Accepted Manuscript

Aggregate productivity and the allocation of resources over the business cycle

Sophie Osotimehin

PII: S1094-2025(18)30270-9
DOI: https://doi.org/10.1016/j.red.2019.02.003
Reference: YREDY 911

To appear in: *Review of Economic Dynamics*

Received date: 5 May 2017
Revised date: 22 January 2019

Highlights

- I propose a new decomposition that measures changes in allocative efficiency.
- Within-sector allocative efficiency changes are counter-cyclical.
- Between-sector allocative efficiency is acyclical.
- Entry and exit contribute little to year-on-year aggregate productivity changes.
- Firm-level productivity changes are highly procyclical.

Aggregate productivity and the allocation of resources over the business cycle☆

Sophie Osotimehin

Université du Québec à Montréal, C.P. 8888, succ. Centre-ville Montréal H3C 3P8, Canada.

Abstract

This paper proposes a novel decomposition of aggregate productivity to evaluate the role of allocative efficiency for the cyclical dynamics of aggregate productivity. The decomposition, which is derived from the aggregation of heterogeneous firm-level production functions, accounts for changes in allocative efficiency, as well as for changes in entry and exit. This approach thereby extends Solow's (1957) growth accounting exercise to a framework with firm heterogeneity and frictions in the allocation of resources across firms. I apply the decomposition to a comprehensive dataset of French manufacturing and service firms, and I find that entry and exit contribute little to the year-on-year variability of sectoral productivity, between-sector allocative efficiency plays a limited role for the volatility of aggregate productivity, whereas within-sector allocative efficiency is countercyclical and tends to reduce the volatility of sectoral productivity.

☆This paper was initially written as part of my 2011 PhD thesis at Paris 1 University-Paris School of Economics, with support from the CREST-INSEE dissertation fellowship. I thank Diego Restuccia (the editor), several anonymous referees, Guy Laroque and Jean-Olivier Hairault for helpful comments and suggestions. This paper also benefited from the comments of Paul Beaudry, Jean-Michel Grandmont, Hugo Hopenhayn, Jean Imbs, Julien Martin, Isabelle Méjean, Toshihiko Mukoyama, Gianluca Violante and seminar participants at CREST, Paris School of Economics, NBER summer institute, Midwest Macro Meeting, Tsinghua Workshop in Macroeconomics, IIES Stockholm, Université de Laval, Université de Montreal, UQAM, University of Pennsylvania, University of Virginia, Federal Reserve Bank of Cleveland, Federal Reserve Board of Governors, Indiana University and University of Edinburgh. I thank Diego Legal for valuable research assistance and Susan Langenkamp for editorial assistance.

Email address: `osotimehin.sophie@uqam.ca` (Sophie Osotimehin)

URL: `https://sites.google.com/site/sosotimehin/` (Sophie Osotimehin)

Keywords: aggregate productivity, allocative efficiency, business cycle, resource allocation, entry and exit, cleansing.
JEL codes: E32, O47, D24.

1. Introduction

Recessions are often considered as a time when the economy is "cleansed"; as the least productive firms are forced to exit the market, resources are reallocated towards more productive uses. This Schumpeterian view of recessions suggests that recessions improve the efficiency of resource allocation. Despite the high interest in this view, little is known about how business cycles affect the efficiency of resource allocation. Moreover, the literature typically focuses on the cyclicality of the firms' exit rate and on the productivity of exiting firms.[1] Resource reallocation, however, mainly involves incumbent firms, as shown by Davis and Haltiwanger (1998), among others. A potentially important contribution to changes in the efficiency of resource allocation has therefore been neglected. Does the efficiency of resource allocation across incumbent firms vary over the business cycle? Is allocative efficiency an important determinant of aggregate productivity over the business cycle?

To answer these questions, I propose a novel decomposition of aggregate total factor productivity (TFP) changes that separates out the variations that are due to firm-level productivity from those due to entry and exit and to the efficiency of resource allocation. I apply this decomposition at the sectoral and economy-wide levels and hence obtain the contribution of allocative efficiency both within and between sectors. The decomposition requires deriving the link between firm-level and aggregate productivity in a framework in which resources are potentially inefficiently allocated. Exploring the micro determinants of aggregate TFP raises a methodological problem: how should we aggregate firm-level TFP? There is no consensus on this question.[2] In

[1] Recent work on the topic includes Caballero and Hammour (1994), Barlevy (2003), Ouyang (2009) and Osotimehin and Pappadà (2017) who study the cleansing effect of recessions through the lens of a model and Baden-Fuller (1989), Bresnahan and Raff (1991), Lee and Mukoyama (2015) and Foster et al. (2016) who analyze the question empirically.

[2] See e.g. Hulten (2000) for a presentation of the challenges faced by productivity analysts in linking the micro and macro productivity.

this paper, I advocate the use of a measure of aggregate productivity, computed from the aggregation of firm-level production functions, and which is therefore consistent with the measure used at the firm level. This paper thus extends Solow's (1957) growth accounting exercise to a framework with firm heterogeneity and allocative inefficiency. Deriving the contribution of allocative efficiency to aggregate productivity changes is not only of theoretical interest, it also has important implications for our understanding of the dynamics of aggregate productivity. I estimate the decomposition on French firm-level data from the manufacturing and service sectors over 1991-2006. I find that between-sector allocative efficiency shows little cyclicality and plays a limited role in the dynamics of aggregate TFP. Entry and exit and within-sector allocative efficiency both tend to be countercyclical but only within-sector allocative efficiency plays a substantial role for the dynamics of aggregate TFP. I find that the movements in within-sector allocative efficiency dampen the fluctuations in aggregate TFP.

To measure changes in allocative efficiency, I follow Chari et al. (2007) and Restuccia and Rogerson (2008) and do not specify the frictions that may distort the allocation of resources. Rather, the distortions are modeled as wedges between the firms' marginal products. The wedges, which measure the distance from the frictionless allocation of resources, encompass various sources of distortions such as adjustment costs, markups, search frictions, financial constraints or distortionary regulations, provided they create a wedge between the firms' marginal products. Distortions that modify the marginal products of all firms in the same proportion do not affect aggregate productivity; only cross-sectional inefficiencies affect the total output that can be produced out of a given quantity of inputs. Within this framework, I show how to aggregate firm-level production functions into an aggregate production function. In the aggregation literature, the traditional approach consists in defining the aggregate production function as the efficient frontier of the production possibilities set (e.g. May, 1946; Fisher, 1969; Houthakker, 1955). This approach is not suited for my purpose since it gives a measure of aggregate productivity under the assumption that resources are efficiently allocated. To derive a measure of aggregate productivity when resources may be inefficiently allocated, I use Malinvaud's (1993) insight and define the aggregate production function as the relation between aggregate output and inputs for a given allocation of resources. This aggregation method can be applied to any production function and only requires a framework that relates firm-level inputs to aggregate inputs. The aggregate production function is specific to the framework and, in particular, to the link between firm-level inputs and aggregate inputs. In this framework, which can be viewed as an accounting framework, the allocation rule is a function of firm-level productivity and distortions, the change in allocative efficiency is measured as the effects of changes in distortions. I measure the contribution of distortions, holding fixed aggregate inputs and productivity, and the contribution of firm-level productivity and aggregate inputs, holding fixed distortions. I show how to derive an exact decomposition of sectoral productivity changes into productivity changes at the firm-level, changes in allocative efficiency among incumbent firms, and changes in entry and exit. Applied across sectors, the decomposition separates the economy-wide aggregate productivity changes into within-sector productivity changes and between-sector allocative efficiency. This approach is, however, silent on the factors behind the change in allocative efficiency and cannot disentangle allocative efficiency changes due to variation in the severity of frictions from those due to aggregate inputs or firm-level productivity changes.

I use the decomposition to document the dynamics of allocative efficiency over the business cycle and quantify its role in the fluctuations of aggregate TFP. I estimate the decomposition on a comprehensive firm-level dataset of French manufacturing and service firms over the period 1991-2006. The data are collected annually by the tax administration and combined with survey data in the INSEE unified system of business statistics (SUSE). I find that, in most sectors, both the contribution of entry and exit and the efficiency of resource allocation within sectors are countercyclical. On the other hand, the efficiency of resource allocation between sectors appears to be acyclical. In line with the cleansing effect of recessions, these results indicate that within-sector allocative efficiency contributes to raising productivity during downturns. Whereas the literature emphasizes the role of entry and exit, I find that the efficiency of resource allocation between incumbent firms is more important to understand the dynamics of aggregate productivity over the business cycle. Entry and exit play a limited role in the cyclical dynamics of aggregate productivity. The countercyclicality of within-sector allocative efficiency is robust to changes in the specification of the firms' production functions. I find that allocative efficiency is countercyclical also when accounting for overhead labor costs and heterogeneity in the firms' factor elasticities. Overall, these results suggest new directions for future theoretical work since little is known on the mechanisms behind the cyclical patterns of allocative efficiency.

A vast literature has documented the importance of resource reallocation for aggregate productivity growth. Following Baily et al. (1992) and Foster et al. (2001), the most common approach is to decompose the weighted average of firm-level TFP into changes in output (or input) shares and changes in firm-level TFP. Although they provide useful insights on the patterns of reallocation, these shift-share decompositions are based on *average* firm-level TFP and may therefore not give information on the role of resource reallocation for *aggregate* TFP. In fact, aggregate TFP depends on how inputs are reallocated across firms with different marginal productivities rather than on the correlation between changes in output shares and firm-level TFP. Several recent papers, building on Basu and Fernald's (2002) work, have advocated the use of a measure of aggregate TFP to study the impact of resource reallocation (e.g., Basu et al., 2009; Petrin et al., 2011; Petrin and Levinsohn, 2012). In the vein of the earlier contribution by Jorgenson et al. (1987), these papers propose a decomposition of the aggregate Solow residual in a framework that allows for distortions and firm heterogeneity.[3] My decomposition mainly differs from these papers in its objective. As the shift-share decompositions, these papers focus on measuring the effects of changes in resources allocation. The recent paper by Baqaee and Farhi (2018) follows this approach as well. By contrast, my objective is to provide a decomposition that accounts for changes in the efficiency of resource allocation. In addition, I propose an alternative aggregation method that relies on the aggregation of firm-level production functions rather than on the Solow residual. With the objective of measuring the effect of allocative efficiency, I derive the link between aggregate output and aggregate inputs, taking distortions as given, whereas the Basu-Fernald decomposition takes the firms' input shares as given. The difference between my decomposition and the Basu-Fernald decomposition is best shown by an example. Consider an economy with a fixed input allocation: firms never vary their inputs, even in the face of productivity shocks. In that economy, the contribution of firm-level inputs to aggregate produc-

[3]These papers acknowledge the aggregation issue that arises in the presence of frictions; they emphasize that the link between aggregate output and inputs depends on how inputs are allocated across firms. Another example in the applied macro literature is the paper by Rotemberg and Woodford (1993).

tivity, measured by the Basu-Fernald decomposition, is equal to zero since there is no change in inputs. The efficiency of resource allocation, however, will vary over time depending on the distribution of firm-level productivity shocks. In particular, allocative efficiency would improve if firms with low marginal productivity experienced an increase in their TFP.
To underline the differences between my approach and the existing literature, I compare the results of my decomposition to those obtained with Foster et al.'s (2001) shift-share decomposition and with a reallocation decomposition in the style of Basu and Fernald (2002). I find that neither the shift-share decomposition nor the reallocation decomposition is a good proxy for changes in allocative efficiency. Whereas changes in allocative efficiency tend to reduce the volatility of aggregate productivity, the contribution of reallocation to aggregate productivity is negligible, and the reallocation component in the shift-share decomposition tends to raise the volatility of average productivity. In the end, these decompositions give complementary information on the patterns of reallocation and its role for the dynamics of aggregate productivity.

This paper is also related to the growing literature that highlights the role of firm-level allocative distortions as a determinant of aggregate productivity.[4] In particular, my paper is related to Oberfield (2013), Sandleris and Wright (2014), and Gopinath et al. (2017) who study the dynamics of allocative efficiency.[5] As these papers, I build on Hsieh and Klenow (2009) and measure allocative efficiency from the wedges between the firms' marginal productivity. My paper contributes to the literature by deriving the contribution of entry and exit and by extending Hsieh and Klenow's (2009)

[4]See for example Restuccia and Rogerson (2008); Guner et al. (2008); Buera et al. (2011); Hsieh and Klenow (2009); Sandleris and Wright (2014); Hopenhayn (2012); Bartelsman et al. (2013).

[5]Sandleris and Wright (2014) focus on the evolution of allocative efficiency during the Argentine crisis of 2001, Oberfield (2013) during the Chilean crisis of 1982, and Gopinath et al. (2017) in Spain after the introduction of the euro. My approach is related but distinct from these three papers. I propose a general method to derive the aggregate production function in the presence of distortions, and I measure the contribution of changes in distortions to aggregate productivity, whereas Sandleris and Wright (2014) use a more restrictive setting that can be only applied within sectors, and Oberfield (2013) and Gopinath et al. (2017) use the aggregate production function implied by the efficient allocation and focus on the gap between the actual and efficient output.

approach to more flexible production functions.[6] I propose a method to aggregate firm-level production functions in the presence of distortions in the allocation of inputs across firms. The aggregation method does not require a specific functional form and therefore permits evaluating the robustness of the results to alternative functional forms. With distortions likely to be sensitive to the specification of the production function, evaluating the robustness of the results to alternative specifications is crucial. I show that the results are robust to allowing for heterogenous factor elasticities and for fixed labor costs of production. Furthermore, while Hsieh and Klenow (2009) focus on within-sector allocative efficiency, the aggregation method allows me to derive the change in the allocative efficiency between sectors.

The paper is organized as follows. Section 2 describes, in a simplified framework, how to aggregate heterogeneous production units and how to derive the decomposition of aggregate productivity. That section also compares the decomposition to the existing literature. Section 3 presents the decomposition of aggregate productivity in the general framework. Section 4 presents the estimation method and the results obtained on French firm-level data. Section 5 concludes.

2. Measuring changes in allocative efficiency: a simplified framework

In this section, I show, in a simplified framework, how to measure the contribution of allocative efficiency to changes in aggregate productivity, and I highlight how my approach differs from the existing decompositions of aggregate productivity changes. In the general framework, presented in Section 3, firms produce using both labor and capital inputs; factor elasticities are heterogenous between sectors; and goods are heterogenous across firms. In the stripped-down framework presented here, labor is the only input; factor elasticities are identical across firms; and goods are homogenous.

[6] The aggregation method leads to the same aggregate production function as Hsieh and Klenow (2009) when firms use a Cobb-Douglas production function and have identical factor elasticities, as is assumed within sectors.

2.1. Framework

Let us consider a simple framework in which firms produce homogenous goods according to the production function

$$Y_{it} = A_{it} L_{it}^{\beta},$$

with $\beta < 1$, where Y_{it} denotes firm i's output, L_{it} its labor and A_{it} its TFP in period t. Aggregate output is $Y_t = \sum_i Y_{it}$, and aggregate input $L_t = \sum_i L_{it}$. Entry and exit are exogenous events, with I_{it} indicating whether the firm is active.[7] The firm is active when $I_{it} = 1$ and inactive when $I_{it} = 0$, and n_t denotes the number of active firms. Because of market frictions, labor may be inefficiently allocated across firms. The firm's marginal productivity can be written

$$\beta A_{it} L_{it}^{\beta - 1} = \mu_t (1 + \tau_{it}), \tag{1}$$

where μ_t is the shadow cost of labor, and τ_{it} is the distortion that captures the effect of market frictions that prevent the marginal productivity of labor from being equalized across firms. The distortion hence measures the deviation from the first-order conditions of the first-best allocation of resources. As in Chari et al. (2007) and Restuccia and Rogerson (2008), this reduced-form approach allows me to remain agnostic about the nature of the frictions that may distort the allocation of resources. The distortion τ_i encompasses different types of frictions such as imperfect competition, adjustment costs, search frictions, financial constraints, and distortionary regulation as they may all generate wedges between the firms' marginal productivity. As will be shown in the next section, distortions of this form reduce aggregate productivity only if they are heterogeneous across firms. Frictions that generate a wedge between the marginal product and the marginal cost of an input are not a source of allocative inefficiency as long as they affect all firms in the same proportion. Hence, imperfect competition distorts the cross-sectional allocation of resources only to the extent that markups differ across firms.[8]

[7] This assumption allows me to measure how entry and exit affect the distribution of firm-level productivity and distortions. By contrast, Bartelsman et al. (2013) consider a related model with endogenous entry that allows them to study the effects of distortions on equilibrium entry.

[8] Note that this would no longer be the case if firms produced with intermediate inputs as shown, for example, by Rotemberg and Woodford (1993).

2.2. Aggregation in an inefficient economy

To assess the contribution of allocative efficiency to aggregate TFP, we first need to derive the aggregate production function. By definition, aggregate TFP is intrinsically related to the production function; it measures the change in real output not accounted for by changes in real inputs. This section shows how to derive the aggregate production function in the presence of distortions in the allocation of resources.

The aggregation of heterogeneous firm-level production functions is a classical problem in macroeconomics.[9] The link between aggregate inputs and output is not straightforward to derive. In fact, the same change in aggregate input may affect aggregate output differently, depending on the way the additional input is allocated across firms. It is well known that, if the firms' inputs are allowed to take any values, the aggregation of firm-level production functions is impossible unless very restrictive conditions are imposed on the production functions.[10] In the aggregation literature, the traditional solution consists in deriving the aggregate production function consistent with an *optimal* allocation of resources (e.g. May, 1946; Houthakker, 1955; Fisher, 1969). As the focus is here on resource misallocation and production inefficiencies, I use instead Malinvaud's (1993) insight and derive the aggregate production function for a *given* allocation rule. The aggregate production function is then the sum of the individual production functions expressed using the allocation rule, which specifies how aggregate input is allocated between firms. More formally, if firms produce with technology $Y_{it} = f_i(L_{it})$, and the allocation rule is $L_{it} = l_{it}(L_t)$, the aggregate production function is simply $Y_t = \sum_i f_i(l_{it}(L_t))$. I apply this method to derive a measure of the contribution of firm-level productivity and allocative efficiency to aggregate productivity. Note that, while Malinvaud's insight is general, the aggregate production function is specific to the allocation rule, and a different allocation rule would lead to a different aggregate production function. I use the allocation rule that allows me to measure changes in allocative efficiency. To do so, I parameterize the allocation rule by the distribution of firm-level productivity and distortions, and measure allocative efficiency as the contribution of distortions to aggregate productivity. With this allocation rule,

[9] The aggregation problem is formalized by Klein (1946).

[10] The individual production functions must be linear and have the same slopes (Nataf, 1948).

the aggregate production function measures the changes in output due to changes in firm-level productivity and changes in aggregate inputs, holding fixed distortions, and the changes in distortions holding fixed firm-level productivity and aggregate inputs. Note that, with this approach, the measured contribution of an increase in aggregate inputs does not take into account the possible accompanying change in allocative efficiency (as would happen, for example, if the additional labor gets allocated to less constrained firms); this effect is captured in the allocative efficiency component. Without a more structural approach and an explicit modeling of the frictions behind the distortions, we cannot separate the variation in allocative efficiency due to changes in the severity of frictions from those due to changes in aggregate inputs or changes in firm-level productivity.

The allocation rule, derived from equation (1) and from the aggregate resource constraint, is such that the firms' labor is proportional to aggregate labor.[11] Substituting the allocation rule in the expression of aggregate output, $Y_t = \sum_i A_{it} L_{it}^{\beta}$, gives the following aggregate production function,

$$Y_t = \mathrm{TFP}_t L_t^{\beta}, \tag{2}$$

where aggregate TFP is a function of firm-level productivity and distortions,

$$\mathrm{TFP}_t = \frac{\sum_i A_{it}^{\frac{1}{1-\beta}} (1+\tau_{it})^{-\frac{\beta}{1-\beta}}}{\left(\sum_i A_{it}^{\frac{1}{1-\beta}} (1+\tau_{it})^{-\frac{1}{1-\beta}}\right)^{\beta}}. \tag{3}$$

Malinvaud's (1993) aggregation method does not require any conditions on the firm-level production functions and can be applied to more flexible production functions than the one considered here. In general, however, the aggregate production function does not share the same functional form as the individual production functions. When the individual production function is Cobb Douglas, I find that, given the proportional allocation rule, the aggregate production function is also Cobb-Douglas if the factor elasticities are identical across firms (see Appendix A.1). The Cobb-Douglas specification leads to a closed form solution for aggregate TFP and facilitates the

[11]The allocation rule is $L_{it} = \left(A_{it}^{\frac{1}{1-\beta}} (1+\tau_{it})^{-\frac{1}{1-\beta}}\right) / \left(\sum_i A_{it}^{\frac{1}{1-\beta}} (1+\tau_{it})^{-\frac{1}{1-\beta}}\right) L_t$.

measure of the contribution of entry and exit.[12]

Aggregate TFP can be measured either from aggregate data as the residual of the aggregate production function, using equation (2), or from firm-level data on productivity and distortions, using equation (3). To see more clearly how aggregate productivity depends on the cross-sectional distribution of firm-level productivity and distortions, let us assume that firm-level productivity and distortions are given by $A_i = \bar{A}(1 + \hat{A}_i)$ and $(1 + \tau_i) = (1 + \overline{\tau})(1 + \hat{\tau}_i)$, where $\hat{A}_i$ and $\hat{\tau}_i$ are jointly distributed, with zero mean, standard deviations σ_A, σ_τ and covariance $\sigma_{A\tau}$. For uniformly small deviations, using a Taylor expansion of equation (3) and the law of large numbers, aggregate productivity can be approximated by:

$$\text{TFP} \simeq n^{1-\beta}\bar{A}\left[1 + \frac{1}{2}\frac{\beta}{1-\beta}\left(\sigma_A^2 - \sigma_\tau^2\right)\right]. \tag{4}$$

The approximation, which can also be derived for the general framework (see Appendix A.2), shows that the dispersion of firm-level distortions reduces aggregate productivity. It also appears that the average level of distortions $\bar{\tau}$ has no impact on aggregate productivity. Indeed, the distortions, defined as the distance to the first-order conditions of the frictionless allocation, affect the cross-sectional efficiency of aggregate production only if they differ across firms. Therefore, fluctuations in the average level of distortions do not affect allocative efficiency; only changes in the relative level of distortions matter. As will be shown later, the irrelevance of the average level of distortions for aggregate productivity will be useful to estimate firm-level distortions. Equation (4) also shows, in line with Hopenhayn (2012), that the covariance between firm-level productivity and distortions has no effect on aggregate productivity. This result contrasts with Restuccia and Rogerson (2008), who underline the role of the correlation between firm-level productivity and distortions for aggregate productivity. Note that equation (4) is an approximation of aggregate productivity; therefore, the contrast with Restuccia and Rogerson (2008) suggests that the correlation between

[12] The measure of the contribution of entry and exit relies on the counterfactual output in the absence of entry or exit. If there were no closed form solution for aggregate TFP, the contribution of entry and exit on aggregate productivity could still be measured but would be more involved; in particular, it would require computing how the labor used by entrants and exiters is reallocated to continuing firms.

firm-level productivity and distortions matters through higher-order components. Finally, the equation shows that aggregate productivity increases with the number of firms when firms have decreasing returns to scale.[13]

2.3. Accounting for changes in aggregate productivity

Let us now decompose TFP changes ($\Delta \ln \mathrm{TFP} = \ln \mathrm{TFP}_t - \ln \mathrm{TFP}_{t-1}$) into changes in technical efficiency (ΔTE), changes in allocative efficiency (ΔAE), and changes in entry and exit (ΔEX):

$$\Delta \ln \mathrm{TFP} = \Delta\mathrm{TE} + \Delta\mathrm{AE} + \Delta\mathrm{EX}. \tag{5}$$

To decompose the change in TFP, I use equation (3) and apply a decomposition similar to those used to separate price and volume changes in the index number literature. The components of the decomposition of equation (5) add up exactly to the change in TFP. For simplicity, I present here an approximation of ΔTE and ΔAE based on the continuous-time decomposition of equation (3). A similar approximation is presented in Section 3 for the general case. The exact decomposition used for the empirical analysis, and the derivation of the approximation presented here, are both described in Appendix A. I use here and in the rest of the paper the notation $\Delta x_t \equiv x_t - x_{t-1}$ for any generic variable x. For simplicity, I drop the time subscript when there is no risk of confusion.

To derive the decomposition, we first need to separate the contribution of continuing firms from that of entering and exiting firms:

$$\Delta \ln \mathrm{TFP} = \Delta \ln \mathrm{TFP}^C + \underbrace{\ln \mathrm{TFP}_t - \ln \mathrm{TFP}_t^C - (\ln \mathrm{TFP}_{t-1} - \ln \mathrm{TFP}_{t-1}^C)}_{\Delta\mathrm{EX}}, \tag{6}$$

where TFP_t^C is the aggregate TFP computed on the set of continuing firms, $\mathcal{C}_t$, that is, firms with $I_{it-1} = 1$ and $I_{it} = 1$.[14] The component ΔEX gives

[13] Note that in the general framework, the number of firms has no effect on aggregate TFP if firms have constant returns to scale and in the absence of variety effects on the demand side, as assumed in the empirical application (see Appendix A.2).

[14] The aggregate TFP of continuing firms is given by

$$\ln \mathrm{TFP}_t^C = \ln\left(\sum_{i\in\mathcal{C}_t} A_{it}^{\frac{1}{1-\beta}}(1+\tau_{it})^{-\frac{\beta}{1-\beta}}\right) - \beta \ln\left(\sum_{i\in\mathcal{C}_t} A_{it}^{\frac{1}{1-\beta}}(1+\tau_{it})^{-\frac{1}{1-\beta}}\right)$$

.

the effect of entry and exit on the distribution of firm-level productivity and distortions. It measures the gap between TFP and the counterfactual TFP in the absence of entry or exit; more precisely, the contribution of entry is measured by the gap between the TFP of continuing firms and the overall TFP at time t, and the contribution of exit is measured by the gap between the TFP of continuing firms and the overall TFP at time $t-1$. Note that ΔEX captures only the contemporaneous and direct effect of entry and exit on aggregate productivity.[15]

Let us now separate the aggregate productivity of continuing firms into the contribution of firm-level productivity and distortions. The decomposition used in the empirical application is an exact decomposition based on index number methods. I present here, for simplicity, an approximation based on the continuous-time decomposition of TFP^C. Although the details of the decomposition will be affected by the particular method used, the logic behind the two methods is identical. The continuous-time decomposition is obtained by taking the total derivative of aggregate productivity:

$$\Delta \ln \text{TFP}^C \simeq \underbrace{\sum_{i \in \mathcal{C}_t} \frac{d \ln \text{TFP}^C}{d A_i} \Delta A_i}_{\Delta\text{TE}} + \underbrace{\sum_{i \in \mathcal{C}_t} \frac{d \ln \text{TFP}^C}{d \tau_i} \Delta \tau_i}_{\Delta\text{AE}} . \tag{7}$$

The first component, ΔTE, measures the impact of changes in firm-level productivity holding fixed the distortions, that is, holding fixed the degree of allocative efficiency. Note that because of measurement issues, this "technical" efficiency component may capture non-technological factors as well, such as variation in input utilization or demand shocks. The allocative efficiency component, ΔAE, measures how changes in firm-level distortions affect aggregate productivity. Combining this equation with equation (6) yields the overall decomposition of aggregate productivity growth, equation (5).

Alternatively, the decomposition can be derived starting from the aggregate production function, equation (2), instead of starting from equation (3). From equation (2), we have $\ln \text{TFP}^C = \ln \sum_{i \in \mathcal{C}_t} A_i L_i^\beta - \beta \ln \sum_{i \in \mathcal{C}_t} L_i$, and the change in the aggregate TFP of continuing firms can then be decomposed into

[15] Furthermore, the framework leaves aside the endogenous determination of entry and exit. Contrary to Bartelsman et al. (2013), I do not measure the impact of distortions on the number of entering or exiting firms. The extensive margin component instead captures the variations in the distribution of distortions and productivity induced by entry and exit.

the changes in firm-level productivity and the changes in firm-level inputs,

$$\Delta \ln \mathrm{TFP}^C \simeq \underbrace{\sum_{i\in\mathcal{C}_t} \frac{Y_i}{Y^C}\frac{\Delta A_i}{A_i}}_{\text{within}} + \underbrace{\beta \sum_{i\in\mathcal{C}_t} \frac{\tau_i - \overline{\tau}^L}{1+\tau_i}\frac{Y_i}{Y^C}\frac{\Delta L_i}{L_i}}_{\text{reallocation}}, \tag{8}$$

with $\overline{\tau}^L \equiv \sum_i \tau_i (L_i/L)$. The contribution of changes in firm-level inputs measures the effects of reallocation on aggregate productivity. To obtain the allocative efficiency decomposition, we need to further separate the reallocation term into the changes in inputs that are due to firm-level productivity and those due to distortions. We obtain

$$\Delta \ln \mathrm{TFP}^C \simeq \underbrace{\sum_{i\in\mathcal{C}_t} \frac{Y_i}{Y^C}\frac{\Delta A_i}{A_i} + \frac{\beta}{1-\beta}\sum_{i\in\mathcal{C}_t} \frac{\tau_i - \overline{\tau}^L}{1+\tau_i}\frac{Y_i}{Y^C}\frac{\Delta A_i}{A_i}}_{\Delta TE} \underbrace{- \frac{\beta}{1-\beta}\sum_{i\in\mathcal{C}_t} \frac{\tau_i - \overline{\tau}^L}{1+\tau_i}\frac{Y_i}{Y^C}\frac{\Delta \tau_i}{1+\tau_i}}_{\Delta AE}, \tag{9}$$

which gives a decomposition identical to equation (7). The derivation in two steps helps clarify the nature of the decomposition and its relation to existing decompositions. The existing literature approach, which is similar to equation (8), focuses on reallocation and consists in separating the effect of firm-level productivity changes, holding fixed the input allocation, and the effect of changes in the input allocation, holding fixed productivity. In my decomposition, given by equation (7) or (9), I instead focus on changes in allocative efficiency, and I separate the effect of firm-level productivity changes, holding fixed distortions, and the effect of changes in distortions, holding fixed firm-level productivity. In the end, the key difference is what is held fixed.

As shown by equation (9), allocative efficiency falls when distortions rise in firms with higher-than-average distortions. Conversely, higher distortions in firms with below-average distortions raise allocative efficiency. As already mentioned, what matters for allocative efficiency is the dispersion of distortions across firms, not the level of distortion per se. When distortions are identical across firms, $\tau_i = \overline{\tau}^L$, inputs are efficiently allocated, and there is therefore no change in allocative efficiency. In that case, $\Delta AE = 0$, and the technical efficiency term simplifies to the weighted average of firm-level productivity changes, that is $\Delta TE = \sum_{i\in\mathcal{C}_t}(Y_i/Y^C)\Delta A_i/A_i$. In the general case where the firms' decisions may be distorted, however, the technical efficiency term has an additional component. The first component captures the effect of the change in firm-level productivity for a given input allocation, and the second component captures the effect of the change in inputs that is required to hold allocative efficiency constant after the change in firm-level productivity. To fix ideas, consider a firm with a higher-than-average distortion hit by a positive productivity shock. If the firm keeps its labor constant, despite the increase in productivity, then the firm's marginal productivity – already too high – rises further, and the firm becomes even more distorted. To maintain allocative efficiency constant, the firm's labor must increase. If the firm's labor does not increase as much as what is required to maintain allocative efficiency constant, then allocative efficiency declines. Conversely, if the firm's labor increases more than what is required to maintain allocative efficiency constant, then allocative efficiency improves. The aggregate effect of the increase in the firm's input depends on the firm's relative distortion. The required increase in the firm's input will have a larger impact on aggregate productivity if the firm's distortion – and hence its marginal productivity – is high. Relative to the reallocation decomposition of equation (8), the firm-level productivity component puts more weight on firms with high distortions and less weight on firms with low distortions.

The allocative efficiency decomposition (equation (7) or (9)) can be rewritten more simply as a function of output and input shares:

$$\Delta TE \simeq \sum_{i \in \mathcal{C}_t} \frac{Y_i}{Y_t^C} \frac{\Delta A_i}{A_i} + \frac{\beta}{1-\beta} \sum_{i \in \mathcal{C}_t} \left(\frac{Y_i}{Y^C} - \frac{L_i}{L^C} \right) \frac{\Delta A_i}{A_i}, \tag{10}$$

$$\Delta AE \simeq \frac{\beta}{1-\beta} \sum_{i \in \mathcal{C}_t} \left(\frac{L_i}{L^C} - \frac{Y_i}{Y^C} \right) \frac{\Delta \tau_i}{1+\tau_i}. \tag{11}$$

Here, the deviation of the firm's distortion from the average distortion is captured by the difference between the firm's output share and input share. In the absence of distortions, the firm size depends only on firm-level productivity, and input and output shares are equal. The presence of distortions creates a wedge between the firm's output and input shares; the larger the wedge is, the more distorted the firm is. Equation (11) also shows more clearly that the change in allocative efficiency is equal to zero if the change in distortions is identical across firms, $\Delta\tau_i/(1+\tau_i) = \Delta\bar{\tau}/(1+\bar{\tau})$; the relative marginal productivities are not modified in that case.

The contribution of the extensive margin can also be written more simply as a function of the output and labor shares of entering and exiting firms:

$$\Delta EX \simeq (1-\beta)\frac{\Delta n_t}{n_{t-1}} + \frac{e_t}{n_t}\left[\left(\frac{\frac{1}{e_t}\sum\limits_{i\in\mathcal{E}_t} Y_{it}}{\frac{1}{c_t}\sum\limits_{i\in\mathcal{C}_t} Y_{it}} - 1\right) - \beta\left(\frac{\frac{1}{e_t}\sum\limits_{i\in\mathcal{E}_t} L_{it}}{\frac{1}{c_t}\sum\limits_{i\in\mathcal{C}_t} L_{it}} - 1\right)\right]$$
$$- \frac{x_t}{n_{t-1}}\left[\left(\frac{\frac{1}{x_t}\sum\limits_{i\in\mathcal{X}_t} Y_{it-1}}{\frac{1}{c_t}\sum\limits_{i\in\mathcal{C}_t} Y_{it-1}} - 1\right) - \beta\left(\frac{\frac{1}{x_t}\sum\limits_{i\in\mathcal{X}_t} L_{it-1}}{\frac{1}{c_t}\sum\limits_{i\in\mathcal{C}_t} L_{it-1}} - 1\right)\right], \quad (12)$$

where $\mathcal{E}_t$ the set of entering firms ($I_{it-1} = 0$ and $I_{it} = 1$), $\mathcal{X}_t$ the set of exiting firms ($I_{it-1} = 1$ and $I_{it} = 0$), and c_t, e_t and x_t denote the number of continuing, entering, and exiting firms. Aggregate productivity rises if entering firms are more productive and exiting firms less productive than incumbent firms, or if entrants are less distorted and exiting firms more distorted than the average incumbent. The relative firm-level productivity and distortions are captured in the relative output and input size of the firms.

2.4. Relation to the literature

Several decompositions have been proposed to measure the contribution of input reallocation to aggregate productivity. My decomposition differs from existing decompositions both in the objective and in the aggregation method. Whereas the literature has focused on measuring the impact of the change in resource allocation, the decomposition derived in this paper aims at measuring the change in the *efficiency* of resource allocation. In addition, I propose an alternative aggregation method that relies on the aggregation of the firms' production functions. In this section, I describe in more detail how my decomposition relates to the literature, and I clarify the interpretation of the existing decompositions. I focus on two types of decompositions: the shift-share decompositions, which are based on the weighted average of firm-level productivity, as in Foster et al. (2001), and the decompositions based on the Solow residual, as in Basu and Fernald (2002). I abstract here from entry and exit, absent from Basu and Fernald's (2002) framework.[16] For simplicity, time subscripts are omitted in this section.

[16] I report the entry and exit component of Foster et al. (2001) in Appendix C.2.2. In contrast to equation (12), the contribution of entry and exit in Foster et al.'s (2001) only depends on the relative productivity of entrants and exiters.

2.4.1. From reallocation to allocative efficiency

The key difference between my approach and the existing decompositions reflects the difference in the objective of the decompositions. Most decompositions in the literature focus on the contribution of input reallocation to aggregate productivity, similarly to equation (8). I pursue a related but distinct objective. I aim to measure the change in the efficiency of resource allocation. To clarify the difference between these two related concepts, I leave aside here the other differences with the literature (coming from the aggregation method) and compare my decomposition, equation (9), to the reallocation decomposition of equation (8).

The contribution of input reallocation, measured by the second component of equation (8), is conceptually distinct from changes in allocative efficiency. Measuring changes in allocative efficiency requires measuring the distance between the data and the efficient allocation of resources. In equation (9), the distance to the efficient allocation is captured by the distortions τ_i, and allocative efficiency is measured by the effects of changes in τ_i on aggregate productivity. The contribution of reallocation and changes in allocative efficiency are identical only when firm-level productivity is constant. In general, however, firm-level productivity changes create a discrepancy between the reallocation component and allocative efficiency, and the patterns of input reallocation are insufficient to infer changes in allocative efficiency. As shown by equation (8), labor reallocation contributes positively to aggregate productivity if labor increases in firms with higher-than-average distortions. This does not imply, however, that allocative efficiency has improved. Equation (9) shows that allocative efficiency improves only if distortions are reduced in firms with higher-than-average distortions. To illustrate the difference between input reallocation and allocative efficiency, let us consider an economy in which firms face productivity shocks but never vary their inputs. In this economy, the contribution of labor reallocation to aggregate productivity changes is equal to zero since there is no change in firm-level inputs. The efficiency of resource allocation, however, will vary over time with the distribution of firm-level productivity shocks. In fact, in this example, the change in firm-level distortions is equal to the change in productivity. Allocative efficiency may thus increase or decrease depending on the correlation between firm-level productivity changes and distortions.[17] Allocative

[17]Using the first order condition of the firm, $\frac{dL_i}{L_i} = 0$ implies $\frac{d\tau_i}{1+\tau_i} = \frac{dA_i}{A_i}$. Replace in

efficiency would stay constant only if inputs were reallocated across firms so as to maintain the same distribution of marginal productivities. Note that the firm-level productivity component also is different in the two decompositions. In the reallocation decomposition, the firm-level productivity component measures the effects of firm-level productivity, holding inputs constant whereas it measures, in the allocative efficiency decomposition, the effects of firm-level productivity, holding allocative efficiency constant. In a nutshell, the two firm-level productivity components differ in what is held fixed, the distortions or the allocation of labor across firms.

All in all, the allocative-efficiency and the reallocation decompositions give complementary information on the role of reallocation for aggregate productivity. Whereas the reallocation decomposition helps us answer the question "Has the reallocation of inputs contributed to increasing aggregate productivity?", the allocative-efficiency decomposition answers the question "Has allocative efficiency improved over the period?". The difference between the allocative-efficiency and reallocation decompositions is akin to the difference between the two decompositions used in the development accounting literature. Using the aggregate production function, cross-country income differences can be decomposed into differences in capital accumulation and differences in aggregate productivity; alternatively, we could view capital accumulation as driven by aggregate productivity and the saving rate, and then decompose output changes into changes in the saving rate and changes induced by aggregate productivity, which includes the endogenous response of capital.[18] Likewise, we can decompose aggregate productivity into changes in firm-level productivity and changes in input allocation, or we can take into account the response of the input allocation to changes in firm-level productivity, and decompose aggregate productivity into changes induced by firm-level productivity, including the endogenous response of input allocation, and changes in the efficiency of input allocation. In both cases, the first approach has the advantage of being simple and of requiring less structure (only the production function); the second approach requires more structure (we also need to specify an allocation rule) but is more useful to describe how changes in behavior or changes in efficiency affect aggregate output.

equation (9) to see the role of the correlation between firm-level productivity changes and distortions.

[18]I thank an anonymous referee for this useful analogy.

In addition to pursuing a different objective, I propose a different aggregation method, which creates additional discrepancies between the allocative efficiency decomposition and the existing decompositions. To shed light on the role of the aggregation method, I compare, in the next sections, the existing decompositions with the reallocation decomposition (equation (8)).

2.4.2. Shift-share decompositions

Following Baily et al. (1992), Griliches and Regev (1995) and Foster et al. (2001), the role of resource reallocation for aggregate productivity has mostly been studied using shift-share decompositions of average firm-level productivity.[19] The shift-share decompositions are based on the weighted average of firm-level productivity, $\ln \bar{A} \equiv \sum_i s_i \ln A_i$. I will consider here the case where the weights are output shares, $s_i \equiv Y_i/Y$, although input shares could be used as well. I abstract from entry and exit and focus on the decomposition of the productivity of continuing firms. The decomposition is of the form[20]

$$\Delta \ln \bar{A} \;\; = \;\; \sum_i s_i \; \Delta \ln A_i + \sum_i \Delta s_i \;\; \ln A_i. \tag{13}$$

The first component of the equation captures the changes in firm-level productivity holding fixed the firms' shares. The second component measures the impact of changes in the shares on average productivity.

As already mentioned, the objective of my decomposition is different from that of the shift-share decomposition. Whereas the shift-share decomposition focuses on the contribution of changes in firms' shares, my decomposition aims at measuring changes in allocative efficiency. Although often interpreted as such, changes in firms' shares do not capture changes in allocative efficiency. The shift-share decompositions also substantially differ in the aggregation method. While the shift-share decompositions focus on *average* firm-level TFP, I analyze the changes in *aggregate* TFP, which I derive from the aggregation of firm-level production functions. Hence, the shift-share decomposition differs not only from the allocative efficiency decomposition, equation (9), but also from the reallocation decomposition, equation

[19] An alternative decomposition of average firm-level productivity has been proposed by Olley and Pakes (1996) and Melitz and Polanec (2015).

[20] There are different variants of the shift-share decomposition, depending on whether the weights used to sum $\Delta \ln A_i$ and Δs_i are the weights of period t, period $t-1$, or the average of the two. See Appendix C.2 for the precise decomposition used in the empirical application (which includes the entry and exit component).

(8), which measures the contribution of changes in input shares to aggregate TFP. The comparison with the reallocation decomposition clarifies the role of the aggregation method. It turns out that the aggregation method does not affect the measure of the within-firm productivity component, which is identical in the shift-share and reallocation decompositions (when the shift-share decomposition is computed using output shares). The measure of aggregate productivity, however, becomes crucial when deriving the role of input reallocation. Indeed, input reallocation has a different impact on average and aggregate TFP.[21] A reallocation of inputs towards high TFP firms unambiguously raises *average* TFP but could, at the same time, reduce *aggregate* TFP if those firms also have a low marginal productivity.

To further illustrate the role of the aggregation method, consider the case where inputs are efficiently allocated. In that case, the marginal productivities are equalized across firms, and the contribution of reallocation to aggregate productivity changes is zero; the distortions are identical across firms, and equation (8) becomes:

$$\Delta \ln \mathrm{TFP} = \underbrace{\sum_i s_i \Delta \ln A_i}_{\text{within}} + \underbrace{0}_{\text{reallocation}} .$$

By contrast, the contribution of reallocation to average firm-level productivity may be positive or negative. Equation (13) can be rewritten:[22]

$$\Delta \ln \bar{A} \simeq \underbrace{\sum_i s_i \; \Delta \ln A_i}_{\text{within}} + \underbrace{\frac{1}{1-\beta} \sum_i s_i (\ln A_i - \ln \bar{A}) \Delta \ln A_i}_{\text{reallocation}} .$$

When inputs are efficiently allocated, the change in the firm's output share is a function of the firm's productivity change, and the reallocation component will therefore depend on the correlation between the firm's level and growth rate of productivity.[23] The shift-share decomposition has the advantage of

[21]Note that the two reallocation components are identical when goods are homogeneous and produced with one input and with constant returns to scale. In the more general case, where goods are heterogeneous or when the marginal productivity of one input is decreasing, reallocation may have different implications for average and aggregate TFP.

[22]See Appendix A.6 for the details of the derivation.

[23]With a mean reverting process of productivity, the reallocation component of the

not requiring any structural assumption, but the decomposition can also be more difficult to interpret. As illustrated by the example above, the contribution of input reallocation to average firm-level productivity may be very different from its contribution to aggregate TFP.

2.4.3. Solow residual decompositions

The second type of decomposition used in the literature is based on the decomposition proposed by Basu and Fernald (2002). Basu and Fernald (2002), along with other papers in the applied macroeconomic literature, acknowledge explicitly the aggregation issue that arises in the presence of frictions and highlight that aggregate productivity depends not only on technology but also on how inputs are allocated across firms. By contrast with Foster et al. (2001), they use the Solow residual as a measure of aggregate productivity.[24] Within the simplified framework, their measure of aggregate productivity growth (denoted APG) is:

$$APG = \frac{\Delta Y}{Y} - \frac{wL}{Y}\frac{\Delta L}{L},$$

shift-share decomposition would be negative if the initial period is used to compute $\ln A_i$ since mean reversion implies a negative correlation between the firms' growth rate and level of productivity.

[24]Another decomposition of the Solow residual has been proposed by Jorgenson et al. (1987). They decompose the growth rate of the aggregate Solow residual into a weighted sum of sectoral productivity growth rates and a component that reflects the contribution of the reallocation of outputs and inputs across sectors. The latter is computed as weighted sums of the growth rate of value added, capital, and labor. The Basu and Fernald's (2002) decomposition reduces to the Jorgenson et al.'s (1987) decomposition in the case where the only source of distortions is input price heterogeneity.

where $wL = \sum_i wL_i$ is the aggregate wage bill.[25] The Solow residual is closely related to my measure of aggregate productivity change, and differs only inasmuch as the factor share of income is different from the output elasticity with respect to that factor. The two measures are equal when average distortions are equal to zero. Note that this condition is sufficient only if factor elasticities are identical across firms. As shown in Section 4.2.2, another source of bias arises in the presence of heterogeneity in the factor elasticities.[26]

Using the firm-level production function, we can express APG as a function of firm-level productivity and inputs:

$$APG = \sum_i \frac{Y_i}{Y}\frac{\Delta A_i}{A_i} + \sum_i \frac{Y_i}{Y}\left(\beta - \frac{wL_i}{Y_i}\right)\frac{\Delta L_i}{L_i}.$$

This is the decomposition used by Petrin et al. (2011) and Petrin and Levinsohn (2012).[27] To facilitate the comparison with equation (8), let us rewrite the decomposition as a function of τ_i:[28]

$$APG = \sum_i \frac{Y_i}{Y}\frac{\Delta A_i}{A_i} + \beta \sum_i \frac{\tau_i}{1+\tau_i}\frac{Y_i}{Y}\frac{\Delta L_i}{L_i}. \tag{14}$$

The first term measures the impact of changes in firm-level productivity for a given level of input, and the second term gives the impact of changes

[25] For simplicity, I consider here the case where input prices are identical across firms. Introducing input price heterogeneity does not change the main point made here; it adds an additional component to Basu and Fernald's (2002) decomposition, which reflects the distortions caused by input price heterogeneity. Note that with input price heterogeneity, the Solow residual correctly measures changes in aggregate productivity under the assumption that average *unobserved* distortions are equal to zero. To see this clearly, consider the first order condition of the firms, aggregated across firms, in the case where input prices can vary across firms: $\beta = (wL/Y)(\sum_i L_i(1+\hat{\tau}_i)/L)$, where $\hat{\tau}_i$ is the unobserved distortion, i.e., not captured in the input price. Hence the factor share correctly measures the factor elasticity if the only source of distortions is the dispersion in factor prices. I thank one anonymous referee for pointing this out.

[26] This additional discrepancy comes from the difference in the underlying allocation rule. While Basu and Fernald (2002) derive the link between aggregate inputs and output taking the input shares as given, I use an allocation rule that takes distortions as given.

[27] See for example equation (9) in Petrin and Levinsohn (2012).

[28] To derive equation (14), use the decentralized version of equation (1), that is $\beta A_{it} L_{it}^{\beta-1} = w_t(1+\tau_{it})$.

in firm-level inputs. The decomposition of Petrin and Levinsohn (2012) is identical to the reallocation decomposition of equation (8) when there is no average distortion.[29] The decomposition, however, becomes difficult to interpret in the presence of average distortions. The presence of average distortions creates a gap between the factor elasticity and the factor income share, which generates a bias in the Solow residual. Because of this bias, movements in the Solow residual will reflect variations in aggregate inputs. This effect, highlighted by Hall (1991) for the specific case of markups, is captured in the second term of equation (14). To illustrate this point, suppose that there is only one firm in the economy. The reallocation component of equation (14) is nonetheless different from zero (and equal to $\tau/(1+\tau)\Delta L/L$), despite the absence of input reallocation. Using the reallocation component of Petrin and Levinsohn (2012) to study the effects of reallocation on aggregate productivity can be misleading since the reallocation component may be capturing variation in aggregate inputs.

Basu and Fernald (2002) instead suggest the following decomposition:[30]

$$APG = \sum_i \frac{Y_i}{Y}\frac{\Delta A_i}{A_i} + \beta \sum_i \frac{\overline{\tau}^Y}{1+\tau_i}\frac{Y_i}{Y}\frac{\Delta L_i}{L_i} + \beta \sum_i \frac{\tau_i - \overline{\tau}^Y}{1+\tau_i}\frac{Y_i}{Y}\frac{\Delta L_i}{L_i}, \qquad (15)$$

where $\overline{\tau}^Y = \sum_i (Y_i/Y)\tau_i$. The first component measures the effect of changes in firm-level productivity, the second component captures the effect of the presence of average distortions, and the third component measures the effect of changes in input across firms with different levels of distortions. The decomposition separates the effect of average distortions from the reallocation component, and hence does not suffer from the same bias as equation (14). In fact, the reallocation component of Basu and Fernald (2002) is very similar to the reallocation component of equation (8) and would be identical if the average distortion was computed using input shares as weights instead of output shares. The fundamental difference between my benchmark decomposition and Basu and Fernald's (2002) hence comes from the difference

[29] In the presence of input price heterogeneity, this would be the case if the only source of distortions is input price heterogeneity.

[30] This corresponds to equation (28) in Basu and Fernald (2002). The first term corresponds to dt, the second term to $(\mu^V - 1)dx^V$ and the last term to R_μ. The components R_M R_K and R_L are not present in this simplified framework as I have abstracted from intermediaries as well as from heterogeneity in input prices.

in the objective. They focus on the consequences of input reallocation on aggregate productivity, which requires less structure than my benchmark decomposition. In particular, their decomposition does not rely on a particular specification of the production function. By contrast, my objective is to provide a measure of changes in allocative efficiency. As explained in Section 2.4.1, these are related but distinct concepts that give complementary information on the role of resource allocation for aggregate productivity.

3. Decomposition of aggregate productivity in the general framework

This section gives the decomposition of aggregate productivity in the general framework. In the general framework, firms produce heterogeneous goods, use both capital and labor inputs, and their factor elasticities are allowed to differ across sectors. I apply the method described in Section 2 to derive the sectoral and economy-wide production functions and the decomposition of aggregate TFP changes at the sectoral and economy-wide levels.

3.1. The general framework

Consider an economy with S sectors. In sector s, firm i produces a differentiated good using a Cobb-Douglas technology[31]

$$Y_{it} = A_{it} K_{it}^{\alpha_s} L_{it}^{\beta_s},$$

where Y_{it} denotes firm's i value added, K_{it} its capital, L_{it} its labor, and A_{it} its TFP in period t. To simplify notations, sector subscripts s are omitted on firm-specific variables.

[31] Note that to abstract from intermediate inputs, I have implicitly assumed, as in Rotemberg and Woodford (1993), that firms produce gross output with a Leontief technology of the form

$$Q_{it} = \min\left[\frac{A_{it} K_{it}^{\alpha_s} L_{it}^{\beta_s}}{1-c}, M_{it}\right],$$

where Q_{it} is gross output and M_{it} is intermediate input. I thank an anonymous Referee for this suggestion. See Appendix B.1 for more details.

The firm faces constant or decreasing returns to scale $\gamma_s \equiv \alpha_s + \beta_s \leq 1$.[32] The factor elasticities are assumed to be identical within sectors but may vary from one sector to another. As in Section 2, entry and exit are exogenous, and $I_{it} = 1$ indicates an active firm, $I_{it} = 0$ an inactive firm and $\mathcal{N}_{st} = \{i | I_{it} = 1\}$ is the set of active firms. The number of active firms in the sector is denoted $n_{st} = \text{Card}(\mathcal{N}_{st})$. Sectoral inputs are denoted by $K_{st} = \sum_{i \in \mathcal{N}_{st}} K_{it}$ and $L_{st} = \sum_{i \in \mathcal{N}_{st}} L_{it}$. Sectoral output is given by the CES aggregate[33]

$$Y_{st} = n_{st}^{\frac{\theta_s - 1}{\theta_s}} \left(\sum_{i \in \mathcal{N}_{st}} Y_{it}^{\theta_s} \right)^{1/\theta_s},$$

where $0 < \theta_s < 1$ and the elasticity of substitution within sector s is equal to $1/(1 - \theta_s)$. For simplicity, I assume that each good has the same weight in the aggregation and hence abstract from firm-specific demand shocks. I show, in Appendix B, that this is without loss of generality. Although the firm's productivity A_i may be a function of the firm-specific demand shock, the derivation and the estimation of the decomposition are not modified in the presence of firm-specific demand shocks.
Aggregate output is given by the CES aggregate of sectoral outputs

$$Y_t = \left(\sum_{s=1}^{S} Y_{st}^{\rho} \right)^{1/\rho},$$

where $0 < \rho < 1$ and the elasticity of substitution between sectors is equal to $1/(1 - \rho)$. Aggregate inputs are denoted by $K_t = \sum_{s=1}^{S} K_{st}$ and $L_t = \sum_{s=1}^{S} L_{st}$.
Inputs may be inefficiently allocated across firms, and the distortions are defined, as in Section 2, as wedges in the first-order conditions of the first-

[32] This condition can be relaxed. Firms can have increasing returns to scale $\gamma_s > 1$ provided $\gamma_s \theta_s < 1$. The framework can also accommodate increasing returns caused by the presence of a fixed cost. In that case, the decomposition would be unchanged if the fixed cost affects the firms' profits but is not reflected in the firms' measured output or inputs. I consider in Appendix B.3 how the decomposition is modified when the fixed cost is a labor cost.

[33] I choose to be agnostic on the size of the variety effect and normalize the CES aggregate by the number of firms n_{st}. This eliminates the variety effect and avoids having total output increase with the number of firms. This is a conservative choice, which is motivated by the lack of evidence on the size of variety effects.

best allocation of resources

$$\alpha_s \frac{p_{it} Y_{it}}{P_{st} K_{it}} = \mu^K_{st}(1+\tau^K_{it}), \tag{16}$$

$$\beta_s \frac{p_{it} Y_{it}}{P_{st} L_{it}} = \mu^L_{st}(1+\tau^L_{it}), \tag{17}$$

where the relative price level for good i in sector s is $p_{it}/P_{st} = (n_{st}(Y_{it}/Y_{st}))^{\theta_s - 1}$ and μ^K_{st} and μ^L_{st} are the Lagrange multipliers of the first-best problem. Sectoral distortions are similarly defined by

$$\alpha_s \frac{P_{st} Y_{st}}{P_t K_{st}} = \lambda^K_t(1+\omega^K_{st}), \tag{18}$$

$$\beta_s \frac{P_{st} Y_{st}}{P_t L_{st}} = \lambda^L_t(1+\omega^L_{st}), \tag{19}$$

where the relative price for sector s is $P_{st}/P_t = (Y_{st}/Y_t))^{\rho-1}$ and λ^K_t and λ^L_t are the Lagrange multipliers of the first-best problem. The dispersion of $(\tau^K_{it}, \tau^L_{it})$ captures the inefficiency of resource allocation within sector and the dispersion of $(\omega^K_{st}, \omega^L_{st})$ captures the inefficiency of resource allocation across sectors.

3.2. The sectoral and economy-wide production functions

Following the aggregation method described in Section 2.2, I first aggregate the firm-level production functions at the sectoral level and then aggregate the sectoral production functions. I report here the sectoral and economy-wide production functions and provide the details of the derivation in Appendix A. Aggregating the production functions in two steps, at the sectoral level and then at the economy-wide level, permits separating the contribution of within and between-sector allocative efficiency. The sectoral production function $Y_{st} = F_s(K_{st}, L_{st}, \mathbf{A}_{st}, \boldsymbol{\tau}_{st})$ has the same functional form as the individual production functions:

$$F_s(K_{st}, L_{st}, \mathbf{A}_{st}, \boldsymbol{\tau}_{st}) = \mathrm{TFP}_{st} K^{\alpha_s}_{st} L^{\beta_s}_{st},$$

with

$$\mathrm{TFP}_{st} = n_{st}^{\frac{\theta_s - 1}{\theta_s}} \left(\sum_{i \in \mathcal{N}_{st}} A^{\theta_s}_{it} \left(\frac{K_{it}}{K_{st}}\right)^{\alpha_s \theta_s} \left(\frac{L_{it}}{L_{st}}\right)^{\beta_s \theta_s} \right)^{1/\theta_s},$$

where K_{it}/K_{st} and L_{it}/L_{st} are functions of the vector of firm-level productivities $\mathbf{A}_{st} = \{A_{it}, i \in \mathcal{N}_{st}\}$ and distortions $\boldsymbol{\tau}_{st} = \{\tau_{it}, i \in \mathcal{N}_{st}\}$.[34]

By contrast, the economy-wide production function $Y_t = F(K_t, L_t, \mathbf{TFP}_t, \boldsymbol{\omega}_t)$ is not separable and does not have a closed form expression. The economy-wide production function is characterized by:

$$F(K_t, L_t, \mathbf{TFP}_t, \boldsymbol{\omega}_t) = \left(\sum_{s=1}^{S} \left(\mathrm{TFP}_{st} K_{st}^{\alpha_s} L_{st}^{\beta_s}\right)^{\rho}\right)^{\frac{1}{\rho}},$$

together with the allocation rules:

$$\begin{aligned} K_{st} &= k_s(\mathbf{TFP}_t, \boldsymbol{\omega}_t, K_t, L_t), \\ L_{st} &= l_s(\mathbf{TFP}_t, \boldsymbol{\omega}_t, K_t, L_t), \end{aligned}$$

where $\mathbf{TFP}_t = \{\mathrm{TFP}_{1t}, ..., \mathrm{TFP}_{St}\}$ is the vector of sectoral-level productivity, and $\boldsymbol{\omega}_t = \{\omega_{1t}, ..., \omega_{St}\}$ is the vector of sectoral-level distortions. The allocation rules are characterized implicitly by the first-order conditions (equations (18) and (19)) and the aggregate inputs constraints. Aggregate productivity is defined by:

$$\Delta \ln \mathrm{TFP}_t = \Delta \ln Y_t - \varepsilon_t^K \Delta \ln K_t - \varepsilon_t^L \Delta \ln L_t, \tag{20}$$

where the elasticities ε_t^K and ε_t^L are both functions of the sectoral labor and capital elasticities and of the elasticity of substitution between sectors.[35]

3.3. Decomposition of sectoral productivity

I propose an exact decomposition of sectoral productivity. Sectoral TFP growth can be decomposed into changes in technical efficiency ($\Delta \mathrm{TE}_s$), changes in allocative efficiency ($\Delta \mathrm{AE}_s$), and changes at the extensive margin ($\Delta \mathrm{EX}_s$):

$$\Delta \ln \mathrm{TFP}_s = \Delta \mathrm{TE}_s + \Delta \mathrm{AE}_s + \Delta \mathrm{EX}_s. \tag{21}$$

The components of the decomposition add up exactly to the change in TFP. The allocative efficiency and technical efficiency components are based on

[34] See Appendix A.1 for the expression of sectoral TFP.

[35] The expressions of the elasticities are given in Section 4.2.2. See Appendix A.5 for the details of the derivation.

Fisher-like indexes.[36] For simplicity, I present here the discrete-time approximation of the continuous-time decomposition. The expressions and derivations of both the exact and the approximate decompositions are described in Appendix A. While the details of the decomposition differ, the exact decomposition, used in the empirical analysis, can also be computed from firm-level input and output shares.

The change in firm-level productivity can be approximated as a combination of weighted averages of the firm-level productivity changes:

$$\Delta TE_s \simeq \frac{1}{1-\gamma_s\theta_s} \sum_{i\in\mathcal{C}_{st}} \frac{\Delta A_{it}}{A_{it-1}} \left[\frac{p_{it-1}Y_{it-1}}{\sum\limits_{i\in\mathcal{C}_{st}} p_{it-1}Y_{it-1}} - \alpha_s\theta_s \frac{K_{it-1}}{\sum\limits_{i\in\mathcal{C}_{st}} K_{it-1}} - \beta_s\theta_s \frac{L_{it-1}}{\sum\limits_{i\in\mathcal{C}_{st}} L_{it-1}} \right], \tag{22}$$

where $\mathcal{C}_{st}$ is the set of continuing firms in sector s at time t. As explained in Section 2, the technical efficiency component includes both the effect of changes in firm-level productivity, holding fixed the firms' input shares as well as the effects of the changes in input shares required to maintain the same allocative efficiency level. In addition, the technical efficiency component will likely reflect other factors than technology such as demand shocks or factor utilization.[37]

The change in allocative efficiency is a combination of weighted averages of the firm-level changes in distortions:

$$\begin{aligned} \Delta AE_s \simeq & \frac{\alpha_s}{1-\gamma_s\theta_s} \sum_{i\in\mathcal{C}_{st}} \frac{\Delta\tau^K_{it}}{1+\tau^K_{it-1}} \left[(1-\beta_s\theta_s) \frac{K_{it-1}}{\sum\limits_{i\in\mathcal{C}_{st}} K_{it-1}} + \beta_s\theta_s \frac{L_{it}}{\sum\limits_{i\in\mathcal{C}_{st}} L_{it-1}} - \frac{p_{it-1}Y_{it-1}}{\sum\limits_{i\in\mathcal{C}_{st}} p_{it-1}Y_{it-1}} \right] \\ + & \frac{\beta_s}{1-\gamma_s\theta_s} \sum_{i\in\mathcal{C}_{st}} \frac{\Delta\tau^L_{it}}{1+\tau^L_{it-1}} \left[\alpha_s\theta_s \frac{K_{it-1}}{\sum\limits_{i\in\mathcal{C}_{st}} K_{it-1}} + (1-\alpha_s\theta_s) \frac{L_{it-1}}{\sum\limits_{i\in\mathcal{C}_{st}} L_{it-1}} - \frac{p_{it-1}Y_{it-1}}{\sum\limits_{i\in\mathcal{C}_{st}} p_{it-1}Y_{it-1}} \right]. \end{aligned} \tag{23}$$

The allocative efficiency component reflects the effects of changes in capital and labor distortions on aggregate productivity. As explained in Section 2,

[36] The exact decomposition is described in Appendix A.4. The Fisher-like indexes are equal to the geometric mean of the Laspeyres-like and Paasche-like indexes.

[37] As explained in Appendix B.2, when goods are heterogenous the firm's contribution to aggregate productivity is also a function of firm-specific demand shocks.

allocative efficiency increases if distortions are reduced in firms with higher-than-average distortions.

The contribution of the extensive margin depends on the average output and input of entering and exiting firms:

$$\Delta EX_s \simeq (1-\gamma_s)\frac{\Delta n_{st}}{n_{st-1}}+\frac{e_{st}}{n_{st}}\left[\frac{1}{\theta_s}\left(\frac{\frac{1}{e_{st}}\sum\limits_{i\in\mathcal{E}_{st}} p_{it}Y_{it}}{\frac{1}{c_{st}}\sum\limits_{i\in\mathcal{C}_{st}} p_{it}Y_{it}}-1\right)-\alpha_s\left(\frac{\frac{1}{e_{st}}\sum\limits_{i\in\mathcal{E}_{st}} K_{it}}{\frac{1}{c_{st}}\sum\limits_{i\in\mathcal{C}_{st}} K_{it}}-1\right)-\beta_s\left(\frac{\frac{1}{e_{st}}\sum\limits_{i\in\mathcal{E}_{st}} L_{it}}{\frac{1}{c_{st}}\sum\limits_{i\in\mathcal{C}_{st}} L_{it}}-1\right)\right]$$
$$-\frac{x_{st}}{n_{st-1}}\left[\frac{1}{\theta_s}\left(\frac{\frac{1}{x_{st}}\sum\limits_{i\in\mathcal{X}_{st}} p_{it-1}Y_{it-1}}{\frac{1}{c_{st}}\sum\limits_{i\in\mathcal{C}_{st}} p_{it-1}Y_{it-1}}-1\right)-\alpha_s\left(\frac{\frac{1}{x_{st}}\sum\limits_{i\in\mathcal{X}_{st}} K_{it-1}}{\frac{1}{c_{st}}\sum\limits_{i\in\mathcal{C}_{st}} K_{it-1}}-1\right)-\beta_s\left(\frac{\frac{1}{x_{st}}\sum\limits_{i\in\mathcal{X}_{st}} L_{it-1}}{\frac{1}{c_{st}}\sum\limits_{i\in\mathcal{C}_{st}} L_{it-1}}-1\right)\right], \quad (24)$$

where $\mathcal{E}_{st}$ is the set of entering firms, $\mathcal{X}_{st}$ is the set of exiting firms, and c_t, e_t and x_t denote the number of continuing, entering and exiting firms. The equation shows that the contribution of entry and exit does not depend on the number of firms when firms face constant returns to scale (as assumed in the empirical analysis).

3.4. Decomposition across sectors

The change in aggregate TFP can be decomposed into the effect of sectoral productivity and that of sectoral distortions:

$$\Delta \ln \mathrm{TFP} \simeq \Delta E + \Delta AE_{\mathrm{between}}, \quad (25)$$

where ΔE denotes the contribution of sectoral productivity, and $\Delta AE_{\mathrm{between}}$ the contribution of between-sector allocative efficiency.[38] The decomposition of aggregate productivity can be obtained from the discrete-time approximation of the continuous-time decomposition. The continuous-time decomposition is similar to the one derived within sectors but is more tedious to derive because there is no closed-form solution for aggregate TFP. For simplicity, I report here a (slightly) simplified version of the decomposition used in the empirical analysis.[39] The details of the derivation are given in Appendix A.

[38] There can be a discrepancy between the two ways of computing aggregate productivity; both are approximations of aggregate productivity changes as there is no closed form solution for the economy-wide aggregate production function.

[39] The decomposition used in the empirical analysis also relies on the discrete-time approximation of the continuous-time decomposition but uses the average of time $t-1$ and t of the output and input shares.

The contribution of sectoral productivity can be written

$$\Delta E = \sum_{s=1}^{S} \frac{1}{1-\gamma_s \rho} \left[\frac{P_{st}Y_{st}}{P_t Y_t} - \rho \varepsilon_t^K \frac{K_{st}}{K_t} - \rho \varepsilon_t^L \frac{L_{st}}{L_t} \right] \frac{\Delta \text{TFP}_{st}}{\text{TFP}_{st-1}}. \tag{26}$$

The change in between-sector allocative efficiency is computed as follows:

$$\begin{aligned} \Delta AE_{\text{between}} &= \sum_{s=1}^{S} \frac{1}{1-\gamma_s \rho} \left[\varepsilon_t^K (1-\beta_s \rho) \frac{K_{st}}{K_t} + \varepsilon_t^L \alpha_s \rho \frac{L_{st}}{L_t} - \alpha_s \frac{P_{st}Y_{st}}{P_t Y_t} \right] \frac{\Delta \omega_{st}^K}{1+\omega_{st-1}^K} \\ &+ \sum_{s=1}^{S} \frac{1}{1-\gamma_s \rho} \left[\varepsilon_t^K \beta_s \rho \frac{K_{st}}{K_t} + \varepsilon_t^L (1-\alpha_s \rho) \frac{L_{st}}{L_t} - \beta_s \frac{P_{st}Y_{st}}{P_t Y_t} \right] \frac{\Delta \omega_{st}^L}{1+\omega_{st-1}^L} \end{aligned} \tag{27}$$

Note that, when the factor elasticities are identical across sectors, the decomposition becomes identical to the within-sector decomposition presented in the previous section.

Combining equations (21), (25), and , we get the full decomposition of aggregate productivity:

$$\Delta \ln \text{TFP} \simeq \Delta \text{TE} + \Delta \text{AE}_{\text{within}} + \Delta \text{EX} + \Delta AE_{\text{between}}, \tag{28}$$

where the aggregate technology efficiency is given by

$$\Delta \text{TE} = \sum_{s=1}^{S} \frac{1}{1-\gamma_s \rho} \left[\frac{P_{st}Y_{st}}{P_t Y_t} - \rho \varepsilon_t^K \frac{K_{st}}{K_t} - \rho \varepsilon_t^L \frac{L_{st}}{L_t} \right] \Delta \text{TE}_s$$

and $\Delta\text{AE}_{\text{within}}$ and ΔEX are similarly computed.

4. Estimation

I estimate the decomposition and study the role of allocative efficiency in the dynamics of aggregate productivity over the business cycle. This section presents the data used, the estimation strategy, and the results of the decomposition.

4.1. Data description

I use a rich dataset collected annually by the French tax administration and combined with survey data in the INSEE unified system for business statistics (SUSE). Whereas data limitation has led most of the literature to focus on the manufacturing industry, this dataset permits studying both manufacturing and service sectors. The dataset includes all firms that file their taxes under the standard procedure (*Bénéfice réel normal*) from 1989 to 2007. In 2003, the dataset accounted for 24.4 % of firms and 94.3% of total sales.[40] Under the standard procedure, firms are required to provide balance sheet data, which includes measures of the firms' value added, investment expenditures, and number of employees. I restrict the analysis to sectors that provide market services and hence exclude the education, health, public administration, and non-profit sectors.

Each firm is assigned an identification number (SIREN), and I detect potential entries and exits by following firms with at least one employee that appear in and disappear from the dataset.[41] One important issue is that firms appear and disappear from the dataset not only when they actually open or shut down their businesses, but also whenever their identification number is modified, as in cases of restructuring or takeover. Entry and exit rates are therefore likely to be overestimated. I limit this bias by excluding firms that appear or disappear with a size above the 99th percentile.[42] The high number of restructuring operations in the finance sector led me to exclude this sector from the analysis.[43] I also correct for missing observations by linear interpolation and I exclude from the sample firms whose productivity changes are in the bottom and top 2 percentiles. Finally, I remove the first

[40]Despite accounting for 24.4% of firms, this dataset is well suited to study the aggregate implications of entry and exit as firms absent from the dataset only represent 5.7% of total sales.

[41]Small firms may appear in the database when they cross the sales threshold of 763 000 euros for manufacturing and 230 000 euros for services above which the *Bénéfice réel normal* regime is mandatory. However, these spurious entry flows are likely to be limited as the "normal" tax regime is widely chosen by firms that are below the threshold: in 2003, 46% of the sample was below the sales threshold for manufacturing and 32% for services. Besides, the dataset covers a large number of small firms: in 2003, 70% of the sample (restricted to firms with at least one employee) had less than 20 employees in the manufacturing sector and less than 9 employees in the service sector.

[42]Firm size is measured in terms of employment, fixed assets, and value added. The employment threshold is 705 employees in the manufacturing sector and 169 employees in the service sector. The threshold is quite high in light of the median size of firms involved in takeover and mergers. Using French data over the period 2000-2004, Bunel et al. (2009) find a median size between 30 and 87 employees, depending on the exact nature of the restructuring operation.

[43]The high number of restructuring operation leads to high volatility in the sector: the standard deviation of value added growth in the finance sector is 40%.

and last years of the sample, and I undertake the analysis over the period 1991-2006.

After the trimming procedure, the dataset contains about 65 000 active firms per year in the manufacturing and 265 000 in the service sector. Appendix C.1 provides more details on the data sources, and Table D.1 provides some descriptive statistics on each sector's size, labor elasticity, and median firm size. Figure D.1 and D.2 compare the aggregate value added obtained from the firm-level data to the official national account series published by INSEE. The figures show that correcting for temporary exit is crucial and that the firm-level data capture quite well the cyclical movements in aggregate output.

4.2. Estimation method

In this section, I describe the method used and the assumptions required to estimate firm-level and sectoral distortions, elasticities, and productivities.

4.2.1. Estimation of firm-level and sectoral distortions

Firm-level distortions are defined by equations (16) and (17) as the wedge between the firms' marginal productivity and their frictionless values. As noted before, the impact of the distortions on aggregate productivity, however, only depends on the relative marginal productivities (cf. equations (4) and (9) and the subsequent discussions). This property simplifies the estimation of firm-level distortions since it implies that the change in allocative efficiency can be computed without measuring the shadow cost of labor and capital. In fact, the change in allocative efficiency can be rewritten as a function of the firms' marginal productivities in nominal terms, $MVPK_{it} \equiv \alpha_s p_{it} Y_{it}/K_{it}$ and $MVPL_{it} \equiv \beta_s p_{it} Y_{it}/L_{it}$. To see this more clearly, write

$$
\begin{aligned}
(1+\tau_{it}^K) &= MVPK_{it}/(\mu_{st}^K P_{st}), \\
(1+\tau_{it}^L) &= MVPL_{it}/(\mu_{st}^L P_{st}).
\end{aligned}
$$

Substitute in equation (23) and notice that the terms common to all firms cancel out:[44]

$$\Delta AE_s \simeq \frac{\alpha_s}{1-\gamma_s\theta_s}\sum_{i\in C_s}\frac{\Delta MVPK_{it}}{MVPK_{it-1}}\left[(1-\beta_s\theta_s)\frac{K_{it}}{\sum\limits_{i\in C_s}K_{it}}+\beta_s\theta_s\frac{L_{it}}{\sum\limits_{i\in C_s}L_{it}}-\frac{p_{it}Y_{it}}{\sum\limits_{i\in C_s}p_{it}Y_{it}}\right]$$
$$+\frac{\beta_s}{1-\gamma_s\theta_s}\sum_{i\in C_s}\frac{\Delta MVPL_{it}}{MVPL_{it-1}}\left[\alpha_s\theta_s\frac{K_{it}}{\sum\limits_{i\in C_s}K_{it}}+(1-\alpha_s\theta_s)\frac{L_{it}}{\sum\limits_{i\in C_s}L_{it}}-\frac{p_{it}Y_{it}}{\sum\limits_{i\in C_s}p_{it}Y_{it}}\right].$$

Likewise, to compute between-sector allocative efficiency changes, we only need sectoral marginal productivities. Equation (27) can be rewritten:

$$\Delta E_{\text{between}} = \sum_{s=1}^{S}\frac{1}{1-\gamma_s\rho}\left(\varepsilon_t^K(1-\beta_s\rho)\frac{K_{st}}{K_t}+\varepsilon_t^L\alpha_s\rho\frac{L_{st}}{L_t}-\alpha_s\frac{P_{st}Y_{st}}{P_tY_t}\right)\frac{\Delta MVPK_{st}}{MVPK_{st-1}}$$
$$+\sum_{s=1}^{S}\frac{1}{1-\gamma_s\rho}\left(\varepsilon_t^K\beta_s\rho\frac{K_{st}}{K_t}+\varepsilon_t^L(1-\alpha_s\rho)\frac{L_{st}}{L_t}-\beta_s\frac{P_{st}Y_{st}}{P_tY_t}\right)\frac{\Delta MVPL_{st}}{MVPL_{st-1}}.$$

It is worth noting that the measure of distortion does not rely on any assumption about the firms' environment or behavior. The measure of distortions is based on the distance between the data and the first-best allocation of resources, and hence captures frictions that may come from various sources. The measure of distortions, however, is likely to be sensitive to the framework specification. In particular, the measure crucially depends on the functional form of the firms' production functions and the assumption that inputs are homogenous. The results of the decomposition hence implicitly rely on the assumption that the dynamics of allocative efficiency over the business cycle is not caused by specification errors. I evaluate the robustness of the results to changes in the specification of the production function in Section 4.3.3.

4.2.2. Estimation of the production functions

Specific issues arise when estimating the production functions in the presence of distortions and heterogeneity in factor elasticities. I first describe the assumptions required to measure firm-level and sectoral TFP. I then describe how to compute the aggregate elasticities and show that aggregate

[44] The terms common across firms cancel out in the exact decomposition as well.

factor shares may give a biased measure of aggregate factor elasticities when the firm-level factor elasticities are heterogeneous.

Firm-level and sectoral TFP

The estimation of production functions is challenging when resources may be inefficiently allocated. Specifically, the firm-level factor shares cannot be used as a measure of factor elasticites because the factor shares depend both on factor elasticities and firm-level distortions.[45] To address this issue, I exploit the assumption that factor elasticities are identical within sectors and estimate the labor elasticities using the first-order condition of each sector's representative firm, equation (19), aggregated over time. I assume that at the sectoral level, input price heterogeneity is the only source of average distortions and use average labor income shares to estimate the labor elasticities:[46]

$$\beta_s = \frac{1}{T}\sum_t \frac{w_{st}L_{st}}{P_{st}Y_{st}}.$$

The labor elasticities estimates are reported in Table D.1. Assuming constant returns to scale in each sector, $\gamma_s = \gamma = 1$, the capital elasticities are then obtained as $\alpha_s = 1 - \beta_s$. Using the sectoral production function, I can then measure sectoral productivity as

$$\mathrm{TFP}_{st} = \frac{Y_{st}}{K_{st}^{\alpha_s} L_{st}^{\beta_s}}.$$

Since the labor elasticity is measured by the labor share, the measure of TFP at the sectoral level is equal to the standard Solow residual. It is important

[45] The presence of allocation distortion also makes difficult the use of standard semi-parametric methods such as Olley and Pakes (1996) or Levinsohn and Petrin (2003). These methods, which consist in using a proxy for productivity, can only accommodate a unique unobservable state variable. In the present framework, firms' decisions depend on their productivity but also on their capital and labor distortions, which raises to three the number of unobservable state variables.

[46] The first-order condition for sector s can be rewritten

$$\beta_s \frac{P_{st}Y_{st}}{L_{st}} = w_{st}(1 + \hat{\omega}_{st}).$$

The distortion arises from differences in the sector wage w_{st} or in the unobserved distortion $\hat{\omega}_{st}$. The estimation of the labor elasticity is based on the assumption that input price heterogeneity is the only source of average cross-sectoral inefficiencies $\frac{1}{T}\sum_t \frac{1}{1+\hat{\omega}_{st}} = 1$.

to note that normalizing average unobserved sectoral distortions to zero has no direct impact on the contribution of firm-level distortions to aggregate productivity. As already mentioned, what matters for productivity is the relative level of distortions and not the average level. As a robustness check, I also estimate the factor elasticities with average distortions (see Section 4.3.3).

To estimate firm-level TFP, we must, in addition, deal with the fact that firm-level prices, and hence firm-level real output, are not observable. I use the assumption about the demand function, $p_{it}/P_{st} = (n_{st}Y_{it}/Y_{st})^{\theta_s - 1}$ and compute the firm's real value added from its nominal value added, the sectoral price index, the sectoral real value added, and the elasticity of demand. The estimate of firm-level TFP is then

$$A_{it} = \frac{(p_{it}Y_{it}/P_{st})^{1/\theta_s}}{K_{it}^{\alpha_s} L_{it}^{\beta_s}} \left(\frac{Y_{st}}{n_{st}}\right)^{(\theta_s - 1)/\theta_s},$$

where P_{st} is measured by the sectoral deflator, and Y_{st} is computed as the sector's nominal value added divided by the sectoral deflator. For simplicity, I assume that the elasticity of substitution is identical across sectors.[47] I set $\theta_s = \theta = 0.66$ and $\rho = 0.5$, which corresponds to an elasticity of substitution of 3 within sectors and of 2 between sectors. These numbers are in line with Broda and Weinstein (2006) estimates.[48] It is important to note that the measure of firm-level productivity A_i does not capture only the technological productivity of the firm. As explained in Appendix B.2, the firm-level productivity may also reflects firm-specific demand shocks. In addition, the measure of productivity suffers from standard limitations associated with measurement issues, and variation in the firm-level productivity may reflect variation in input utilization.

Aggregate elasticities and the heterogeneity bias

Because factor elasticities are heterogeneous across sectors, the factor income shares cannot be used to estimate the elasticities of the aggregate

[47]Note that despite this assumption, the measure of distortion includes inefficiencies due to differences in markups across firms and/or sectors. In fact, the measure of distortion is based on the distance to the social planner's allocation and does not hinge on the markup implied by the structure of demand.

[48]They report a mean of 4 and a median of 2.2 over the period 1990-2001 between goods within 3-digit industries.

production function ε^K and ε^L. Contrary to the case where input elasticities are identical, assuming that input price heterogeneity is the only source of distortion is not sufficient to use factor shares as a measure of factor elasticities. When the decreasing-to-scale parameter is identical across sectors and equal to one – as assumed here, the aggregate elasticities are given by:

$$\varepsilon^K = \frac{\alpha^Y - \rho\alpha^L}{1-\rho(\alpha^L+\beta^K)} \quad \text{and} \quad \varepsilon^L = \frac{\beta^Y - \rho\beta^K}{1-\rho(\alpha^L+\beta^K)},$$

where $\alpha^Y = \sum_s \frac{P_sY_s}{PY}\alpha_s$, $\beta^Y = \sum_s \frac{P_sY_s}{PY}\beta_s$, $\alpha^X = \sum_s \frac{X_s}{X}\alpha_s$ and $\beta^X = \sum_s \frac{X_s}{X}\beta_s$, for $X = K, L$. The derivation of the elasticities and their expressions in the general case are given in Appendix A.5. The aggregate labor share is equal to $\sum_s w_sL_s/(PY) = \beta^Y$, which is different from the expression of ε^L given above.[49] The Solow residual, computed using the aggregate labor income share, is then a biased measure of aggregate productivity changes when factor elasticities are heterogeneous. In practice, I find, however, that the Solow residual gives a close approximation of aggregate TFP. The heterogeneity bias only has a small effect on the measure of aggregate TFP.[50]

4.3. Empirical results

I first examine the cross-sectional characteristics of firm-level distortions and document the role of allocative efficiency in the dynamics of aggregate productivity. I then evaluate the robustness of the results to alternative specifications of the production function. Finally, I compare my results to those obtained using the shift-share and the reallocation decompositions. All the equations used in the empirical analysis are reported in Appendix C.

4.3.1. Firm-level distortions and allocative inefficiency

The measure of allocative efficiency relies on the assumption that the wedges in the firms' first-order conditions reflect distortions in the allocation of resources. To check whether this is a reasonable assumption, I relate capital and labor distortions to various firms' characteristics that are likely to be correlated with the level of frictions faced by the firms.

[49]The expression of the labor share comes from the decentralized version of the first-order condition of each sector's representative firm (equation (19)), aggregated across sectors under the assumption of zero unobserved heterogeneity.

[50]The difference between the two measures is at most of the order of magnitude 10^{-4}.

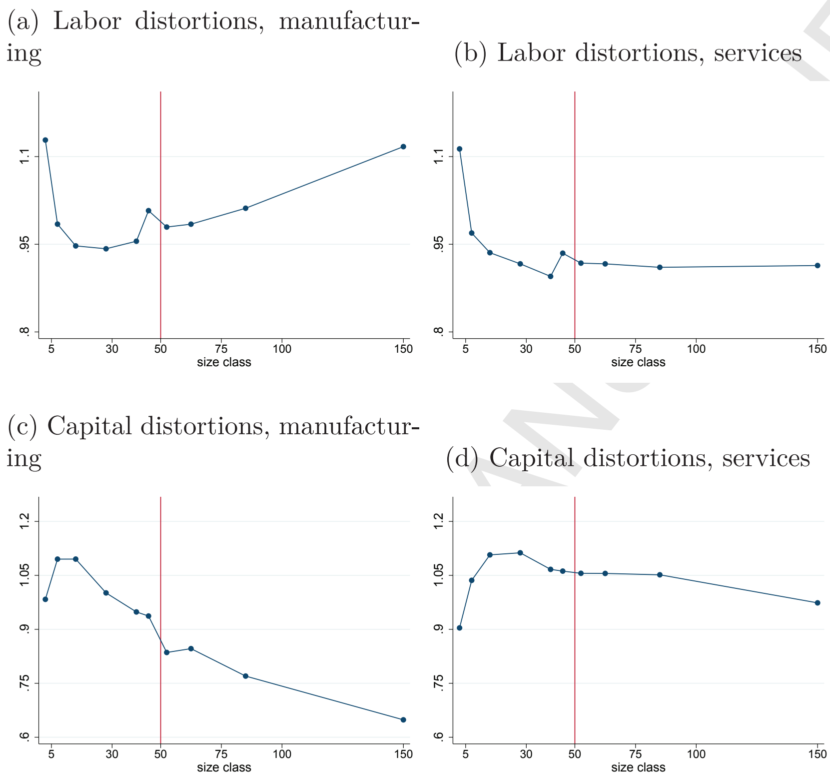
Figure 1 and 2 display the distortions as a function of the firm's size and age. The figures reveal that the capital distortion is lower for larger and older firms, which is consistent with the fact that those firms are likely to face less financial frictions. The patterns of labor distortions are quite different. Whereas firms with less than 5 employees face high labor distortions, the largest and oldest firms do not seem to face lower labor distortions than the average firm. Figure 1 also shows that firms right below 50 employees have higher labor distortions than those right above 50 employees. In France, many labor regulations come into effect when the firm reaches 50 employees, and this leads to higher adjustment costs for firms right below the threshold.[51] These higher adjustment costs are captured by the measure of labor distortion. As shown in Figure 3, the capital distortion maps well with the firm's financial condition. The figure indicates that firms that are likely to be more financially constrained, such as firms facing high interest rates or having high debt-to-asset ratios, tend to have high capital distortions.

I report, in Appendix D, additional statistics on labor and capital distortions, which give some reassurance for using wedges as a measure of frictions. Figure D.3 shows that the dispersion of capital and labor distortions is higher among younger firms. This suggests that the allocation of inputs improves over time within cohorts, in line with the fact that the severity of financial frictions is likely to decline with age. As shown in Table D.3, the dispersion of capital and labor distortions is also higher among firms that do not adjust their employment or their capital, which is consistent with the presence of non-convex adjustment costs.[52] In addition, I find that the changes in firm-level distortions are strongly correlated to the changes in firm-level productivity. This strong correlation suggests that firms adjust their inputs less than what they would do in a frictionless environment. All in all, these patterns indicate that the capital and labor distortions capture frictions faced by firms in their input decisions.

[51] See Gourio and Roys (2014) and Garicano et al. (2016) for a study of how the these size-dependent regulations affects productivity and employment.

[52] See Asker et al. (2014) for further evidence on the link between capital distortions and adjustment costs.

Figure 1: Capital and labor distortions, by firm size (2005)



(a) Labor distortions, manufacturing

(b) Labor distortions, services

(c) Capital distortions, manufacturing

(d) Capital distortions, services

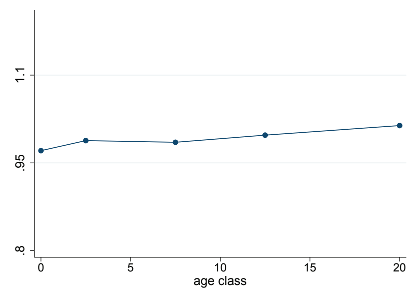
Note: This figure gives the median of the relative capital and labor distortions in 2005 by firm size. The size classes, which refer to the number of employees, are the following: 1-4, 5-9, 10-19, 20-34, 35-44, 45-49, 50-54, 55-69, 70-99 - 100+. Each size class is represented on the graph by the midpoint of the class, with the size class 100+ represented by the "midpoint" 150. The distortions are relative to the 2-digit industry medians.

4.3.2. The cyclical dynamics of allocative efficiency

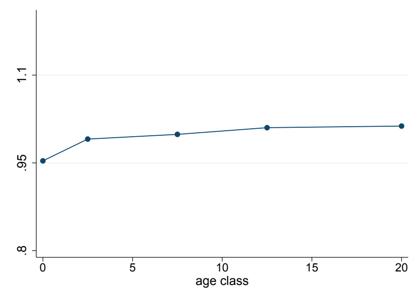
I first report, in Figure 4, aggregate TFP growth and its components in France over the period 1991-2006. The decomposition is obtained, from equation (28), by combining the within-sector and between-sector decompositions. The Figure shows that technical efficiency increases and within-sector allocative efficiency declines when the growth rate of aggregate TFP is high. By contrast, between-sector allocative efficiency does not show a clear correlation with aggregate TFP changes. The entry and exit component seems

Figure 2: Capital and labor distortions, by firm age (2005)

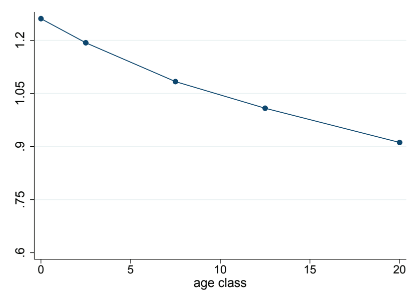
(a) Labor distortions, manufacturing



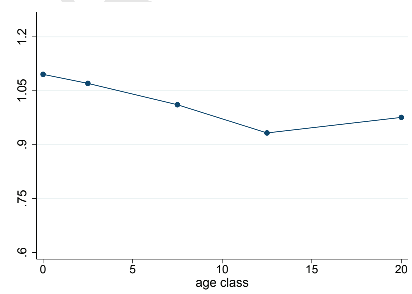
(b) Labor distortions, services



(c) Capital distortions, manufacturing



(d) Capital distortions, services



Note: This figure gives the median of the relative capital and labor distortions in 2005 by firm age. The age classes are the following: 0, 1-4, 5-9, 10-14, 15+. Each age class is represented by the midpoint of the class, with firms aged 15 and older represented by the "midpoint" 20. The distortions are relative to the 2-digit industry medians.

to be somewhat negatively correlated to aggregate TFP changes, but its volatility is smaller than that of within-sector allocative efficiency. Finally, within-sector allocative efficiency is on average negative, indicating that the efficiency of resource allocation has deteriorated over the period. I focus here on the measure of allocative efficiency and its cyclicality and leaves the analysis of the downward trend in allocative efficiency for further research.

To complement the patterns suggested by Figure 4 and to better assess the cyclicality of the different components of aggregate productivity, I com-

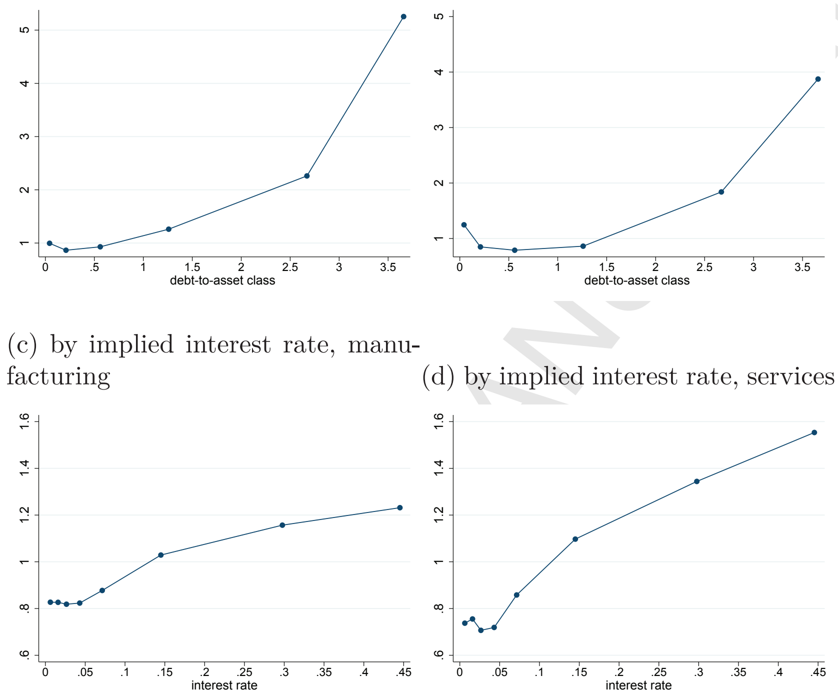
Figure 3: Capital distortions, by financial condition (2005)

(a) by debt-to-asset ratio, manufacturing

(b) by debt-to-asset ratio, services

(c) by implied interest rate, manufacturing

(d) by implied interest rate, services



Note: This figure gives the median of relative capital distortions in 2005 by the firm's financial condition. The financial condition is measured by the debt-to-asset ratio and the implied interest rate. The size classes are delimited by the 5th, 10th, 25th, 50th, 75th, 90th, 95th percentile of each variable. Each size class is represented by the midpoint of the class, with the highest class midpoint set arbitrarily. The distortions are relative to the 2-digit industry medians.

pute, for each sector, the correlation between real value added growth and the components of TFP changes.[53] The results, reported in Table 1, show

[53]The results on the overall change in productivity and allocative efficiency over the period are reported in Table D.4 (Appendix D). The table shows that allocative efficiency decreased over the period, while entry and exit played only a small role for aggregate productivity. To provide some insights on these findings, I have related the average change

that TFP growth is procyclical in all sectors, with an average correlation of 0.82 for manufacturing sectors and of 0.64 for service sectors. Table 1 indicates that the within-firm productivity component is procyclical. Note that because of standard measurement issues, procyclical within-firm productivity is not necessarily a sign of procyclical technology and could instead be the result of unobserved variation in input utilization. Disentangling technology shocks from variations in input utilization is, however, outside the scope of this paper. The results of the table also show that both changes in the efficiency of resource allocation and the contribution of entry and exit are countercyclical in most sectors, especially in service sectors. During downturns, the allocation of resources between incumbents improves and the contribution of entering and exiting firms increases. Lower entry rates and higher exit rates contribute to raising aggregate TFP during downturns because entering and exiting firms tend to be less productive than the average incumbent firm. The decline in allocative efficiency between incumbent firms during expansions could be due to the tightness of the labor market. With the reduction in the number of job seekers, firms face more frictions in their hiring decisions but also in their firing decisions (since hiring again is more difficult). The firms' labor decision therefore becomes less aligned with their productivity, which reduces allocative efficiency. Another possible mechanism is that, in periods of boom, more credit gets allocated to firms that are already less financially constrained, which reduces allocative efficiency.

Is the cyclicality of allocative efficiency and that of entry and exit important for the volatility of sectoral productivity? To answer this question, I decompose the variance of sectoral productivity changes as follows:

$$\begin{aligned} V(\Delta \ln \mathrm{TFP}_s) = {} & Cov(\Delta \ln \mathrm{TFP}_s, \Delta TE_s) \\ & + Cov(\Delta \ln \mathrm{TFP}_s, \Delta AE_s) + Cov(\Delta \ln \mathrm{TFP}_s, \Delta EX_s). \end{aligned}$$

The first covariance term gives the contribution of within-firm productivity changes to the volatility of annual TFP changes; this term captures the variations in $\Delta \ln \mathrm{TFP}$ due to variations in within-firm productivity ΔTE, both directly and through ΔTE's correlation with ΔEX and ΔAE. The second term gives the contribution of within-sector allocative efficiency and the last

in allocative efficiency to a variety of sectoral characteristics. I find that the average change in allocative efficiency is correlated negatively to the persistence of firm-level TFP and positively to the average change in inputs.

Table 1: Correlation with sectoral real value added growth, 1991-2006

	$\Delta \ln \mathrm{TFP}_s$	$\Delta \mathrm{TE}_s$	$\Delta \mathrm{AE}_s$	$\Delta \mathrm{EX}_s$
MANUFACTURING				
Food products	0.88	0.69	0.00	0.04
Wearing apparel and leather products	0.94	0.86	-0.18	0.05
Printing and reproduction	0.90	0.79	-0.30	0.31
Pharmaceutical products and perfumes	0.94	0.90	-0.68	0.18
Household equipment	0.88	0.77	-0.06	-0.29
Motor vehicles	0.84	0.65	-0.11	-0.62
Other transportation equipment	0.96	0.96	-0.38	0.38
Mechanical equipment	0.82	0.84	-0.23	-0.17
Electric equipment	0.81	0.71	-0.56	0.23
Mineral products	0.94	0.86	-0.42	0.11
Textiles	0.66	0.60	0.14	-0.34
Wood and paper products	0.93	0.90	0.08	-0.12
Chemical, rubber and plastics products	0.92	0.79	0.05	-0.01
Fabricated metal products	0.76	0.77	0.22	-0.86
Electronic products	0.69	0.55	0.47	-0.10
SERVICES				
Trade and repair of motor vehicles	0.88	0.87	-0.51	0.28
Wholesale trade	0.88	0.89	-0.33	-0.39
Retail trade	0.66	0.72	-0.43	-0.49
Transportation	0.90	0.84	-0.03	-0.03
Real estate	0.77	0.76	-0.54	-0.25
Telecommunications and postal services	0.77	0.72	-0.22	-0.11
Professional and technical services	0.34	0.52	-0.46	0.32
Administrative and support services	0.47	0.49	-0.10	0.06
Accommodation and food services	0.80	0.88	-0.46	-0.20
Arts, entertainment and recreation	0.53	0.65	-0.41	-0.71
Personal and domestic services	0.38	0.28	0.12	-0.23

Note: This table presents, for each sector, the correlation of productivity growth and its components with real value added growth. The first column gives the correlation of sectoral productivity growth ; the second column gives the correlation of firm-level efficiency ; the third column gives the correlation of allocative efficiency and the last column that of the extensive margin.

term that of the extensive margin.

The contribution of each component is reported in Table 2. Firm-level productivity appears to be the main driver of sectoral productivity fluctuations. By contrast, fluctuations in allocative efficiency tend to dampen the movements of sectoral productivity. In fact, the volatility of sectoral productivity growth would have been higher if the level of allocative efficiency had

Table 2: Contribution to sectoral volatility, 1991-2006

	$\Delta \mathrm{TE}_s$	$\Delta \mathrm{AE}_s$	$\Delta \mathrm{EX}_s$
MANUFACTURING			
Food products	1.02	-0.08	0.06
Wearing apparel and leather products	1.09	-0.14	0.05
Printing and reproduction	1.09	-0.21	0.11
Pharmaceutical products and perfumes	1.37	-0.41	0.04
Household equipment	1.15	-0.10	-0.06
Motor vehicles	1.33	-0.31	-0.02
Other transportation equipment	1.03	-0.06	0.03
Mechanical equipment	1.00	-0.03	0.03
Electric equipment	1.87	-0.92	0.04
Mineral products	1.24	-0.25	0.01
Textiles	1.06	-0.13	0.07
Wood and paper products	0.99	0.01	0.01
Chemical, rubber and plastics products	1.06	-0.08	0.01
Fabricated metal products	1.15	0.06	-0.21
Electronic products	0.75	0.22	0.03
SERVICES			
Trade and repair of motor vehicles	1.12	-0.21	0.09
Wholesale trade	1.16	-0.09	-0.07
Retail trade	1.36	-0.33	-0.03
Transportation	1.04	-0.04	0.00
Real estate	1.67	-0.64	-0.04
Telecommunications and postal services	0.98	-0.14	0.16
Professional and technical services	0.79	0.15	0.05
Administrative and support services	0.58	0.42	0.00
Accommodation and food services	1.13	-0.21	0.08
Arts, entertainment and recreation	1.57	-0.52	-0.05
Personal and domestic services	1.05	-0.09	0.04

Note: This table presents the contribution of each components to the volatility of sectoral productivity growth. The first column gives the contribution of firm-level efficiency to the volatility of sectoral productivity growth. The second column gives the contribution of allocative efficiency, and the last column gives the contribution of the extensive margin. The contributions are computed as described in the text. The sum of the contribution of firm-level efficiency, allocative efficiency and the extensive margin equals 1.

stayed constant. The variations in allocative efficiency play a sizable role for the dynamics of sectoral productivity; on average, allocative efficiency reduces the volatility of sectoral TFP by 16%. The role of the extensive margin for the volatility of sectoral productivity changes is substantially less important.

Table 3 presents the results for the between-sector decomposition. The decomposition separates the role of between-sector allocative efficiency from within-sector productivity (which includes firm-level productivity changes, within-sector allocative efficiency and the entry/exit component). Table 3 indicates that, in line with the results of Figure 4, allocative efficiency has a distinct dynamics within and between sectors. In fact, changes in between-sector allocative efficiency are acyclial and contribute little to the variance of aggregate TFP changes. Within-sector allocative efficiency is more important than between-sector allocative efficiency to understand the cyclical dynamics of aggregate TFP.

Table 3: Within- and between-sector efficiency, 1991-2006

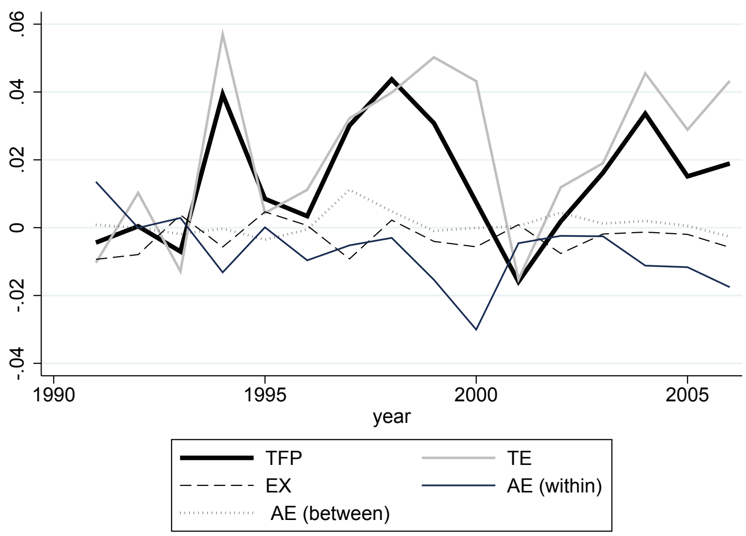
	$\Delta \ln \mathrm{TFP}$	ΔE	$\Delta AE_{\mathrm{between}}$
correlation with real value added	0.85	0.86	-0.01
contribution to TFP volatility	1	0.93	0.07

Note: The first line gives the correlation of aggregate productivity changes $\Delta \ln \mathrm{TFP}$, within-sector productivity changes ΔE and between-sector allocative efficiency $\Delta AE_{\mathrm{between}}$ with real value added growth in the aggregate economy. The second line gives the contribution of each of these variables to the volatility of aggregate productivity changes.

The cyclicality of allocative efficiency has interesting implications for the literature that aims at disentangling technology from measured productivity. Basu et al. (2006) show how to construct a measure of technology from measured productivity, controlling for aggregation effects, variation in input utilization, nonconstant returns, and imperfect competition. They find that technology varies about half as much as measured productivity. Moreover, their results indicate that, while measured productivity is procyclical, technology does not comove with output.[54] They do not, however, control for changes in allocative efficiency. The results of my decomposition suggest that, if allocative efficiency were countercyclical in the U.S. as well, technology could be more procyclical than what Basu et al.'s (2006) results suggest. Since allocative efficiency is countercyclical, measures of technology that do

[54]They also find that technology improvement is associated with lower hours and lower input use. In a related paper, Gali (1999) finds that, while hours are not correlated to labor productivity, they are correlated negatively to technology.

Figure 4: The allocative efficiency decomposition



Note: This figure shows aggregate TFP changes, measured from equation (20), and the components of the allocative efficiency decomposition, given in equation (28). The components do not add up exactly to aggregate TFP changes because both measures of TFP are approximations. The gap between aggregate TFP changes and the sum of the components is at most 0.4 percentage point.

not control for allocative efficiency are likely to underestimate the procyclicality of technology. Allocative efficiency is an additional factor that should be controlled for when disentangling technology from measured productivity.

4.3.3. Robustness checks

Because the measure of allocative efficiency proposed in this paper relies on the wedges between the data and the first-best allocation of resources, the decomposition is likely to be sensitive to the model specification. In particular, an important concern with the wedge approach is that firm-level distortions may capture specification errors in addition to frictions. To address this issue, I evaluate the robustness of the results to alternative specifications of the production function, as well as to the assumption used to estimate the factor elasticities.

I first consider the case where firms face a fixed labor cost of production. In the presence of overhead labor costs, using the expression of the marginal product given in equation (17) would lead to a biased measure of distortions. When incorrectly measured with equation (17), the measured marginal productivity may differ across firms even when labor is efficiently allocated. With fixed costs, firms have increasing returns to scale, and larger firms therefore have a higher measured marginal productivity. The measure of marginal productivity must therefore be corrected for the presence of overhead labor costs. In the second robustness exercise, I allow factor elasticities to vary within sectors. Here as well, differences in the measured marginal productivity of inputs are not necessarily a sign of misallocation. If firms differ in their factor elasticities, some of the measured dispersion in marginal productivities could be the result of the heterogeneity in factor elasticity. I describe in more detail the two alternative specifications in Appendix B and explain how to estimate the decomposition under these assumptions in Appendix C. The results, reported in Figures D.4 to D.6, show that, while the value of the correlation between allocative efficiency and output is sensitive both to the presence of overhead labor costs and to the heterogeneity in factor elasticities, the sign of the correlation is robust to these two alternative specifications. As in the benchmark specification, allocative efficiency is countercyclical in most sectors.

I also examine the robustness of the results to the assumption used to estimate the factor elasticities. Contrary to the benchmark estimation, I allow average unobserved sectoral distortions to differ from zero. To do so, I assume that capital and labor distortions are equal and follow the method described by Hall (1991). Appendix C provides more detail on the estimation. I find that allocative efficiency is countercyclical when allowing for average unobserved distortions (see Figures D.7 and D.8).

4.3.4. Comparison with alternative decompositions

To illustrate the difference between changes in allocative efficiency and the patterns of input reallocation, I compare the results of my decomposition with both the reallocation decomposition and the shift-share decomposition, presented in Section 2.4. I estimate the version of the shift-share decomposition proposed by Foster et al. (2001). The expressions used for the estimation are reported in Appendix C, and the tables of results are reported in Appendix D.

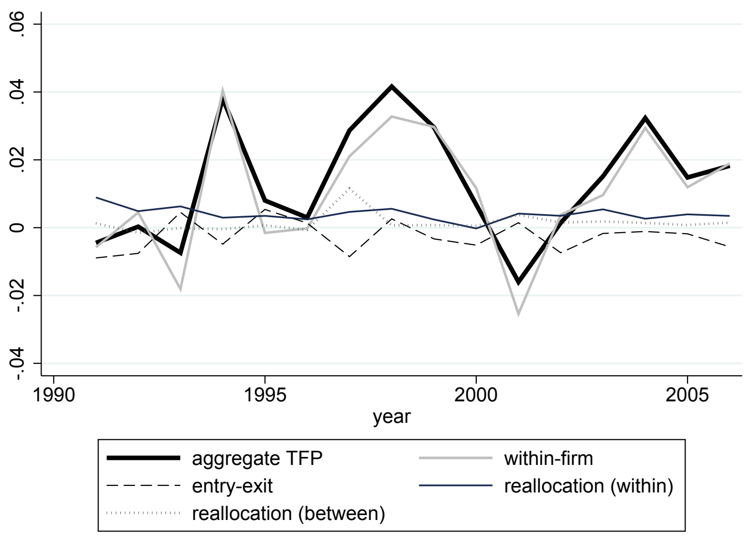
Figure 5 and 6 show the components of the reallocation decomposition and the shift-share decomposition. In both cases, the decomposition is obtained by applying and combining the decompositions within sectors and between sectors, as done for the allocative efficiency decomposition. As shown in the figures, the cyclical patterns of the technology component are similar across the different decompositions. The main difference is in the volatility, which appears higher in the allocative efficiency decomposition. The differences between the decompositions are starker for the reallocation/allocative efficiency components. The dynamics of the within-sector reallocation component, computed using either the shift-share decomposition or the reallocation decomposition, differs substantially from the dynamics of within-sector allocation efficiency. Whereas allocative efficiency is negatively correlated to aggregate TFP growth, the reallocation component of the shift-share decomposition is positively correlated, and that of the reallocation decomposition is essentially flat over the period. To better contrast my decomposition with the alternative decompositions, I also report, in Figures 7 and 8, the correlation of the reallocation components with sectoral real value added growth, as well as their contribution to the volatility of sectoral productivity. The correlation with value added growth confirms the patterns already seen at the aggregate level. Whereas allocative efficiency is countercyclical in most sectors, the reallocation component of the shift-share decomposition is procyclical, and that of the reallocation decomposition does not have a clear cyclical pattern. These striking differences are reflected in their contributions to the volatility of sectoral productivity. The reallocation component of the shift-share decomposition tends to amplify the volatility of average productivity whereas that of the reallocation decomposition contributes little to the volatility of aggregate productivity changes. By contrast, allocative efficiency tends to reduce the volatility of sectoral productivity growth. The two alternative decompositions therefore do not provide a good proxy for changes in allocative efficiency.

Comparing the allocative efficiency decomposition to the two alternative decompositions provides some insight on the patterns behind the changes in allocative efficiency. The stark difference with the reallocation decomposition underlines the role of firm-level productivity changes. As already noted, the reallocation and allocative efficiency decompositions would be identical in the absence of fluctuations in firm-level productivity. When firm-level productivity varies, allocative efficiency depends on the way firms adjust their inputs in response to productivity changes. This dimension, which is absent from the reallocation decomposition, is captured by the distortions. As already mentioned, fluctuations in distortions are highly correlated to fluctuations in firm-level productivity (cf. Table D.2), which implies that firms are not adjusting their inputs as much as what they would do in the absence of frictions; allocative efficiency changes are then closely related to how firm-level productivity varies across firms with different levels of distortions. The comparison with the shift-share decomposition also yields some interesting insight. The fluctuations of allocative efficiency are diametrically opposed to that of the shift-share reallocation component. Here as well, the fluctuations in firm-level productivity and the close connection between productivity and distortions can shed some light on this contrasting pattern. When the productivity of a distorted firm rises, the firm's distortion and its output both tend to increase. Since distorted firms are on average more productive, the increase in output raises the shift-share reallocation component, whereas the increase in distortion leads to a decline in allocative efficiency. All in all, these comparisons indicate that the dynamics of firm-level productivity is crucial to understanding the fluctuations in allocative efficiency.

5. Conclusion

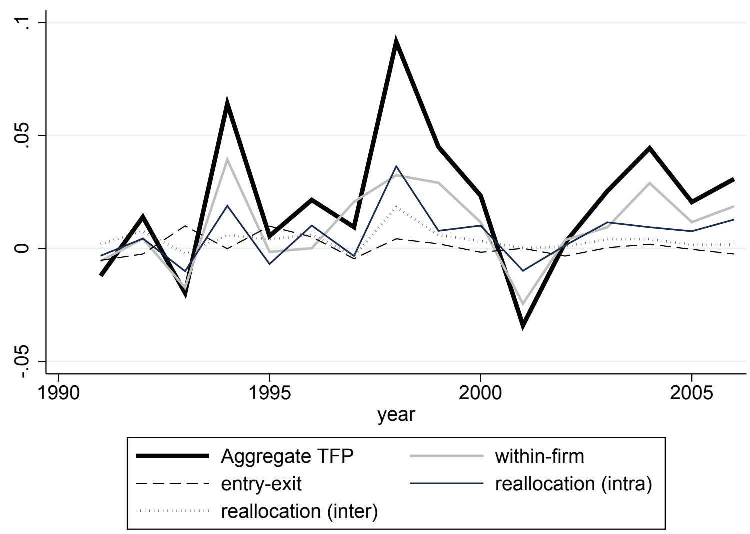
This paper proposes a novel decomposition to document the microeconomic determinants of aggregate productivity changes. After deriving aggregate productivity from the aggregation of firm-level production functions, I show how to decompose sectoral productivity into changes in firm-level efficiency, changes in allocative efficiency, and changes in entry and exit. I then decompose the economy-wide productivity into within-sector productivity and between-sector allocative efficiency. I use this decomposition to study the evolution of allocative efficiency over the business cycle with French firm-level data on manufacturing and service sectors over 1991-2006. In contrast with the literature on the cleansing effect of recessions, which emphasizes the role of entry and exit, this paper highlights the contribution of changes in the efficiency of resource allocation across incumbent firms. I find that within-sector allocative efficiency is countercyclical and reduces the volatility of aggregate productivity. By contrast, between-sector allocative efficiency shows little cyclicality. Furthermore, the extensive margin plays a negligible role in the cyclical dynamics of sectoral productivity for most sectors. It should be noted, however, that the decomposition measures only the contemporaneous impact of entry and exit on aggregate productivity and ignores the effects of entry and exit on long-run productivity growth, as well as their potential effects on the incumbents' innovation. Moreover, the decomposition allows me to document the patterns of allocative efficiency but is silent on the underlying mechanisms. A more structural approach would be needed to identify the frictions that generate the countercyclical allocative efficiency.

Figure 5: The reallocation decomposition



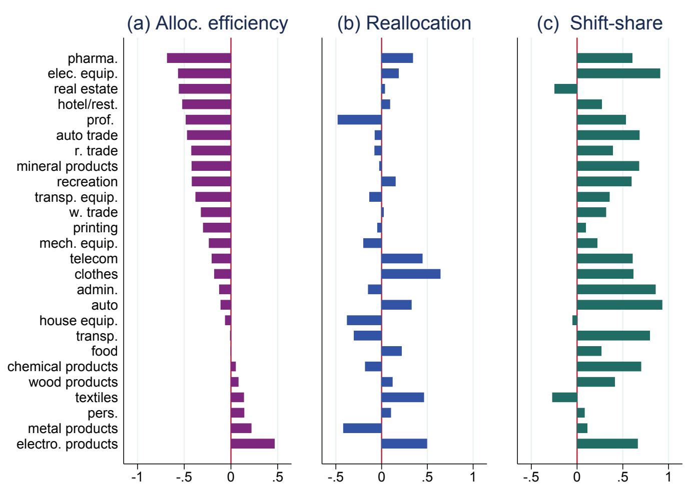
Note: This figures shows aggregate TFP changes, measured from equation (20), and the components of the reallocation decomposition, which are obtained by combining the within-sector and between sector decompositions. The component "reallocation (between)" refers to between-sector reallocation, and "reallocation (within)" refers to within-sector reallocation. The components do not add up exactly to aggregate TFP changes because the between-sector decomposition is an approximation. The gap between aggregate TFP changes and the sum of the components is less than 0.4 percentage point.

Figure 6: The shift-share decomposition



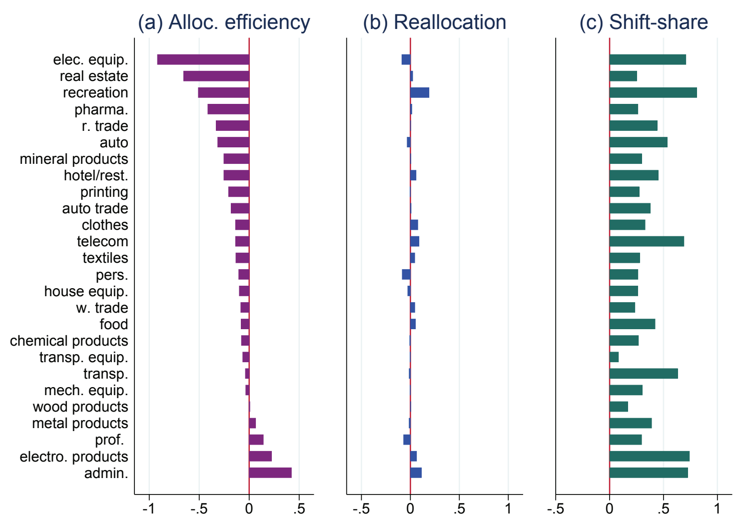
Note: This figure shows aggregate TFP changes, defined as the weighted average of firm-level productivity changes, and the components of the shift-share decomposition. Note that aggregate TFP is thus different from Figure 4 and 5. The components here add up exactly to aggregate TFP changes.

Figure 7: Comparison with alternative decompositions: correlation with value added changes



Note: This figure compares the allocative efficiency decomposition to the reallocation and to the shift-share decompositions. The figure shows for each sector, the correlation with value added growth of (a) the change in allocative efficiency; (b) the contribution of input reallocation to aggregate TFP; and (c) the contribution of reallocation to average productivity.

Figure 8: Comparison with alternative decompositions: contribution to the volatility of aggregate productivity



Note: This figure compares the allocative efficiency decomposition to the reallocation and to the shift-share decompositions. The figure shows for each sector, the share of the volatility of aggregate productivity changes accounted for by (a) the change in allocative efficiency; (b) the contribution of input reallocation to aggregate; and (c) the contribution of reallocation to average productivity.

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

Baden-Fuller, C. (1989), 'Exits from declining industries and the case of steel castings', *Economic Journal* **99**, 949–961.

Baily, M. N., C. Hulten and D. Campbell (1992), 'Productivity dynamics in manufacturing plants', *Brookings Papers on Econmic Activity (Microeconomics)* pp. 187–249.

Baqaee, D. and E. Farhi (2018), Productivity and misallocation in general equilibrium.

Barlevy, G. (2003), 'Credit market frictions and the allocation of resources over the business cycle', *Journal of Monetary Economics* **50**(8), 1795–1818.

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.

Basu, S., J. Fernald and M. Kimball (2006), 'Are Technology Improvements Contractionary?', *American Economic Review* **96**(5), 1418–1448.

Basu, S and J. G. Fernald (2002), 'Aggregate productivity and aggregate technology', *European Economic Review* **46**(6), 963–991.

Basu, S., L. Pascali, F. Schiantarelli and L. Serven (2009), Productivity, welfare and reallocation: Theory and firm-level evidence, NBER Working Papers 15579.

Bresnahan, T. F. and D. M. G. Raff (1991), 'Intra-industry heterogeneity and the great depression: The american motor vehicles industry, 1929 - 1935', *The Journal of Economic History* **51**(02), 317–331.

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

Buera, F., K. Kaboski and Y. Shin (2011), 'Finance and development: A tale of two sectors', *American Economic Review* **101**(5), 1964–2002.

Bunel, M., R. Duhautois and L. Gonzalez (2009), 'Types de fusions-acquisitions et évolution de l'emploi des entreprises restructurées', *Travail et Emploi* **117**.

Caballero, R. J. and M. L. Hammour (1994), 'The cleansing effect of recessions', *American Economic Review* **84**(5), 1350–68.

Chari, V., P. Kehoe and E. McGrattan (2007), 'Business cycle accounting', *Econometrica* **75**(3), 781–836.

Davis, S. and J. Haltiwanger (1998), Measuring gross worker and job flows, *in* 'Labor Statistics Measurement Issues', NBER Chapters, National Bureau of Economic Research, Inc, pp. 77–122.

Fisher, F. (1969), 'The existence of aggregate production functions', *Econometrica* **37**(4), 553–577.

Foster, L., C. Grim and J. Haltiwanger (2016), 'Reallocation in the Great Recession: Cleansing or Not?', *Journal of Labor Economics* **34**(S1), S293 – S331.

Foster, L., J. C. Haltiwanger and C. J. Krizan (2001), Aggregate productivity growth. lessons from microeconomic evidence, *in* 'New Developments in Productivity Analysis', NBER, pp. 303–372.

Gali, J. (1999), 'Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?', *American Economic Review* **89**(1), 249–271.

Garicano, L., C. Lelarge and J. Van Reenen (2016), 'Firm size distortions and the productivity distribution: Evidence from france', *American Economic Review* **106**(11), 3439–3479.

Gopinath, G., S. Kalemli-Ozcan, L. Karabarbounis and C. Villegas-Sanchez (2017), 'Capital Allocation and Productivity in South Europe', *The Quarterly Journal of Economics* **132**(4), 1915–1967.

Gourio, F. and N. Roys (2014), 'Size-dependent regulations, firm size distribution, and reallocation', *Quantitative Economics* **5**, 377–416.

Griliches, Z. and H Regev (1995), 'Productivity and firm turnover in israeli industry: 1979-1988', *Journal of econometrics* **65**, 175–203.

Guner, N., G. Ventura and Yi Xu (2008), 'Macroeconomic implications of size-dependent policies', *Review of Economic Dynamics* **11**(4), 721–744.

Hall, Robert E. (1991), Invariance properties of solow's productivity residual, NBER Working Papers 3034.

Hopenhayn, H. (2012), On the measure of distortion, Technical report.

Houthakker, H. (1955), 'The pareto distribution and the cobb-douglas production function in activity analysis', *Review of Economic Studies* .

Hsieh, C. and P. Klenow (2009), 'Misallocation and manufacturing tfp in china and india', *Quarterly Jounral of Economics* **124**(4).

Hulten, Charles R. (2000), Total factor productivity: A short biography, Nber working papers.

Jorgenson, D., F. Gollop and B. Fraumeni (1987), *Productivity and U.S. Economic Growth*, Harvard University Press.

Klein, L. (1946), 'Macroeconomics and the theory of rational behavior', *Econometrica* **14**(02), 93–108.

Lee, Y. and T. Mukoyama (2015), 'Entry and exit of manufacturing plants over the business cycle', *European Economic Review* **77**(C), 20–27.

Levinsohn, J. and A Petrin (2003), 'Estimating production functions using inputs to control for unobservables', *Review of Economic Studies* **70**(2), 317–341.

Malinvaud, E. (1993), 'A framework for aggregation theories', *Ricerche Economiche* **47**, 107–135.

May, K. (1946), 'The aggregation problem for a one-industry model', *Econometrica* **14**(04), 285–298.

Melitz, M. and S Polanec (2015), 'Dynamic Olley-Pakes productivity decomposition with entry and exit', *RAND Journal of Economics* **46**(2), 362–375.

Nataf, A. (1948), 'Sur la possibilité de construction de certains macromodèles', *Econometrica* **16**, 232–244.

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

Olley, G. S. and A. Pakes (1996), 'The dynamics of productivity in the telecommunication equipment industry', *Econometrica* **64**(6), 1263–1297.

Osotimehin, S. and F. Pappadà (2017), *Economic Journal* **127**(602), 1153–1187.

Ouyang, M. (2009), 'The scarring effect of recessions', *Journal of Monetary Economics* **56**(2), 184–199.

Petrin, A. and J. Levinsohn (2012), 'Measuring aggregate productivity growth using plant-level data', *The RAND Journal of Economics* **43**(4), 705–725.

Petrin, A., J. Reiter and K. White (2011), 'The impact of plant-level resource reallocations and technical progress on u.s. macroeconomic growth', *Review of Economic Dynamics* **14**(1), 3–26.

Restuccia, D. and R. Rogerson (2008), 'Policy distortions and aggregate productivity with heterogeneous plants', *Review of Economic Dynamics* **11**(4), 707–720.

Rotemberg, Julio and Michael Woodford (1993), Dynamic General Equilibrium Models with Imperfectly Competitive Product Markets, Working papers, National Bureau of Economic Research.

Sandleris, G. and M. 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.

Solow, R (1957), 'Technical change and the aggregate production function', *The Review of Economics and Statistics* **39**(3), 312–320.