than $\widetilde{d}_{ab}^{L}$. But it is not possible to simultaneously place a lower bound (or an upper bound) on both differentials $d_{ab}^{L}$ and $d_{ab}^{K}$. The intuition can be understood graphically by noting how changes to ($d_{ab}^{L}$,$d_{ab}^{K}$) shift the straight lines in figure 1. There are many ($d_{ab}^{L}$,$d_{ab}^{K}$) that shift the shaded area such that it encompasses some observed ($\frac{y_b}{y_a}$, $\frac{k_b}{k_a}$) combination, say for example observation $A$ in figure 2. The exact set of possible ($d_{ab}^{L}$,$d_{ab}^{K}$) combinations is characterised by the boundary terms ($\widetilde{d}_{ab}^{L}$,$\widetilde{d}_{ab}^{K}$), which represent the values of ($d_{ab}^{L}$,$d_{ab}^{K}$) yielding an intersection of the horizontal and the upward-sloping line in figure 1 exactly at the observed ($\frac{y_b}{y_a}$, $\frac{k_b}{k_a}$) combination. The shape of the set of possible marginal product differentials is illustrated by the shaded area in figure 3 for the two examples of observed allocations $A$ and $B$ considered in figure 2.

Figure 3: Illustration of Corollary 1




Notes: Each graph refers to one of the hypothetical observations $A$ and $B$ in figure 2. The shaded area represents the marginal product differentials which may underly the respective observation ($\frac{y_b}{y_a}$, $\frac{k_b}{k_a}$) under the basic assumptions and given specific values of $\lambda_a$ and $\lambda_b$.

There are several important insights. First, if one observes an allocation ($\frac{y_b}{y_a}$, $\frac{k_b}{k_a}$) above the non-rejection region in figure 2 like observation $A$ then the boundary terms $\widetilde{d}_{ab}^{L}$ and $\widetilde{d}_{ab}^{K}$ are both necessarily larger than 1. Accordingly, at least one of the marginal product differentials $d_{ab}^{L}$ and $d_{ab}^{K}$ also needs to be larger than 1 such that at least for the respective factor the marginal product must be higher in unit $b$ than in $a$. Conversely, if one observes an allocation below the non-rejection region then $\widetilde{d}_{ab}^{L}$ and $\widetilde{d}_{ab}^{K}$ are necessarily both smaller than 1 such that at least for one factor the marginal product needs to be higher in unit $a$ than in $b$.

Second, the point $d_{ab}^{L} = 1$ and $d_{ab}^{K} = 1$ representing an equalization of marginal products (marked by "+") is of course not part of the shaded area for observations where the test rejects efficiency like observation $A$. In contrast, this point is included in the admissible set for observations where the test does not reject efficiency like observation $B$. However such an observation is in principle also consistent with marginal product differentials substantially different from 1, even though the test could not reject the null hypothesis of an efficient factor allocation. This illustrates the point raised earlier that not being able to reject the null hypothesis does not imply that the allocation is necessarily efficient.

Third, the boundary term $\widetilde{d}_{ab}^L$ depends positively on the observed $\frac{y_b}{y_a}$ and $\widetilde{d}_{ab}^K$ depends positively on the observed ratio $\frac{y_b}{y_a}/\frac{k_b}{k_a}$. Thus a larger vertical distance of the observed $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ from the non-rejection region in figure 2 implies a larger $\widetilde{d}_{ab}^L$, and if the ray from the origin to the observed $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ has a larger slope then $\widetilde{d}_{ab}^K$ is larger. A simultaneous increase of $\widetilde{d}_{ab}^L$ and $\widetilde{d}_{ab}^K$ then implies that for at least one factor the minimally possible deviation from marginal product equalization also increases. In this sense a higher distance of the observed allocation from the non-rejection region of figure 2 is informative on the magnitude of the possible underlying marginal product differentials.

## 2.5 Stronger Assumptions on Output Elasticities

In many potential applications researchers may want to make stronger assumptions on output elasticities based on prior information or beliefs. This section presents an extension that allows to impose such assumptions and explains how this tightens the set of allocations that may be consistent with equalised marginal products. This increases the risk of incorrectly rejecting an efficient allocation, but it decreases the risk of failing to reject efficiency when the allocation is not efficient. Thus, this extension allows to trade off the probability of making these two types of errors.

The only modification to the framework is that instead of requiring output elasticities to only be larger than zero I now introduce general lower bounds for each elasticity given by $\varepsilon_{Li} > \theta_{Li}$ and $\varepsilon_{Ki} > \theta_{Ki}$ where the parameters $\theta_{Li}$ and $\theta_{Ki}$ represent lower bounds on the output elasticities of the respective factors. These lower bounds need to be consistent with the degree of homogeneity $\lambda_i$ and equation (8). This implies that the parameters $\theta_{Li}$ and $\theta_{Ki}$ need to satisfy $\theta_{Li} \geq 0$, $\theta_{Ki} \geq 0$ and $\theta_{Li} + \theta_{Ki} < \lambda_i$. The parameters $\theta_{Li}$ and $\theta_{Ki}$ for each production unit need to be set by the researcher based on prior information about output elasticities. Together with equation (8) the bounds on output elasticities are then given by $\varepsilon_{Li} \in (\theta_{Li}, \lambda_i - \theta_{Ki})$ and $\varepsilon_{Ki} \in (\theta_{Ki}, \lambda_i - \theta_{Li})$.

The resulting specification nests the case considered in sections 2.2 and 2.3 when all parameters $\theta_{Li}$ and $\theta_{Ki}$ are set to zero, and allows to flexibly tighten the test procedure. For any specified values of $\theta_{La}$, $\theta_{Ka}$, $\theta_{Lb}$ and $\theta_{Kb}$ for a pair of production units $(a, b)$ and specific values of the marginal product differentials $d^L_{ab}$ and $d^K_{ab}$ one can then again characterise the set of $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combinations which are in principle consistent with such a situation. This is formalised in the following proposition, which is the equivalent to proposition 1 for the more general formulation with lower bounds on output elasticities.

**Proposition 2.** *If two production units $a$ and $b$ operate with output elasticities of labour and capital bounded from below by $\theta_{La}$, $\theta_{Ka}$, $\theta_{Lb}$ and $\theta_{Kb}$ respectively and their production functions are homogenous of degree $\lambda_a$ and $\lambda_b$, and the factor allocation exhibits marginal product differentials of labour $d^L_{ab}$ and capital $d^K_{ab}$ between the production units then the average product of labour ratio $\frac{y_b}{y_a}$ and capital intensity ratio $\frac{k_b}{k_a}$ at this allocation satisfy either*

$$\frac{k_b}{k_a} > \frac{d^L_{ab}}{d^K_{ab}} \quad \text{and} \quad \max\{\overline{\phi}, \overline{\psi}\} < \frac{y_b}{y_a} < \min\{\underline{\phi}, \underline{\psi}\}, \quad \text{or}$$
$$\frac{k_b}{k_a} < \frac{d^L_{ab}}{d^K_{ab}} \quad \text{and} \quad \max\{\underline{\phi}, \underline{\psi}\} < \frac{y_b}{y_a} < \min\{\overline{\phi}, \overline{\psi}\}, \quad \text{or}$$
$$\frac{k_b}{k_a} = \frac{d^L_{ab}}{d^K_{ab}} \quad \text{and} \quad \frac{y_b}{y_a} = \frac{\lambda_a}{\lambda_b} d^L_{ab}$$

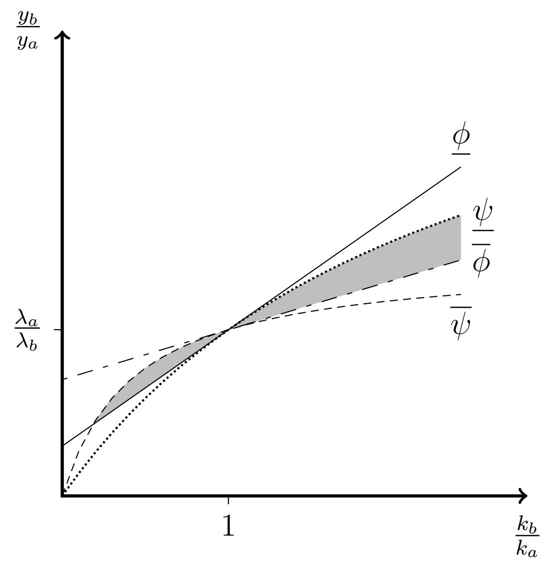
*where*

$$\overline{\phi} = \frac{\lambda_a - \theta_{Ka}}{\lambda_b} d^L_{ab} + \frac{\theta_{Ka}}{\lambda_b}\frac{k_b}{k_a} d^K_{ab}\,, \qquad \underline{\phi} = \frac{\theta_{La}}{\lambda_b} d^L_{ab} + \frac{\lambda_a - \theta_{La}}{\lambda_b}\frac{k_b}{k_a} d^K_{ab}\,,$$
$$\overline{\psi} = \frac{\lambda_a \frac{k_b}{k_a} d^K_{ab}}{\theta_{Kb} + (\lambda_b - \theta_{Kb})\frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}}}\,, \qquad \underline{\psi} = \frac{\lambda_a \frac{k_b}{k_a} d^K_{ab}}{\lambda_b - \theta_{Lb} + \theta_{Lb}\frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}}}\,.$$

Note that proposition 2 is identical to proposition 1 when all parameters $\theta_{La}$, $\theta_{Ka}$, $\theta_{Lb}$ and $\theta_{Kb}$ are set to zero.

Given the assumed values of $\lambda_a$, $\lambda_b$, $\theta_{La}$, $\theta_{Ka}$, $\theta_{Lb}$ and $\theta_{Kb}$ the test for an efficient factor allocation then proceeds as before, but now consists in checking whether an observed $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combination satisfies the conditions of proposition 2 for $d^L_{ab} = 1$ and $d^K_{ab} = 1$. When at least one of the parameters $\theta_{La}$, $\theta_{Ka}$, $\theta_{Lb}$ or $\theta_{Kb}$ is unequal to zero the set of $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combinations that are consistent with an efficient allocation changes. Figure 4 illustrates this by presenting a case where all these parameters are set to a positive value. Unsurprisingly introducing the lower bounds on output elasticities shrinks the set of observations that may be consistent with an efficient factor allocation (the shaded region). The reason is simply that with lower bounds $\theta_{Li}$ and $\theta_{Ki}$ certain combinations of elasticities and thus elasticity ratios $\frac{\varepsilon_{La}}{\varepsilon_{Lb}}$ and $\frac{\varepsilon_{Ka}}{\varepsilon_{Kb}}$ are ruled out compared to the previous case. Thus the $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combinations associated with these elasticity combinations are no longer possible.

Figure 4: The Efficiency Test with Lower Bounds on Output Elasticities




Notes: The graph shows a situation where all the lower bounds on output elasticities are set to a positive value. The shaded area represents the non-rejection region of the test, where one cannot reject an efficient factor allocation.

The set of marginal product differentials $d_{ab}^L$ and $d_{ab}^K$ that are consistent with an observed $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combination takes a more complicated shape in this case with lower bounds on output elasticities. Thus I do not provide the equivalent to corollary 1 here. However it is simple to solve for these sets computationally. One can specify a grid consisting of combinations of $d_{ab}^L$ and $d_{ab}^K$ and check which of these grid points satisfy the conditions of proposition 2 for the observed $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combination. With densely spaced grid points this provides a good approximation to the boundaries of the true set. The resulting set is a subset of the one characterised in corollary 1 and allows to draw sharper conclusions on the magnitude of marginal product differentials.

# 3 Bounds on Potential Output Gains

Factor misallocation implies that the economy could produce more output in total with the given factor endowments. This section shows how one can compute bounds on the potential output gains associated with an elimination of misallocation. This requires additional assumptions. But as in the previous main section the framework continues to not assume exact knowledge of the output elasticities of capital and labour and instead exploits knowledge on the possible range of output elasticities for homogeneous production functions.

## 3.1 Concept

The aim is to investigate the potential output gain associated with moving from the current allocation to an efficient allocation where the distortions to the factor allocation are lifted. I focus on static output gains in the sense that the total endowment of factors and the technology levels of different production units are kept constant. It is of course possible that lifting the distortions also induces faster factor accumulation and technological progress, but investigating such additional dynamic gains is left for future research. The potential output gain $G$ expressed as a fraction of current output reads as

$$G = \frac{Y^* - Y}{Y} \tag{9}$$

where $Y = \sum_{i=1}^{N} p_i Y_i$ denotes the total value of output across the $N$ production units for the actually observed allocation and $Y^* = \sum_{i=1}^{N} p_i Y_i^*$ for the hypothetical efficient allocation. A variable without a star refers to the observed current allocation and a variable with a star to the hypothetical allocation in an equilibrium without distortions. Note that in order to calculate a "real" total output gain, I evaluate both $Y$ and $Y^*$ at the set of prices observed for the current allocation.

## 3.2 Additional Assumptions

Both the current and hypothetical efficient allocation are viewed as the result of an equilibrium model. Determining the counterfactual efficient factor allocation without distortions then requires further assumptions in two respects. First one now needs to make more specific functional form assumptions on production functions. Here I rely on the standard assumption of Cobb-Douglas production functions with degrees of homogeneity $\lambda_i$ for each production unit given by

$$Y_i = A_i K_i^{\alpha_i} L_i^{\lambda_i - \alpha_i} \tag{10}$$

where $\alpha_i$ represents the output elasticity of capital of production unit $i$. The output elasticity of labour is then $\lambda_i - \alpha_i$. $A_i$ represents total factor productivity of a production unit, which includes for example the effect of all factors of production other than capital and labour. It is assumed that $A_i$ is fixed and remains unchanged when moving to the optimal allocation. Importantly, as in the previous main section the magnitude of output elasticities and hence the parameters $\alpha_i$ are not assumed to be known. Instead one again only has knowledge about the range in which output elasticities may fall based on the homogeneity assumption. Thus one can only place upper and lower bounds on the potential output gain.

Second one needs to make assumptions on the demand for the different goods. These assumptions together with those concerning the supply side will then determine the new equilibrium factor allocation when one lifts the distortions. One important property of the demand side is the degree to which consumers are willing to substitute different goods. This affects to what extent resources can be reallocated from production units with low marginal value products to those with high marginal value products at the current allocation. In other words it determines how strong any opposing relative price changes are that may limit the potential for factor reallocation. The demand side is modelled as just one representative consumer here. As a benchmark I assume that preferences are represented by a Cobb-Douglas utility function $U(C_1, \ldots, C_N)$ given by

$$U(C_1, \ldots, C_N) = \prod_{i=1}^{N} C_i^{\beta_i} \tag{11}$$

where $C_i$ is the number of consumed goods of production unit $i$ and $\beta_i > 0$ are parameters of the utility function satisfying $\sum_{i=1}^{N} \beta_i = 1$. The specification implicitly assumes that the elasticity of substitution is the same for any pair of goods and takes a common value of 1. This implies that the share of the total budget spent on each good is unaffected by the presence or absence of factor misallocation. Though this is unlikely to hold perfectly in reality such a specification still seems like a useful benchmark. In the empirical application a production unit is a manufacturing industry and Hsieh and Klenow (2009) also use a Cobb-Douglas specification to determine demand for each industry. Another advantage of this specification is that the potential output gain for given output elasticities then exhibits an analytical solution. However the effect of alternative values of the elasticity of substitution is explored in section 7.2.

Furthermore, it is assumed that the market for the different output goods is undistorted and exhibits market-clearing such that $Y_i = C_i$ for all $i$. Thus at the current allocation the marginal rates of substitution of the representative consumer are equal to the current relative prices between goods. The preference parameters $\beta_i$ can then be backed out from the observed allocation because the demand functions for Cobb-Douglas preferences and market-clearing imply $\beta_i = \frac{p_i Y_i}{Y}$. Thus $\beta_i$ is determined by the share of the value of output of production unit $i$ in the total value of output across all units as observed at the current allocation.

## 3.3 Potential Output Gain for Given Output Elasticities

The overall aim is to place bounds on the potential output gain which do not require knowledge on the exact magnitude of output elasticities and hence the parameters $\alpha_i$. However in order to derive these bounds step by step this subsection first derives the potential output gain as a function of some given parameters $\alpha_i$, which will be denoted as $G(\alpha)$ where $\alpha = (\alpha_1, \ldots, \alpha_N)$. This in turn requires to characterise the hypothetical efficient allocation $Y^*$ for a given vector $\alpha$.

Define the share of total capital and labour employed in production unit $i$ by $\kappa_i \equiv \frac{K_i}{\overline{K}} \in [0,1]$ and $\ell_i \equiv \frac{L_i}{\overline{L}} \in [0,1]$, respectively. Using these definitions the production function (10) of each production unit $i$ may be written as $Y_i = A_i \overline{K}^{\alpha_i} \overline{L}^{\lambda_i - \alpha_i} \kappa_i^{\alpha_i} \ell_i^{\lambda_i - \alpha_i}$. The potential output gain $G(\alpha)$ is then given by

$$G(\alpha) = \left( \sum_{i=1}^{N} \omega_i \; (\kappa_i^*)^{\alpha_i} (\ell_i^*)^{\lambda_i - \alpha_i} \right) - 1 \tag{12}$$

where $\omega_i = p_i A_i \overline{K}^{\alpha_i} \overline{L}^{\lambda_i - \alpha_i} / Y$ is a weighting term that captures for each $i$ how the Cobb-Douglas aggregate of fractions of total labour and capital map into the potential output gain. Note that $\omega_i$ is directly implied by the observed factor allocation because by the production function $\omega_i = s_i / (\kappa_i^{\alpha_i} \ell_i^{\lambda_i - \alpha_i})$ where $s_i = \frac{p_i Y_i}{Y}$ is the share of the value of output of production unit $i$ in the total value of output across all units at the current allocation.

The values of $\kappa_i^*$ and $\ell_i^*$ at the efficient allocation are determined by solving the problem of a social planner who maximises utility subject to the physical constraints. For the assumed production and utility functions this reads as

$$\max_{\{\kappa_i^*,\ell_i^*\}_{i=1}^N} \quad \prod_{i=1}^{N} \left((\kappa_i^*)^{\alpha_i}(\ell_i^*)^{\lambda_i-\alpha_i}\right)^{\beta_i} \tag{13}$$

subject to the resource constraints $\sum_{i=1}^N \kappa_i^* = 1$ and $\sum_{i=1}^N \ell_i^* = 1$, and non-negativity constraints $\kappa_i^* \geq 0$ and $\ell_i^* \geq 0$ for all production units $i$. Note that for Cobb-Douglas preferences the terms $A_i\overline{K}^{\alpha_i}\overline{L}^{\lambda_i-\alpha_i}$ can be omitted in equation (13) without affecting the result. The social planner problem exhibits an analytical solution given by

$$\kappa_i^* = \frac{\alpha_i\beta_i}{\sum_{j=1}^N \alpha_j\beta_j} \tag{14}$$

$$\ell_i^* = \frac{(\lambda_i-\alpha_i)\beta_i}{\sum_{j=1}^N(\lambda_j-\alpha_j)\beta_j} \tag{15}$$

for all production units $i$, which follows directly from the first-order conditions. Plugging this solution into equation (12) one obtains the potential output gain $G(\alpha)$ for a given vector of output elasticities $\alpha$.

## 3.4 Bounds on Potential Output Gains

The previous subsection involves the potential output gain for observed $(s_i, \kappa_i, \ell_i)$ and specific given values of $\alpha_i$ for each production unit. However given the great difficulty to reliably determine output elasticities, the aim is to investigate potential output gains without requiring knowledge of their exact magnitude. Instead I exploit the possible range of output elasticities implied by homogeneous production functions, which amounts to $\alpha_i \in (0, \lambda_i)$. Bounds on the potential output gain for the observed values of $(s_i, \kappa_i, \ell_i)$ then simply consist of the lowest and highest output gains $G(\alpha)$ when searching over these admissible values $\alpha \in A$. Here $A$ is the set for which the $i$-th element of $\alpha$ satisfies $\alpha_i \in (0, \lambda_i)$. Note that given the observed $(s_i, \kappa_i, \ell_i)$ combination for each production unit each possible vector $\alpha$ is associated with a specific combination of marginal product differentials for each pair of production units. Formally the lower bound on the potential output gain $\underline{G}$ from moving to an efficient allocation is then given by $\underline{G} = \inf_{\alpha\in A} G(\alpha)$ and the upper bound $\overline{G}$ by $\overline{G} = \sup_{\alpha\in A} G(\alpha)$.

In the application these bounds are solved numerically.[4] The computed bounds

[4] In the numerical solution the search on open intervals is approximated by searching over closed intervals where the endpoints are shifted inside the original interval by a small value. In the

show what conclusions on potential output gains one can already draw without exact knowledge of output elasticities and instead exploiting only their possible range as implied by the homogeneity assumption. In the application I also compare these bounds to the potential output gains obtained for a traditional approach that sets output elasticities exactly equal to observed factor income shares of each production unit. This allows to better understand how robust such traditional point estimates are.

### 3.5 Stronger Assumptions on Output Elasticities

The bounds on the potential output gains from eliminating misallocation can also be computed under stronger assumptions on output elasticities. One can again specify lower bounds on each elasticity represented by the parameters $\theta_{Li}$ and $\theta_{Ki}$ as explained in section 2.5. The computation of the bounds then evolves as in the previous subsection with a small modification. The admissible values of $\alpha$ for the infimum and supremum of $G(\alpha)$ are now such that $\alpha_i \in (\theta_{Ki}, \lambda_i - \theta_{Li})$. Imposing such stronger assumptions tightens the bounds on potential output gains. A suitable choice of these parameters allows the researcher to trade off generality and prior beliefs on output elasticities, and to flexibly investigate the robustness of point estimates of potential output gains.

## 4 Data on U.S. Manufacturing Industries

The rest of the paper applies the framework to study the allocation of labour and capital between different manufacturing industries in the United States. Section 5 analyses the year 2005 and section 6 investigates the dynamics of misallocation during 2005-2009 with a focus on the Great Recession years 2008 and 2009.

The main data set for the application is the NBER-CES Manufacturing Industry Database (the April 2016 version), which provides annual data on different industries of the United States manufacturing sector between 1958 and 2011.[5] This data set is in turn based on data from Manufacturing Censuses. The database contains information on 473 industries at the six-digit level defined according to

application I use a shift of the size 0.0001 such that $\alpha_i \in (0, \lambda_i)$ is replaced by $\alpha_i \in [0.0001, \lambda_i - 0.0001]$. The solution to these optimization problems may well occur at the boundaries. Thus I use two different numerical methods to solve for the optimum: a standard derivative based optimization algorithm and a hill climbing algorithm that exclusively searches on the boundaries. Each method uses several different starting values. The algorithm then picks the best solution found by these two methods.

[5] The data set can be obtained through the NBER website: http://www.nber.org/nberces/

the North American Industrial Classification System (NAICS). I only use the data for the years 2005-2009.

The database contains information on the value added and real capital stock of each industry $i$ in each of the years $t$ (variables *vadd* and *cap*), which are used directly as measures of the value of output $p_{it}Y_{it}$ and the capital input $K_{it}$. In order to accurately measure the effective labour input I adjust the number of employees for differences between industries in their average human capital and hours of work. The aim is to address an important measurement concern in work on misallocation. The labour input $L_{it}$ in industry $i$ and year $t$ is then measured as $L_{it} = h_{it}n_{it}N_{it}$ where $N_{it}$ denotes the number of employees, $n_{it}$ denotes their average yearly hours and $h_{it}$ their average human capital. The number of employees $N_{it}$ is directly taken from the database (variable *emp*). In contrast the database only allows to compute average hours of production workers, but not of all employees. This is based on total production worker hours and the number of production workers (variables *prodh* and *prode*). Though this is a limitation of the data, this is used as the measure of average hours $n_{it}$. Unfortunately, the database contains no information on the human capital of employees in different industries.

In order to control for potential differences in human capital between industries I rely on the IPUMS-USA database of the Minnesota Population Center.[6] This data is based on repeated cross-sections from census records and contains information at an individual level on wage income, education, demographic characteristics and the industry where an individual is employed.

Individual wages are modeled using a Mincerian regression as

$$\log w_{jt} = \gamma + \delta x_{jt} + \zeta z_{jt} + \epsilon_{jt} \tag{16}$$

where $w_{jt}$ is the hourly wage, $x_{jt}$ is a vector of variables determining human capital and $z_{jt}$ are further controls for individual $j$ and year $t$. Here I model human capital as a function of gender, education and experience. Accordingly, $x_{jt}$ contains a gender dummy, educational attainment dummies with 28 categories in total and a cubic polynomial in an individual's age to capture the effect of experience. The control variables $z_{jt}$ contain year dummies and industry dummies. The industry dummies control for industry wage differentials which are not driven by human capital differences, but by potential distortions between industries. The ability to control for the presence of such wage differentials is an important advantage in the context of an analysis of misallocation. In contrast using the total payroll as

[6] The data is available online at: https://usa.ipums.org/usa/

a measure of the labour input as implemented for example by Hsieh and Klenow (2009) may confound true human capital differences with wage differentials which are driven by distortions. I then construct an estimate of each individual's human capital stock as a function of their gender, education and experience. For this purpose I first estimate regression (16) using data for the years 2005-2009 and all individuals who work in one of the manufacturing industries, which yields about $770,000$ total observations. The estimated coefficient vector of the variables determining human capital $\widehat{\delta}$ is then used to construct a measure of the human capital stock of each individual as $\widehat{h}_{jt} = \exp(\widehat{\delta} x_{jt})$.

Finally, these estimates of individual human capital stocks $\widehat{h}_{jt}$ are used to compute the average human capital $h_{it}$ of employees in each industry $i$ and year $t$. One limitation of this construction is that the IPUMS data does not contain the industry code at the six-digit level as the NBER-CES data does, but usually only at the four-digit or an even coarser level. As a consequence about 60% of the 473 six-digit industries in the main data set are assigned the average human capital stock of their corresponding four-digit industry (or in rare cases a finer level). The remaining about 40% of all six-digit industries are assigned the average human capital stock of an aggregate of two or three four-digit industries. In other words the constructed human capital stocks usually only capture potential human capital differences across industries at the four-digit level, but not within the group of six-digit industries belonging to the same four-digit industry.

# 5 Analysis of the Factor Allocation in 2005

This section presents the assumptions and results of the analysis of the capital and labour allocation across U.S. manufacturing industries in the year 2005.

## 5.1 Assumptions and Scenarios

First one needs to specify the degree of homogeneity $\lambda_i$ with respect to capital and labour of the production function of each industry $i$. I follow most of the literature and assume constant returns to scale such that $\lambda_i$ is set to 1 for all industries. The sensitivity of the results to this choice is discussed in section 7.1.

The framework also allows to set the parameters $\theta_{Ki}$ and $\theta_{Li}$ representing lower bounds on the output elasticities of capital and labour in each industry. A suitable choice of these parameters enables the researcher to trade-off generality of assumptions and prior beliefs on these elasticities. There are of course many potentially informative ways to set these parameters. In order to keep the analysis focussed I restrict the investigation to two scenarios here.

In the first scenario I set $\theta_{Ki}$ and $\theta_{Li}$ to zero for all industries such that output elasticities are only required to be positive. This scenario shows what conclusions one can draw from the data using only the weak basic assumptions on production functions and no further restrictions. I refer to this as the "most general" scenario.

The second scenario is referred to as one with "stronger assumptions" and strikes a compromise between generality and prior beliefs on output elasticities. In the literature it is common practice to determine the value of output elasticities by the observed income share of the respective factor, which relies on factor prices being equal to marginal products. Here I set $\theta_{Ki}$ and $\theta_{Li}$ such that the resulting admissible range of output elasticities is closer to the values obtained by this traditional approach, but still considerably more general. Specifically, I set $\theta_{Ki}$ and $\theta_{Li}$ such that the admissible range of the output elasticity of labour of an industry includes values that are up to 0.2 higher or lower than the observed labour income share of that industry in 2005. For example for a labour income share of about 0.7 the output elasticity of labour could vary between 0.5 and 0.9, and the output elasticity of capital between 0.1 and 0.5. Of course output elasticities still need to sum to the assumed degree of homogeneity. The resulting intervals seem large enough to allow for considerable deviations of marginal products from factor prices and problems in the measurement of factor income shares, which cause a deviation of output elasticities from measured factor income shares. But the intervals are still centred around this traditional approach. Formally, denoting the labour income share of industry $i$ by $\tau_i$ the parameters are set such that $\theta_{Li} = \tau_i - 0.2$ and $\theta_{Ki} = \lambda_i - (\tau_i + 0.2)$ for each industry. In case this yields a negative value for one of the parameters this parameter is reset to zero, and in case it yields a value above $\lambda_i - 0.2$ the parameter is reset to $\lambda_i - 0.2$.

The labour income share of industry $i$ is computed by dividing the total payroll by the value added of each industry (using the variables *pay* and *vadd* in the NBER-CES database). The total payroll variable *pay* in the NBER-CES database omits employer payments for social security or fringe benefits. Thus, Hsieh and Klenow (2009) adjust the observed industry labour income shares in this data by a factor of 3/2 to scale them up to the labour income share of manufacturing observed in the National Income and Product Accounts. I follow their approach.[7]

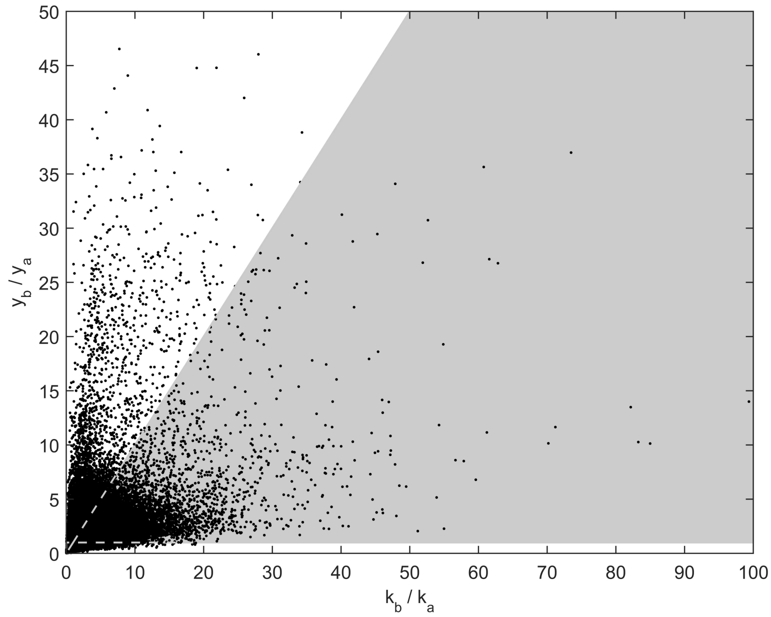
[7] Measured labour income shares vary widely across industries with a value of 0.26 at the 10th and 0.69 at the 90th percentile. Though observed factor income shares are an imperfect measure of output elasticities, this suggests that output elasticities may also vary substantially across industries. This provides another motivation for applying robust methods in this context.

## 5.2 Results

This subsection presents the results for the capital and labour allocation across the 473 U.S. manufacturing industries in 2005. The test for an efficient factor allocation across industries is first conducted for the "most general" scenario. This is illustrated by figure 5 which plots the observed average product of labour and capital intensity ratios $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ for each of the $111,628$ industry pairs along with the shaded non-rejection region of the test. Note that for this scenario the non-rejection region is identical for all industry pairs. The test rejects an efficient factor allocation across the 473 industries. In other words there is no combination of constant returns to scale production functions across industries that can rationalise the observed allocation as one with equalised marginal products. Moreover it is a strong rejection because almost half of all pairs (48.7%) are observed in the rejection region as reported in column 3 of table 1. Furthermore figure 5 shows that many observations are located very far away from the non-rejection region. Of course the test then also rejects an efficient allocation for the scenario with "stronger assumptions". But now the vast majority of all industry pairs (80.2%) are observed in the rejection region associated with this scenario. For this scenario the non-rejection region is specific to each industry pair such that the test cannot be illustrated in one common graph. But a visual inspection of examples of these non-rejection regions shows that they are much smaller than for the most general scenario. This means that even relatively weak restrictions on the range of output elasticities strongly affect which $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combinations can be rationalised as an efficient allocation. Overall these results provide strong and robust evidence against the efficiency hypothesis. They show that deviations from equalised marginal products must be frequent and of a sizeable magnitude even for the most general assumptions.

The framework also allows to characterise for each industry pair the set of all combinations of marginal product differentials $d^L_{ab}$ and $d^K_{ab}$ that may underly the observed allocation for the given assumptions on output elasticities. These sets provide several insights on the magnitude and type of wedges between marginal products for each industry pair. But since these sets are by nature high dimensional objects for each pair, it is somewhat difficult to summarise this information in detail across pairs. One insight is for example that for the $54,310$ industry pairs for which the test rejects efficiency under the most general assumptions the boundary values $(\widetilde{d}^L_{ab}, \widetilde{d}^K_{ab})$ of corollary 1 imply that 50% (20%) of these industry pairs must have either a marginal product differential of labour or capital in ex-

Figure 5: Test for an Efficient Factor Allocation (Most General Scenario)




Notes: The graph plots the $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combinations for each pair of industries observed in the data. The shaded area (also visualised by the grey dashed lines) represents the non-rejection region of the test. If an observed $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combination is not an element of the shaded region then one rejects the null hypothesis of an efficient factor allocation (for this pair and overall).

cess of 1.28 (1.72).[8] If one further narrows down the range of output elasticities as in the scenario with "stronger assumptions" then one can conclude that about 55% of these industry pairs, i.e. 29,709 pairs, necessarily exhibit a wedge between marginal products of capital. Within this more narrow group the median lower bound on this wedge takes a value of about 1.6. In contrast for these still relatively general assumptions there is only a tiny fraction of rejected pairs for which one can definitely rule out an equalization of marginal products of labour.

Columns 4 and 5 of table 1 report the bounds on the potential output gains associated with moving to an efficient allocation for the two scenarios. For both scenarios the lower bound of these potential output gains takes a value of about 22% of current output. In contrast the results for the upper bound differ markedly between scenarios. While the upper bound is 1541% for the "most general" scenario, it is only about 64% for the scenario with "stronger assumptions". In comparison a traditional approach that sets output elasticities exactly equal to

[8]For the statistics in this paragraph the units $a$ and $b$ of a certain pair are suitably ordered such that all marginal product differentials are larger than 1 and thus comparable across pairs.

observed factor income shares of each industry yields a point estimate for potential output gains of 28% here.[9]

Table 1: Results for 2005

| | Test Results | | Potential Output Gains (in %) | |
|---|---|---|---|---|
| Scenario | Reject | Rejected Pairs (in %) | Lower Bound | Upper Bound |
| Most General | YES | 48.7 | 21.5 | 1540.5 |
| Stronger Assumptions | YES | 80.2 | 22.0 | 63.5 |

Notes: "Most General": Output elasticities are only required to be positive. "Stronger Assumptions": Output elasticities are required to be positive and to deviate only up to 0.2 from observed factor income shares, cf. section 5.1. "Reject": Does the test reject an overall efficient factor allocation? "Rejected Pairs": Share of all industry pairs for which a pairwise efficient factor allocation can be rejected (in % of total pairs).

These results suggest several conclusions. First misallocation between industries is an economically significant phenomenon because even under the most general assumptions the lower bound on potential output gains already takes a value of 22%, which is certainly a large number in this context. This challenges the conventional wisdom that the United States is very close to an efficient allocation. Furthermore, this challenge depends only on relatively general assumptions. Second how much the true potential output gain could exceed this lower bound is less clear and does indeed depend more on the exact assumptions on output elasticities. In this application the point estimate of 28% of the traditional approach turns out to be surprisingly close to the lower bound. But if one considers a more general and still reasonable range of output elasticities as in the scenario with "stronger assumptions" then the potential output gains could be considerably higher and up to 64%. Only for output elasticities outside such a range does one obtain very large (and implausible) values of the upper bound on potential output gains as it is the case for the "most general" scenario here. Thus these results confirm that the exact value of output elasticities is in principle of first-order importance for the precise quantitative role of factor misallocation. However at least in this application even restrictive assumptions as in the prior literature do not necessarily imply a huge overstatement of potential output gains. This is encouraging for the prior misallocation literature.

[9] However a stronger version of the traditional approach assumes that the output elasticity is identical in all production units. If one uses such an assumption and sets all output elasticities equal to observed aggregate factor income shares then one obtains a higher potential output gain of about 35% in this application.

# 6 Dynamics of Misallocation

This section investigates the dynamics of misallocation during the Great Recession of 2008 and 2009. The question is whether the strong fall in economic activity during this time period was associated with a measurable increase in factor misallocation across industries and how large these economic effects were. In order to put the following results into perspective it is helpful to remember the magnitude of the observed contraction of the whole U.S. manufacturing sector during this time period. Data from the World Development Indicators shows that during 1997-2007 real manufacturing value added rose on average by about 3.5% per year, followed by a growth rate of about $-3\%$ in 2008 and $-8\%$ in 2009. Thus a very rough estimate of the size of the contraction is that in 2009 manufacturing output was about 18% below trend.

Though the term Great Recession typically refers to 2008 and 2009, I apply the theoretical framework to all years between 2005-2009 to also get a better picture of the time period before the Great Recession. The assumptions and the two scenarios are the same as in the previous main section, cf. section 5.1. There is only one minor difference. For the scenario with "stronger assumptions" I keep the values of $\theta_{Li}$ and $\theta_{Ki}$ constant over time at their 2005 values in order to isolate the effect of a changing observed factor allocation. However using year-specific values of $\theta_{Li}$ and $\theta_{Ki}$ defined in relation to year-specific factor income shares yields almost identical results. Also note that the application of the framework to different years does not impose any restrictions across years. Thus output elasticities, prices, technology levels and preference parameters may fully vary between years.

For brevity I only report the most informative results for the two parts of the framework and the two scenarios. Specifically, for the test procedure I focus on the results of the "most general" scenario. These are reported in columns two and three of table 2 for the different years. The test rejects an efficient factor allocation across industries in all years between 2005 and 2009. Furthermore the fraction of industry pairs for which the test rejects an efficient factor allocation is increasing over time from about 49% in 2005 to about 60% in 2009 at the height of the Great Recession. While there is already a weak positive trend in this share of rejected pairs during 2005-2007 the increase is much stronger in the Great Recession years of 2008 and 2009. Obviously these results imply that an efficient factor allocation is also rejected for the scenario with "stronger assumptions" and that at least some of the observed $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combinations move further away from the non-rejection region of that scenario. These results provide a first indication that factor misallocation may have worsened during the Great Recession.

The computed sets of marginal product differentials suggest that deviations from marginal product equalization have not only become more frequent, but also larger in magnitude. For example for the industry pairs for which the test rejects efficiency in 2009 under the most general assumptions the boundary values ($\widetilde{d}^L_{ab}$, $\widetilde{d}^K_{ab}$) imply that 50% (20%) of these industry pairs must have either a marginal product differential of labour or capital in excess of 1.4 (2.0). In contrast these values were 1.28 (1.72) in 2005. Several other statistics point in the same direction.

Table 2: Results for Different Years between 2005 and 2009

| | Test Results for Most General Scenario | | Potential Output Gains (in %) with Stronger Assumptions | |
|---|---|---|---|---|
| Year | Reject | Rejected Pairs (in %) | Lower Bound | Upper Bound |
| 2005 | YES | 48.7 | 22.0 | 63.5 |
| 2006 | YES | 50.2 | 21.5 | 61.3 |
| 2007 | YES | 51.9 | 22.9 | 62.2 |
| 2008 | YES | 55.4 | 24.2 | 64.0 |
| 2009 | YES | 59.7 | 28.1 | 71.8 |

Notes: "Most General": Output elasticities are only required to be positive. "Stronger Assumptions": Output elasticities are required to be positive and to deviate only up to 0.2 from observed factor income shares. "Reject": Does the test reject an overall efficient factor allocation? "Rejected Pairs": Share of all industry pairs for which a pairwise efficient factor allocation can be rejected (in % of total pairs).

Next I investigate the potential output gains for the different years. Here I focus on the scenario with "stronger assumptions" because section 5.2 showed that the upper bound of potential output gains is so implausibly large for the most general assumptions that it provides very little information. The bounds on potential output gains under the scenario with "stronger assumptions" are reported in columns 4 and 5 of table 2 for the different years. One observes that the range in which potential output gains need to fall shifts upwards over time. While the possible range of the potential output gain from eliminating misallocation is about $22-64\%$ in 2005, it is about $28-72\%$ in 2009. Thus the lower bound increases by about 6 percentage points and the upper bound by 8 percentage points. One observes that these ranges are fairly constant during 2005-2007, followed by a relatively weak shift in 2008 and a strong shift in 2009. These results provide a second piece of evidence for a worsening of factor misallocation during the Great Recession.

However these results do not directly imply that potential output gains necessarily increased between 2005 and 2009 because the two ranges still have considerable overlap. If potential output gains were relatively low in 2005 between about 22 and 28% then potential output gains must have increased because this interval is no longer part of the possible range in 2009. However if potential output gains were above 28% in 2005 then it is theoretically possible that they stayed constant or even decreased over time.

In order to shed more light on this issue I also compute bounds on the change in potential output gains. Specifically, define the change in potential output gains between year $s$ and year $t > s$ to be $\triangle G(\alpha) = G_t(\alpha) - G_s(\alpha)$ where $G_s$ and $G_t$ are the potential output gains in years $s$ and $t$. I then compute the lower bound $\underline{\triangle G}$ of the change in potential output gains by $\underline{\triangle G} = \inf_{\alpha \in A} \triangle G(\alpha)$ and the upper bound $\overline{\triangle G}$ by $\overline{\triangle G} = \sup_{\alpha \in A} \triangle G(\alpha)$.

Note that these calculations assume constant output elasticities and hence constant parameters $\alpha_i$ in the two time periods. This means that these calculations are prone to a similar critique as explained in the introduction for keeping elasticities constant across countries. A defense is that the fundamental technologies of an industry and hence the output elasticities are less likely to change strongly within five years than they are to differ between different countries that operate at very different technological levels. Nevertheless this major caveat remains. An alternative specification could allow at least some changes to the output elasticities of an industry across time periods when computing bounds on the change to potential output gains. This would be more in the general spirit of the paper, but is left for future research. However note that the levels of output elasticities of each industry are still only restricted to be within the assumed ranges and not fixed to some specific values. Variables such as prices, technology levels and preference parameters are still allowed to vary between the two years.

The first row of table 3 reports the bounds on the change to potential output gains between 2005 and 2009 for the standard scenario with "stronger assumptions". These results indicate that between 2005 and 2009 potential output gains changed between about $-4$ and $+16$ percentage points. Thus even by assuming that output elasticities are constant over time one cannot rule out that potential output gains decreased between 2005 and 2009. However if such a decrease occurred its maximum size was only 4 percentage points. In contrast it is possible that potential output gains increased by as much as 16 percentage points which would be a substantial worsening of factor misallocation. Faced with these findings it is interesting to see how much one needs to further strengthen the assumptions to identify a clear increase in potential output gains between 2005 and 2009. Thus the second row of table 3 presents the results for an alternative scenario with “stronger assumptions” which only allows deviations of output elasticities from observed factor income shares up to an absolute value of 0.1 (referred to as “Stronger Assumptions ($\tau_i \pm 0.1$)” in contrast to the standard scenario with $\tau_i \pm 0.2$). Under such assumptions the lower bound becomes strictly positive and the change in potential output gains must then be between about 2 and 11 percentage points. If output elasticities are set perfectly equal to observed factor income shares then one obtains an increase of potential output gains of about 6.4 percentage points.

Table 3: Changes to Potential Output Gains between 2005 and 2009

| Scenario | Lower Bound | Upper Bound |
|---|---|---|
| Stronger Assumptions ($\tau_i \pm 0.2$) | -4.2 | 15.9 |
| Stronger Assumptions ($\tau_i \pm 0.1$) | 1.5 | 11.2 |

Notes: Changes to Potential Output Gains between 2005 and 2009 are expressed in percentage points. “Stronger Assumptions $\tau_i \pm 0.2$ (0.1)”: Output elasticities are required to be positive and to deviate only up to 0.2 (0.1) from observed factor income shares.

Overall the section provides several pieces of evidence for an increase in factor misallocation during the Great Recession. The analysis shows that this increase in misallocation may play a quantitatively important role for explaining the fall in manufacturing output, but is unlikely to be the sole explanation. If one assumes that output elasticities are exactly equal to factor income shares then the contribution of increased misallocation is about 35% of the observed about 18 percentage point fall of manufacturing output below trend. For the more general scenario where output elasticities may deviate up to 0.1 from observed factor income shares the increase in misallocation explains between about 10 and 60% of this fall. Thus the possible quantitative contribution of an increase in misallocation for explaining the fall in manufacturing output is somewhat sensitive to the exact assumptions on output elasticities. The findings also suggest that to the extend one can mitigate the underlying frictions in the economy one can both permanently increase output and reduce the depth of economic crises such as the Great Recession.

# 7 Robustness Checks

This section provides robustness checks of the main results. It investigates the role of the assumed degrees of homogeneity of different industries and the preference specification underlying the demand for the output of different industries.

## 7.1 Degrees of Homogeneity $\lambda_i$

The parameters $\lambda_i$ of the different industries representing the degree of homogeneity of their production functions are the only input into the test procedure that needs to be specified by the researcher. Furthermore these parameters also enter the bounds on potential output gains. In the application I have set these parameters $\lambda_i$ to a value of 1 for all industries based on a constant returns to scale assumption. This section discusses how this choice affects the results.

First, consider setting these parameters to another value (below 1), but still using a common value for all industries. This leaves the test results for the "most general" scenario completely unaffected. The reason is that the shape of the non-rejection region of the test only depends on ratios $\frac{\lambda_a}{\lambda_b}$ of these parameters. These ratios remain constant when changing all parameters $\lambda_i$ by the same factor. But since the test continues to reject an efficient factor allocation for the "most general" scenario, it will also continue to reject for the scenario with "stronger assumptions" even though the exact shape of the non-rejection area may indeed change a bit for the latter scenario. Thus the main test results are unaffected by such alternative values of $\lambda_i$.

In contrast the bounds on potential output gains may be affected. In order to investigate the strength of this effect I recompute the bounds using an alternative value of $\lambda_i = 0.9$. This is motivated by the fact that the literature sometimes also assumes a mild degree of decreasing returns to scale, e.g. Restuccia and Rogerson (2008) or Sandleris and Wright (2014). Table 4 reports the resulting bounds for the years 2005 and 2009 for the scenario with "stronger assumptions" (with $\tau_i \pm 0.2$). One observes that all bounds are a bit lower for the case of $\lambda_i = 0.9$ compared to the benchmark. In 2005 one then obtains a range of potential output gains of about $19 - 61\%$ compared to $22 - 64\%$ for the benchmark, and in 2009 the range is $23 - 68\%$ relative to $28 - 72\%$ for the benchmark. Again one only finds a strictly positive change to potential output gains between 2005 and 2009 if one only allows deviations of output elasticities from observed factor income shares up to an absolute value of 0.1. However for $\lambda_i = 0.9$ the resulting range for the change to potential output gains is then $0-9$ percentage points such that this range is also a bit lower than the one obtained for the benchmark. If one sets output elasticities exactly equal to factor income shares then the change in potential output gains is 5.0 percentage points compared to 6.4 percentage points for the benchmark. Overall these results show that though the computed bounds are generally a bit lower, the broad conclusions on the presence of significant potential output gains and their increase over time are unaffected by such an alternative value of $\lambda_i = 0.9$.

Table 4: Bounds on Potential Output Gains in 2005 and 2009 for the Scenario with "Stronger Assumptions" for Alternative Degrees of Homogeneity $\lambda_i$

| Specification | 2005 | | 2009 | |
|---|---|---|---|---|
| | Lower | Upper | Lower | Upper |
| $\lambda_i = 1$ (Benchmark) | 22.0 | 63.5 | 28.1 | 71.8 |
| $\lambda_i = 0.9$ | 18.5 | 61.4 | 23.4 | 68.0 |

Finally, one may of course also consider setting different values of $\lambda_i$ for different industries. In practice investigating such a specification is hampered by the lack of strong evidence on how this parameter differs across the 473 industries. Nevertheless, one can still think about whether a reasonable variation of $\lambda_i$ across industries could make the observations consistent with an efficient factor allocation for the "most general" scenario. In order to do this, consider the data for the year 2005 in figure 5 again. There are for example many observations outside the non-rejection region with values of $\frac{k_b}{k_a}$ around 2 and $\frac{y_b}{y_a}$ around 5. Such an observation can only be efficient if $\frac{\lambda_a}{\lambda_b}$ is at least 5/2 and at most 5 (remember how the non-rejection region shifts when $\frac{\lambda_a}{\lambda_b}$ changes, cf. figure 2). Assuming that $\lambda_a = 1$ this would require $0.2 \leq \lambda_b \leq 0.4$. In other words explaining such an observation as part of an efficient factor allocation requires large (but not too large) differences in $\lambda_i$ between industries and strongly decreasing returns to scale in some industries. There are other observations on that figure for which the differences in $\lambda_i$ would need to be even larger. For the most extreme industry pair this would require one of the industries to have a value of $\lambda_b$ below about 0.034 given the other industry has $\lambda_a = 1$. Such large differences in returns to scale are inconsistent with empirical evidence and the common belief that production exhibits constant or mildly decreasing returns to scale. For example Gao and Kehrig (2017) report estimates of returns to scale for 82 different four-digit NAICS manufacturing industries. Their estimates of returns to scale of capital and labour in value added production tend to be below but not too far away from 1. Estimated returns to scale exhibit only a weak variation across industries with a value of about 0.79 for the 10th and 1.06 for the 90th percentile industry.[10]

[10]Their estimates refer to gross output, while this paper analyses value added. Thus I report a transformation of their estimates given by $(\beta_k + \beta_l)/(1 - \beta_m - \beta_e)$ here.

## 7.2 Preference Specification

The benchmark specification assumes Cobb-Douglas preferences, which implies that the elasticity of substitution is the same for any pair of goods and equal to 1. This assumption only affects the bounds on potential output gains, but not the test procedure. Conducting useful robustness checks on the preferences specification is difficult due to a lack of detailed empirical estimates of the elasticity of substitution for all possible industry pairs. Thus I only investigate the role of different values of the elasticity of substitution, but need to maintain the unrealistic assumption that this value is common for all industry pairs. In order to vary this parameter, I specify a Constant Elasticity of Substitution (CES) utility function given by

$$U(C_1, \dots, C_N) = \left[ \sum_{i=1}^{N} \beta_i C_i^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{17}$$

where $\sigma$ is the elasticity of substitution with $\sigma \neq 1$, and $\beta_i$ are share parameters of the utility function. The efficient allocation for given output elasticities is determined by maximizing this utility function subject to the same resource and non-negativity constraints as in the benchmark. Importantly, for a given value of $\sigma$ the parameters $\beta_i$ can again be determined using information from the current observed allocation, which allows to conveniently formulate the social planner problem (details in the appendix). Otherwise the bounds on potential output gains are computed as in the benchmark with the exception that the social planner problem for given values of $\alpha_i$ has to be solved numerically now.

Relative to the benchmark I investigate alternative specifications where the elasticity of substitution is either 50% higher or lower. This amounts to the values of $\sigma = 1.5$ and $\sigma = 0.5$.[11] Table 5 presents the resulting bounds on potential output gains for the standard scenario with "stronger assumptions" (with $\tau_i \pm 0.2$) and the years 2005 and 2009. One observes that generally the bounds on potential output gains depend positively on the elasticity of substitution. Both the lower and upper bound are larger for $\sigma = 1.5$ and smaller for $\sigma = 0.5$ relative to the benchmark specification. This is intuitive because a high willingness to substitute

[11] Broda and Weinstein (2006) estimate elasticities of substitution at the SITC 3-digit level (256 categories, in contrast to 473 categories here) in excess of these values with a median estimate of 2.2 for the most recent time period. However their estimates refer to elasticities of substitution among different goods *within* each of these 256 categories. In contrast for this paper the elasticity of substitution *across* such categories is relevant. Though I do not have direct estimates of these elasticities, it seems plausible that it is considerably more difficult to substitute across categories than within category. This motivates the investigation of a range of elasticities of substitution below these empirical estimates.

different goods allows to shift resources to their most productive use with only limited price changes in the opposite direction. However the broad patterns are the same in all specifications. For instance the lower bound on potential output gains is of an economically significant magnitude even for $\sigma = 0.5$, where it is about 10% in 2005 compared to 22% in the benchmark. Furthermore, the range of potential output gains also shifts up between 2005 and 2009 for all specifications, though the effect is less pronounced for a low elasticity of substitution.

Table 5: Bounds on Potential Output Gains (in %) in 2005 and 2009 for the Scenario with "Stronger Assumptions" for Alternative Preference Specifications

| | 2005 | | 2009 | |
|---|---|---|---|---|
| Preference Specification | Lower | Upper | Lower | Upper |
| CES with $\sigma = 1.5$ | 38.7 | 92.7 | 47.1 | 100.9 |
| Cobb-Douglas (Benchmark) | 22.0 | 63.5 | 28.1 | 71.8 |
| CES with $\sigma = 0.5$ | 9.8 | 41.0 | 12.7 | 46.1 |

Regarding the change to potential output gains between 2005 and 2009 one finds the following pattern for the two alternative preference specifications. For the scenario with "Stronger Assumptions ($\tau_i \pm 0.1$)" where output elasticities may deviate from observed factor income shares up to an absolute value of 0.1 one finds a range for the change to potential output gains of 1 to 6 ($-0.6$ to 16) percentage points for $\sigma = 0.5$ ($\sigma = 1.5$) relative to about 2 to 11 percentage points for the benchmark. Thus in contrast to the benchmark one cannot rule out a very mild decrease of potential output gains for $\sigma = 1.5$ here. If one sets output elasticities exactly equal to observed factor income shares then one finds an increase of potential output gains of about 4 (8) percentage points for $\sigma = 0.5$ ($\sigma = 1.5$) relative to about 6 percentage points for the benchmark. Accordingly, identifying a strictly positive increase in potential output gains requires sufficiently strong assumptions on output elasticities as in the benchmark specification.

Overall this subsection shows that the exact quantitative magnitudes of the bounds on potential output gains and their changes over time do indeed depend on the preference specification and specifically the elasticity of substitution. Thus more detailed demand estimates and specifications are an important area for refining such calculations in future research. However the broad pattern for the considered alternative preference specifications is the same as for the benchmark.

# 8 Conclusions

This paper has developed a novel theoretical framework to measure the extent of factor misallocation under more general assumptions than the prior literature. Such an approach extends the range of possible applications in which one can credibly measure factor misallocation and can in its current or an extended form be applied in many contexts.

An application shows that the labour and capital allocation across 473 six-digit manufacturing industries in the United States is inconsistent with an efficient allocation. Misallocation is an economically significant phenomenon with potential output gains from an efficient reallocation exceeding 22% of actual output. The analysis also provides evidence for an increase in misallocation during the Great Recession, which contributed about $10-60\%$ to the observed decline in manufacturing output. These results suggest that mitigating the underlying frictions in the economy may both permanently increase output and reduce economic fluctuations during episodes like the Great Recession.

As in the rest of the literature employing an "indirect approach" these measurements are only informative on the overall level of misallocation and not on its sources or the causes for an increase of misallocation during the Great Recession. The literature using a "direct approach" offers many potential explanations, but unfortunately their relative contribution to the overall level of misallocation is still unknown. It is sometimes even unclear what features of reality these broad explanations really capture. For example the fact that an investment model with adjustment costs can better explain the data than one without such costs as shown by Asker, Collard-Wexler, and Loecker (2014) is not informative on the precise nature of these frictions. These adjustment costs may represent some true physical costs of adjusting capital, but they could also reflect information problems on the second hand market for capital or the effect of some regulation or tax policy. Depending on the exact sources it is possible that only a certain part of misallocation can be lifted, while another part is unavoidable and constitutes a natural level of misallocation. It may also not even be desirable to lift all avoidable distortions for example because certain policy distortions are in place to correct some other imperfections like externalities. Thus a better understanding of the economic mechanisms causing the measured level of misallocation is key for drawing policy implications and an important topic for future research.

# References


Asker, J., A. Collard-Wexler, and J. D. Loecker (2014). Dynamic Inputs and Resource (Mis)Allocation. *Journal of Political Economy 122*(5), 1013–1063.

Banerjee, A. and E. Duflo (2005). Growth Theory through the Lens of Development Economics. In P. Aghion and S. N. Durlauf (Eds.), *Handbook of Economic Growth*, pp. 473–552. Amsterdam: Elsevier.

Bartelsman, E., J. Haltiwanger, and S. Scarpetta (2013). Cross-Country Differences in Productivity: The Role of Allocation and Selection. *American Economic Review 103*(1), 305–334.

Broda, C. and D. E. Weinstein (2006). Globalization and the Gains from Variety. *Quarterly Journal of Economics 121*(2), 541–585.

Buera, F. J., J. P. Kabowski, and Y. Shin (2011). Finance and Development: A Tale of Two Sectors. *American Economic Review 101*(5), 1964–2002.

Caselli, F. and J. Feyrer (2007). The Marginal Product of Capital. *Quarterly Journal of Economics 122*(2), 535–568.

Caselli, F. and N. Gennaioli (2013). Dynastic Management. *Economic Inquiry 51*(1), 971–996.

Chen, K. and A. Irarrazabal (2015). The Role of Allocative Efficiency in a Decade of Recovery. *Review of Economic Dynamics 18*(3), 523–550.

David, J. M., H. A. Hopenhayn, and V. Venkateswaran (2016). Information, Misallocation, and Aggregate Productivity. *Quarterly Journal of Economics 131*(2), 943–1005.

Dias, D. A., C. R. Marques, and C. Richmond (2016). Misallocation and Productivity in the lead up to the Eurozone Crisis. *Journal of Macroeconomics 49*, 46–70.

Gao, W. and M. Kehrig (2017). Returns to Scale, Productivity and Competition: Empirical Evidence from U.S. Manufacturing and Construction Establishments. *Working Paper*.

García-Santana, M., E. Moral-Benito, J. Pijoan-Mas, and R. Ramos (2016). Growing like Spain: 1995-2007. *Working Paper*.

García-Santana, M. and J. Pijoan-Mas (2014). The Researvation Laws in India and the Misallocation of Production Factors. *Journal of Monetary Economics 66*, 193–209.

Gollin, D. (2002). Getting Income Shares Right. *Journal of Political Economy 110*(2), 458–474.

Gopinath, G., S. Kalemli-Özcan, L. Karabarbounis, and C. Villegas-Sanchez (2015). Capital Allocation and Productivity in South Europe. *Working Paper*.

Guner, N., G. Ventura, and Y. Xu (2008). Macroeconomic Implications of Size-Dependent Policies. *Review of Economic Dynamics 11*(4), 721–744.

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

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and Manufacturing TFP in China and India. *Quarterly Journal of Economics 124*(4), 1403–1448.

Lagos, R. (2006). A Model of TFP. *Review of Economic Studies 73*(4), 983–1007.

Larrain, M. and S. Stumpner (2015). Capital Account Liberalization and Aggregate Productivity: The Role of Firm Capital Allocation. *Working Paper*.

Mian, A. and A. Sufi (2015). *House of Debt: How They (and You) Caused the Great Recession, and How We Can Prevent It from Happening Again.* University of Chicago Press.

Midrigan, V. and D. Y. Xu (2014). Finance and Misallocation: Evidence from Plant-Level Data. *American Economic Review 104*(2), 422–458.

Moll, B. (2014). Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation? *American Economic Review 104*(10), 3186–3221.

Oberfield, E. (2013). Productivity and Misallocation during a Crisis: Evidence from the Chilean Crisis of 1982. *Review of Economic Dynamics 16*(1), 100–119.

Peters, M. (2013). Heterogeneous Mark-Ups, Growth and Endogenous Misallocation. *Working Paper*.

Rajan, R. G. and L. Zingales (1998). Financial Dependence and Growth. *American Economic Review 88*(3), 559–586.

Restuccia, D. and R. Rogerson (2008). Policy Distortions and Aggregate Productivity with Heterogeneous Establishments. *Review of Economic Dynamics 11*(4), 707–720.

Restuccia, D. and R. Rogerson (2013). Misallocation and Productivity. *Review of Economic Dynamics 16*(1), 1–10.

Restuccia, D. and R. Rogerson (2017). The Causes and Costs of Misallocation. *Journal of Economic Perspectives 31*(3), 151–174.

Rosenzweig, M. R. (1988). Labor Markets in Low-Income Countries. In H. Chenery and T. N. Srinivasan (Eds.), *Handbook of Development Economics*, Volume I, pp. 713–762. Amsterdam: North-Holland.

Sandleris, G. and M. L. J. Wright (2014). The Costs of Financial Crises: Resource Misallocation, Productivity, and Welfare in the 2001 Argentine Crisis. *Scandinavian Journal of Economics 116*(1), 87–127.

Senga, T. (2016). A New Look at Uncertainty Shocks: Imperfect Information and Misallocation. *Working Paper*.

Song, Z. and G. L. Wu (2015). Identifying Capital Misallocation. *Working Paper*.

Vollrath, D. (2009). How Important are Dual Economy Effects for Aggregate Productivity? *Journal of Development Economics 88*(2), 325–334.

Ziebarth, N. L. (2015). Misallocation and Productivity in the Great Depression. *Working Paper*.

# Appendix

# A Proofs and Derivations

## A.1 Proof of Proposition 1

The general strategy of the proof is to show that equations (6), (7) and (8) together with the bounds on output elasticities and given values of $d^L_{ab}$ and $d^K_{ab}$ imply the restrictions on the quantities $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ stated in the proposition.

As a preliminary step, note that equation (6) implies $\frac{\varepsilon_{La}}{\varepsilon_{Lb}} = \frac{\frac{y_b}{y_a}}{d^L_{ab}}$ and $\varepsilon_{La} = \frac{\frac{y_b}{y_a}}{d^L_{ab}}\varepsilon_{Lb}$. Using these expressions and equation (8) for units $a$ and $b$ one can then substitute for $\varepsilon_{La}$, $\varepsilon_{Ka}$ and $\varepsilon_{Kb}$ in equation (7) such that it reads as

$$
\begin{aligned}
\frac{k_b}{k_a} &= \frac{\frac{y_b}{y_a}}{d^L_{ab}} \frac{\lambda_b - \varepsilon_{Lb}}{\lambda_a - \frac{\frac{y_b}{y_a}}{d^L_{ab}}\varepsilon_{Lb}} \frac{d^L_{ab}}{d^K_{ab}} \\
\iff \frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}}\left[\lambda_a - \frac{\frac{y_b}{y_a}}{d^L_{ab}}\varepsilon_{Lb}\right] &= \frac{\frac{y_b}{y_a}}{d^L_{ab}}[\lambda_b - \varepsilon_{Lb}] \\
\iff \frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}}\lambda_a - \frac{\frac{y_b}{y_a}}{d^L_{ab}}\lambda_b &= \frac{\frac{y_b}{y_a}}{d^L_{ab}}\left[\frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}} - 1\right]\varepsilon_{Lb} \qquad (18)
\end{aligned}
$$

If $\frac{k_b}{k_a} = \frac{d^L_{ab}}{d^K_{ab}}$ then equation (18) directly implies $\frac{y_b}{y_a} = \frac{\lambda_a}{\lambda_b}d^L_{ab}$, which is the last line of proposition 1.

However if $\frac{k_b}{k_a} \neq \frac{d^L_{ab}}{d^K_{ab}}$ then equation (18) can be solved for $\varepsilon_{Lb}$ as

$$
\varepsilon_{Lb} = \frac{\lambda_a \frac{k_b}{k_a}\frac{\frac{d^K_{ab}}{\frac{y_b}{y_a}}} - \lambda_b}{\frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}} - 1} \qquad (19)
$$

Given the value of $\varepsilon_{Lb}$ the other elasticities $\varepsilon_{La}$, $\varepsilon_{Ka}$ and $\varepsilon_{Kb}$ are implied by equations (6) and equation (8) for units $a$ and $b$.

Now impose the restrictions $\varepsilon_{La} \in (0, \lambda_a)$, $\varepsilon_{Ka} \in (0, \lambda_a)$, $\varepsilon_{Lb} \in (0, \lambda_b)$ and $\varepsilon_{Kb} \in (0, \lambda_b)$. Note that if $\varepsilon_{La} \in (0, \lambda_a)$ and $\varepsilon_{Lb} \in (0, \lambda_b)$ then this directly implies $\varepsilon_{Ka} \in (0, \lambda_a)$ and $\varepsilon_{Kb} \in (0, \lambda_b)$ due to equation (8). Thus it suffices to impose the restrictions $\varepsilon_{La} \in (0, \lambda_a)$ and $\varepsilon_{Lb} \in (0, \lambda_b)$.

First consider the case $\frac{k_b}{k_a} > \frac{d^L_{ab}}{d^K_{ab}}$. Note that the denominator of the RHS of equation (19) is positive in this case. The restrictions in the first line of equations of proposition 1 are the collection of the following restrictions:

- $\varepsilon_{Lb} > 0$ requires that the numerator of the RHS of equation (19) is positive which implies $\frac{y_b}{y_a} < \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a}d^K_{ab}$.

- $\varepsilon_{Lb} < \lambda_b$ requires that the RHS of equation (19) is smaller than $\lambda_b$ which implies
$$\lambda_a \frac{k_b}{k_a}\frac{d^K_{ab}}{\frac{y_b}{y_a}} - \lambda_b < \lambda_b \left[\frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}} - 1\right]$$
and hence $\frac{y_b}{y_a} > \frac{\lambda_a}{\lambda_b}d^L_{ab}$.

- $\varepsilon_{La} > 0$ does not generate further constraints because it is always satisfied when $\varepsilon_{Lb} > 0$ because $\varepsilon_{La} = \frac{\frac{y_b}{y_a}}{d^L_{ab}}\varepsilon_{Lb}$.

- $\varepsilon_{La} < \lambda_a$ requires due to $\varepsilon_{La} = \frac{\frac{y_b}{y_a}}{d^L_{ab}}\varepsilon_{Lb}$ that
$$\frac{\frac{y_b}{y_a}}{d^L_{ab}}\left[\lambda_a \frac{k_b}{k_a}\frac{d^K_{ab}}{\frac{y_b}{y_a}} - \lambda_b\right] < \lambda_a \left[\frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}} - 1\right]$$
and hence $\frac{y_b}{y_a} > \frac{\lambda_a}{\lambda_b}d^L_{ab}$. This is the same constraint as imposed by $\varepsilon_{Lb} < \lambda_b$.

Second consider the case $\frac{k_b}{k_a} < \frac{d^L_{ab}}{d^K_{ab}}$. Note that the denominator of the RHS of equation (19) is negative in this case. When imposing the restrictions $\varepsilon_{Lb} > 0$, $\varepsilon_{Lb} < \lambda_b$, $\varepsilon_{La} > 0$ and $\varepsilon_{La} < \lambda_a$, all inequalities are reversed compared to the previous case. This generates the restrictions in the second line of equations of proposition 1.

## A.2 Proof of Corollary 1

Corollary 1 follows from proposition 1. The strategy to prove corollary 1 is to show that the ($d^L_{ab}$,$d^K_{ab}$) combinations stated in the corollary are consistent with proposition 1, but all other ($d^L_{ab}$,$d^K_{ab}$) combinations lead to a contradiction with proposition 1.

As a preliminary step, note that by definition of $\widetilde{d}^L_{ab}$ and $\widetilde{d}^K_{ab}$ it holds that $\frac{\widetilde{d}^L_{ab}}{\widetilde{d}^K_{ab}} = \frac{k_b}{k_a}$.

First consider the case of $d^L_{ab} = \widetilde{d}^L_{ab}$:

- It directly follows from the definition of $\widetilde{d}^L_{ab}$ that $\frac{y_b}{y_a} = \frac{\lambda_a}{\lambda_b}d^L_{ab}$. If also $d^K_{ab} = \widetilde{d}^K_{ab}$ as stated in the proposition then $\frac{d^L_{ab}}{d^K_{ab}} = \frac{\widetilde{d}^L_{ab}}{\widetilde{d}^K_{ab}} = \frac{k_b}{k_a}$. This is consistent with the last equation of proposition 1.

- Now confirm that any other $d_{ab}^K \neq \widetilde{d}_{ab}^K$ does not satisfy proposition 1. Note that for the case $\frac{y_b}{y_a} = \frac{\lambda_a}{\lambda_b} d_{ab}^L$ proposition 1 requires $\frac{k_b}{k_a} = \frac{d_{ab}^L}{d_{ab}^K}$ which implies

$$\frac{k_b}{k_a} = \frac{d_{ab}^L}{d_{ab}^K} = \frac{\widetilde{d}_{ab}^L}{d_{ab}^K} = \frac{\frac{y_b}{y_a}}{\frac{\lambda_a}{\lambda_b} d_{ab}^K} \Longleftrightarrow d_{ab}^K = \frac{\frac{y_b}{y_a}}{\frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a}} \equiv \widetilde{d}_{ab}^K.$$

  Thus $\frac{k_b}{k_a} = \frac{d_{ab}^L}{d_{ab}^K}$ can only be satisfied for $d_{ab}^K = \widetilde{d}_{ab}^K$. Instead any $d_{ab}^K \neq \widetilde{d}_{ab}^K$ leads to a contradiction with proposition 1.

Second consider the case of $d_{ab}^L > \widetilde{d}_{ab}^L$:

- If as stated in the proposition $d_{ab}^K < \widetilde{d}_{ab}^K$ then $\frac{k_b}{k_a} = \frac{\widetilde{d}_{ab}^L}{\widetilde{d}_{ab}^K} < \frac{d_{ab}^L}{d_{ab}^K}$ and

$$\frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a} d_{ab}^K < \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a}\widetilde{d}_{ab}^K = \frac{y_b}{y_a} = \frac{\lambda_a}{\lambda_b}\widetilde{d}_{ab}^L < \frac{\lambda_a}{\lambda_b} d_{ab}^L$$

  which is consistent with the second equation of proposition 1.

- Now confirm that any other $d_{ab}^K \geq \widetilde{d}_{ab}^K$ does not satisfy proposition 1. Note that $d_{ab}^L > \widetilde{d}_{ab}^L$ implies $\frac{y_b}{y_a} < \frac{\lambda_a}{\lambda_b} d_{ab}^L$. In this case proposition 1 requires $\frac{y_b}{y_a} > \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a} d_{ab}^K$ which can be written as

$$\frac{y_b}{y_a} > \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a} d_{ab}^K = \frac{\lambda_a}{\lambda_b}\frac{\widetilde{d}_{ab}^L}{\widetilde{d}_{ab}^K} d_{ab}^K = \frac{y_b}{y_a}\frac{d_{ab}^K}{\widetilde{d}_{ab}^K} \Longleftrightarrow d_{ab}^K < \widetilde{d}_{ab}^K.$$

  Thus $\frac{y_b}{y_a} > \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a} d_{ab}^K$ can only be satisfied if $d_{ab}^K < \widetilde{d}_{ab}^K$. Instead any $d_{ab}^K \geq \widetilde{d}_{ab}^K$ leads to a contradiction with proposition 1.

Third consider the case of $d_{ab}^L < \widetilde{d}_{ab}^L$:

- If as stated in the proposition $d_{ab}^K > \widetilde{d}_{ab}^K$ then $\frac{k_b}{k_a} = \frac{\widetilde{d}_{ab}^L}{\widetilde{d}_{ab}^K} > \frac{d_{ab}^L}{d_{ab}^K}$ and

$$\frac{\lambda_a}{\lambda_b} d_{ab}^L < \frac{\lambda_a}{\lambda_b}\widetilde{d}_{ab}^L = \frac{y_b}{y_a} = \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a}\widetilde{d}_{ab}^K < \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a} d_{ab}^K$$

  which is consistent with the first equation of proposition 1.

- Now confirm that any other $d_{ab}^K \leq \widetilde{d}_{ab}^K$ does not satisfy proposition 1. Note that $d_{ab}^L < \widetilde{d}_{ab}^L$ implies $\frac{y_b}{y_a} > \frac{\lambda_a}{\lambda_b} d_{ab}^L$. In this case proposition 1 requires $\frac{y_b}{y_a} < \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a} d_{ab}^K$ which can be written as

$$\frac{y_b}{y_a} < \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a} d_{ab}^K = \frac{\lambda_a}{\lambda_b}\frac{\widetilde{d}_{ab}^L}{\widetilde{d}_{ab}^K} d_{ab}^K = \frac{y_b}{y_a}\frac{d_{ab}^K}{\widetilde{d}_{ab}^K} \Longleftrightarrow d_{ab}^K > \widetilde{d}_{ab}^K.$$

Thus $\frac{y_b}{y_a} < \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a}d^K_{ab}$ can only be satisfied if $d^K_{ab} > \widetilde{d}^K_{ab}$. Instead any $d^K_{ab} \leq \widetilde{d}^K_{ab}$ leads to a contradiction with proposition 1.

## A.3 Proof of Proposition 2

The initial steps and the general setup of the proof are identical to the one in section A.1. Again $\frac{k_b}{k_a} = \frac{d^L_{ab}}{d^K_{ab}}$ directly implies $\frac{y_b}{y_a} = \frac{\lambda_a}{\lambda_b}d^L_{ab}$ because of equation (18). For the case of $\frac{k_b}{k_a} \neq \frac{d^L_{ab}}{d^K_{ab}}$ one now needs to impose the restrictions $\varepsilon_{La} \in (\theta_{La}, \lambda_a - \theta_{Ka})$ and $\varepsilon_{Lb} \in (\theta_{Lb}, \lambda_b - \theta_{Kb})$ on equation (19).

First consider the case $\frac{k_b}{k_a} > \frac{d^L_{ab}}{d^K_{ab}}$. Note that the denominator of the RHS of equation (19) is positive in this case. The restrictions in the first line of equations of proposition 2 are the collection of the following restrictions:

- $\varepsilon_{Lb} > \theta_{Lb}$ requires that

$$\lambda_a \frac{k_b}{k_a}\frac{d^K_{ab}}{\frac{y_b}{y_a}} - \lambda_b > \theta_{Lb}\left[\frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}} - 1\right] \iff \frac{y_b}{y_a} < \frac{\lambda_a \frac{k_b}{k_a} d^K_{ab}}{\lambda_b - \theta_{Lb} + \theta_{Lb}\frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}}}$$

- $\varepsilon_{Lb} < \lambda_b - \theta_{Kb}$ requires that

$$\lambda_a \frac{k_b}{k_a}\frac{d^K_{ab}}{\frac{y_b}{y_a}} - \lambda_b < (\lambda_b - \theta_{Kb})\left[\frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}} - 1\right] \iff \frac{y_b}{y_a} > \frac{\lambda_a \frac{k_b}{k_a} d^K_{ab}}{\theta_{Kb} + (\lambda_b - \theta_{Kb})\frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}}}$$

- $\varepsilon_{La} > \theta_{La}$ requires that

$$\frac{\frac{y_b}{y_a}}{d^L_{ab}}\left[\lambda_a \frac{k_b}{k_a}\frac{d^K_{ab}}{\frac{y_b}{y_a}} - \lambda_b\right] > \theta_{La}\left[\frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}} - 1\right]$$
$$\iff \frac{y_b}{y_a} < \frac{\theta_{La}}{\lambda_b}d^L_{ab} + \frac{\lambda_a - \theta_{La}}{\lambda_b}\frac{k_b}{k_a}d^K_{ab}$$

- $\varepsilon_{La} < \lambda_a - \theta_{Ka}$ requires that

$$\frac{\frac{y_b}{y_a}}{d^L_{ab}}\left[\lambda_a \frac{k_b}{k_a}\frac{d^K_{ab}}{\frac{y_b}{y_a}} - \lambda_b\right] > (\lambda_a - \theta_{Ka})\left[\frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}} - 1\right]$$
$$\iff \frac{y_b}{y_a} < \frac{\lambda_a - \theta_{Ka}}{\lambda_b}d^L_{ab} + \frac{\theta_{Ka}}{\lambda_b}\frac{k_b}{k_a}d^K_{ab}$$

Second consider the case $\frac{k_b}{k_a} < \frac{d^L_{ab}}{d^K_{ab}}$. Note that the denominator of the RHS of equation (19) is negative in this case. When imposing the restrictions $\varepsilon_{Lb} > \theta_{Lb}$,

$\varepsilon_{Lb} < \lambda_b - \theta_{Kb}$, $\varepsilon_{La} > \theta_{La}$ and $\varepsilon_{La} < \lambda_a - \theta_{Ka}$, all inequalities are reversed compared to the previous case. This generates the restrictions in the second line of equations of proposition 2.

## A.4 Social Planner Problem with CES Preferences

For CES preferences the social planner problem can be written as

$$\max_{\{\kappa_i^*, \ell_i^*\}_{i=1}^N} \left[ \sum_{i=1}^N \frac{s_i}{\left(\kappa_i^{\alpha_i} \ell_i^{\lambda_i - \alpha_i}\right)^{\frac{\sigma-1}{\sigma}}} \left( (\kappa_i^*)^{\alpha_i} (\ell_i^*)^{\lambda_i - \alpha_i} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{20}$$

subject to the resource constraints $\sum_{i=1}^N \kappa_i^* = 1$ and $\sum_{i=1}^N \ell_i^* = 1$, and non-negativity constraints $\kappa_i^* \geq 0$ and $\ell_i^* \geq 0$ for all production units $i$.

In order to derive equation (20) first note that at the current observed allocation the marginal rate of substitution between two goods needs to be equal to the relative price, which for a CES utility function can be written as

$$\frac{p_i Y_i}{p_j Y_j} = \left(\frac{\beta_i}{\beta_j}\right)^{\sigma} \left(\frac{p_i}{p_j}\right)^{1-\sigma} \iff \beta_i = (p_i Y_i)^{\frac{1}{\sigma}} p_i^{\frac{\sigma-1}{\sigma}} \frac{\beta_j}{(p_j Y_j)^{\frac{1}{\sigma}} p_j^{\frac{\sigma-1}{\sigma}}}$$

for two goods $i$ and $j$.

Substituting the production functions into the utility function involved in the social planner problem, noting that $A_i \overline{K}^{\alpha_i} \overline{L}^{\lambda_i - \alpha_i} = \frac{Y_i}{\kappa_i^{\alpha_i} \ell_i^{\lambda_i - \alpha_i}}$ and substituting $\beta_i$ for all $i$ using the expression above for some fixed unit $j$ and rearranging terms yields

$$\max_{\{\kappa_i^*, \ell_i^*\}_{i=1}^N} \left[ Y \frac{\beta_j}{(p_j Y_j)^{\frac{1}{\sigma}} p_j^{\frac{\sigma-1}{\sigma}}} \sum_{i=1}^N \frac{s_i}{\left(\kappa_i^{\alpha_i} \ell_i^{\lambda_i - \alpha_i}\right)^{\frac{\sigma-1}{\sigma}}} \left( (\kappa_i^*)^{\alpha_i} (\ell_i^*)^{\lambda_i - \alpha_i} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}$$

where $s_i = \frac{p_i Y_i}{Y}$. Finally, note that the term $Y \frac{\beta_j}{(p_j Y_j)^{\frac{1}{\sigma}} p_j^{\frac{\sigma-1}{\sigma}}}$ can be omitted without altering the maximization problem. This gives equation (20).