No. 87

UNIVERSITY OF COLOGNE

WORKING PAPER SERIES IN ECONOMICS

GENERAL METHODS FOR MEASURING FACTOR MISALLOCATION

THOMAS SCHELKLE

Department of Economics
University of Cologne
Albertus-Magnus-Platz
D-50923 Köln
Germany

http://www.wiso.uni-koeln.de

General Methods for Measuring Factor Misallocation*

Thomas Schelkle†

November 28, 2016

Abstract

The paper develops novel methods to measure the extend of factor misallocation. These rely on production functions being homogeneous, but not on specific parameterizations and partly not even on specific functional forms. This reduces the risk to incorrectly reject an efficient allocation. In an empirical application these general methods strongly reject an efficient capital and labor allocation across 473 six-digit U.S. manufacturing industries. Potential output gains of efficiently reallocating factors are between 22 and 64% of observed output. There is also evidence that misallocation increased substantially during the Great Recession with a sizeable contribution to the observed fall in manufacturing output.

JEL codes: E23, D61, O11
Keywords: Misallocation, factor allocation, test, bounds

*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

The allocation of scarce production factors to different production units is a central topic in economics. An "efficient" factor allocation in the sense that it maximizes the total value of output at current prices, requires the equalization of marginal value products of factors across production units. Under certain ideal conditions such a situation may arise in a market economy. However determining whether an empirically observed factor allocation exhibits this property or involves a misallocation of resources is extremely challenging. The main reason is that marginal products are not directly observable. In recent years there is a strong renewed interest in factor misallocation following Hsieh and Klenow (2009) and Restuccia and Rogerson (2008) in order to explain income and productivity differences between countries or over time. 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 strong assumptions on production functions or on being able to perfectly control for differences in output elasticities by observed differences in factor income shares. This important and fast-growing literature may therefore often incorrectly reject an efficient factor allocation simply because these assumptions are too restrictive and misspecified.

This paper addresses this concern by developing a framework for the identification of misallocation under more general assumptions. It consists of a novel test for an efficient allocation and in placing bounds on the potential output gains associated with an efficient reallocation of factors. If these more general methods still detect significant misallocation then one can be much more confident that the factor allocation is truly not efficient. Such an outcome would strongly support the prior literature on misallocation because its findings are then not just an artefact of its restrictive assumptions. In contrast, if one only finds misallocation using restrictive assumptions, but not using more general methods then this raises doubts on the validity of such results.

The developed test for an efficient factor allocation only requires that production functions are homogeneous of a known degree and satisfy standard regularity conditions. But it does not depend on a specific functional form or common values of output elasticities across production units. The paper then studies the allocation of two production factors, say capital and labor, among several production units. The theoretical analysis establishes that these general assumptions imply restrictions on the joint variation of capital intensities and average products of labor across pairs of production units for any given marginal product differentials. If one production unit has a higher average product than another unit then by definition this either reflects a higher marginal product or a lower output elasticity of the respective production factor. But for homogeneous production functions a relatively low output elasticity of one factor is always associated with a relatively high output elasticity of the other factor conditional on potential differences in the degree of homogeneity between units. Thus in a situation of equalized marginal products a high average product of labor of a production unit relative to another unit should be accompanied by a sufficiently low average product of capital and therefore a sufficiently high capital intensity. If an observed factor allocation does not exhibit such a pattern then it cannot possibly be characterized by equalized marginal products for any combination of homogeneous production functions with the assumed degrees of homogeneity. This theoretical insight allows to test and possibly refute the null hypothesis that the factor allocation is efficient. The main advantage of this test is that its relatively general assumptions reduce the risk to incorrectly reject an efficient allocation due to a misspecification of the underlying production model. Thus a rejection of efficiency with this test is very informative and should induce a high confidence that the null hypothesis is indeed false.

A natural measure of the economic significance of misallocation is by how much total output could potentially be increased by eliminating the distortions and reallocating resources efficiently. Computing this measure requires to impose additional functional form assumptions on production functions and preferences as in previous work. However here the contribution of the paper is to not depend on exact knowledge of the magnitude of output elasticities, which are key quantities for these computations. It is shown how one can then place upper and lower bounds on the potential output gains from factor reallocation exploiting knowledge on the possible range of output elasticities for homogeneous production functions. These bounds are informative on how much potential output gains depend on the value of output elasticities and on the restrictive assumptions on these objects in the prior literature. An extension allows to flexibly strengthen the assumptions on output elasticities, which translates into a more demanding test procedure and tighter bounds on potential output gains.

The developed framework can be used to study misallocation in many contexts and at various levels of aggregation, for example between firms, industries, sectors, regions or countries. This paper presents an empirical application to the allocation of capital and labor across 473 manufacturing industries at the six-digit level in the

United States using data from the NBER-CES Manufacturing Industry Database. One motivation for this application is that many observers consider the U.S. to be a relatively undistorted economy. If such a belief is correct then it should be straightforward to rationalize this data as an efficient factor allocation when using the general methods developed here. However if one can refute efficiency in this application then it seems likely that misallocation is also present in many other countries and applications, though of course only a case by case analysis can definitely confirm this. Contrary to the conventional wisdom, the analysis provides strong evidence for significant misallocation across industries. In 2005 the test rejects an efficient factor allocation for almost half of all industry pairs 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 rationalize 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 centered 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%. This is surprisingly close to the lower bound here. Overall the analysis of this data shows that there is evidence for economically significant misallocation between U.S. manufacturing industries even under considerably more general assumptions than used in the prior literature.

Furthermore I investigate the dynamics of misallocation during 2005-2009 to shed light on whether there was an increase in misallocation during the Great Recession. The analysis provides several pieces of evidence that this was the case. In 2009 the test rejects an efficient allocation for 60% of all industry pairs 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. While this range is $22-64\%$ in 2005, it is $28-72\%$ in 2009. Thus the lower bound increases by about 6 percentage points and the upper bound by about 8 percentage points. But it is also true that these ranges overlap such that for these relatively general assumptions it is theoretically possible that potential output gains remained constant or even decreased between these two years. Only if one further strengthens the assumptions on output elasticities, can one indeed conclude that the change to potential output gains was positive. If one assumes output elasticities to remain constant across years and one also narrows the range within which they may fall then one obtains an increase in potential output gains between about 2 and 11 percentage points. The increase in misallocation then contributes between about 10 and 60% to the observed about 18 percentage point negative deviation of manufacturing output from its 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, which amounts to a 35% contribution to the observed fall in manufacturing output. Thus an increase in misallocation during the Great recession can be detected under more general assumptions than are typically used in the prior literature. Nevertheless, the documented need for sufficiently strong assumptions implies that this result is not as robust as the one on the general presence of misallocation.

The paper relates to various strands of the prior literature. First, it is related to the literature on partial identification, which attempts to draw inferences from data using only weak, but credible assumptions (Manski 2007). Such an approach may help to establish a wide consensus among researchers on certain questions. The paper applies this general idea to the topic of factor misallocation. However the developed test is a theoretical test and not a statistical test.

Second, the paper is related to an active recent literature on factor misallocation. Surveys of recent work in this area are provided by Restuccia and Rogerson (2013) and Hopenhayn (2014). The present paper fits into what Restuccia and Rogerson 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.[1] The prior literature using the "indirect approach" has applied two main methods to identify misallocation.

One set of papers use Cobb-Douglas production functions and assume that their parameters are constant either across the production units under study, or across countries, or both. Examples are the studies of Hsieh and Klenow (2009) and Bartelsman, Haltiwanger, and Scarpetta (2013) on misallocation between firms within an industry or Vollrath (2009) on misallocation between sectors. Under such assumptions these papers draw conclusions from observed differences

[1]The paper is also related to a large literature that employs the "direct approach" and studies the effects of specific imperfections and distortionary policies, cf. the surveys of Restuccia and Rogerson (2013) and Hopenhayn (2014). Examples are financial frictions (Buera, Kabowski, and Shin 2011; Caselli and Gennaioli 2013; Midrigan and Xu 2014; Moll 2014), frictional labor markets (Lagos 2006), size-dependent policies and regulation (Guner, Ventura, and Xu 2008; García-Santana and Pijoan-Mas 2014) or imperfect output markets (Peters 2013), among others.

in average products to differences in marginal products between production units, and from differences in average product dispersion between countries to differences in marginal product dispersion and hence differences in the degree of misallocation. But the drawback of this approach is that if output elasticities vary between different production units then this fact could entirely explain observed differences in average products without any differences in marginal products. Furthermore if different countries have a different dispersion of output elasticities then differences in average product dispersion across countries could be explained without any differences in marginal product dispersion. In general output elasticities are not constant, but fully depend on the functional form of the production function and the amount of factors used. A main advantage of the approach in this paper is that it is robust to the presence of such fundamental technological differences between production units or countries, and can still point to observations which are necessarily inconsistent with an efficient factor allocation.

Another strand of work using the "indirect approach" for identifying misallocation assumes that factor prices are equal to marginal products. Some papers then take factor price differentials as a direct indication of misallocation. For instance Banerjee and Duflo (2005) review extensive evidence on a great heterogeneity of rates of return to the same factor in developing countries. Based on the same basic assumption on factor prices, a related approach combines information on factor income shares and average products to calculate marginal products. An example is the analysis of Caselli and Feyrer (2007) of the cross-country capital allocation. Furthermore, the approach mentioned above using Cobb-Douglas production functions becomes equivalent to this approach if it fixes the parameters of the Cobb-Douglas according to the factor income shares of each individual production unit instead of setting it to some average value or using information from a reference country. In contrast the framework developed in this paper is independent of any direct assumptions and data requirements concerning factor prices. This is an important advantage. One reason is the possible departure of factor prices from marginal value products, which at least for developing countries has traditionally played an important role in the literature as discussed in the survey by Rosenzweig (1988). Another reason is that reported labor income shares tend to underestimate true labor income shares because labor compensation of the self-employed is often treated as capital income as argued by Gollin (2002).

One paper that also addresses concerns on heterogeneity of output elasticities among other factors for the identification of misallocation is Song and Wu (2015). Thus they pursue a similar aim to this paper, 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 on misallocation can be drawn using only very general assumptions and minimal economic structure such that one needs only cross-sectional data. Therefore, these two approaches are complementary.

Finally, the empirical application of the paper complements the literature by providing evidence on misallocation across industries in the United States, while most of the prior literature following Hsieh and Klenow (2009) focusses on misallocation between firms within an industry. The analysis of the Great Recession is also related to a growing literature that documents changes to misallocation over time particularly during economic crises or periods of institutional change. There is for example evidence for an increase in misallocation 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 European countries around and after 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 European countries during a period of capital account liberalization (Larrain and Stumpner 2015) and in Chile in the period following the 1982 crisis (Chen and Irarrazabal 2015). All these papers employ the methodology of Hsieh and Klenow (2009) or a close variant of it. The present paper contributes to this literature by providing evidence on an increase in misallocation during another period of large economic fluctuations, namely the Great Recession in the United States. Furthermore, it uses a different methodology based on weaker assumptions. 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 the literature. Similar robustness issues may be present in other comparisons of misallocation between time periods or across countries, but have so far gone unnoticed because the prior literature does not conduct suitable robustness checks in this respect. In the light of this paper such checks seem important.

The paper is structured as follows. The test for 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 for the empirical application. 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 Test for an Efficient Factor Allocation

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 properties allow to test the hypothesis that an observed factor allocation exhibits equalized marginal products and to place bounds on the unobserved marginal product differentials.

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. The paper concentrates on a situation where the production units use two common production factors which are called labor L and capital K here. The total amount of factors that can be allocated between all production units is exogenous and for labor denoted by $\overline{L}$ and for capital by $\overline{K}$. There may also be other factors of production which are not directly part of the factor allocation problem, for example because they are only used by one of the units. The factor allocation of labor and capital across the production units is then given by the values of L_i and K_i for each production unit. 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 λ_i is allowed to differ between the production units.[2]

In the following I allow for the presence of distortions that drive a wedge between the marginal value products of the production units. The labor wedge

[2]The theoretical properties of allocations with equalized marginal value products and the test procedure 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 equalized marginal value products is not necessarily a situation where the value of total output is maximized. Accordingly equalization of marginal value products would not be desirable and the developed test would not be very interesting then. This is the reason for restricting attention to production functions which are at a maximum linearly homogeneous.

for a pair of production units (a, b) is denoted by d^L_{ab} and the capital wedge by d^K_{ab}. 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

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

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

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^L_{ab} and d^K_{ab}. Hence given known differentials d^L_{ab} and d^K_{ab} 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 labor differentials satisfy $d^L_{ba} = 1/d^L_{ab}$ and $d^L_{ac} = d^L_{ab} \times d^L_{bc}$ 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^L_{ab} 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 labor (capital) wedge d^L_{ab} (d^K_{ab}) above one indicates that the marginal value product of labor (capital) is higher in production unit b than in a, and vice versa. If the wedges d^L_{ab} and d^K_{ab} are not equal to one for all pairs of production units $(a, b) \in P$ then marginal value products are not equalized across all production units and accordingly total income is not maximized at this allocation. I refer to such a situation as "factor misallocation" and the allocation being "inefficient". In contrast an "efficient" allocation is one where marginal value products are equalized ($d^L_{ab} = d^K_{ab} = 1$ for all pairs $(a, b) \in P$) such that the total value of output is maximized. 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.[3]

[3]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

2.2 Observable Restrictions on Factor Allocations for given Marginal Product Differentials

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

$$\frac{d_{ab}^L}{d_{ab}^K}\frac{\frac{\partial Y_a}{\partial L_a}}{\frac{\partial Y_a}{\partial K_a}} = \frac{\frac{\partial Y_b}{\partial L_b}}{\frac{\partial Y_b}{\partial K_b}}. \tag{3}$$

It is more convenient to work with equations (1) and (3), which together with the resource constraints also determine the factor allocation. The key step now is to apply two simple "multiply and divide" tricks to equations (1) and (3). These read as

$$d_{ab}^L \, p_a \frac{Y_a}{L_a}\frac{\partial Y_a}{\partial L_a}\frac{L_a}{Y_a} = p_b \frac{Y_b}{L_b}\frac{\partial Y_b}{\partial L_b}\frac{L_b}{Y_b} \tag{4}$$

$$\frac{d_{ab}^L}{d_{ab}^K}\frac{K_a \frac{\partial Y_a}{\partial L_a}\frac{L_a}{Y_a}}{L_a \frac{\partial Y_a}{\partial K_a}\frac{K_a}{Y_a}} = \frac{K_b \frac{\partial Y_b}{\partial L_b}\frac{L_b}{Y_b}}{L_b \frac{\partial Y_b}{\partial K_b}\frac{K_b}{Y_b}}. \tag{5}$$

Rearranging and denoting the average value product of labor 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 labor 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

$$\frac{y_b}{y_a} = \frac{\varepsilon_{La}}{\varepsilon_{Lb}} d_{ab}^L \tag{6}$$

$$\frac{k_b}{k_a} = \frac{\varepsilon_{La}}{\varepsilon_{Lb}}\frac{\varepsilon_{Kb}}{\varepsilon_{Ka}}\frac{d_{ab}^L}{d_{ab}^K}. \tag{7}$$

where $\frac{y_b}{y_a}$ is the ratio of average value products of labor between two production units a and b, which is sometimes also called "relative labor 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) are key equations of the paper and provide a number of insights. First, they show that there is indeed a meaningful relationship be-

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.

tween the average product of labor ratio $\frac{y_b}{y_a}$ and the marginal product of labor 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 labor. Note that 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. This shows that 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 the equation shows that a difference in average products could 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 and possibly even all of it may simply be caused by variation in output elasticities across countries. Such a variation in output elasticities may result from different functional forms of production functions or from different amounts of used factors. A second insight from equations (6) and (7) is that in principle suitable values of d_{ab}^L and d_{ab}^K can rationalize 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[4] 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 to-

[4]Throughout the paper I refer to the key assumption as being one of homogeneity of production functions with known λ_i due to the importance and wide use of this assumption in both theoretical and applied work. However strictly speaking the derived restrictions and test require only a weaker assumption which is that equation (8) holds for known λ_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.

gether 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).

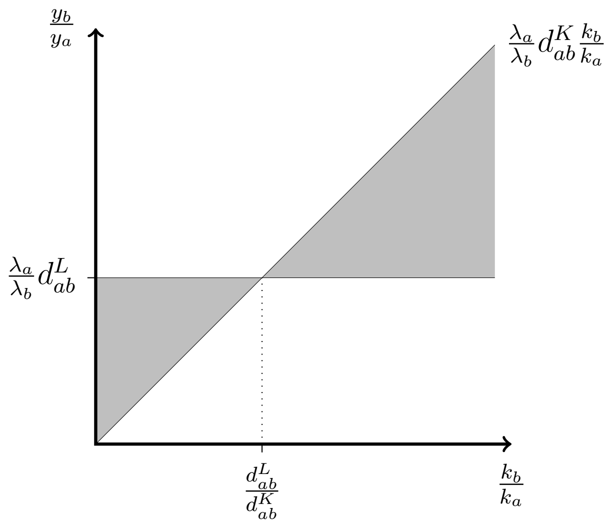
Proposition 1. *If two production units (a, b) operate with well-behaved and homogenous production functions of degree λ_a and λ_b and the factor allocation exhibits marginal product differentials of labor d^L_{ab} and capital d^K_{ab} between these production units then the average product of labor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

$$
\begin{aligned}
&\frac{k_b}{k_a} > \frac{d^L_{ab}}{d^K_{ab}} && \textit{and} && \frac{\lambda_a}{\lambda_b} d^L_{ab} < \frac{y_b}{y_a} < \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a} d^K_{ab}, \quad \textit{or} \\
&\frac{k_b}{k_a} < \frac{d^L_{ab}}{d^K_{ab}} && \textit{and} && \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a} d^K_{ab} < \frac{y_b}{y_a} < \frac{\lambda_a}{\lambda_b} d^L_{ab}, \quad \textit{or} \\
&\frac{k_b}{k_a} = \frac{d^L_{ab}}{d^K_{ab}} && \textit{and} && \frac{y_b}{y_a} = \frac{\lambda_a}{\lambda_b} d^L_{ab}.
\end{aligned}
$$

Proposition 1 shows that for any pair of production units $(a, b) \in P$ the maintained assumptions on production functions and given values of λ_a, λ_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. This forms the basis for a hypothesis test proposed in the next subsection.

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 labor compared to unit a $(\frac{y_b}{y_a} > \frac{\lambda_a}{\lambda_b} d^L_{ab})$. For a given marginal product of labor differential d^L_{ab} this indicates a relatively low output elasticity of labor of unit b $(\frac{\varepsilon_{La}}{\varepsilon_{Lb}} > \frac{\lambda_a}{\lambda_b})$. But for homogeneous production functions a relatively low output elasticity of labor 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

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 λ_a, λ_b, d^L_{ab}, d^K_{ab} as described in proposition 1.

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 labor 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 labor differential d^L_{ab} the value of $\frac{y_b}{y_a}$ only requires a certain ratio of output elasticities of labor $\frac{\varepsilon_{La}}{\varepsilon_{Lb}}$. There are many possible combinations of ε_{La} and ε_{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 The Test for an Efficient Factor Allocation

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 equalized such that $d^L_{ab} = 1$ and $d^K_{ab} = 1$ 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 λ_i of the production function of each production unit. For example in many applications there may be good reasons to assume constant returns to scale and accordingly choose a value of 1. In other applications the researcher may want to use a value below 1 because a fixed factor other than capital or labor like land or managerial skills is also key for production and there are only constant returns to scale to all factors.

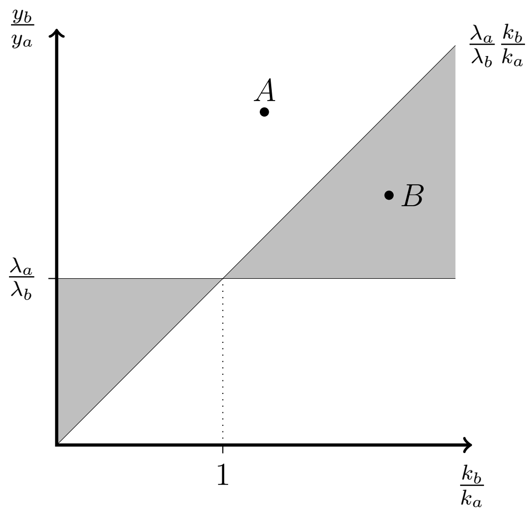
The test requires observations on the average product of labor ratio $\frac{y_b}{y_a}$ and capital intensity ratio $\frac{k_b}{k_a}$ for all pairs (a, b) at the current allocation. Given the assumed values of λ_i the test of the null hypothesis then simply consists in checking whether for each pair $(a, b) \in P$ the observed combination $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ satisfies the conditions of proposition 1 for $d^L_{ab} = d^K_{ab} = 1$, 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 $d^L_{ab} = d^K_{ab} = 1$. 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 labor 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 equalized 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 equalized 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 strong the rejection of overall efficiency is.

Figure 2: The Test for an Efficient Factor Allocation



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 pros and cons of the proposed test can be explained using analogies from statistical hypothesis testing. The main advantage of the test is that it relies on relatively weak assumptions. 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. The following subsection provides more details on these power issues and shows what degree of misallocation may still be present in cases where one cannot reject the null hypothesis. An extension presented in section 2.5 allows to tighten the basic assumptions and to trade off the probability of making these two types of 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 may also be helpful to directly compare the test developed here with the implicit rejection and non-rejection region of the approach in the previous literature that assumes Cobb-Douglas production functions with some given values for the elasticity parameters. In that case the non-rejection region, where one does not reject an efficient factor allocation, shrinks to one single point in this graph. This shows that the approach of the prior literature will basically always reject 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 ($d_{ab}^L = 1$ and $d_{ab}^K = 1$ for all pairs) one can of course also test other hypotheses on specific values of d_{ab}^L and d_{ab}^K. 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_{ab}^L and d_{ab}^K that one wishes to test.

2.4 Bounds on Marginal Product Differentials

This section provides bounds on the magnitude of marginal product differentials d_{ab}^L and d_{ab}^K 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 can be achieved without adding any further assumptions. The form of these bounds 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 λ_a and λ_b and the factor allocation involves an average product of labor 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 labor d_{ab}^L and capital d_{ab}^K between the production units satisfy either*

$$\begin{array}{lcll} d_{ab}^L > \widetilde{d}_{ab}^L & \textit{and} & d_{ab}^K < \widetilde{d}_{ab}^K, & \textit{or} \\ d_{ab}^L < \widetilde{d}_{ab}^L & \textit{and} & d_{ab}^K > \widetilde{d}_{ab}^K, & \textit{or} \\ d_{ab}^L = \widetilde{d}_{ab}^L & \textit{and} & d_{ab}^K = \widetilde{d}_{ab}^K, & \end{array}$$

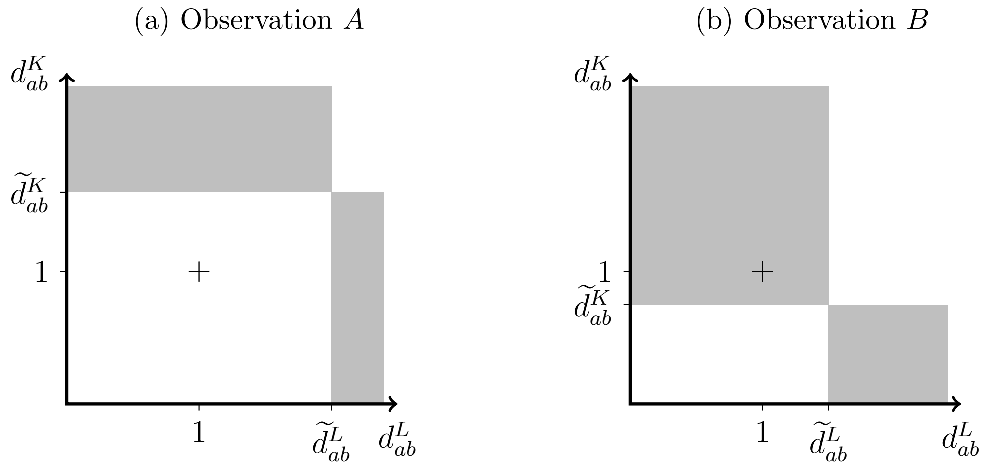
where

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

The intuition behind the corollary can be understood using graphical arguments and figures 1 and 2. Consider for example observation A in figure 2. This observation is not part of the shaded area in this graph and thus inconsistent with a situation of equalized marginal products ($d_{ab}^L = 1$ and $d_{ab}^K = 1$). One can then ask which values of d_{ab}^L and d_{ab}^K would be consistent with this observation or in other words how does the shaded area need to be shifted such that observation A becomes part of it? The answer is provided by noting how changes to d_{ab}^L and d_{ab}^K shift the straight lines in figure 1. For example a sufficiently high d_{ab}^L keeping d_{ab}^K at a value of 1 or a sufficiently high d_{ab}^K keeping d_{ab}^L equal to 1 will shift the shaded area such that it encompasses observation A. In fact many different combinations of d_{ab}^L and d_{ab}^K achieve this aim. The exact set of possible (d_{ab}^L,d_{ab}^K) combinations is provided by corollary 1 and characterized by the two boundary terms $\widetilde{d}_{ab}^L$ and $\widetilde{d}_{ab}^K$. Graphically the values of $\widetilde{d}_{ab}^L$ and $\widetilde{d}_{ab}^K$ are those that yield 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 shaded area in figure 3 shows the possible marginal product differentials as described by corollary 1 for the two examples A and B considered in figure 2. These graphs illustrate that each $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combination can in principle be generated by many different (d_{ab}^L,d_{ab}^K) combinations. It is not possible to place a bound on each marginal product differential separately. Instead one can only place bounds on combinations of d_{ab}^L and d_{ab}^K.

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 in each graph 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 λ_a and λ_b.

First consider observation A, for which the test rejects an efficient factor allocation. The point $d^L_{ab} = 1$ and $d^K_{ab} = 1$ (marked by "+"), where marginal products are equalized, is not part of the shaded area for these observations. This is just a different way to state that one was able to reject $d^L_{ab} = 1$ and $d^K_{ab} = 1$ here. Also observe that for observation A at least one of the marginal product differentials needs to be larger than 1. If A had been observed in the bottom part of the rejection area in figure 2 then at least one marginal product differential would have needed to be smaller than 1.

In contrast for observation B where the test does not reject the null hypothesis of an efficient factor allocation, the point $d^L_{ab} = 1$ and $d^K_{ab} = 1$ is part of the shaded area. However this 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.

2.5 Stronger Assumptions on Output Elasticities

The framework presented so far relies on the fact that mild assumptions on production functions already imply bounds on output elasticities. These bounds on output elasticities in turn generate restrictions on the ratios of average products of labor and capital intensities one may observe for given marginal product ratios. However in many potential applications the researcher may want to make stronger assumptions on output elasticities based on some prior information. This section extends the framework to allow imposing such assumptions and explains how this tightens the set of allocations that may be consistent with equalized marginal products. Imposing such stronger assumptions on output elasticities will increase the risk of committing a type 1 error (rejecting efficiency when the allocation is efficient) and decrease the risk of committing a type 2 error (failing to reject efficiency when the allocation is not efficient). Thus, this extension allows to trade off the probability of these two types of errors.

Though the general logic of the theoretical framework remains unchanged, there are some modifications to the details. Specifically, instead of requiring that output elasticities are only 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 θ_{Li} and θ_{Ki} represent lower bounds on the output elasticities of the respective factors. These lower bounds need to be consistent with the degree of homogeneity λ_i and equation (8). This implies that the parameters θ_{Li} and θ_{Ki} need to satisfy $\theta_{Li} \geq 0$, $\theta_{Ki} \geq 0$ and $\theta_{Li} + \theta_{Ki} < \lambda_i$. The parameters θ_{Li} and θ_{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 θ_{Li} and θ_{Ki} are set to zero, and allows to flexibly tighten the test procedure. For any specified values of θ_{La}, θ_{Ka}, θ_{Lb} and θ_{Kb} for a pair of production units (a, b) and specific values of the marginal product differentials d_{ab}^L and d_{ab}^K one can then again characterize 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 formalized 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 labor and capital bounded from below by θ_{La}, θ_{Ka}, θ_{Lb} and θ_{Kb} respectively and their production functions are homogenous of degree λ_a and λ_b, and the factor allocation exhibits marginal product differentials of labor d_{ab}^L and capital d_{ab}^K between the production units then the average product of labor ratio $\frac{y_b}{y_a}$ and capital intensity* *ratio* $\frac{k_b}{k_a}$ *at this allocation satisfy either*

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

where

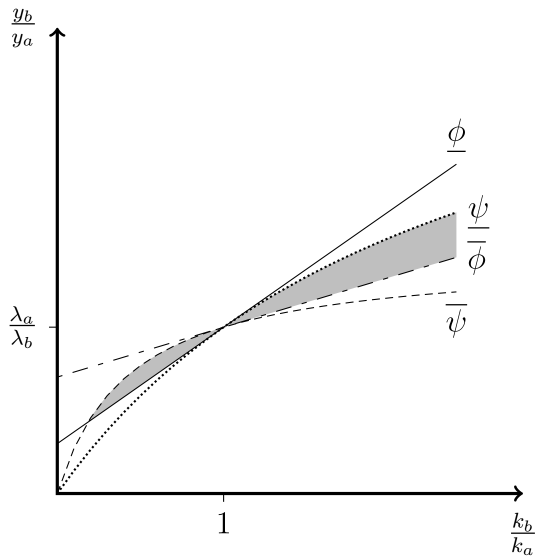
$$\begin{aligned}
\overline{\phi} &= \frac{\lambda_a - \theta_{Ka}}{\lambda_b} d^L_{ab} + \frac{\theta_{Ka}}{\lambda_b}\frac{k_b}{k_a} d^K_{ab}\ , \qquad \underline{\phi} = \frac{\theta_{La}}{\lambda_b} d^L_{ab} + \frac{\lambda_a - \theta_{La}}{\lambda_b}\frac{k_b}{k_a} d^K_{ab}\ , \\
\overline{\psi}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

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

Given the assumed values of λ_a, λ_b, θ_{La}, θ_{Ka}, θ_{Lb} and θ_{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 θ_{La}, θ_{Ka}, θ_{Lb} or θ_{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 θ_{Li} and θ_{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 general case. Thus the $(\frac{y_b}{y_a}, \frac{k_b}{k_a})$ combinations associated with these elasticity combinations are no longer possible.

The set of marginal product differentials d^L_{ab} and d^K_{ab} 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 if desired it is simple to solve for these sets computationally. One can simply specify a grid consisting of combinations of d^L_{ab} and d^K_{ab} 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.

Figure 4: The Test for an Efficient Factor Allocation 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.

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 labor such that one can only place an upper and lower bound on the potential output gain.

3.1 Concept

The general aim is to compute upper and lower bounds on 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. Placing bounds on 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. Here and below 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 resulting from an equilibrium 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 λ_i for each production unit given by

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

where α_i is a parameter governing the output elasticity of capital of production unit i. The output elasticity of labor 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 labor. It is assumed that A_i is fixed and remains unchanged when moving to the optimal allocation. Importantly, as in the previous main section that developed the test procedure the magnitude of output elasticities and hence the parameters α_i are not assumed to be known. Thus one can only place upper and lower bounds on the potential output gain because as before one only has knowledge about the range in which output elasticities may fall.

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 in the application I also explore the effect of alternative values of the elasticity of substitution 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. This in turn implies that the preference parameters β_i can be backed out from the observed allocation. The demand functions for Cobb-Douglas preferences and market-clearing then imply $\beta_i = \frac{p_i Y_i}{Y}$. This means that β_i can be 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

As explained already the magnitude of output elasticities and hence the parameters α_i are assumed to be unknown and are not fixed to some specific values. Accordingly one can only place bounds on the potential output gains. 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 α_i, which will be denoted as $G(\alpha)$ where $\alpha = (\alpha_1, \ldots, \alpha_N)$. This in turn requires to characterize the hypothetical efficient allocation Y^* for given values of α_i.

Define the share of total capital and labor employed in production unit i by $\widetilde{k}_i \equiv \frac{K_i}{\overline{K}} \in [0,1]$ 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 $Y_i = A_i \overline{K}^{\alpha_i} \overline{L}^{\lambda_i - \alpha_i} \widetilde{k}_i^{\alpha_i} \ell_i^{\lambda_i - \alpha_i}$. The potential output gain $G(\alpha)$ is then given by

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

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 labor and capital map into the potential output gain. Note that ω_i is directly implied by the observed factor allocation because by the production function

$$\omega_i = \frac{s_i}{\widetilde{k}_i^{\alpha_i} \ell_i^{\lambda_i - \alpha_i}} \tag{13}$$

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 $\widetilde{k}_i^*$ and ℓ_i^* at the efficient allocation can be determined by solving the problem of a social planner who maximizes utility subject to the physical constraints. For the assumed production and utility functions this can be written as

$$\max_{\{\widetilde{k}_i^*, \ell_i^*\}_{i=1}^N} \quad \prod_{i=1}^{N} \left((\widetilde{k}_i^*)^{\alpha_i} (\ell_i^*)^{\lambda_i - \alpha_i} \right)^{\beta_i} \tag{14}$$

subject to the resource constraints $\sum_{i=1}^N \widetilde{k}_i^* = 1$ and $\sum_{i=1}^N \ell_i^* = 1$, and non-negativity constraints $\widetilde{k}_i^* \geq 0$ 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 (14) without affecting the result. An advantage of the Cobb-Douglas assumption on preferences is that the social planner problem then exhibits an analytical solution given by

$$\widetilde{k}_i^* = \frac{\alpha_i \beta_i}{\sum_{j=1}^N \alpha_j \beta_j} \tag{15}$$

$$\ell_i^* = \frac{(\lambda_i - \alpha_i)\beta_i}{\sum_{j=1}^N (\lambda_j - \alpha_j)\beta_j} \tag{16}$$

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 α.

3.4 Bounds on Potential Output Gains

The derivation in the previous subsection involves the potential output gain for observed $(s_i, \widetilde{k}_i, \ell_i)$ and specific given values of α_i for each production unit. However I continue to assume that the specific value of output elasticities, and thus the parameters α_i, are not known. Instead only the possible range of output elasticities is known, which for homogeneous production functions amounts to $\alpha_i \in (0, \lambda_i)$. The bounds on the potential output gain for the observed values of $(s_i, \widetilde{k}_i, \ell_i)$ consist of the lowest and highest output gains $G(\alpha)$ when searching over these admissible values $\alpha \in A$ where A is the set of values for which the i-th element of α satisfies $\alpha_i \in (0, \lambda_i)$. Note that given the observed $(s_i, \widetilde{k}_i, \ell_i)$ combination for each production unit each possible vector α is associated with a specific combination of marginal product differentials for each pair of production units. Furthermore the values of α_i affect how these marginal product differentials map into potential output gains. Formally the lower bound on the potential output gain $\underline{G}$ from moving to an efficient allocation is then given by

$$\underline{G} = \inf_{\alpha \in A} G(\alpha) \tag{17}$$

and the upper bound $\overline{G}$ by

$$\overline{G} = \sup_{\alpha \in A} G(\alpha). \tag{18}$$

In the application these bounds are solved numerically.[5] The computed bounds show what conclusions on potential output gains one can draw without assuming a specific value of output elasticities and instead exploiting only knowledge on 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

[5] 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 application I used 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.

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 θ_{Li} and θ_{Ki} as explained in section 2.5. The computation of the bounds can then evolve as in the previous subsection with a small modification. The admissible values of α for the infimum in equation (17) and the supremum in equation (18) 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 more flexibly investigate the robustness of point estimates of potential output gains.

4 Data on U.S. Manufacturing Industries

The rest of the paper presents an application of the theoretical framework to study the allocation of labor and capital between different manufacturing industries in the United States. This section describes the data set used for this investigation. Section 5 then provides an analysis for 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.[6] 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 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 labor input I adjust the number of employees

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

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 labor 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 data base 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.[7] 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

$$\log w_{jt} = \gamma + \delta x_{jt} + \zeta z_{jt} + \epsilon_{jt} \tag{19}$$

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 latter controls for industry wage differentials which are not driven by human capital differences, but by potential distortions between industries. 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 (19) 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

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

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 applies the developed theoretical framework to analyse the capital and labor allocation across the 473 manufacturing industries in the United States in the year 2005. First I specify the assumptions and then present the test results and bounds on potential output gains.

5.1 Assumptions and Scenarios

The developed framework requires to first specify the degree of homogeneity λ_i with respect to capital and labor of the production function of each industry i. I follow most of the literature and assume constant returns to scale such that λ_i is set to 1 for all industries. However I also discuss the robustness of the results with respect to this choice in section 7.1.

The framework also allows to set the parameters θ_{Ki} and θ_{Li} representing lower bounds on the output elasticities of capital and labor 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 θ_{Ki} and θ_{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 analysed 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 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 θ_{Ki} and θ_{Li} such that the resulting admissible range of output elasticities is closer to the values that would be obtained by this traditional approach, but still considerably more general. Specifically, I set θ_{Ki} and θ_{Li} such that the admissible range of the output elasticity of labor of an industry includes values that are up to 0.2 higher or lower than the observed labor income share of that industry. For example for a labor income share of about 0.7 as often observed in aggregate data and used at the macro level, the output elasticity of labor could then vary between 0.5 and 0.9. Of course the sum of elasticities needs to sum to the assumed degree of homogeneity such that in this example the output elasticity of capital could vary between 0.1 and 0.5. Thus the resulting intervals seem large enough to allow for considerable deviations of marginal products from factor prices and hence of output elasticities from factor income shares. But at the same time the intervals are still centered around this traditional approach. Formally, denoting the labor income share of industry i by τ_{Li} the parameters are set such that $\theta_{Li} = \tau_{Li} - 0.2$ and $\theta_{Ki} = \lambda_i - (\tau_{Li} + 0.2)$ 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 $\lambda_i - 0.2$. I refer to this scenario as the one with "Stronger Assumptions".

The labor 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) who also use this data to construct labor income shares adjust the observed industry labor income shares by a factor of 3/2 to scale them up to the labor income share of manufacturing observed in the National Income and Product Accounts. I follow their approach.

5.2 Results

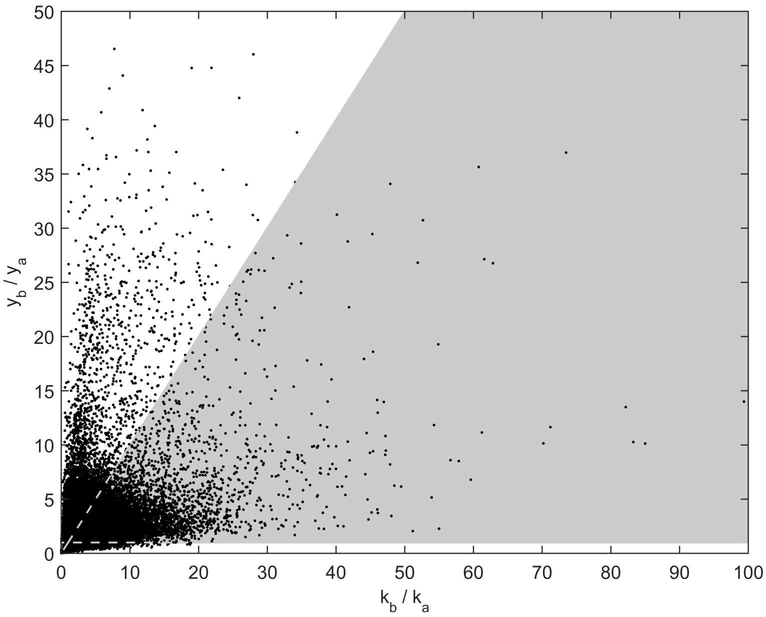
This subsection presents the results for the efficiency test and the bounds on potential output gains for the factor allocation across the 473 U.S. manufacturing industries in 2005. The analysis is conducted for the two scenarios defined in the previous subsection.

The test for an efficient factor allocation across the 473 industries is first conducted for the "most general" scenario. This is illustrated by figure 5 which plots the observed average product of labor 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 this means that there is no combination of constant returns to scale production functions across industries that can rationalize the observed allocation as one with equalized 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 rationalized as an efficient allocation. Overall these results are striking and show that even for the most general assumptions a very large share of the observations is inconsistent with an efficient factor allocation across industries. This is strong evidence against such a hypothesis.

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 observed factor income shares of each industry yields a point estimate for potential output gains of 28% here.[8] These results suggest several conclusions. First misallocation between industries is an economically significant phenomenon as indicated by the fact that potential output gains must be in excess of 22% even for the most general and lenient assumptions. Second how much the true potential

[8]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.

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 visualized 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).

output gain could exceed this lower bound does indeed depend more on the exact assumptions on output elasticities. In this application the point estimate of the traditional approach of 28% turns out to be surprisingly close to this 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. Overall these results show that the exact value of output elasticities is in principle of first-order importance for the 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.

Finally, I would like to emphasize that of course not all empirical patterns witnessed in this application are general features of the framework that necessarily carry over to other contexts. For example the fact that imposing stronger

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).

assumptions on output elasticities tightens the non-rejection region of the test and the bounds on potential output gains is a general feature of the framework. In contrast in other applications the test may not reject an efficient factor allocation for the most general assumptions and only rejects for stronger assumptions. It could also be that for the most general assumptions the upper bound on potential output gains takes a value which is large, but judged as being of a reasonable magnitude in that context. Furthermore, it does not have to be the case that a traditional approach that sets output elasticities exactly equal to factor income shares always yields a point estimate close to the lower bound.

6 Dynamics of Misallocation during the Great Recession

This section investigates the dynamics of misallocation during the Great Recession of 2008 and 2009. The question is whether the fall in economic activity during this time period was associated with a measurable increase in factor misallocation between the 473 manufacturing industries. 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 θ_{Li} and θ_{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 θ_{Li} and θ_{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. This means that objects such as output elasticities, prices, technology levels and preference parameters can 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 test results provide a first indication that factor misallocation may have worsened during the Great Recession.

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

$$\underline{\triangle G} = \inf_{\alpha \in A} \triangle G(\alpha) \tag{20}$$

and the upper bound $\overline{\triangle G}$ by

$$\overline{\triangle G} = \sup_{\alpha \in A} \triangle G(\alpha). \tag{21}$$

Note that these calculations assume constant output elasticities and hence constant parameters α_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 level of output elasticities of each industry are still only restricted to be within the assumed ranges and not fixed to some specific values. Furthermore, other quantities 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 it was only a mild decrease. 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 this section provides several pieces of evidence that suggest an increase in factor misallocation during the Great Recession. The analysis shows that this increase in misallocation may play a quantitatively significant 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.

However the analysis also reveals that identifying a strictly positive increase in potential output gains during the Great Recession does require sufficiently strong assumptions on output elasticities. Though these assumptions are still more general than traditionally used in the literature, this qualification should be firmly kept in mind. Similarly, the exact quantitative contribution of an increase in misallocation for explaining the fall in manufacturing output is sensitive to the exact assumptions on output elasticities.

The role of an increase in misallocation during the Great Recession is therefore an example where only relatively strong assumptions yield a clear answer and where the quantitative magnitude is sensitive to these assumptions. Revealing whether the answer to a certain question is or is not surrounded by such a potential ambiguity is an important strength of the presented framework. Similar robustness issues may well be present in other comparisons of misallocation across time periods or also across countries. So far these are completely unnoticed because the literature does not conduct suitable robustness checks in this respect. But the results in this section suggest that such checks are important to better understand how much confidence we should attach to the results of such analyses.

7 Robustness Checks

This section provides robustness checks of the main results. Specifically, 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 λ_i

The parameters λ_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 λ_i to a value of 1 for all industries based on a constant returns to scale assumption. Accordingly, it is an important question how this choice affects the results. This is discussed in this section.

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 λ_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 λ_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 λ_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 λ_i for different industries. In practice investigating such a specification is hampered by the lack of strong evidence on how this parameter differs across industries. Nevertheless, one can still think about whether a reasonable variation of λ_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, 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 λ_i between industries and strongly decreasing returns to scale in some industries. There are other observations on that figure for which the differences in λ_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 λ_b below about 0.034 given the other industry has $\lambda_a = 1$. Views on whether such strong differences in λ_i are reasonable may of course differ. But at the very least this discussion shows that the common beliefs that production exhibits constant or mildly decreasing returns to scale are inconsistent with the differences in λ_i required to understand this data as an efficient factor allocation.

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 does not enter the test procedure, but only affects the bounds on potential output gains. However it is clear that the ease of substituting different goods plays an important role for computing these bounds because it affects to what extent resources can be reallocated from production units with currently low to those with currently high marginal value products. In other words it determines how strong the opposing relative price changes are which limit the potential for factor reallocation.

However, 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 can 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 now specify a Constant Elasticity of Substitution (CES) utility function given by

$$U(C_1, \ldots, C_N) = \left[\sum_{i=1}^{N} \beta_i C_i^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{22}$$

where σ is the elasticity of substitution with $\sigma \neq 1$, and β_i are share parameters of the utility function. The efficient allocation for given output elasticities is then 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 σ the parameters β_i can be determined using information from the current observed allocation (derivation in the appendix) such that the social planner problem can be written as

$$\max_{\{\widetilde{k}_i^*, \ell_i^*\}_{i=1}^N} \quad \left[\sum_{i=1}^{N} \frac{s_i}{\left((\widetilde{k}_i)^{\alpha_i}(\ell_i)^{\lambda_i-\alpha_i}\right)^{\frac{\sigma-1}{\sigma}}} \left((\widetilde{k}_i^*)^{\alpha_i}(\ell_i^*)^{\lambda_i-\alpha_i}\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} \tag{23}$$

subject to the resource constraints $\sum_{i=1}^{N} \widetilde{k}_i^* = 1$ and $\sum_{i=1}^{N} \ell_i^* = 1$, and non-negativity constraints $\widetilde{k}_i^* \geq 0$ and $\ell_i^* \geq 0$ for all production units i. Otherwise the computation of the bounds on potential output gains proceeds as in the benchmark with the exception that the social planner problem for given values of α_i needs to also 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 $\sigma = 0.5$. 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 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 and more pronounced for a high 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 it was the case 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 general pattern for the considered alternative preference specifications is the same as for the benchmark results.

8 Conclusions

This paper has developed a novel theoretical framework to identify factor misallocation under more general assumptions than the prior literature. The test procedure relies on homogeneity, but not on specific functional forms of production functions. Thus it faces a lower risk of incorrectly rejecting an efficient factor allocation due to a misspecification of the basic assumptions. The test also has the virtue of being very simple and transparent. The bounds on potential output gains exploit knowledge on the range of output elasticities implied by homogeneity, but do not require information on their specific values.

In an application to the labor and capital allocation across 473 six-digit manufacturing industries in the United States the test strongly rejects an efficient allocation in 2005. The bounds show that 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 contributes substantially to the observed fall in output of the manufacturing sector. However the results on the Great Recession depend more on the exact assumptions on output elasticities and are thus not as robust as those on the general presence of misallocation. Overall these results show that one can detect misallocation in this application under much more general assumptions than used by the literature. This also suggests that the findings on misallocation in other contexts by the prior literature may not just be driven by its restrictive assumptions. But of course only a case by case analysis using more general methods can confirm this conjecture.

As in the rest of the misallocation literature employing the "indirect approach" the aim of the paper was to measure the overall degree of misallocation. This leaves the question which specific frictions, policies or institutions are responsible for misallocation across industries or the possible increase in misallocation during the Great Recession as an important topic for future research. But the paper at least shows that explanations of these phenomena which completely abstract from frictions are inconsistent with the evidence.

However reliable methods for measuring misallocation as developed in this paper may also be useful for research on specific sources of misallocation. First, they allow to put estimates of the effect of a certain distortion computed by the quantitative model of a study using the "direct approach" into perspective. One can then quantify the contribution of this distortion to the overall level of misallocation in the economy. Second, one can use such methods for conducting standard empirical analyses that compare the overall level of misallocation across regions or time periods exhibiting variation in the presence of a certain distortion.

Future work could also extend and improve the theoretical framework of this paper. One could for example modify some of the basic assumptions on production functions or preferences underlying the bounds on potential output gains. It would also be possible to search over a whole set of such specifications when computing these bounds. In contrast, one can also envisage that other researchers may want to continue using more specific and restrictive assumptions when computing potential output gains at least for their benchmark specification. But they could then still use the relatively general test procedure and suitably defined bounds to lend more credence to the general presence of misallocation in the analyzed situation. Finally, the theoretical framework in its present or a modified form could be applied to misallocation in many other contexts for example between firms, industries, sectors or regions, and for comparisons of the degree of misallocation across countries and time periods.

References

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.

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.

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.

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*.

Manski, C. F. (2007). *Identification for Prediction and Decision*. Harvard University 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*.

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.

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.

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 ε_{La}, ε_{Ka} and ε_{Kb} in equation (7) such that it reads as

$$\begin{aligned}
\frac{k_b}{k_a} &= \frac{\frac{y_b}{y_a}}{d^L_{ab}} \frac{\lambda_b - \varepsilon_{Lb}}{\lambda_a - \frac{\frac{y_b}{y_a}}{d^L_{ab}}\varepsilon_{Lb}} \frac{d^L_{ab}}{d^K_{ab}} \\
\iff \frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}}\left[\lambda_a - \frac{\frac{y_b}{y_a}}{d^L_{ab}}\varepsilon_{Lb}\right] &= \frac{\frac{y_b}{y_a}}{d^L_{ab}}[\lambda_b - \varepsilon_{Lb}] \\
\iff \frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}}\lambda_a - \frac{\frac{y_b}{y_a}}{d^L_{ab}}\lambda_b &= \frac{\frac{y_b}{y_a}}{d^L_{ab}}\left[\frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}} - 1\right]\varepsilon_{Lb} \qquad (24)
\end{aligned}$$

If $\frac{k_b}{k_a} = \frac{d^L_{ab}}{d^K_{ab}}$ then equation (24) 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 (24) can be solved for ε_{Lb} as

$$\varepsilon_{Lb} = \frac{\lambda_a \frac{k_b}{k_a}\frac{d^K_{ab}}{\frac{y_b}{y_a}} - \lambda_b}{\frac{k_b}{k_a}\frac{d^K_{ab}}{d^L_{ab}} - 1} \qquad (25)$$

Given the value of ε_{Lb} the other elasticities ε_{La}, ε_{Ka} and ε_{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 (25) 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 (25) 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 (25) is smaller than λ_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 (25) 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 $d^K_{ab} \neq \widetilde{d}^K_{ab}$ does not satisfy proposition 1. Note that for the case $\frac{y_b}{y_a} = \frac{\lambda_a}{\lambda_b} d^L_{ab}$ proposition 1 requires $\frac{k_b}{k_a} = \frac{d^L_{ab}}{d^K_{ab}}$ which implies

$$\frac{k_b}{k_a} = \frac{d^L_{ab}}{d^K_{ab}} = \frac{\widetilde{d}^L_{ab}}{d^K_{ab}} = \frac{\frac{y_b}{y_a}}{\frac{\lambda_a}{\lambda_b} d^K_{ab}} \iff d^K_{ab} = \frac{\frac{y_b}{y_a}}{\frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a}} \equiv \widetilde{d}^K_{ab}.$$

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

Second consider the case of $d^L_{ab} > \widetilde{d}^L_{ab}$:

- If as stated in the proposition $d^K_{ab} < \widetilde{d}^K_{ab}$ then $\frac{k_b}{k_a} = \frac{\widetilde{d}^L_{ab}}{\widetilde{d}^K_{ab}} < \frac{d^L_{ab}}{d^K_{ab}}$ and

$$\frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a} d^K_{ab} < \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a}\widetilde{d}^K_{ab} = \frac{y_b}{y_a} = \frac{\lambda_a}{\lambda_b}\widetilde{d}^L_{ab} < \frac{\lambda_a}{\lambda_b} d^L_{ab}$$

which is consistent with the second equation of proposition 1.

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

$$\frac{y_b}{y_a} > \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a} d^K_{ab} = \frac{\lambda_a}{\lambda_b}\frac{\widetilde{d}^L_{ab}}{\widetilde{d}^K_{ab}} d^K_{ab} = \frac{y_b}{y_a}\frac{d^K_{ab}}{\widetilde{d}^K_{ab}} \iff d^K_{ab} < \widetilde{d}^K_{ab}.$$

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 $d^K_{ab} \geq \widetilde{d}^K_{ab}$ leads to a contradiction with proposition 1.

Third consider the case of $d^L_{ab} < \widetilde{d}^L_{ab}$:

- If as stated in the proposition $d^K_{ab} > \widetilde{d}^K_{ab}$ then $\frac{k_b}{k_a} = \frac{\widetilde{d}^L_{ab}}{\widetilde{d}^K_{ab}} > \frac{d^L_{ab}}{d^K_{ab}}$ and

$$\frac{\lambda_a}{\lambda_b} d^L_{ab} < \frac{\lambda_a}{\lambda_b}\widetilde{d}^L_{ab} = \frac{y_b}{y_a} = \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a}\widetilde{d}^K_{ab} < \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a} d^K_{ab}$$

which is consistent with the first equation of proposition 1.

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

$$\frac{y_b}{y_a} < \frac{\lambda_a}{\lambda_b}\frac{k_b}{k_a} d^K_{ab} = \frac{\lambda_a}{\lambda_b}\frac{\widetilde{d}^L_{ab}}{\widetilde{d}^K_{ab}} d^K_{ab} = \frac{y_b}{y_a}\frac{d^K_{ab}}{\widetilde{d}^K_{ab}} \iff d^K_{ab} > \widetilde{d}^K_{ab}.$$

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 (24). 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 (25).

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 (25) 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

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

Second consider the case $\frac{k_b}{k_a} < \frac{d_{ab}^L}{d_{ab}^K}$. Note that the denominator of the RHS of equation (25) 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 Derivation of Equation (23)

In order to derive equation (23) 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 reads as

$$\max_{\{\widetilde{k}_i^*, \ell_i^*\}_{i=1}^N} \left[\sum_{i=1}^N \beta_i \left(A_i \overline{K}^{\alpha_i} \overline{L}^{\lambda_i - \alpha_i} (\widetilde{k}_i^*)^{\alpha_i} (\ell_i^*)^{\lambda_i - \alpha_i}\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}.$$

Noting that $A_i \overline{K}^{\alpha_i} \overline{L}^{\lambda_i - \alpha_i} = \frac{Y_i}{(\widetilde{k}_i)^{\alpha_i} (\ell_i)^{\lambda_i - \alpha_i}}$, substituting β_i for all i using the expression above for some fixed unit j and rearranging terms yields

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

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 (23) in the main text.