Financial frictions, capital structure, and aggregate productivity

Alexandros Fakos*

May 4, 2024

Abstract

I develop a general equilibrium model of heterogeneous firms to study the importance of financial frictions on aggregate productivity. Frictions lead to lower aggregate productivity by inducing suboptimal firm capital structure and leading to misallocation of capital and labor across firms. Financial frictions are the only cause of suboptimal capital structure by driving a wedge between the price of debt and equity. Yet, they are one of the possible sources of the dispersion in the marginal revenue product of capital (MRPK) which leads to the misallocation of real resources. Model estimates from European public firms imply that suboptimal capital structure leads to a 2% loss in aggregate productivity, which is a lower bound on the importance of financial frictions, and represents 1/10 of the total loss from all frictions. The quantities of interest are precisely estimated, and I obtain their standard errors using the influence function approach, which provides a general inferential framework for multi-step estimators of structural models in finance.

Keywords: Capital structure, financial frictions, aggregate productivity, influence functions.

JEL classification codes: G30, E23, C14.

*Department of Business Administration, ITAM, Mexico City, México; email: alexfakos@icloud.com. I am grateful to Toni Whited for introducing me to the concept of influence functions during the Mitsui Center Summer School on Structural Estimation in Corporate Finance she organized in 2019 at the University of Michigan. I thank Ignacio Lobato and Lukas Schmid for helpful suggestions and comments. Any errors or omissions are my responsibility. Declarations of interest: none.

1 Introduction

How much do frictions in financial markets reduce aggregate total factor productivity (TFP)? I address this question by developing a general equilibrium framework featuring heterogeneous firms, a production function that depends on capital structure (debt–equity ratio), frictions in financial asset markets that lead to suboptimal firm capital structure, and a host of other frictions (financial or otherwise) reflected in a firm-specific capital wedge that prevents the efficient allocation of capital and labor across firms. I quantify the model using data from European public firms and find that suboptimal capital structure due to financial frictions lowers aggregate TFP by 2% in a typical year in 1989–2022, which is a lower bound on the importance of financial frictions in the economy, and represents 1/10 of the total loss from all frictions.

In this model, firms are heterogeneous in terms of their productivity and choose the amount of labor and physical capital, along with the liabilities that finance capital (debt and equity), to maximize profit. Debt and equity are not perfect substitutes in production: there is an optimal capital structure (debt share), with a suboptimal capital structure manifesting as lower firm productivity. This modeling choice is motivated by the literature in finance that views firms as more than an exogenous cashflow process (Jensen and Meckling, 1976; Hart, 1988; Aghion and Bolton, 1992; Hart, 1995; Whited and Zhao, 2021).[1] Given a productivity distribution, a first-best allocation of capital, labor, and firm capital structure maximizes output (hence aggregate TFP, given the aggregate supply of resources). The price of equity and debt vary across firms, reflecting distortions in financial and other markets, leading to lower aggregate TFP than in the first-best allocation.

I decompose the distortions in the model into two wedges: a wedge between the price of equity and debt for each firm (which I call the capital structure distortion) and a

[1] For a detailed discussion on how modeling choices are motivated from research on the theory of the firm, see the end of Section 2.1.

wedge between the overall cost of capital faced by a firm and the economy-wide price of capital that reflects the aggregate scarcity of the physical resource. The capital structure distortion reduces aggregate TFP through a purely financial mechanism: when the relative price of debt is different than 1, the capital structure is suboptimal, leading to lower firm productivity, which results in lower aggregate TFP. Notably, this mechanism leads to lower aggregate TFP even if all firms face the same relative price of debt, as long as this is different than 1. The reason is that investors supplying the financial resources are indifferent about how the firm cash flow is split between debt and equity because money is fungible; therefore, in a competitive asset market, debt and equity are elastically supplied at a relative price of 1. Thus, in a competitive first-best equilibrium, the perfectly elastic relative supply of debt to equity determines the price, and the demand side (the firm's production function) determines the relative quantity.[2]

All distortions affect the allocation of capital and labor across firms. The capital wedge generates dispersion in the marginal revenue product of capital (MRPK) across firms, leading to misallocation of capital and labor and lower aggregate TFP. The capital structure distortion alters a firm's effective productivity, making it endogenous. Hence, firm productivity has an exogenous component, which I call frontier productivity, and an endogenous component that depends on capital structure. The capital structure distortion reduces firm productivity, which determines a firm's demand for capital, thus altering the allocation of capital and labor across firms relative to the first-best. The effect of this channel on aggregate TFP depends on the joint distribution of frontier productivity and capital structure distortion.

The model's novelty is that it embeds the capital structure misallocation model of Whited and Zhao (2021) into the Hsieh and Klenow (2009) model of misallocation of

[2]For a detailed discussion on how the elimination of the capital structure distortion relates to the Modigliani-Miller environment, see the end of Section 2.1.

capital and labor. The advantage of my framework is that it relates firm total factor productivity to capital structure and distortions affecting real (non-financial) resources, which allows measuring the importance of financial frictions on aggregate TFP and their importance relative to all frictions.

I quantify the model using data on European public firms, which include information on labor expenditures (unlike data on US public firms). Data on labor allow the measurement of firms' value-added and productivity from capital and labor (firm total factor productivity), which allows for a precise estimate of the macroeconomic importance of distortions. After estimating the production function parameters, I recover firm productivity, the capital wedge, and the capital structure distortion that is a function of the debt-to-equity ratio.

I use the estimated firm productivity, capital wedge, and capital structure distortion to quantify the TFP loss from frictions. I find that suboptimal capital structure is responsible for a 2% loss in aggregate TFP, which is substantial. This loss represents roughly 1/10 of the total loss from all frictions. This exercise implies that even if a policy eliminates dispersion in MRPK, it will not eliminate the 2% loss from suboptimal capital structure. Because financial frictions are the sole driver of the capital structure distortion but can also be a driver of the capital wedge, the 2% estimate represents a lower bound to the importance of financial frictions in the total TFP loss from all distortions.

I use the influence function approach to evaluate the statistical precision of the results and conduct inference on the model parameters. Inference on the counterfactual quantities of interest is challenging because the TFP loss is a function of the data and estimated parameters in multiple stages. In addition, the TFP loss is not a standard statistic; therefore, its asymptotic properties are unknown and must be derived. The starting point of the approach is the von Mises expansion of an estimator/statistic (Fernholz, 1983), which for a broad class of estimators (e.g., MLE, GMM, functions of means such as the TFP loss) implies that we can study their asymptotic properties from the derivative of the estimator/statistic with respect to the probability distribution generating the data. This derivative representation allows for building the asymptotic distribution of a multistep estimator using concepts familiar from calculus, like the chain rule or obtaining the total derivative from partial derivatives.

This approach simplifies inference because the derivative in the von Mises expansion can be expressed as an integral of a specific random variable: the influence function, which has the same asymptotic distribution as the statistic/estimator of interest. This representation means that to calculate the standard error of the statistic/estimator, we need to calculate the standard error of the mean of the influence function, which is trivial once the influence function is known. Expressions for the influence function of the typical classes of estimators MLE, GMM, quantiles, and moments are well-studied in the literature (Newey and McFadden, 1994; Van der Vaart, 1998; Tsiatis, 2006). In addition, I show that the influence function of any statistic that is a function of means, no matter how complicated (as is the TFP loss), can be numerically approximated by finite differences, sidestepping the need for analytical derivations.

The influence function is a general inferential framework particularly useful for conducting inference on structural models that usually involve many steps and different classes of estimators in each step. For example, this paper combines a quantile, a GMM estimator, and a novel statistic: the TFP loss, a function of averages. As long as the influence function of each step is obtained, I can add all influence functions and calculate the standard errors of the quantity of interest. An alternative approach to inference on multistep estimators is the bootstrap, whose disadvantage is computational inefficiency if an estimation step involves a computationally expensive estimator. The influence function approach facilitates a hybrid approach to inference: use analytical expressions for computationally expensive steps (usually a Simulated Minimum Distance of a dynamic model) and approximate numerically the computationally cheap steps (like the TFP loss formula). This hybrid approach simplifies the development of the computer code used to carry out the calculations and reduces computing time.

Related literature. This paper contributes to the literature on misallocation (seminal papers include Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009) that studies how firm-specific distortions impact aggregate TFP when firms are heterogeneous in their productivity (for surveys see Hopenhayn, 2014; Restuccia and Rogerson, 2017). A large body of work has focused on how MRPK dispersion arises from different market frictions (e.g., information frictions in David et al., 2016) and financial frictions (Gopinath et al., 2017; Karabarbounis and Macnamara, 2021). This paper, instead, focuses on a different channel: the effect of financial frictions on aggregate TFP through its direct impact on firm-specific productivity by distorting capital structure. The role of frictions in capital structure has first been studied in the seminal paper of Whited and Zhao (2021)—I extend that model by adding real variables that allow for the estimation of firm-specific total factor productivity. In this model, firm productivity is endogenous as it depends on capital structure—therefore, it is related to the work on endogenous firm productivity such as Aghion et al. (2023), who study the aggregate implications of misallocation in R&D.

This paper also contributes to corporate finance by providing a quantitative model of the firm's capital structure to assess the aggregate implications of suboptimal capital structure. Several theoretical papers have motivated optimal capital structure Jensen and Meckling (1976); Hart (1988); Aghion and Bolton (1992); Hart (1995), mainly in environments with conflicts of interest between managers and investors and contract incompleteness. In these environments, optimal capital structure is a solution to the problem of contracting frictions, and any distortion leading to suboptimal capital structure worsens allocations. My model provides a reduced-form specification of the optimal capital structure similar to

Whited and Zhao (2021) and quantitatively evaluates the importance of capital structure frictions relative to all frictions that reduce aggregate productivity.

Lastly, this paper contributes to the methodology of structural estimation in finance by providing a general framework for inference on multistep multiclass estimators using influence functions. Bazdresch et al. (2018) show how to use influence functions for inference in one-step estimators of dynamic corporate finance models targeting values (benchmarks) of one class of estimators (either moments or regression coefficients of policy functions). This paper extends their methodology by allowing the target benchmarks to come from multistep hybrid (including different estimator classes, e.g., MLE, GMM, quantile, or functions of means such as standard deviation, correlation, or TFP loss) estimators. This paper is also related to recent methodological studies on causal inference using influence functions to study semiparametric estimators with a non-parametric first step (see Cattaneo et al., 2013; Chernozhukov et al., 2018; Ichimura and Newey, 2022; Kennedy, 2022). This paper, instead, focuses on parametric multistep estimators and exploits the conceptual simplicity and generality of influence functions to present a general framework for inference in multistep hybrid estimation of structural models.

The remainder of the paper is organized as follows. Section 2 develops the model, Section 3 presents the dataset, Section 4 describes the parameter estimation, Section 5 presents the results and develops the inferential framework using influence functions, and Section 6 concludes. An online appendix includes formula derivations, additional results, and an overview of the influence function approach to inference.

2 Model

This section presents a model of a monopolistically competitive sector with heterogeneous firms and capital market frictions that reduce sectoral aggregate productivity through two channels: by inducing misallocation of capital across firms as in Hsieh and Klenow (2009) and by leading to suboptimal capital structure (debt and equity financing) similar to Whited and Zhao (2021) that leads to lower firm productivity. The online appendix contains details about the model and additional formulas.

Consider a sector at time t populated by N_t firms. Each firm i at time t faces productivity A_{it} and requires capital K_{it} and labor L_{it} to produce a differentiated product Y_{it}. The capital input can be financed through debt D_{it} or equity E_{it}, with $K_{it} = D_{it} + E_{it}$. Capital structure is characterized by the debt share $d_{it} = \frac{D_{it}}{K_{it}}$, which determines the equity share $\frac{E_{it}}{K_{it}} = 1 - d_{it}$. The capital structure enters the production function together with productivity, capital, and labor:

$$Y_{it} = A_{it} \left[d_{it}^{\gamma} (1 - d_{it})^{1-\gamma} K_{it} \right]^{\alpha} L_{it}^{1-\alpha}, \ \alpha, \gamma \in (0, 1). \tag{1}$$

Let $F_{it} = \left[d_{it}^{\gamma} (1 - d_{it})^{1-\gamma} \right]^{\alpha}$ represent the contribution of capital structure in the production function. The quantity F_{it} is bounded above by $\overline{F} = \left[\gamma^{\gamma} (1 - \gamma)^{1-\gamma} \right]^{\alpha}$ and is maximized at capital structure $d = \gamma$, which I call the first-best capital structure. Therefore, the firm-level productivity residual $\widetilde{A}_{it} = \frac{Y_{it}}{K_{it}^{\alpha} L_{it}^{1-\alpha}}$ has an exogenous technological efficiency component A_{it} and an endogenous financial component F_{it}.

As in Hsieh and Klenow (2009) and Whited and Zhao (2021), each monopolistic competitor i sells its output to a single competitive aggregator firm that produces the sector's final good using a CES production function $\mathbf{Y}_t = \left[\sum_{i=1}^{N_t} Y_{it}^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}}$, $\eta > 1$. This market structure leads to downward-sloping isoelastic demand $p(Y_{it})$ for each product.

Firms choose capital structure d_{it}, capital K_{it}, and labor L_{it} to maximize period t profits in a monopolistic competitive market subject to a firm-specific price of debt $r_{d_{it}}$ and price of equity $r_{d_{it}}$, and a common wage w_t.

$$\max_{d_{it}, K_{it}, L_{it}} p(Y_{it}) Y_{it} - r_{d_{it}} d_{it} K_{it} - r_{e_{it}} (1 - d_{it}) K_{it} - w_t L_{it}. \tag{2}$$

Market frictions manifest as within-period across-firm dispersion in the price of debt and equity and the relative price of equity and debt $r_{e_{it}}/r_{d_{it}}$ as in Whited and Zhao (2021). The profit-maximizing capital structure d^*_{it} is

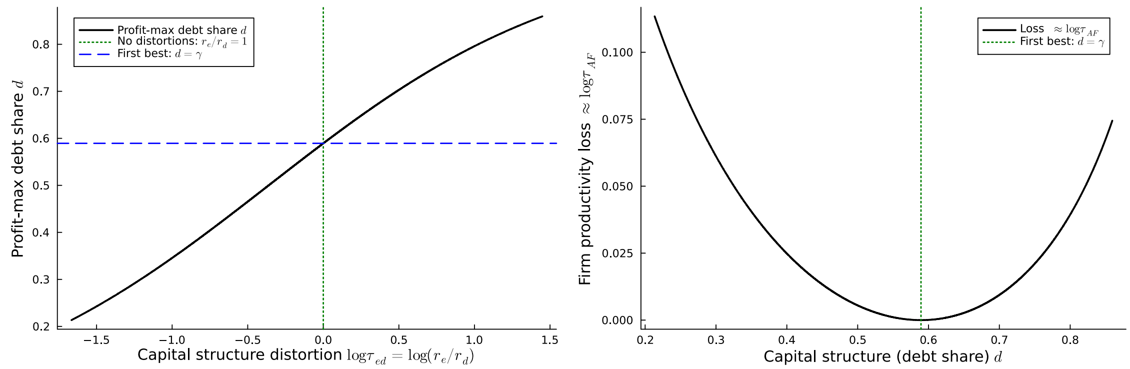
$$d^*_{it} = \frac{\gamma \frac{r_{e_{it}}}{r_{d_{it}}}}{(1-\gamma) + \gamma \frac{r_{e_{it}}}{r_{d_{it}}}}. \tag{3}$$

Note that the profit-maximizing capital structure depends only on the relative price of equity and debt but not their levels $r_{e_{it}}, r_{d_{it}}$. Thus, at the profit-maximizing capital structure, the contribution of capital structure to output F_{it} is a function of $r_{e_{it}}/r_{d_{it}}$, which I denote $f(r_e/r_d)$. When the price of equity is equal to the price of debt, the firm chooses the first-best capital structure $d^*_{it} = \gamma$. If the price of debt is different than that of equity $r_{e_{it}} \neq r_{d_{it}}$, then the contribution of capital structure to output $F_{it} = f(r_e/r_d)$ is less than it's potential $\overline{F} = f(1)$, leading to lower output. Therefore, if, for some firms, the price of debt is different than that of equity $r_{e_{it}}/r_{d_{it}} \neq 1$, there is an output loss in the economy from suboptimal capital structure. (See also Figure 1 for a graphical presentation of the above argument.)

Substituting $f(r_e/r_d)$ in the production function we have $Y_{it} = A_{it} f(\frac{r_e}{r_d}) K^{\alpha}_{it} L^{1-\alpha}_{it}$. The contribution of capital structure $f(r_e/r_d)$ appears in the production function as another productivity term. I discuss later in this section why the quantity F is more like a productivity rather than a resource. Substituting the profit-maximizing d^*_{it} in (2) leads to the reduced-form profit function $p(Y_{it})Y_{it} - \frac{r_{e_{it}}}{(1-\gamma)+\gamma \frac{r_{e_{it}}}{r_{d_{it}}}} K_{it} - w_t L_{it}$. At the profit-maximizing capital K^*_{it}, the first-order optimality conditions imply that the marginal revenue product of capital $MRPK_{it}$ is equal to the cost of capital

$$MRPK_{it} = \frac{r_{e_{it}}}{(1-\gamma) + \gamma \frac{r_{e_{it}}}{r_{d_{it}}}}. \tag{4}$$

Figure 1: The loss in firm-specific productivity from the capital structure distortion.



The left panel of this figure demonstrates how the capital structure distortion $\tau_{ed_{it}} \equiv \frac{r_{e_{it}}}{r_{d_{it}}}$ leads to suboptimal profit-maximizing capital structure (debt share) d_{it}. The right panel demonstrates how suboptimal capital structure leads to a loss in firm productivity. Suboptimal d_{it} manifests as the financial productivity wedge $\tau_{AF_{it}}$, which brings firm-specific productivity from its frontier $\overline{A}_{it}$ down to $\frac{\overline{A}_{it}}{\tau_{AF_{it}}}$, with $\tau_{AF_{it}} > 1$. Therefore the productivity loss for firm i is $1 - 1/\tau_{AF_{it}} \approx \tau_{AF_{it}} - 1 \approx \log(\tau_{AF_{it}})$. When there is no distortion (the price of debt is equal to the price of equity), the capital structure is at its first-best value $d_{it} = \gamma$ (where the green dotted line intersects the blue dashed line in the left panel), and there is no productivity loss (at the green dashed line of the right panel). The parameters α, γ are set at their estimated values (Table 1). The domain of the x-axis of the right panel is restricted to values between the 5th and 95th percentiles in the data, while the domain of the left panel's x-axis is determined from the domain of d and parameter γ.

Dispersion in the MRPK leads to misallocation and reduced aggregate sectoral TFP as in Hsieh and Klenow (2009). This model nests their model as the two models are identical when the price of equity is equal to the price of debt $r_{e_{it}} = r_{d_{it}} = r_{it}$.

An insight from this model is that frictions in asset markets leading to different prices of equity and debt $r_e/r_d \neq 1$ reduce aggregate output and aggregate sectoral TFP without necessarily generating dispersion in the MRPK across firms. More specifically, consider an environment where there is dispersion in $r_{e_{it}}/r_{d_{it}}$ while all firms exhibit the same MRPK=c, which is possible when the following relationship holds $r_{e_{it}} = c[(1-\gamma) + \gamma r_{e_{it}}/r_{d_{it}}]$. In that stylized economy, aggregate output is less than in the first best because firms have suboptimal capital structure.

2.1 Equilibrium

Let the $\mathbf{K}_t, \mathbf{L}_t$ denote the aggregate capital and labor supply, respectively. Let $\tau_{ed_{it}}$ represent the capital structure distortion: the wedge between the price of equity and debt. I reparameterize the flow price of capital so that it consists of a common component R_t and an idiosyncratic component $\tau_{K_{it}}$

$$\tau_{ed_{it}} \equiv \frac{r_{e_{it}}}{r_{d_{it}}}, \quad \frac{r_{e_{it}}}{(1-\gamma)+\gamma\tau_{ed_{it}}} = R_t \tau_{K_{it}}. \tag{5}$$

In general equilibrium, the prices of capital and labor R_t, w_t are such that the aggregate demand for capital and labor resulting from the profit maximization problem (2) given the distribution of productivities and wedges $\{A_{it}, \tau_{ed_{it}}, \tau_{K_{it}}\}_{i=1}^{N_t}$ equals their aggregate supply $\mathbf{K}_t, \mathbf{L}_t$. Such equilibrium exists for every economy $\{A_{it}, \tau_{ed_{it}}, \tau_{K_{it}}\}_{i=1}^{N_t}$ since the production possibility set of each producer is convex. Note that the aggregate resource constraint is in terms of capital and labor but not in terms of aggregate debt or equity. The total debt plus equity by construction equals the aggregate capital, so there is no need for an additional constraint on financial resources. In addition, the aggregate debt and equity are promises like money, and hence fungible, so the aggregate debt and equity can change without any cost to the economy (see also the discussion below about the Modigliani-Miller benchmark).

Aggregate TFP is the ratio between aggregate output and a Cobb-Douglas aggregator of aggregate capital and labor $TFP_t = \frac{\mathbf{Y}_t}{\mathbf{K}_t^{\alpha}\mathbf{L}_t^{1-\alpha}}$. In this model, the aggregate TFP admits a closed-form expression. What matters for TFP is the manifestation of wedge $\tau_{ed_{it}}$ on wedge τ_{AF_i} between the capital structure value added of the firm $f(\tau_{ed_{it}})$ and the capital structure value added at the first best $f(1)$.

$$\tau_{AF_{it}} \equiv \frac{\left(\gamma^{\gamma}[1-\gamma]^{1-\gamma}\right)^{\alpha}}{\left(d_{it}^{\gamma}[1-d_{it}]^{1-\gamma}\right)^{\alpha}} = \frac{f(1)}{f(\tau_{ed_{it}})} \geq 1 \tag{6}$$

This parameterization implies that the production function can be expressed as $Y_{it} = \frac{\overline{A}_{it}}{\tau_{AF_{it}}} K_{it}^{\alpha} L_{it}^{1-\alpha}$, where $\overline{A}_{it} = A_{it} f(1)$ represents productivity under the first-best capital structure. This parameterization of the model shows that a wedge between the price of equity and debt manifests as a factor that reduces a firm's productivity. (See also Figure 1 for a graphical presentation of the this mechanism.) Therefore, I call $\tau_{AF_{it}}$ the financial productivity wedge.

By improving the debt and equity mix, each firm's productivity can reach $\overline{A}_{it}$. In addition, increasing a firm's productivity by bringing $\tau_{AF_{it}}$ down to 1 does not require reducing the productivity of another firm. This property is in stark contrast with a physical resource like capital or labor, where to increase the capital of one firm, another firm's capital has to decrease. In other words, capital or labor are rival resources, while the capital structure value added F_{it} is non-rival, and this is the reason that it is more like a productivity component instead of a third resource, and this is why I call $\tau_{AF_{it}}$ a financial productivity wedge.

The TFP formula is:

$$TFP_t(\{\overline{A}_{it}, \tau_{AF_{it}}, \tau_{K_{it}}\}_{i=1}^{N_t}) = \left[\sum_{i=1}^{N}\left(\frac{\overline{A}_{it}}{\tau_{AF_{it}}\tau_{K_{it}}^{\alpha}}\right)^{\eta-1}\right]^{\frac{1}{\eta-1}+\alpha} \Bigg/ \left[\sum_{i=1}^{N}\frac{\left(\frac{\overline{A}_{it}}{\tau_{AF_{it}}}\right)^{\eta-1}}{\tau_{K_{it}}^{1+\alpha(\eta-1)}}\right]^{\alpha} \tag{7}$$

Note that aggregate TFP is scale-invariant with respect to the capital wedge $\tau_{K_{it}}$ implying that multiplying the capital wedge of every firm by a scalar $c > 0$ leaves the TFP unchanged. This property is common in models of misallocation (see Fakos, 2023). But TFP is not scale invariant with respect to the financial productivity wedge $\tau_{AF_{it}}$. In fact, TFP is homogeneous of degree -1 with respect to the productivity wedge, implying that multiplying $\tau_{AF_{it}}$ by a scalar $c > 0$ is equivalent to multiplying TFP by $1/c$.

The TFP formula (7) helps quantify the effect of frictions on aggregate TFP by comparing TFP under different distributions $\{\tau_{AF_{it}}, \tau_{K_{it}}\}_i$. A quantity of interest is the loss in aggregate TFP from the capital structure distortion $\tau_{ed_{it}}$ characterizing the financial productivity wedge—see expression (6).

There are two approaches to isolating the effect of $\tau_{ed_{it}}$ on aggregate TFP. One way to isolate this effect is to eliminate the financial productivity wedge in an environment without capital wedges $1 - TFP(\{\overline{A}_{it}, \tau_{AF_{it}}, \tau_{K_{it}} = 1\}_i)/TFP(\{\overline{A}_{it}, \tau_{AF_{it}} = 1, \tau_{K_{it}} = 1\}_i)$, which I call the inframarginal effect of the capital structure distortion.[3] The other way to measure this loss is to compare the TFP of an economy at its current state $\{\overline{A}_i, \tau_{AF_i}, \tau_{K_i}\}_i$ to an economy without capital structure distortions $\tau_{ed} = \tau_{AF_i} = 1$. The issue with the second approach is that changing τ_{ed} also changes the capital wedge—see equation (4). The capital wedge changes because $\tau_{ed} = 1$ implies that the tradeoff between reducing the cost of financing and increasing the capital structure productivity disappears, leading to lower cost of capital $\tau_{K_{it}}$. To isolate the capital structure effect, I keep $\tau_{K_{it}}$ constant by setting $r_{e_{it}} = R\tau_{K_i}$. This counterfactual can be calculated by $1 - TFP(\{\overline{A}_i, \tau_{AF_i}, \tau_{K_i}\}_i)/TFP(\{\overline{A}_i, \tau_{AF_i} = 1, \tau_{K_i}\}_i)$ which I call the compensated marginal effect (analogous to the compensated/Hicksian demand that isolates the substitution effect). These two counterfactual quantities should be close but not identical. In the special case where three variables are distributed jointly lognormal, the inframarginal and compensated marginal loss are identical, and the TFP loss from the capital structure wedge takes the following approximate expression.

$$\text{inframarginal TFP loss } \approx \mathbf{E}(\log \tau_{AF}) + (\eta - 1)\sigma_{A,F} - \frac{\eta - 1}{2}\sigma_F^2, \tag{8}$$

where $\sigma_{A,F}$ is the covariance between a demeaned $\log \overline{A}$ and a demeaned $\log \tau_{AF}$ and σ_F^2 is the variance of demeaned $\log \tau_{AF}$. While expression (8) is not general, it helps build intuition about how the capital structure distortion affects aggregate TFP: the average capital structure distortion $\mathbf{E}(\log \tau_{AF})$ decreases TFP, and the covariance between the financial productivity wedge and productivity $\sigma_{A,F}$, increases loss because high-productivity firms face higher distortions. A high variance of the capital structure distortion reduces

[3] I borrow the terminology marginal and inframarginal effect from Asker et al. (2022) who also study an environment with multiple types of frictions in the oil-producing industry.

TFP loss because the effect of capital structure distortions on output can be dampened by reallocating resources from the firms facing high capital structure distortions to the ones facing lower distortions—a general equilibrium channel.

All counterfactual quantities of interest in this paper are functions of the ratio of TFP under different distributions of frictions, where the distribution of productivity $\overline{A}$ is kept constant. Note that the ratio of two TFPs from two economies with the same productivity distribution is scale-invariant with respect to $\overline{A}_{it}$. The scale invariance with respect to $\overline{A}_{it}$ and $\tau_{K_{it}}$ (discussed earlier) implies that the counterfactual quantities are independent of the scale or the unit of measurement of $\overline{A}_{it}$ or $\tau_{K_{it}}$. Unit invariance facilitates the analysis of data coming from multiple years since demeaning the logarithm of $\overline{A}_{it}$ or $\tau_{K_{it}}$ in each year makes the observations comparable across time. In addition, since prices do not appear in the TFP formula, any aggregate shocks to capital, labor, or technology that affect the scale of firm-specific productivity do not affect the results of the counterfactual analysis (see also Fakos, 2023 and the references therein). Moreover, the financial productivity wedge depends only on capital structure d_{it}, which, as the debt share, is unit-free and is also comparable across time.

The financial productivity wedge and the Modigliani-Miller benchmark. In the counterfactuals above, eliminating the financial productivity wedge implies setting the price of equity equal to the price of debt for each firm $r_{e_{it}} = r_{d_{it}}$. This price equality is because investors are indifferent between receiving their return through debt or equity due to the fungibility of money. Therefore, the marginal rate of transforming a dollar of debt to a dollar of equity in this environment is $\frac{r_{e_{it}}}{r_{d_{it}}} = 1$. An efficient allocation should reflect this marginal rate of transformation. This efficient, frictionless benchmark is proposed in the seminal contribution of Modigliani and Miller (1958). That paper argues that the cost of capital is independent of the capital structure. In my model, the average and marginal cost of capital is $r_{d_{it}} d_{it} - r_{e_{it}}(1 - d_{it})$. For it to be independent of capital structure, it must be that it is the same for any d_{it}, d'_{it} with $d_{it} \neq d'_{it}$. $r_{d_{it}} d_{it} - r_{e_{it}}(1 - d_{it}) = r_{d_{it}} d'_{it} - r_{e_{it}}(1 - d'_{it})$ can only hold for $d_{it} \neq d'_{it}$ if and only if $r_{e_{it}} = r_{d_{it}}$.

Loosely speaking, the Modigliani-Miller argument implies that the price of debt equals the price of equity. Still, it leaves the quantity of debt and equity undetermined. The argument involves two steps. The first step is that investors only care about the return on their asset, not the type of asset because money is fungible. In other words, they care about the claims to the firm's cash flow and are indifferent to how this cash flow is split between debt and equity. Therefore, debt and equity are supplied at the same price; that is, their relative price is 1. Whether that relative price prevails in equilibrium depends on the market structure. The second step of the argument is that if markets are competitive, prices are determined by supply (opportunity cost), not by demand. Therefore, arbitrage pricing implies that the cost of capital doesn't depend on capital structure, which in this model is equivalent to $r_{e_{it}} = r_{d_{it}}$ (see also Ross, 1988). As Merton Miller explains in Miller (1988), the Modigliani and Miller (1958) argument involves a Fisherian firm, a black box characterized by a cash flow process, further implying that capital structure is indeterminate.

Theories of the firm and optimal capital structure. Suppose a firm is more than a cash flow process. In that case, the Modigliani-Miller result that the price of debt equals the price of equity in competitive equilibrium still holds, but other considerations determine the capital structure. In their seminal paper, Jensen and Meckling (1976) argue that capital structure is chosen to solve agency problems. Hart (1988), Aghion and Bolton (1992), and Hart (1993) provide the theoretical foundations for the debt and equity tradeoffs in environments with a conflict of interest between investors and managers where contracts are incomplete. Loosely speaking, in such environments, one of the parties can cheat on the original agreement, but a court cannot verify the cheating despite it being common knowledge by both parties, so the cheater cannot be held accountable, which leads to inefficiencies. Grossman and Hart (1986) solve this problem by optimally allocating residual control rights. But in most corporate settings, this solution is infeasible because no single party is wealthy enough to own the firm. The following example from Hart (1995) is relevant for the debt and equity tradeoff in public firms, which is the focus of this paper.

Suppose managers don't care only about their monetary compensation but derive private benefits from managing the firm and keeping their cronies employed, even if that decreases firm value. In this case, to implement an efficient allocation, investors should have control. However, this is impossible in large public companies because each investor is not wealthy enough to have a large enough share of the company's equity to exercise control. A solution to this problem is debt issuance that promises interest payments and default if payments cannot be met. If managers reduce the company's value too much, interest payments cannot be delivered, which triggers bankruptcy, and managers lose their private benefits forever. In sum, debt disciplines managers. Still, debt cannot be too high because it may trigger inefficient bankruptcy during a temporary adverse shock, so there is an optimal capital structure. Therefore, too little debt implies low productivity, as the company leadership can mismanage the company to gain private benefits. Too much debt leads to a high likelihood of bankruptcy that distorts the acquisition of talent through fewer and lower quality applicants (Brown and Matsa, 2016), which can lead to a mismatch between firms and workers, leading to lower labor productivity. A high likelihood of bankruptcy can also lower the value of a durable product as part of its value comes from the ability of the firm to service it in the future (Hortaçsu et al., 2013). In sum, there is an optimal capital structure, and if it is suboptimal, it leads to lower productivity.

The capital structure component $g(d) = d^{\gamma}[1-d]^{1-\gamma}$ in the production function (1) captures this tradeoff in a reduced form fashion. Essentially, the function g summarizes the contracting frictions that require capital structure to deal with conflicts of interest between the firm's stakeholders, implying that there is an optimal capital structure at $d = \gamma$ that attenuates such conflicts, leading to more output. Additional frictions in the asset markets manifest as the relative price of debt being different than that of equity $r_e/r_d \neq 1$, which prevents the capital structure from being at its optimal level, reducing output. This paper quantifies the effect of asset market frictions on aggregate output given existing contracting frictions that characterize g.

Comparison with the model of Whited and Zhao. The seminal paper of Whited and Zhao (2021) presents a model of production in which capital structure affects output directly, like in my model. The models are different in several dimensions.

One difference between the two models is the functional form of the capital structure component of value-added output. The capital structure component in their paper has a CES functional form, while in this paper, it has the Cobb-Douglas form of function g. [4] The advantage of the CES functional form is that as the elasticity of substitution tends to infinity, capital structure doesn't have real effects (capital structure irrelevance). Since their paper was the first to put capital structure in a production model, their model needed to nest the capital structure irrelevance case so that they could test the irrelevance hypothesis in the data. Using data for US and Chinese firms, they reject the irrelevance hypothesis and estimate a CES elasticity of substitution between debt and equity around 1.5. The functional form g in my model is equivalent to a CES with an elasticity of substitution of 1, which is not far from the 1.5 estimate of Whited and Zhao (2021). The advantage of g is analytical tractability that allows for simple formulas for crucial model quantities, like (7)

[4] In Whited and Zhao (2021) output takes the form $A_{it}\left[\gamma D_{it}^{\frac{\sigma-1}{\sigma}} + (1-\gamma)E_{it}^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$. Factoring out $K_{it} = D_{it} + E_{it}$ and using my definition of $d_{it} = \frac{D_{it}}{K_{it}}$, output can equivalently be written as $A_{it}\left[\gamma d_{it}^{\frac{\sigma-1}{\sigma}} + (1-\gamma)(1-d_{it})^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}} K_{it}$. The limit of this function as the elasticity of substitution σ goes to 1 is the Cobb-Douglas function $A_{it}d_{it}^{\gamma}[1-d_{it}]^{1-\gamma}K_{it}$.

or (8), that help us gain intuition about the model mechanics.

Another difference between the two models is the inclusion of real factors in the production function. My model includes capital and labor, while their model only includes capital. Therefore, the two models coincide when my model has no labor, the capital share α is one, and their capital structure elasticity of substitution is one. One advantage of having two factors of production is that the estimates of the effect of capital market frictions on aggregate TFP are conservative since capital is only one factor of production. Another advantage is that firm-level productivity is total factor productivity, as it is calculated using both factors of production. Calculating firm-level TFP is possible in this paper because I use data on European firms that report labor expenditure, while they use data for US public firms, most of which do not report the wage bill. Moreover, firm-level TFP includes labor productivity, which can be a crucial channel through which suboptimal capital structure has real effects (see the discussion above about the theory of the firm).

3 Data

Data on European public firms in 1989–2022 come from the Compustat Global dataset. The advantage of data on European firms over data on US firms is that they contain information on the expenditure on labor that is necessary for recovering total factor productivity at the firm level and using a production function with capital and labor, like the one used in this paper. For the analysis I use the same four European countries as David et al. (2023), France, Germany, Sweden, and the UK, which have enough observations with non-missing data.

Dataset construction. I measure labor expenditure $w_t L_{it}$ by the variable `XLR`. As in Hsieh and Klenow (2009), I assume that workers are paid their marginal product, and hence, the firm's labor expenditure reflects the amount of labor in efficiency units, which allows for comparing productivity across firms with different levels of human capital per worker. Therefore, for the quantitative analysis, within a period, we can normalize $w_t = 1$ so that $w_t L_{it} = L_{it}$. Yearly normalizations with respect to prices, firms-specific productivity, or capital distortions do not affect the results since all counterfactual quantities are scale-invariant with respect to those quantities (see the discussion in Section 2.1).

I construct nominal value-added output $p_{it}Y_{it}$as the sum of operating profit before taxes, depreciation or interest (OIBDP) plus the labor expenditure (XLR) plus the change in inventory of finished goods ($\texttt{INVFG}_{it}$-$\texttt{INVFG}_{it-1}$) as in David et al. (2023). I measure the capital stock K_{it} by the balance sheet variable property plant and equipment (PPENT) as in David et al. (2016). Value added, capital, and labor are necessary for constructing the production function residual $\widetilde{A}_{it}$. I measure debt by total liabilities (LT) and equity by the difference (AT-LT) as in Whited and Zhao (2021). Therefore, the capital structure measure d_{it} is debt over debt plus equity, which is the ratio LT/AT.

To arrive at the dataset used in the analysis, I first eliminate observations with missing or non-positive values of capital, labor, value-added output, assets, or liabilities. I also eliminate observations with a debt share outside the $(0, 1)$ interval. I also drop firms in the finance/insurance or public administration sectors (NAICS codes 52 and 92). To reduce the impact of outlier observations on the results I trim the dataset by eliminating observations lying in the top or bottom 1% of the distribution of $MRPK_{it} = p_{it}Y_{it}/K_{it}$ or $MRPL_{it} = p_{it}Y_{it}/K_{it}$ in each country/year which leads to the deletion of 4.2% of the observations. I trim the $MRPK$ variable because it is proportional to the capital wedge $\tau_{K_{it}}$, a component of the aggregate TFP formula. I trim the $MRPL$ variable to eliminate observations with outlier labor share (which is the inverse of MRPL), which is the variable used to estimate the parameter α.

The dataset used in the analysis contains 43,389 observations from 4,933 firms. The mean and median debt share are 55.5% and 56.7%, respectively. The 25th and 75th percentile are 42.4% and 69.7%, respectively, indicating substantial variation.

4 Estimation

To estimate the capital elasticity α in the production function, I use the first-order optimality conditions for profit maximization that imply that the labor expenditure as a share of value added equals the labor elasticity $1-\alpha$. I estimate the value $1-\alpha$ as the median of the labor share $\frac{w_{it}L_{it}}{p_{it}Y_{it}}$ in the sample like Asker et al. (2014). Then α is one minus the median of the labor share and is reported in Table 1. The parameter α has a value of 0.342 and is estimated precisely, reflected in a small standard error. Its estimated value is close to a value of 1/3 typically used in macro models of misallocation (see David et al., 2016; Gopinath et al., 2017). I use the median to estimate the labor elasticity instead of the mean to avoid outliers affecting the estimate. More specifically, approximately 10% of the sample's observations have a labor share higher than 1, which happens when profit is negative, and the inventory change is insufficient to keep value-added output greater or equal to the labor expenditure. This static production model cannot accommodate negative profit, and considering such observations by using the mean as an estimator would overestimate the labor elasticity (mean labor share is 0.71). The median is not affected by such outliers and is the better estimator in this context.

I estimate the capital structure parameter γ using the fixed-effects estimator of Whited and Zhao (2021). The starting point is the relationship between capital structure, productivity, and the production function residual $\widetilde{A}_{it}$ after taking logs: $\log \widetilde{A}_{it} = \alpha\gamma \log d_{it} + \alpha(1-\gamma)\log(1-d_{it}) + \log A_{it}$. In that equation, d_{it} is observable, A_{it} is unobservable, and $\widetilde{A}_{it}$ can be recovered from the data. I calculate the residual in the production function as in Hsieh and Klenow (2009): $\widetilde{A}_{it} = \frac{(p_{it}Y_{it})^{\frac{\eta}{\eta-1}}}{K_{it}^{\alpha}L_{it}^{1-\alpha}}$, where $p_{it}Y_{it}$ is the observed nominal

Table 1: Model parameters.

	Value	S.E.
Capital elasticity α in production	0.342	0.0016
Debt exponent γ in capital structure	0.589	0.0196
Elasticity of substitution in demand η	3	
Number of observations	43,389	

This table presents the model parameters from the entire sample of public firms in France, Germany, Sweden, and the UK. The elasticity of substitution η in the CES demand is preset to 3 as in Hsieh and Klenow (2009). Standard errors are heteroskedasticity robust, incorporate variation from all estimation steps, and are calculated using the influence function approach described in Section 5.1.

value-added output, which, exponentiated in $\frac{\eta}{\eta-1}$ is the value-added quantity Y_{it}. To recover the production function residual, I use the estimated capital elasticity α and a demand elasticity $\eta = 3$ as in Hsieh and Klenow (2009) and Gopinath et al. (2017). The demand elasticity is challenging to estimate without data on quantities and prices and a static model, and this is the reason many studies (Hsieh and Klenow, 2009; David et al., 2016; Gopinath et al., 2017) set it to a specific number, lying within bounds reported in the industrial organization literature. I use a fixed-effects model to estimate the coefficients of $\log d_{it}$ and $\log(1 - d_{it})$. I demean $\widetilde{A}_{it}$ at the yearly level before using it in the fixed-effect estimator to remove aggregate nominal shocks from the productivity measure. I recover γ from the two coefficients by dividing them to get $(1-\gamma)/\gamma$. The identifying assumption behind the fixed effects estimator is that shocks to a firm's productivity are orthogonal to shocks to the friction $\tau_{ed_{it}}$ that drives the capital structure d_{it}. This identifying assumption is reasonable, as frictions originate in the asset market while productivity shocks are technology shocks.

The estimated value of γ is 0.589, with the parameter estimated precisely, reflected in a small standard error (Table 1). The estimated γ value implies that the optimal capital structure is a debt share of 58.9%, which is close to the median debt share in the data. To gain intuition about the implications of the estimated γ I calculate the productivity loss from suboptimal capital structure $1 - 1/\tau_{AF}$ at various points on the capital structure distribution. More specifically, the productivity loss of a firm with the average (median) capital structure is 0.08% (0.04%), a negligible loss. In contrast, the productivity loss of a firm with a debt share in the 25th (75th) percentile of the debt share distribution is 1.91% (0.9%), which is substantial given that it translates to an equivalent loss in the firm's output. Aggregating these firm-level losses to an economy-wide loss in output requires calculating the aggregate TFP loss, which I do in Section 5.

Model fit. The estimated coefficients in the productivity capital structure equation used to estimate γ can also be used to evaluate the fit of the functional form of the capital structure component F_{it} in the production function. More specifically, the model implies that the coefficient on $\log d_{it}$ is the product of two parameters $\alpha\gamma$. So, given the estimate of γ, we can recover the value for α implied by the fixed effects regression, which is 0.377, with a standard error of 0.0356. This value is close to and statistically indistinguishable from the estimated value of α from the labor share, indicating that this parsimonious model that restricts the exponent of the capital structure to be the same as the capital share fits the data well, lending credence to the quantitative results of this paper.

5 Results and inference

This section quantifies the aggregate effects of the two types of frictions and develops a framework to derive their standard errors using the influence function approach to inference. Tables 2 and 3 present the results.

To quantify the importance of frictions, I estimate how changing the distribution of frictions affects aggregate TFP by comparing the TFP of pairs of economies with different distributions of frictions: $\{\overline{A}_{it}, \tau_{AF_{it}}, \tau_{K_{it}}\}_{it}$ and $\{\overline{A}_{it}, \tau'_{AF_{it}}, \tau'_{K_{it}}\}_{it}$ keeping each firm's productivity frontier $\overline{A}_{it}$ constant. Therefore, all results are a function of a ratio of two aggregate TFPs: TFP/TFP'. I calculate the productivity frontier using the formula $\overline{A}_{it} = \widetilde{A}_{it}\tau_{AF_{it}}$ and applying equation (6) at the estimated parameters. In addition, I demean $\overline{A}$ and $\tau_{K_{it}}$ at the yearly level to avoid nominal effects having an impact on the across-year comparison of the TFP loss. The financial productivity wedge $\tau_{AF_{it}}$ does not require yearly demeaning because it is a share and, hence, is unit-free.

Table 2: TFP loss from distortions

Type of distortion considered	TFP loss in % Value	S.E.
(1) All distortions (total): $\{\tau_{AF_{it}}, \tau_{K_{it}}\}_{it} \to \{1,1\}_{it}$	26.2	1.39
(2) Capital structure (inframarginal): $\{\tau_{AF_{it}}, 1\}_{it} \to \{1,1\}_{it}$	2.01	0.24
(3) Capital structure (compensated marginal): $\{\tau_{AF_{it}}, \tau_{K_{it}}\}_{it} \to \{1, \tau_{K_{it}}\}_{it}$	2.11	0.32

This table presents results from three counterfactual exercises using data from public firms in France, Germany, Sweden, and the UK. The TFP loss is calculated as $1 - TFP/TFP'$, where TFP' reflects an economy with less distortions and is expressed in percent. Each row of the table represents counterfactuals involving eliminating different types of distortions. Standard errors are heteroskedasticity robust, incorporate variation from all estimation steps, and are calculated using the influence function approach described in Section 5.1.

Table 2 presents the results from three counterfactual exercises. Row (1) shows the loss in TFP from both types of frictions by comparing the TFP in the data using the distribution $\{\tau_{AF_{it}}, \tau_{K_{it}}\}_{it}$ to the TFP in the absence of distortions TFP'. The loss is then $1 - TFP/TFP'$, is expressed in percent, and is estimated at 26%, implying that aggregate output could be 1/3 higher each year with the same aggregate resources (capital and labor) if distortions disappeared. This estimate is in line with estimates from datasets from other countries (see Fakos, 2023 and the references therein). The loss is precisely estimated, reflected in the small standard error (Table 2).

The last two columns of Table 2 report the quantitative importance of the capital structure distortion, reflected in the financial productivity wedge $\tau_{AF_{it}}$, on aggregate productivity. Column (2) measures how aggregate productivity would change if we eliminated the financial productivity wedge in an economy with no other frictions (inframarginal effect). The estimated TFP loss is 2% with a small standard error, implying that it is statistically different from zero. This magnitude is large compared to, for example, the cost of business cycles, which is estimated to be below 1% (see Alvarez and Jermann, 2004 and the references therein). Column (3) measures how aggregate productivity would change if we eliminated the financial productivity wedge while keeping the capital distortions $\tau_{K_{it}}$ at their existing level (compensated marginal effect). The compensated marginal effect is nearly identical to the inframarginal effect. This reassuring since the compensated marginal effect is designed to isolate the capital structure distortion and should be close to the inframarginal effect, lending credence the results of the analysis.

Table 3: The importance of the capital structure distortions

	Share of total loss in %				
	Inframarginal		Compens. marginal		
Market definition	Value	S.E.	Value	S.E.	Obs.
(1) France, Germany, Sweden, and UK (baseline)	7.67	1.01	8.06	1.30	43,389
(2) France and Germany	12.51	1.99	12.11	2.11	19,930
(3) France	13.43	2.49	13.25	2.57	8,939
(4) Germany	15.22	4.06	14.69	4.34	10,991
(5) Sweden	8.85	4.18	9.65	4.55	5,438
(6) UK	7.51	1.22	7.59	2.01	18,021

This table decomposes the total TFP loss into the effects of the financial productivity wedge τ_{AF} (from the capital structure distortion) and the capital wedge τ_K on TFP loss. The inframarginal component is the ratio of the quantities in rows (2) and (1) from Table 2, while the compensated marginal component is the ratio of row (3) to row (1). The first row of this table reports results from the sample used throughout this paper, which includes public firms from France, Germany, Sweden, and the UK. The subsequent rows report the decomposition from re-estimating the model from each subsample. Standard errors are heteroskedasticity robust, incorporate variation from all estimation steps, and are calculated using the influence function approach described in Section 5.1.

TFP loss decompositions. Table 3 presents results on the relative importance of the capital structure distortions among all distortions by decomposing the total TFP loss into the loss from the financial productivity wedge and the loss from the capital wedge. There are two ways to perform this decomposition. One way is first to eliminate the capital wedge τ_K and then eliminate the financial productivity wedge τ_{AF}, the last step of which measures the inframarginal TFP loss, which is 7.67% of the total loss (the ratio of column (2) to column (1) from Table 2). The other way is first to eliminate the financial productivity wedge and then the capital wedge, the first step of which measures the compensated marginal TFP loss, which is 8.06% of the total loss (the ratio of column (3) to column (1) from Table 2). Given that the marginal effect is compensated (changing r_e to keep the capital wedge constant), these two decompositions are expected to give similar results. In this dataset, the two decompositions give statistically identical results, lending credence to the analysis.

To explore the robustness of the results, I re-estimate the model parameters and distortion distributions in several subsamples and report the decomposition in the remaining rows of Table 3. More specifically, I re-estimated the model for each country separately and the France-Germany sample since both countries are geographically close, part of the EU and the Euro monetary union. All samples exhibit two robust patterns: capital structure distortions are responsible for approximately 1/10 of the overall TFP loss, and the marginal effect is indistinguishable from the inframarginal effect.

The share of TFP loss emanating from the capital structure wedge provides a lower bound on the effect of financial frictions on aggregate productivity. Financial frictions drive a wedge between the price of debt and equity, leading to the financial productivity wedge τ_{AF}. Financial frictions are also reflected in the capital wedge τ_K (Gopinath et al., 2017) along with other frictions such as information frictions (David et al., 2016) or resale market frictions (Chen et al., 2023). While decomposing the loss from τ_K into financial and other sources is challenging, the capital structure wedge τ_{AF} is a pure financial wedge. The finding that the capital structure wedge is responsible for approximately 1/10 of the TFP loss sets a lower bound on the importance of financial frictions, which is substantial, providing evidence that financial frictions have significant real effects.

5.1 Inference

This section describes the influence function approach to inference used to calculate the standard errors in Tables 1–3. The online appendix contains a more detailed exposition of the methodology, formula derivations, and additional references.

To obtain the standard errors and confidence intervals of the counterfactual quantities in Tables 2 and 3, I need to derive the asymptotic distribution of the ratio of two aggregate TFPs, as all results depend on such a ratio. Since the ratio of two TFPs is not a well-studied statistic, the first step is to verify that it can be expressed as a statistical functional so that standard asymptotic inference can be applied. Given a sample $X_1, \ldots, X_n$, a statistical functional is a map from an empirical cumulative distribution function $\mathbb{F}_n$ (or the empirical probability measure $\mathbb{P}_n$) to the real numbers, implying that the functional should not depend directly on the sample size n but only on the distribution of the data (Fernholz, 1983). For instance, the TFP formula in equation (7) cannot be expressed as a statistical functional because TFP increases as the number of firms increases. To see this, note that if we duplicate every observation, the TFP is larger by a factor of $2^{\frac{1}{\eta-1}}$. As a result, as the sample size grows ($n \to \infty$), the TFP does not converge but goes to infinity, and hence, the asymptotic statistical theory is not applicable.

The ratio of two TFPs, in contrast, can be written as a statistical functional and asymptotic theory applies. To see this, note that the ratio of two TFPs can be expressed only in terms of averages:

$$S_n = \frac{\left[\sum_{j=1}^{n} \frac{1}{n} \left(\frac{\overline{A}_j}{\tau_{AF_j}\tau_{K_j}^{\widehat{\alpha}}}\right)^{\eta-1}\right]^{\frac{1}{\eta-1}+\widehat{\alpha}} \Bigg/ \left(\sum_{j=1}^{n} \frac{1}{n} \frac{\left(\frac{\overline{A}_j}{\tau_{AF_j}}\right)^{\eta-1}}{\tau_{K_j}^{1+\widehat{\alpha}(\eta-1)}}\right)^{\widehat{\alpha}}}{\left[\sum_{j=1}^{n} \frac{1}{n} \left(\frac{\overline{A}_j}{\tau'_{AF_j}\tau'^{\widehat{\alpha}}_{K_j}}\right)^{\eta-1}\right]^{\frac{1}{\eta-1}+\widehat{\alpha}} \Bigg/ \left(\sum_{j=1}^{n} \frac{1}{n} \frac{\left(\frac{\overline{A}_j}{\tau'_{AF_j}}\right)^{\eta-1}}{\tau'^{1+\widehat{\alpha}(\eta-1)}_{K_j}}\right)^{\widehat{\alpha}}}. \tag{9}$$

Inference on estimator S_n is challenging as it is a function of data and estimated parameters $\widehat{\alpha}, \widehat{\gamma}$, and is not a statistic with well-known asymptotic properties. If estimator S_n depended only on estimated parameters, as in studies where counterfactuals use simulated data like Fakos (2023), then the standard error of S_n could be calculated using the delta method. I conduct inference on S_n using the influence function approach, which is a general framework for conducting inference as it can be applied to settings where the statistic of interest depends on parameters estimated using any combination of the GMM, quantile, Maximum Likelihood (MLE), or Simulated Minimum Distance (SMD) frameworks because they are asymptotically linear estimators (Newey and McFadden, 1994; Van der Vaart, 1998; Tsiatis, 2006).

The von Mises expansion. An asymptotically linear estimator ϕ can be analyzed by studying the asymptotic properties of the first term of its von Mises expansion:

$$\sqrt{n}[\phi(\mathbb{P}_n) - \phi(\mathbf{P})] \approx \sqrt{n}\phi'_{\mathbf{P}}(\mathbb{P}_n - \mathbf{P}). \tag{10}$$

Where $\mathbb{P}_n$ denotes the empirical probability measure, $\mathbf{P}$ the true probability measure, and $\phi'_{\mathbf{P}}(\mathbb{P}_n - \mathbf{P})$ the Hadamard derivative at $\mathbf{P}$ in the direction $\mathbb{P}_n - \mathbf{P}$ (see Van der Vaart, 1998).

The influence function. The asymptotic limit ($n \to \infty$) of derivative $\phi'_{\mathbf{P}}$ can be expressed as the expectation $\int \psi(x) d\mathbf{P}(x)$ of random variable $\psi(x)$ that is called the influence function of estimator ϕ. Therefore, we can study the asymptotic properties of estimator ϕ by analyzing the sample average of the influence function $\frac{1}{n}\sum_{i=1}^{n} \psi(x_i)$. For example, suppose the estimated parameter is a $\phi_{K\times 1}$ vector. In that case, the standard errors of the estimates can be calculated from the influence function: $\sqrt{\frac{1}{n^2}\text{diag}(\sum_{i=1}^{n} \psi(x_i)\psi'(x_i))}$. Formulas for the influence function of the GMM, MLE, SMD, quantiles, and the mean are available in, e.g., Newey and McFadden (1994); Van der Vaart (1998); Tsiatis (2006). The influence function of the mean is $x - \int x d\mathbf{P}(x)$.

The influence function of multi-step estimators and the chain rule. Estimator S_n is a function of the first stage median $\widehat{\alpha}$ and GMM coefficients characterizing $\hat{\gamma}$, and several means, which I denote μ. To derive the influence function of S_n, I first must establish that it is Hadamard differentiable. The median, means, and GMM estimators are Hadamard differentiable, and $\hat{\gamma}$ is a continuous differentiable function of the GMM coefficients, rendering it Hadamard differentiable. This is because continuous partial differentiability implies Hadamard differentiability in finite dimensions. S_n is continuously differentiable with respect to $\widehat{\alpha}, \hat{\gamma}$ and the several means in formula (9) so it is Hadamard differentiable.

Express the estimator S_n as $H(\mu(\mathbf{P}), \alpha(\mathbf{P}), \gamma(\mathbf{P}, \alpha(\mathbf{P})))$. It depends on the probability measure $\mathbf{P}$ directly through the various means and indirectly through $\widehat{\alpha}, \hat{\gamma}$.[5] To derive the influence function of H, I use the chain rule for Hadamard differentiable functions, which extends to influence functions. Therefore the influence function $\psi_S(x)$ of S is

$$\psi_S(x) = \frac{\partial H}{\partial \mu}\psi_\mu(x) + \frac{\partial H}{\partial \alpha}\psi_\alpha(x) + \frac{\partial H}{\partial \gamma}\left[\psi_\gamma(x) + \frac{\partial \gamma}{\partial \alpha}\psi_\alpha(x)\right] \tag{11}$$

[5] The estimator S_n also depends on the estimated yearly means of τ_{AF}, τ_K used to demean these variables at the annual level. I suppress this dependence here for the sake of brevity, but I take it into account when conducting inference. The online appendix contains details about inference on yearly means.

The influence function at each x_i is straightforward to obtain since there are formulas for $\psi_\mu(x)$, $\psi_\alpha(x)$, and the GMM coefficients, while the partial derivatives admit a numerical approximation.

Approximating the influence function of estimators that are functions of means. Deriving the influence function $\frac{\partial H}{\partial \mu}\psi_\mu(x)$ of a new statistic analytically can be tedious and prone to error. Fortunately, the influence function of any statistic $h(\mathbb{P}_n)$ that is a continuous differentiable function of means can be numerically approximated by perturbing the empirical distribution $\mathbb{P}_n$ as follows

$$\psi(x_i) = \lim_{\epsilon \downarrow 0} \frac{h([1-\epsilon]\mathbb{P}_n + \epsilon\delta_{x_i}) - h(\mathbb{P}_n)}{\epsilon}, \quad \epsilon \in (0,1), \tag{12}$$

where δ_{x_i} is the probability measure with all mass at x_i. I demonstrate in the online appendix that the numerical approximation for spacing $\epsilon = 10^{-7}$ or less results in an influence function nearly identical for all practical purposes to the one derived analytically, leading to identical standard errors. Estimators that are smooth functions of averages include misallocation statistics like S_n, moments, variance, standard deviation, and the correlation between two random variables.

Advantages of the influence function approach to inference. The influence function approach to multi-step estimators has several advantages compared to the resampling alternative (subsampling or bootstrap).[6] It can be computationally cheaper when some estimation step involves a computationally challenging estimator like a simulated method of moments of a dynamic model as in Bazdresch et al. (2018). In addition, the numerical approximation described above allows for a hybrid approach: use analytical formulas for

[6]The online appendix demonstrates that the standard errors in Table 2 line up with the bootstrapped ones.

computationally intensive steps and the numerical approach for analytically complex but computationally trivial statistics such as S_n. Moreover, the influence function approach is a deterministic function of a sample that facilitates replication of the results compared to a resampling method that introduces resampling error.

6 Conclusions

I develop a general equilibrium model with heterogeneous firms and market frictions that distort each firm's capital structure (debt–equity mix) and the allocation of capital and labor across firms, reducing aggregate productivity. Suboptimal capital structure manifests as lower firm productivity, is reflected in a firm-specific financial productivity wedge that prevents the firm from attaining its productivity frontier, and is driven by the cost of debt being different from that of equity due to financial frictions. Lower firm-specific productivity leads to lower aggregate productivity. Aggregate productivity is also reduced by the misallocation of capital and labor across firms driven by a firm-specific wedge on the cost of capital, reflecting financial or other market frictions.

I use data on European public firms to quantify the effect of the capital structure distortion and the capital wedge on aggregate productivity. I find that the capital structure wedge leads to a permanent TFP loss of 2%, which implies an equivalent loss in yearly aggregate output, which is substantial. The capital structure distortion is responsible for roughly 1/10th of the overall TFP loss from all distortions. The TFP loss from the capital structure distortion is a lower bound on the overall TFP loss from financial frictions, which are solely responsible for the capital structure distortion but can also affect the capital wedge. Identifying policies that alleviate or deteriorate the capital structure distortion is an avenue for future research.

References

Aghion, P., A. Bergeaud, T. Boppart, P. Klenow, and H. Li (2023). Good rents versus bad rents: R&D misallocation and growth. Technical report, Working Paper, Stanford.

Aghion, P. and P. Bolton (1992). An incomplete contracts approach to financial contracting. *The Review of Economic Studies 3*(3), 473–494.

Alvarez, F. and U. J. Jermann (2004). Using asset prices to measure the cost of business cycles. *Journal of Political economy 112*(6), 1223–1256.

Asker, J., A. Collard-Wexler, and J. De Loecker (2014). Dynamic inputs and resource (mis)allocation. *Journal of Political Economy 122*(5), 1013–1063.

Asker, J., A. Collard-Wexler, and J. De Loecker (2022). The Welfare Impact of Market Power: The OPEC Cartel. Presentation at OWA available on Youtube at https://youtu.be/zuwm7rBCvew?si=mhqq9bl8UxIvEJP8.

Bazdresch, S., R. J. Kahn, and T. M. Whited (2018). Estimating and testing dynamic corporate finance models. *The Review of Financial Studies 31*(1), 322–361.

Brown, J. and D. A. Matsa (2016). Boarding a sinking ship? an investigation of job applications to distressed firms. *The Journal of Finance 71*(2), 507–550.

Cattaneo, M. D., R. K. Crump, and M. Jansson (2013). Generalized jackknife estimators of weighted average derivatives. *Journal of the American Statistical Association 108*(504), 1243–1256.

Chen, Z., X. Jiang, Z. Liu, J. C. S. Serrato, and D. Y. Xu (2023). Tax policy and lumpy investment behaviour: Evidence from China's VAT reform. *The Review of Economic Studies 90*(2), 634–674.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal 21*.

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

David, J. M., R. Ranciere, and D. Zeke (2023). International diversification, reallocation, and the labor share. Technical report, NBER working paper No. 31168.

Fakos, A. (2023). Subsidies with misallocation. *SSRN working paper No. 3483259*.

Fernholz, L. T. (1983). *von Mises Calculus for Statistical Functionals*, Volume 19. Springer-Verlag.

Gopinath, G., Ş. Kalemli-Özcan, L. Karabarbounis, and C. Villegas-Sanchez (2017). Capital allocation and productivity in south Europe. *The Quarterly Journal of Economics 132*(4), 1915–1967.

Grossman, S. J. and O. Hart (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy 94*(4), 691–719.

Hart, O. (1993). Theories of optimal capital structure: a managerial discretion perspective.

In M. Blair (Ed.), *The Deal Decade: What Takeovers and Leveraged Buyouts Mean for Corporate Governance*. Washington: Brookings Institution.

Hart, O. (1995). *Firms, contracts, and financial structure*. Clarendon Press.

Hart, O. D. (1988). Incomplete contracts and the theory of the firm. *The journal of Law, Economics, and Organization 4*(1), 119–139.

Hopenhayn, H. A. (2014). Firms, misallocation, and aggregate productivity: A review. *Annual Review of Economics 6*(1), 735–770.

Hortaçsu, A., G. Matvos, C. Syverson, and S. Venkataraman (2013). Indirect costs of financial distress in durable goods industries: The case of auto manufacturers. *The Review of Financial Studies 26*(5), 1248–1290.

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

Ichimura, H. and W. K. Newey (2022). The influence function of semiparametric estimators. *Quantitative Economics 13*(1), 29–61.

Jensen, M. C. and W. H. Meckling (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics 3*(4), 305–360.

Karabarbounis, M. and P. Macnamara (2021). Misallocation and financial frictions: The role of long-term financing. *Review of Economic Dynamics 40*, 44–63.

Kennedy, E. H. (2022). Semiparametric doubly robust targeted double machine learning: A review. *arXiv preprint arXiv:2203.06469*.

Miller, M. H. (1988). The Modigliani-Miller propositions after thirty years. *Journal of Economic perspectives 2*(4), 99–120.

Modigliani, F. and M. H. Miller (1958). The cost of capital, corporation finance and the theory of investment. *The American economic review 48*(3), 261–297.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engel and D. L. McFadden (Eds.), *Handbook of econometrics*, Volume 4, pp. 2111–2245. Elsevier.

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

Restuccia, D. and R. Rogerson (2017). The causes and costs of misallocation. *Journal of Economic Perspectives 31*(3), 151–74.

Ross, S. A. (1988). Comment on the Modigliani-Miller propositions. *Journal of Economic perspectives 2*(4), 127–133.

Tsiatis, A. A. (2006). *Semiparametric theory and missing data*. Springer New York.

Van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge university press.

Whited, T. M. and J. Zhao (2021). The misallocation of finance. *The Journal of Finance 76*(5), 2359–2407.

Internet Appendix for
Financial frictions, capital structure, and aggregate productivity

Alexandros Fakos[1]

This document is an internet appendix posted online but not intended for publication. Its sections roughly correspond to the paper's sections that are numbered in Arabic numerals.

A Model

Consider a sector at time populated by N firms. Each firm i faces productivity A_i and requires capital K_i and labor L_i to produce a differentiated product Y_i. Capital needs to be financed either by debt D_i or equity E_i, with $D_i + E_i = K_i$. Capital structure is characterized by the debt share $d_{it} = \frac{D_{it}}{K_{it}}$, which determines the equity share $\frac{E_{it}}{K_{it}} = 1 - d_{it}$. The production function is:

$$Y_i = A_i \left(D_i^{\gamma} E_i^{1-\gamma}\right)^{\alpha} L_i^{1-\alpha}, \alpha \in (0,1], \gamma \in (0,1), K_i = D_i + E_i.$$
$$Y_i = \overbrace{A_i \overbrace{\left(d_i^{\gamma}(1-d_i)_i^{1-\gamma}\right)^{\alpha}}^{F_i}}^{\widetilde{A}_i} K_i^{\alpha} L_i^{1-\alpha}, \alpha \in (0,1], \gamma \in (0,1), d_i = \frac{D_i}{K_i} \in [0,1]. \tag{A.1}$$

$F_i = \left[d_i^{\gamma}(1-d_i)_i^{1-\gamma}\right]^{\alpha}$ which represents the contribution of capital structure in the production function. The firm-level productivity residual $\widetilde{A}_i = \frac{Y_i}{K_i^{\alpha} L_i^{1-\alpha}}$ has a technological

[1]Fakos (alexfakos@icloud.com) is from the Department of Business Administration at ITAM.
Date: May 4, 2024.

efficiency component A_i and a financial component F_i.

Demand. A competitive firm produces the final good $\mathbf{Y} = \left[\sum_{i=1}^{N} Y_i^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}}$, and therefore, the firm's problem can be solved by cost minimization $\min_{\{Y_i\}_i} \sum p_i Y_i$ for a given level of output $\mathbf{Y} - \left[\sum_{i=1}^{N} Y_i^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}} = 0$. The Lagrangian function for this problem is $\mathfrak{L} = \sum p_i Y_i + \lambda\left(\mathbf{Y} - \left[\sum_{i=1}^{N} Y_i^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}}\right)$. The first order sufficient conditions (FOCs) for each i are $p_i = \lambda \mathbf{Y}^{\frac{1}{\eta}} Y_i^{-\frac{1}{\eta}}$. Solving the FOCs for Y_i we get $Y_i = p_i^{-\eta}\lambda^{\eta}\mathbf{Y}$. Substituting in the constraint we get $\lambda = [\sum p_i^{1-\eta}]^{\frac{1}{1-\eta}}$. Therefore, the demand function is $Y_i = p_i^{-\eta}[\sum p_i^{1-\eta}]^{\frac{\eta}{1-\eta}}\mathbf{Y}$. Since the firm is competitive, revenue equals to cost $\mathbf{PY} = \sum p_i Y_i$. Substituting for the demand we get $\mathbf{PY} = \sum p_i^{1-\eta}[\sum p_i^{1-\eta}]^{\frac{\eta}{1-\eta}}\mathbf{Y}$. By simplifying the equation we get $\mathbf{P} = [\sum p_i^{1-\eta}]^{\frac{1}{1-\eta}} = \lambda$. Since $\mathbf{Y}$ is the numeraire, we set $\mathbf{P}$ to 1. The demand for firm i's good comes from the final good producer and is $p_i = Y_i^{-\frac{1}{\eta}}\mathbf{Y}^{\frac{1}{\eta}}$.

Revenue function. The revenue is

$$
\begin{aligned}
p_i Y_i &= \overbrace{\Omega_i d_i^{\beta_d}(1-d_i)^{\beta_e}}^{\widetilde{\Omega}_i} K_i^{\beta_k} L_i^{\beta_l} \mathbf{Y}^{\frac{1}{\eta}} \\
\beta_k = \alpha(1-1/\eta), \beta_l &= (1-\alpha)(1-1/\eta), \Omega_i = A_i^{1-1/\eta}, \widetilde{\Omega}_i = \widetilde{A}_i^{1-1/\eta} \\
\beta_d = \alpha\gamma(1-1/\eta) &= \gamma\beta_k,\ \beta_e = \alpha(1-\gamma)(1-1/\eta) = (1-\gamma)\beta_k\,.
\end{aligned}
\tag{A.2}
$$

Revenue productivity $\widetilde{\Omega}_i = A^{1-1/\eta} d_i^{\beta_d}(1-d_i)^{\beta_e}$ has an invariant component $A^{1-1/\eta}$ and a distortion-dependent component $d_i^{\beta_d}(1-d_i)^{\beta_e}$.

Profit maximization. Firms maximize profit given the price of debt r_{d_i}, the price of equity r_{e_i}, and the labor wage w.

$$\max_{d_i,K_i,L_i} \Omega_i d_i^{\beta_d}(1-d_i)^{\beta_e} K_i^{\beta_k} L_i^{\beta_l} \mathbf{Y}^{\frac{1}{\eta}} - r_{d_i} d_i K_i - r_{e_i}(1-d_i)K_i - wL_i \tag{A.3}$$

$$D_i = d_i K_i, \quad E_i = (1-d_i)K_i, \quad d_i \in [0,1]$$

If $r_{e_i} = r_{d_i}$ then there are no frictions in the supply of financing. The first order optimality condition for capital implies that

$$MRPK_i = \beta_K \frac{p_i Y_i}{K_i} = r_{d_i} d_i + r_{e_i}(1-d_i). \tag{A.4}$$

The first order optimality condition with respect to the debt share d_i is

$$\begin{aligned}
MRPd_i = \beta_d \frac{p_i Y_i}{d_i} - \beta_e \frac{p_i Y_i}{1-d_i} &= r_{d_i} K_i - r_{e_i} K_i \\
\gamma \beta_k \frac{p_i Y_i}{d_i} - (1-\gamma)\beta_k \frac{p_i Y_i}{1-d_i} &= (r_{d_i} - r_{e_i})K_i \\
\beta_k p_i Y_i \frac{\gamma(1-d_i) - (1-\gamma)d_i}{d_i(1-d_i)} &= (r_{d_i} - r_{e_i})K_i \\
\beta_k \frac{p_i Y_i}{K_i} \frac{\gamma - d_i}{d_i(1-d_i)} &= r_{d_i} - r_{e_i}
\end{aligned} \tag{A.5}$$

The profit-maximizing debt share d_i. Combining the first order conditions with respect to capital and the debt share we have

$$\begin{aligned}
MRPK_i \frac{\gamma - d_i}{d_i(1-d_i)} &= r_{d_i} - r_{e_i} \\
[r_{d_i} d_i + r_{e_i}(1-d_i)] \frac{\gamma - d_i}{d_i(1-d_i)} &= r_{d_i} - r_{e_i}
\end{aligned} \tag{A.6}$$

Equation (A.6) is only in terms of d_i which lives in $(0,1)$ as any of the two extrema of the interval will give zero profit which is never optimal. However, $\gamma - d_i$ can be zero, and so can $r_{d_i} - r_{e_i}$. Now solve equation (A.6) for d_i.

$$\begin{aligned}
[r_{d_i} d_i + r_{e_i}(1-d_i)]\frac{\gamma - d_i}{d_i(1-d_i)} - r_{d_i} - r_{e_i} &= 0 \\
\frac{[(r_{d_i} - r_{e_i})d_i + r_{e_i}](\gamma - d_i) - (r_{d_i} - r_{e_i})d_i(1-d_i)}{d_i(1-d_i)} &= 0 \Rightarrow \\
[(r_{d_i} - r_{e_i})d_i + r_{e_i}](\gamma - d_i) - (r_{d_i} - r_{e_i})d_i(1-d_i) &= 0 \\
(r_{d_i} - r_{e_i})d_i(\gamma - d_i - 1 + d_i) + r_{e_i}(\gamma - d_i) &= 0 \\
(r_{d_i} - r_{e_i})d_i(\gamma - 1) + r_{e_i}(\gamma - d_i) &= 0 \\
d_i[(r_{d_i} - r_{e_i})(\gamma - 1) - r_{e_i}) + r_{e_i}\gamma &= 0 \\
[(1-\gamma)r_{d_i} + \gamma r_{e_i}]d_i = r_{e_i}\gamma &\overset{r_{d_i}, r_{e_i} > 0}{\Rightarrow} \\
d_i = \frac{\gamma r_{e_i}}{(1-\gamma)r_{d_i} + \gamma r_{e_i}} &\overset{r_{d_i} > 0}{\Rightarrow} \\
d_i = \frac{\gamma \frac{r_{e_i}}{r_{d_i}}}{(1-\gamma) + \gamma \frac{r_{e_i}}{r_{d_i}}} &
\end{aligned} \tag{A.7}$$

Note that the profit-maximizing debt share d_i depends only on relative prices, not their level. The profit-maximizing d_i has the following properties:

$$d_i \begin{cases} \nearrow 1 & \text{as } \frac{r_{e_i}}{r_{d_i}} \to \infty \\ \in (\gamma, 1) & \text{if } \frac{r_{e_i}}{r_{d_i}} > 1 \\ = \gamma & \text{if } \frac{r_{e_i}}{r_{d_i}} = 1 \\ \in (0, \gamma) & \text{if } \frac{r_{e_i}}{r_{d_i}} < 1 \\ \searrow 0 & \text{as } \frac{r_{e_i}}{r_{d_i}} \searrow 0 \end{cases} \tag{A.8}$$

This is because

$$
\begin{aligned}
d_i = \frac{\gamma \frac{r_{e_i}}{r_{d_i}}}{(1-\gamma) + \gamma \frac{r_{e_i}}{r_{d_i}}} < \gamma \Rightarrow \\
\frac{\frac{r_{e_i}}{r_{d_i}}}{(1-\gamma) + \gamma \frac{r_{e_i}}{r_{d_i}}} - 1 < 0 \\
\frac{\frac{r_{e_i}}{r_{d_i}} - (1-\gamma) - \gamma \frac{r_{e_i}}{r_{d_i}}}{(1-\gamma) + \gamma \frac{r_{e_i}}{r_{d_i}}} < 0 \\
(1-\gamma)\left(\frac{r_{e_i}}{r_{d_i}} - 1\right) < 0 \\
\frac{r_{e_i}}{r_{d_i}} < 1
\end{aligned}
\tag{A.9}
$$

The formula for the profit-maximizing share of equity is

$$
1 - d_i = 1 - \frac{\gamma \frac{r_{e_i}}{r_{d_i}}}{(1-\gamma) + \gamma \frac{r_{e_i}}{r_{d_i}}} = \frac{1-\gamma}{(1-\gamma) + \gamma \frac{r_{e_i}}{r_{d_i}}}
\tag{A.10}
$$

The cost of capital and the capital wedge τ_{K_i}. Substituting the profit-maximizing d_i into the first order condition with respect to capital we get that the cost of capital takes the following expression.

$$
\begin{aligned}
r_{d_i} d_i + r_{e_i}(1-d_i) = \frac{r_{e_i}}{(1-\gamma) + \gamma \frac{r_{e_i}}{r_{d_i}}} \equiv R\tau_{K_i} \\
\tau_{ed_i} = \frac{r_{e_i}}{r_{d_i}}, \quad \frac{r_{e_i}}{(1-\gamma) + \gamma \tau_{ed_i}} = R_t \tau_{K_i}
\end{aligned}
\tag{A.11}
$$

Note that the cost of capital depends only on exogenous prices and from the first order condition we know that $MRPK_i = R\tau_{K_i}$. The cost of capital depends both on the level of r_{e_i} and its relative price, which I call the capital structure distortion τ_{ed_i}. Equation (A.11) implies that we can perform counterfactual exercises by keeping one of the τ_{K_i}, τ_{ed_i} quantities constant and eliminating the distortion in the other. For example, if we eliminate the relative price friction, $\tau_{ed_i} = 1$, we can set $r_{e_i} = R\tau_{K_i}$ so that the capital wedges are unchanged. Or, if we eliminate the capital wedge $\tau_{K_i} = 1$, we can set $r_{e_i} = R\left[(1-\gamma) + \gamma\tau_{ed_i}\right]$ so that all firms have the same profit-maximizing MRPK but the capital structure distortions τ_{ed_i} are unchanged.

Recovering the capital structure distortion τ_{ed_i} from the data. From the profit-maximizing equations for d_i, e_i equations (A.7) and (A.10), we get an expression for the relative prices that depends only on the debt share d_i.

$$\tau_{ed_i} \equiv \frac{r_{e_i}}{r_{d_i}} = \frac{1-\gamma}{\gamma}\frac{d_i}{1-d_i} \tag{A.12}$$

The above equations show that the capital structure distortion τ_{ed_i} can be recovered using only data on capital structure d_i, given parameter γ.

The financial productivity wedge τ_{AF_i}. At the profit-maximizing capital structure, the contribution of capital structure to output F_i is a function of τ_{ed_i}, which I denote $f(\tau_{ed_i})$.

$$\begin{aligned}
F_i = f(\tau_{ed_i}) &= \left[d_i^{\gamma}(1-d_i)^{1-\gamma}\right]^{\alpha} && \text{(A.13)}\\
&= \left[\left(\frac{\gamma\tau_{ed_i}}{1-\gamma+\gamma\tau_{ed_i}}\right)^{\gamma}\left(\frac{1-\gamma}{(1-\gamma)+\gamma\tau_{ed_i}}\right)^{1-\gamma}\right]^{\alpha} && \text{(A.14)}\\
&= \left[\gamma^{\gamma}(1-\gamma)^{1-\gamma}\right]^{\alpha}\left[\frac{\tau_{ed_i}^{\gamma}}{1-\gamma+\gamma\tau_{ed_i}}\right]^{\alpha} && \text{(A.15)}\\
&= f(1)\left[\frac{\tau_{ed_i}^{\gamma}}{1-\gamma+\gamma\tau_{ed_i}}\right]^{\alpha} && \text{(A.16)}
\end{aligned}$$

When the price of equity is equal to the price of debt, the firm chooses the first-best capital structure $d^*_{it} = \gamma$. If the price of debt is different than that of equity $r_{e_{it}} \neq r_{d_{it}}$, then the contribution of capital structure to output $F_{it} = f(r_e/r_d)$ is less than it's potential $\overline{F} = f(1)$, leading to lower output. Therefore, if, for some firms, the price of debt is different than that of equity $r_{e_{it}}/r_{d_{it}} \neq 1$, there is an output loss in the economy from suboptimal capital structure.

Define the financial productivity wedge as the factor difference between the first best F and the actual F_i:

$$\tau_{AF_i} \equiv \frac{f(1)}{f(\tau_{ed_i})} = \frac{1}{\left[\frac{\tau_{ed_i}^{\gamma}}{1-\gamma+\gamma\tau_{ed_i}}\right]^{\alpha}} \geq 1 \tag{A.17}$$

I call τ_{AF_i} the financial productivity wedge because a suboptimal capital structure manifests as lower firm-specific total factor productivity in the production function (expression (A.1)):

$$\widetilde{A}_i = A_i f(\tau_{ed_i}) = A_i \frac{f(1)}{\tau_{AF_i}} = \frac{\overline{A}_i}{\tau_{AF_i}}, \quad \overline{A}_i \equiv A_i f(1) \tag{A.18}$$

I consider $\overline{A}_i = A_i f(1)$ the productivity frontier for firm i, which is achieved in the absence of capital structure distortions, and is invariant to changes in frictions τ_{ed_i}, τ_{K_i} since it depends only on productivity A_i and the parameters γ, α.

A.1 Equilibrium

In this model, an economy is characterized by the distribution of productivities and distortions $\{\overline{A}_i, \tau_{AF_i}, \tau_{K_i}\}_i$ and the aggregate resource constraint in the economy in terms of capital $\mathbf{K}$ and labor $\mathbf{L}$. In equilibrium, a price of capital R and a price of labor w bring the demand for capital $\sum_i K_i$ and labor $\sum_i L_i$ equal to their respective supply.

$$K_i = \left[\frac{\widetilde{\Omega}_i}{\tau_{K_i}^{1-\beta_l}}\right]^{\frac{1}{1-\beta_k-\beta_l}} \left[\frac{\beta_k}{R}\left(\frac{\beta_l R}{b_k w}\right)^{\beta_l} \mathbf{Y}^{1/\eta}\right]^{\frac{1}{1-\beta_k-\beta_l}} \tag{A.19}$$

$$L_i = \left[\frac{\widetilde{\Omega}_i}{\tau_{K_i}^{\beta_k}}\right]^{\frac{1}{1-\beta_k-\beta_l}} \left[\frac{\beta_l}{w}\left(\frac{\beta_k w}{\beta_l R}\right)^{\beta_k} \mathbf{Y}^{1/\eta}\right]^{\frac{1}{1-\beta_k-\beta_l}} \tag{A.20}$$

Market-clearing prices R, w

$$\left[\frac{\beta_k}{R}\left(\frac{\beta_l R}{\beta_k w}\right)^{b_l} \mathbf{Y}^{1/\eta}\right]^{\frac{1}{1-\beta_k-\beta_l}} \sum_i \left[\frac{\widetilde{\Omega}_i}{\tau_{K_i}^{1-\beta_l}}\right]^{\frac{1}{1-\beta_k-\beta_l}} = \mathbf{K} \tag{A.21}$$

$$\left[\frac{\beta_l}{w}\left(\frac{\beta_k w}{\beta_l R}\right)^{\beta_k} \mathbf{Y}^{1/\eta}\right]^{\frac{1}{1-\beta_k-\beta_l}} \sum_i \left[\frac{\widetilde{\Omega}_i}{\tau_{K_i}^{\beta_k}}\right]^{\frac{1}{1-\beta_k-\beta_l}} = \mathbf{L} \tag{A.22}$$

Therefore firm-level demand of capital and labor at market-clearing prices can be expressed as

$$K_i = \frac{\frac{\widetilde{A}_i^{\eta-1}}{\tau_{K_i}^{1+\alpha(\eta-1)}}}{\sum \frac{\widetilde{A}_i^{\eta-1}}{\tau_{K_i}^{1+\alpha(\eta-1)}}}\mathbf{K} = \frac{\frac{\left(\frac{\overline{A}_i}{\tau_{AF_i}}\right)^{\eta-1}}{\tau_{K_i}^{1+\alpha(\eta-1)}}}{\sum \frac{\left(\frac{\overline{A}_i}{\tau_{AF_i}}\right)^{\eta-1}}{\tau_{K_i}^{1+\alpha(\eta-1)}}}\mathbf{K} \equiv \frac{\left(\frac{\overline{A}_i}{\tau_{AF_i}}\right)^{\eta-1}}{\tau_{K_i}^{1+\alpha(\eta-1)}} \frac{\mathbf{K}}{\mathbf{\Phi}_K} \tag{A.23}$$

$$L_i = \frac{\frac{\widetilde{A}_i^{\eta-1}}{\tau_{K_i}^{\alpha(\eta-1)}}}{\sum \frac{\widetilde{A}_i^{\eta-1}}{\tau_{K_i}^{\alpha(\eta-1)}}}\mathbf{L} = \frac{\left(\frac{\overline{A}_i}{\tau_{AF_i}\tau_{K_i}^{\alpha}}\right)^{\eta-1}}{\sum \left(\frac{\overline{A}_i}{\tau_{AF_i}\tau_{K_i}^{\alpha}}\right)^{\eta-1}}\mathbf{L} \equiv \left(\frac{\overline{A}_i}{\tau_{AF_i}\tau_{K_i}^{\alpha}}\right)^{\eta-1} \frac{\mathbf{L}}{\mathbf{\Phi}_L} \tag{A.24}$$

Firm output is

$$Y_i = \frac{\overline{A}_i}{\tau_{AF_i}} \left[\frac{\left(\frac{\overline{A}_i}{\tau_{AF_i}}\right)^{\eta-1}}{\tau_{K_i}^{1+\alpha(\eta-1)}} \frac{\mathbf{K}}{\mathbf{\Phi}_K}\right]^{\alpha} \left[\left(\frac{\overline{A}_i}{\tau_{AF_i}\tau_{K_i}^{\alpha}}\right)^{\eta-1} \frac{\mathbf{L}}{\mathbf{\Phi}_L}\right]^{1-\alpha} = \left(\frac{\overline{A}_i}{\tau_{AF_i}\tau_{K_i}^{\alpha}}\right)^{\eta} \frac{\mathbf{K}^{\alpha}\mathbf{L}^{1-\alpha}}{\mathbf{\Phi}_K^{\alpha}\mathbf{\Phi}_L^{1-\alpha}} \tag{A.25}$$

Aggregate output is

$$\mathbf{Y} = \left[\sum_{i=1}^{N} Y_i^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}} = \left[\sum_{i=1}^{N}\left(\frac{\overline{A}_i}{\tau_{AF_i}\tau_{K_i}^{\alpha}}\right)^{\eta-1}\right]^{\frac{\eta}{\eta-1}} \frac{\mathbf{K}^{\alpha}\mathbf{L}^{1-\alpha}}{\mathbf{\Phi}_K^{\alpha}\mathbf{\Phi}_L^{1-\alpha}} \tag{A.26}$$

Aggregate TFP. Total factor productivity is defined as $\frac{\mathbf{Y}}{\mathbf{K}^{\alpha}\mathbf{L}^{1-\alpha}}$, which takes the following form.

$$\begin{aligned} TFP &= \frac{\left[\sum_{i=1}^{N}\left(\frac{\overline{A}_i}{\tau_{AF_i}\tau_{K_i}^{\alpha}}\right)^{\eta-1}\right]^{\frac{\eta}{\eta-1}}}{\mathbf{\Phi}_K^{\alpha}\mathbf{\Phi}_L^{1-\alpha}} \\ &= \frac{\left[\sum_{i=1}^{N}\left(\frac{\overline{A}_i}{\tau_{AF_i}\tau_{K_i}^{\alpha}}\right)^{\eta-1}\right]^{\frac{\eta}{\eta-1}}}{\left[\sum \frac{\left(\frac{\overline{A}_i}{\tau_{AF_i}}\right)^{\eta-1}}{\tau_{K_i}^{1+\alpha(\eta-1)}}\right]^{\alpha}\left[\sum\left(\frac{\overline{A}_i}{\tau_{AF_i}\tau_{K_i}^{\alpha}}\right)^{\eta-1}\right]^{1-\alpha}} \\ &= \frac{\left[\sum_{i=1}^{N}\left(\frac{\overline{A}_i}{\tau_{AF_i}\tau_{K_i}^{\alpha}}\right)^{\eta-1}\right]^{\frac{1}{\eta-1}+\alpha}}{\left[\sum_{i=1}^{N} \frac{\left(\frac{\overline{A}_i}{\tau_{AF_i}}\right)^{\eta-1}}{\tau_{K_i}^{1+\alpha(\eta-1)}}\right]^{\alpha}} \end{aligned} \tag{A.27}$$

The TFP formula (A.27) is a function of only wedge $\tau_{AF_i}\tau_{K_i}$ and the productivity frontier $\overline{A}_i$. It doesn't depend on prices as we have already solved for the equilibrium. Therefore, it is appropriate for conducting counterfactual experiments on the distortions by changes the wedges $\tau_{AF_i}\tau_{K_i}$ while keeping $\overline{A}_i$ constant.

First-best TFP. At the first-best allocation $\{\tau_{AF_i} = 1, \tau_{K_i} = 1\}$, the TFP is

$$TFP_0 = \left(\sum_{i=1}^{N} \overline{A}_i^{\eta-1} \right)^{\frac{1}{\eta-1}}. \tag{A.28}$$

A.2 Counterfactual analysis

There are two approaches to quantifying the importance of the capital structure distortion τ_{ed}. The first approach is to compare the TFP of an economy without capital wedges to the first best $1 - TFP(\{\overline{A}_i, \tau_{AF_i}, \tau_{K_i} = 1\}_i)/TFP(\{\overline{A}_i, \tau_{AF_i} = 1, \tau_{K_i} = 1\}_i)$, which I call the inframarginal effect. The other is to compare the TFP of an economy at its current state $\{\overline{A}_i, \tau_{AF_i}, \tau_{K_i}\}_i$ to an economy without capital structure distortions $\tau_{ed} = \tau_{AF_i} = 1$. The issue with the second approach is that changing τ_{ed} changes also the capital wedge—see equation (A.11). This is because $\tau_{ed} = 1$ implies no tradeoff between reducing the cost of financing and increasing the capital structure productivity. To isolate the capital structure effect, I keep τ_{K_i} constant by increasing $r_{e_i} = R\tau_{K_i}$. This counterfactual then can be calculated by $1 - TFP(\{\overline{A}_i, \tau_{AF_i}, \tau_{K_i}\}_i)/TFP(\{\overline{A}_i, \tau_{AF_i} = 1, \tau_{K_i}\}_i)$ which I call the compensated marginal effect (akin to the compensated/Hicksian demand that isolates the substitution effect by keeping utility constant). These two counterfactual quantities should be close but not identical. In the special case where three variables are distributed jointly lognormal, the inframarginal and compensated marginal loss are identical—see Section A.3 below for an analytical derivation.

A.3 TFP loss under log normality

To gain intuition about how the joint distribution of $\overline{A}_i, \tau_{AF_i}, \tau_{K_i}$ determines the TFP loss, I derive the TFP loss formula for the case where the the log deviation of the mean of each variable is normally distributed. More specifically, let $\mu_{\overline{A}} = e^{\mathbf{E}(\log \overline{A})}, \mu_{\tau_{AF}} = e^{\mathbf{E}(\log \tau_{AF})}$

and $\mu_{\tau_K} = e^{\mathbf{E}(\log \tau_K)}$. Let $\widehat{\overline{A}} = \overline{A}/\mu_{\overline{A}}$, $\widehat{\tau_{AF}} = \tau_{AF}/\mu_{\tau_{AF}}$, and $\widehat{\tau_K} = \tau_K/\mu_{\tau_K}$. Then $\mathbf{E}(\widehat{\overline{A}}) = \mathbf{E}(\widehat{\tau_{AF}}) = \mathbf{E}(\widehat{\tau_K}) = 0$. Let $\widehat{\overline{A}}, \widehat{\tau_{AF}}, \widehat{\tau_K}$ be log-normally distributed. The the ratio of the TFP with distortions to the TFP without distortions is

$$\frac{TFP_1}{TFP_0} = \frac{\left[\sum_{i=1}^{N}\left(\frac{\overline{A}_i}{\tau_{AF_i}\tau_{K_i}^{\alpha}}\right)^{\eta-1}\right]^{\frac{1}{\eta-1}+\alpha}}{\left[\sum_{i=1}^{N}\frac{\left(\frac{\overline{A}_i}{\tau_{AF_i}}\right)^{\eta-1}}{\tau_{K_i}^{1+\alpha(\eta-1)}}\right]^{\alpha}} \Bigg/ \left(\sum_{i=1}^{N}\overline{A}_i^{\eta-1}\right)^{\frac{1}{\eta-1}}$$

$$= \frac{\left[\sum_{i=1}^{N}\frac{1}{N}\left(\frac{\overline{A}_i}{\tau_{AF_i}\tau_{K_i}^{\alpha}}\right)^{\eta-1}\right]^{\frac{1}{\eta-1}+\alpha}}{\left[\sum_{i=1}^{N}\frac{1}{N}\frac{\left(\frac{\overline{A}_i}{\tau_{AF_i}}\right)^{\eta-1}}{\tau_{K_i}^{1+\alpha(\eta-1)}}\right]^{\alpha}} \Bigg/ \left(\sum_{i=1}^{N}\frac{1}{N}\overline{A}_i^{\eta-1}\right)^{\frac{1}{\eta-1}}$$

$$\overset{\text{Law of large numbers}}{=} \frac{\left[\mathbf{E}\left(\frac{\overline{A}}{\tau_{AF}\tau_K^{\alpha}}\right)^{\eta-1}\right]^{\frac{1}{\eta-1}+\alpha}}{\left[\mathbf{E}\frac{\left(\frac{\overline{A}}{\tau_{AF}}\right)^{\eta-1}}{\tau_K^{1+\alpha(\eta-1)}}\right]^{\alpha}} \Bigg/ \left(\mathbf{E}\overline{A}^{\eta-1}\right)^{\frac{1}{\eta-1}} \tag{A.29}$$

Equation (A.29) expresses the ratio f the two TFPs in terms of expectations. To apply the expectations formulas of the log-normal distribution to expression (A.29), we need to express it in terms of $\widehat{\overline{A}}, \widehat{\tau_{AF}}, \widehat{\tau_K}$. The numerator of (A.29) is

$$
\begin{aligned}
Q(\overline{A},\tau_{AF},\tau_K) &\equiv \frac{\left[\mathbf{E}\left(\frac{\overline{A}}{\tau_{AF}\tau_K^{\alpha}}\right)^{\eta-1}\right]^{\frac{1}{\eta-1}+\alpha}}{\left[\mathbf{E}\frac{\left(\frac{\overline{A}}{\tau_{AF}}\right)^{\eta-1}}{\tau_K^{1+\alpha(\eta-1)}}\right]^{\alpha}} \\
&= \frac{\mu^1_{\overline{A}}}{\mu^1_{\tau_{AF}}\mu^0_{\tau_K}} \frac{\left[\mathbf{E}\left(\frac{\widehat{\overline{A}}}{\widehat{\tau_{AF}}\widehat{\tau_K}^{\alpha}}\right)^{\eta-1}\right]^{\frac{1}{\eta-1}+\alpha}}{\left[\mathbf{E}\frac{\left(\frac{\widehat{\overline{A}}}{\widehat{\tau_{AF}}}\right)^{\eta-1}}{\widehat{\tau_K}^{1+\alpha(\eta-1)}}\right]^{\alpha}} \\
&= \frac{\mu_{\overline{A}}}{\mu_{\tau_{AF}}} \frac{\left[\mathbf{E}e^{(\eta-1)\left(\log\widehat{\overline{A}}-\log\widehat{\tau_{AF}}-\alpha\log\widehat{\tau_K}\right)}\right]^{\frac{1}{\eta-1}+\alpha}}{\left[\mathbf{E}e^{(\eta-1)\left(\log\widehat{\overline{A}}-\log\widehat{\tau_{AF}}\right)-[1+\alpha(\eta-1)]\log\widehat{\tau_K}}\right]^{\alpha}} \\
&= \frac{\mu_{\overline{A}}}{\mu_{\tau_{AF}}} \frac{\left[e^{\frac{1}{2}\mathbf{Var}(\eta-1)\left(\log\widehat{\overline{A}}-\log\widehat{\tau_{AF}}-\alpha\log\widehat{\tau_K}\right)}\right]^{\frac{1}{\eta-1}+\alpha}}{\left[e^{\frac{1}{2}\mathbf{Var}\left[(\eta-1)\left(\log\widehat{\overline{A}}-\log\widehat{\tau_{AF}}\right)-[1+\alpha(\eta-1)]\log\widehat{\tau_K}\right]}\right]^{\alpha}} \\
&= \frac{\mu_{\overline{A}}}{\mu_{\tau_{AF}}} \frac{\left[e^{\frac{(\eta-1)^2}{2}\left(\sigma_A^2+\sigma_F^2+\alpha^2\sigma_K^2-2\sigma_{A,F}-2\alpha\sigma_{A,K}+2\alpha\sigma_{F,K}\right)}\right]^{\frac{1}{\eta-1}+\alpha}}{\left[e^{\frac{1}{2}\left[(\eta-1)^2(\sigma_A^2+\sigma_F^2-2\sigma_{A,F})+[1+\alpha(\eta-1)]^2\sigma_K^2+2[1+\alpha(\eta-1)](\eta-1)(\sigma_{F,K}-\sigma_{A,K})\right]}\right]^{\alpha}} \\
&= \frac{\mu_{\overline{A}}}{\mu_{\tau_{AF}}} e^{\frac{\eta-1}{2}(\sigma_A^2+\sigma_F^2-2\sigma_{A,F})-\frac{\alpha}{2}[\alpha(\eta-1)+1]\sigma_K^2} \quad ,
\end{aligned} \tag{A.30}
$$

where $\sigma_A^2 = \mathbf{Var}(\log\widehat{\overline{A}})$, $\sigma_F^2 = \mathbf{Var}(\log\widehat{\tau_{AF}})$, $\sigma_K^2 = \mathbf{Var}(\log\widehat{\tau_K})$, $\sigma_{A,F} = \mathbf{Cov}(\log\widehat{\overline{A}}, \log\widehat{\tau_{AF}})$, $\sigma_{A,K} = \mathbf{Cov}(\log\widehat{\overline{A}}, \log\widehat{\tau_K})$, $\sigma_{F,K} = \mathbf{Cov}(\log\widehat{\tau_{AF}}, \log\widehat{\tau_K})$. Note that $\sigma_{F,K}, \sigma_{A,K}$ disappear. The denominator of (A.29) is

$$
Q_0 \equiv \left(\mathbf{E}\overline{A}^{\eta-1}\right)^{\frac{1}{\eta-1}} = \mu_{\overline{A}}\left(\mathbf{E}\widehat{\overline{A}}^{\eta-1}\right)^{\frac{1}{\eta-1}} = \mu_{\overline{A}}\left(e^{\frac{(\eta-1)^2}{2}\sigma_A^2}\right)^{\frac{1}{\eta-1}} = \mu_{\overline{A}}e^{\frac{\eta-1}{2}\sigma_A^2} \tag{A.31}
$$

Combining expressions (A.30) and (A.31), the expression for the TFP ratio in (A.29) is

$$\frac{TFP_1}{TFP_0} = \frac{Q(\overline{A}, \tau_{AF}, \tau_K)}{Q_0} = \frac{1}{\mu_{\tau_{AF}}} e^{\frac{\eta-1}{2}(\sigma_F^2 - 2\sigma_{A,F}) - \frac{\alpha}{2}[\alpha(\eta-1)+1]\sigma_K^2}. \tag{A.32}$$

Note that the TFP loss is $1 - \frac{TFP_1}{TFP_0} \approx -\log\frac{TFP_1}{TFP_0} = -\log\frac{Q(\overline{A},\tau_{AF},\tau_K)}{Q_0}$. Substituting expression (A.32) we have

$$\begin{aligned} \text{TFP loss total} &\approx -\log\left[\frac{1}{\mu_{\tau_{AF}}} e^{\frac{\eta-1}{2}(\sigma_F^2 - 2\sigma_{A,F}) - \frac{\alpha}{2}[\alpha(\eta-1)+1]\sigma_K^2}\right] \\ &= \log\mu_{\tau_{AF}} - \frac{\eta-1}{2}(\sigma_F^2 - 2\sigma_{A,F}) + \frac{\alpha}{2}[\alpha(\eta-1)+1]\sigma_K^2 \\ &= \mathbf{E}(\log\tau_{AF}) + (\eta-1)\sigma_{A,F} - \frac{\eta-1}{2}\sigma_F^2 + \frac{\alpha}{2}[\alpha(\eta-1)+1]\sigma_K^2. \end{aligned} \tag{A.33}$$

Note that the higher the average capital structure distortion ($\mathbf{E}(\log\tau_{AF})$), the higher the TFP loss. The higher the covariance between the financial productivity wedge and productivity $\sigma_{A,F}$, the higher the TFP loss because high-productivity firms face higher distortions. A high variance of the capital structure distortion reduces TFP loss because the effect of capital structure distortions on output can be dampened by reallocating resources from the firms facing high capital structure distortions to the ones facing lower distortions (the general equilibrium effect). As in Hsieh and Klenow (2009); David and Venkateswaran (2019), the higher the variance of the capital wedge σ_K^2, the higher the TFP loss.

Expression (A.30) can also be used to derive the inframarginal loss of the capital structure distortion

$$\text{TFP loss inframarginal} = \frac{Q(\overline{A}, \tau_{AF}, \tau_K = 1)}{Q(\overline{A}, \tau_{AF} = 1, \tau_K = 1)} \tag{A.34}$$

$$\approx \mathbf{E}(\log\tau_{AF}) + (\eta-1)\sigma_{A,F} - \frac{\eta-1}{2}\sigma_F^2 \tag{A.35}$$

Expression (A.30) can also be used to derive the compensated marginal loss of the capital structure distortion

$$\text{TFP loss inframarginal} = \frac{Q(\overline{A}, \tau_{AF}, \tau_K)}{Q(\overline{A}, \tau_{AF} = 1, \tau_K)} \tag{A.36}$$

$$\approx \mathbf{E}(\log \tau_{AF}) + (\eta - 1)\sigma_{A,F} - \frac{\eta - 1}{2}\sigma_F^2 \tag{A.37}$$

which in this case is identical to the inframarginal loss.

B Data

The data come from the dataset Compustat Global and the text includes all the necessary details. Before trimming the baseline dataset for France, Germany, Sweden, and the UK, there are 45,273 and after trimming outliers there are 43,389 remaining observations used for the analysis.

C Estimation

I estimate γ using the fixed-effects estimator of Whited and Zhao (2021). The starting point is the relationship between the residual $\widetilde{A}_{it}$ in the production function that depends on productivity A_{it} and capital structure F_{it}, which in logs is:

$$\log \widetilde{A}_{it} = \overbrace{\alpha\gamma}^{a_d} \log d_{it} + \overbrace{\alpha(1-\gamma)}^{a_e} \log(1 - d_{it}) + \overbrace{\log A_{it}}^{\epsilon_i} \tag{C.1}$$

I estimate equation (C.1) using a within-firm estimator after demeaning $\log \widetilde{A}_{it}$ at the annual level (a year fixed-effect). The second panel of Table C.1 presents the results. The estimated values for a_d, a_e are precisely estimated and lower than 1, which is consistent with $\gamma, \alpha < 1$ in the model. I estimate γ using the following formula resulting from

Table C.1: Parameter estimates from the baseline sample.

	Value	S.E.
α from median	0.342	0.0016
GMM fixed effect estimator		
a_d	0.222	0.0247
a_e	0.155	0.0137
γ	0.589	0.0196
$\widetilde{\alpha}$ (implied α)	0.377	0.0356
Number of observations	43,389	

The baseline sample includes public firms from France, Germany, Sweden, and the UK. Standard errors are heteroskedasticity robust, incorporate variation from all estimation steps, and are calculated using the influence function approach described in Section 5.1.

equation (C.1):

$$a_e/a_d = \frac{1-\gamma}{\gamma} \Rightarrow \gamma = \frac{1}{1+a_e/a_d} \tag{C.2}$$

The estimated value of γ is precisely estimated and less than 1, consistent with the model.

I can also use the estimated γ and the estimated coefficient a_d to back out the implied α: $\widetilde{\alpha} = a_d/\gamma$. The implied $\widetilde{\alpha}$ is 0.377 which is close to the α estimated using the median (Table C.1). In fact, the estimated α using the median is 0.342 and is included in the 95% confidence interval of $\widetilde{\alpha}$, implying that we cannot reject the hypothesis that the two estimates are equal, which gives credence to the model and the estimation approach. For the analysis, I use the median estimate of α which is more precisely estimated than $\widetilde{\alpha}$ and relies on less assumptions than the fixed-effects estimator.

D Results and Inference

Table D.1 presents results from re-estimating the model in subsamples, which are used to generate Table 3 in the text. There are several common patterns across subsamples. The capital share is less than half, the optimal debt share is more than half, the TFP loss from the financial productivity wedge is approximately 2% and the compensated marginal loss has similar magnitude to the inframarginal loss.

Table D.1: Estimates from subsamples

Market definition	α	γ	TFP loss from τ_{AF} in % Inframarginal	Compens. marginal	Obs.
(1) Baseline sample	0.342 (0.002)	0.589 (0.02)	2.0 (0.24)	2.1 (0.32)	43,389
(2) France and Germany	0.302 (0.002)	0.530 (0.03)	1.8 (0.25)	1.7 (0.27)	19,930
(3) France	0.303 (0.003)	0.541 (0.05)	1.8 (0.28)	1.8 (0.30)	8,939
(4) Germany	0.302 (0.003)	0.526 (0.04)	1.7 (0.50)	1.7 (0.54)	10,991
(5) Sweden	0.255 (0.005)	0.459 (0.06)	1.2 (0.56)	1.3 (0.61)	5,438
(6) UK	0.424 (0.003)	0.577 (0.04)	3.3 (0.58)	3.3 (0.92)	18,021

The first row of this table reports results from the sample used throughout this paper, which includes public firms from France, Germany, Sweden, and the UK. Standard errors (in parentheses) are heteroskedasticity robust, incorporate variation from all estimation steps, and are calculated using the influence function approach described in Section 5.1.

D.1 Inference

To obtain the standard errors and confidence intervals of the quantities of interest, I need to derive the asymptotic distribution of the ratio of two aggregate TFPs, as all results depend on such a ratio.

Since the ratio of two TFPs is not a well-studied statistic, the first step is to verify that it can be expressed as a statistical functional so that standard asymptotic inference can be applied. Given a sample $X_1, \ldots, X_n$, a statistical functional is a map from an empirical cumulative distribution function $\mathbb{F}_n$ (or the empirical probability measure $\mathbb{P}_n$) to the real numbers, implying that the functional should not depend directly on the sample size n but only on the distribution of the data (Fernholz, 1983). For instance, the TFP formula in equation (7) cannot be expressed as a statistical functional because TFP increases as the number of firms increases. To see this, note that if we duplicate every observation, the TFP is larger by a factor of $2^{\frac{1}{\eta-1}}$. As a result, as the sample size grows ($n \to \infty$), the TFP does not converge but goes to infinity, and hence, the asymptotic statistical theory is not applicable. The ratio of two TFP's, in contrast, can be written as a statistical functional and asymptotic theory applies. To see this note that the ratio of two TFP's can be expressed only in terms of averages:

$$S_n = \frac{\left[\sum_{j=1}^{n} \frac{1}{n} \left(\frac{\overline{A}_j}{\tau_{AF_j} \tau_{K_j}^{\widehat{\alpha}}} \right)^{\eta-1} \right]^{\frac{1}{\eta-1}+\widehat{\alpha}} \Big/ \left(\sum_{j=1}^{n} \frac{1}{n} \frac{\left(\frac{\overline{A}_j}{\tau_{AF_j}}\right)^{\eta-1}}{\tau_{K_j}^{1+\widehat{\alpha}(\eta-1)}} \right)^{\widehat{\alpha}}}{\left[\sum_{j=1}^{n} \frac{1}{n} \left(\frac{\overline{A}_j}{\tau'_{AF_j} \tau'^{\widehat{\alpha}}_{K_j}} \right)^{\eta-1} \right]^{\frac{1}{\eta-1}+\widehat{\alpha}} \Big/ \left(\sum_{j=1}^{n} \frac{1}{n} \frac{\left(\frac{\overline{A}_j}{\tau'_{AF_j}}\right)^{\eta-1}}{\tau'^{1+\widehat{\alpha}(\eta-1)}_{K_j}} \right)^{\widehat{\alpha}}}. \tag{D.1}$$

Inference on estimator S_n is challenging as it is a function of data and estimated parameters $\widehat{\alpha}, \widehat{\gamma}$, and is not a statistic with well-known asymptotic properties. If estimator S_n depended only on estimated parameters, as in studies where counterfactuals use simulated data like Fakos (2023), then the standard error of S_n could be calculated using the delta method. I conduct inference on S_n using the influence function approach, which is a general framework for conducting inference as it can be applied to settings where the statistic of interest depends on parameters estimated using any combination of the GMM, quantile, Maximum Likelihood (MLE), or Simulated Minimum Distance (SMD) frameworks.

D.1.1 von Mises expansion and the influence function

A large class of estimators can be analyzed using linear approximation around the probability measure, which admits a calculus of estimators that allows for building the asymptotic distribution of complicated multi-step estimators from a sequence of simple asymptotic distributions. Such estimators are often called asymptotically linear (Newey and McFadden, 1994; Tsiatis, 2006). They can be expressed as statistical functionals that map the sample probability measure $\mathbb{P}_n$ into some space (a finite-dimensional Euclidean space for our purposes). The regular versions of GMM, SMD, MLE, quantiles, and smooth functions of means like S_n are such estimators. Below, I use notation close to (Van der Vaart, 1998, ch. 20) whose exposition I follow closely. Let $\phi(\mathbb{P}_n)$ be an asymptotically linear estimator, $\mathbf{P}$ the true probability distribution, and let ϕ be Hadamard differentiable at $\mathbf{P}$. Let $\phi'_{\mathbf{P}}(\sqrt{n}[\mathbb{P}_n - \mathbf{P}])$ be the derivative of ϕ using a perturbation in the direction of $\sqrt{n}[\mathbb{P}_n - \mathbf{P}]$. Then:

$$\phi(\mathbb{P}_n) - \phi(\mathbf{P}) \approx \frac{1}{\sqrt{n}}\phi'_{\mathbf{P}}(\sqrt{n}[\mathbb{P}_n - \mathbf{P}]) = \frac{1}{\sqrt{n}}\phi'_{\mathbf{P}}(\mathbb{G}_n), \tag{D.2}$$

which is called a von Mises expansion and can be used to study the asymptotic properties of the estimator ϕ. To see this, note that if we multiply both sides of equation (D.3) by $\sqrt{n}$, the left-hand side has the asymptotic distribution of the estimator. Since the derivative is a linear operator we have that $\phi'_{\mathbf{P}}(\sqrt{n}[\mathbb{P}_n - \mathbf{P}]) = \sqrt{n}\phi'_{\mathbf{P}}(\mathbb{P}_n - \mathbf{P})$. Substituting this expression into equation (D.2) we have that

$$\sqrt{n}[\phi(\mathbb{P}_n) - \phi(\mathbf{P})] \approx \sqrt{n}\phi'_{\mathbf{P}}(\mathbb{P}_n - \mathbf{P}). \tag{D.3}$$

The above equation implies that to study the asymptotic properties of the estimator ϕ, we need only analyze the asymptotic behavior of its derivative $\phi'_{\mathbf{P}}$. From the definition of Hadamard differentiability (Van der Vaart, 1998, ch. 20), ϕ is Hadamard differentiable at $\mathbf{P}$ if there exists a continuous, linear map $\phi'_{\mathbf{P}}(\widetilde{h})$ such that

$$\left\| \frac{\phi(\mathbf{P} + \epsilon \widetilde{h}_\epsilon) - \phi(\mathbf{P})}{\epsilon} - \phi'_{\mathbf{P}}(\widetilde{h}) \right\| \to 0, \quad \epsilon \downarrow 0, \tag{D.4}$$

for all $\widetilde{h}_\epsilon \to \widetilde{h}$ in a neighborhood of $\widetilde{h}$ where ϕ is well-defined. In this case, $\widetilde{h}_\epsilon = \mathbb{P}_n - \mathbf{P}$, and Hadamard differentiability guarantees that the limit of $\phi'_{\mathbf{P}}$ as $n \to \infty$ is well defined even if the direction of approach is not fixed as $\mathbb{P}_n - \mathbf{P}$ changes with n, since it depends on the empirical measure $\mathbb{P}_n$. By substituting $\widetilde{h}_\epsilon = \mathbb{P}_n - \mathbf{P}$ into (D.4) we obtain a more informative expression

$$\left\| \frac{\phi[(1-\epsilon)\mathbf{P} + \epsilon \mathbb{P}_n] - \phi(\mathbf{P})}{\epsilon} - \phi'_{\mathbf{P}}(\mathbb{P}_n - \mathbf{P}) \right\| \to 0, \quad \epsilon \downarrow 0,\ \mathbb{P}_n \to \mathbf{P}, \tag{D.5}$$

for all $\mathbb{P}_n$ where $\phi[(1-\epsilon)\mathbf{P} + \epsilon\mathbb{P}_n]$ is well-defined. We expect that in most cases $\phi[(1-\epsilon)\mathbf{P} + \epsilon\mathbb{P}_n]$ is well defined for any empirical measure $\mathbb{P}_n$ when $\epsilon \in (0, 1)$.

Chain rule and the delta method. An advantage of Hadamard differentiability is that the differentiation operations in Euclidean spaces ($\mathbb{R}^N$) like the chain rule, still apply. For example, if an estimator θ is a function of another estimator $\phi(\mathbb{P}_n)$ such that $\theta = h(\phi(\mathbb{P}_n))$ and h is Hadamard differentiable then $\theta'_{\mathbf{P}}(\mathbb{P}_n - \mathbf{P}) = h'(\phi(\mathbf{P}))\phi'_{\mathbf{P}}(\mathbb{P}_n - \mathbf{P})$. This chain rule leads to the functional delta method (Van der Vaart, 1998, §20.8), which is the extension of the delta method to estimators with infinite-dimensional parameters like an empirical distribution. Even though the estimator used in this paper is finite-dimensional, von Mises calculus can simplify the derivation of the asymptotic distribution of S_n. Hadamard differentiation in $\mathbb{R}^N$ is equivalent to the definition of a derivative of a function from $\mathbb{R}^N$ to $\mathbb{R}$ (see van der Vaart and Wellner, 1996, ch. 3.9). A sufficient condition for differentiability in $\mathbb{R}^N$ is that the function has continuous partial derivatives that can easily be verified for our estimator S_n. For example, the partial derivatives $\partial S_n / \partial \widehat{\alpha}$ and $\partial S_n / \partial \widehat{\gamma}$ exist and are continuous.

I can express estimator S as a function of the median $\widehat{\alpha}$ and GMM coefficients characterizing $\hat{\gamma}$, and several means, which I denote μ.

$$S(\mathbf{P}) = H(\mu(\mathbf{P}), \alpha(\mathbf{P}), \gamma(\mathbf{P}, \alpha(\mathbf{P}))). \tag{D.6}$$

In (D.6) the dependence of the estimator γ on the first stage estimate α is explicit. Here, I have suppressed the partial derivatives with respect to the yearly demeaning of the distortions for the sake of brevity. Regarding the influence function of yearly means, see further in this section.

By the mean-value theorem, we can express S as a linear function of its arguments

$$
\begin{aligned}
&H(\mu(\mathbb{P}_n), \alpha(\mathbb{P}_n), \gamma(\mathbb{P}_n, \alpha(\mathbb{P}_n))) - H(\mu(\mathbf{P}), \alpha(\mathbf{P}), \gamma(\mathbf{P}, \alpha(\mathbf{P}))) = \\
&+ \frac{\partial H}{\partial \mu}(\bar{\mu}, \bar{\alpha}, \bar{\gamma})[\mu(\mathbb{P}_n) - \mu(\mathbf{P})] \\
&+ \frac{\partial H}{\partial \alpha}(\bar{\mu}, \bar{\alpha}, \bar{\gamma})[\alpha(\mathbb{P}_n) - \alpha(\mathbf{P})] \\
&+ \frac{\partial H}{\partial \gamma}(\bar{\mu}, \bar{\alpha}, \bar{\gamma})[\gamma(\mathbb{P}_n, \alpha(\mathbb{P}_n)) - \gamma(\mathbf{P}, \alpha(\mathbf{P}))]
\end{aligned}
\tag{D.7}
$$

For some point $(\bar{\mu}, \bar{\alpha}, \bar{\gamma})$ between $(\mu(\mathbf{P}), \alpha(\mathbf{P}), \gamma(\mathbf{P}))$ and $(\mu(\mathbb{P}_n), \alpha(\mathbb{P}_n), \gamma(\mathbb{P}_n))$. Since the latter point converges to the former, $(\bar{\mu}, \bar{\alpha}, \bar{\gamma})$ also converges to both points as $n \to \infty$, and all three points are interchangeable in the continuous partial derivatives above. We can further expand the last term using the mean-value theorem to explicitly account for the effect of the first stage estimation of α on γ.

$$
\begin{aligned}
&\gamma(\mathbb{P}_n, \alpha(\mathbb{P}_n)) - \gamma(\mathbf{P}, \alpha(\mathbf{P})) = \\
&= \gamma(\mathbb{P}_n, \alpha(\mathbb{P}_n)) - \gamma(\mathbb{P}_n, \alpha(\mathbf{P})) + \gamma(\mathbb{P}_n, \alpha(\mathbf{P})) - \gamma(\mathbf{P}, \alpha(\mathbf{P})) = \\
&= \gamma(\mathbb{P}_n, \alpha(\mathbf{P})) - \gamma(\mathbf{P}, \alpha(\mathbf{P})) + \frac{\partial \gamma}{\partial \alpha}(\mathbb{P}_n, \alpha^*)[\alpha(\mathbb{P}_n) - \alpha(\mathbf{P})]
\end{aligned}
\tag{D.8}
$$

for some point α^* between $\alpha(\mathbf{P})$ and $\alpha(\mathbb{P}_n)$. Since $\alpha^* \to \alpha(\mathbf{P})$, and $(\bar{\mu}, \bar{\alpha}, \bar{\gamma}) \to (\mu(\mathbf{P}), \alpha(\mathbf{P}), \gamma(\mathbf{P}))$, we can substitute expression (D.8) for $\gamma(\mathbf{P}, \alpha(\mathbf{P})) - \gamma(\mathbb{P}_n, \alpha(\mathbb{P}_n))$ in equation (D.7) to study the asymptotic behavior of H.

To apply the von Mises expansion to S, we need to check whether it is Hadamard differentiable with respect to the $\mathbf{P}$ measure, where we can use the chain rule. The estimator α is Hadamard differentiable as a quantile (Van der Vaart, 1998, corollary 21.5). The estimator γ is also Hadamard differentiable as a continuously differentiable function of GMM estimator coefficients. The estimator γ depends on α through the construction of $\widetilde{A}$, which is a continuous differentiable function of α. The formula S_n depends on averages

μ which are Hadamard differentiable estimators because they are linear functions of the probability measure (the von Mises expansion of means is exact, not approximate Van der Vaart, 1998, example 20.2). Therefore, using the von Mises expansion of each component of S and abusing the partial derivative notation to refer to the Hadamard derivative of each argument of H with respect to the probability measure, we have

$$\begin{aligned} S(\mathbb{P}_n) - S(\mathbf{P}) = &\frac{\partial H}{\partial \mu}\frac{\partial \mu}{\partial \mathbf{P}_n}(\mathbb{P}_n - \mathbf{P}) \\ &+ \frac{\partial H}{\partial \gamma}\left[\frac{\partial \gamma}{\partial \mathbf{P}_n}(\mathbb{P}_n - \mathbf{P}) + \frac{\partial \gamma}{\partial \alpha}\frac{\partial \alpha}{\partial \mathbf{P}_n}(\mathbb{P}_n - \mathbf{P})\right] \\ &+ \frac{\partial H}{\partial \alpha}\frac{\partial \alpha}{\partial \mathbf{P}_n}(\mathbb{P}_n - \mathbf{P}). \end{aligned} \tag{D.9}$$

The von Mises expansion (D.9) is often used to construct efficient first step estimators (see Robins et al., 2009), but here I use it to derive the asymptotic variance of S_n. To study the asymptotic properties of an estimator ϕ using the von Mises expansion, we need to study the limit $\widetilde{\phi}_{\mathbf{P}} = \lim_{n\to\infty} \phi'_{\mathbf{P}}(\mathbb{P}_n - \mathbf{P})$ of the derivative.

The influence function. While the derivative in the von Mises expansion helps to break down the estimator to a sum of its components, it does not help derive an estimator of the asymptotic variance from the sample. To obtain a formula for the estimator's variance, we would like to express $\widetilde{\phi}_{\mathbf{P}}$ as an expectation so that we can approximate it by an average in finite samples. Expressing $\widetilde{\phi}_{\mathbf{P}}$ in terms of an expectation is possible since the Riesz representation theorem establishes that a linear functional (such as a derivative) can be expressed as an integral of a measurable function (Hines et al., 2022). More specifically, there exists a measurable function $\psi(x)$ such that

$$\widetilde{\phi}_{\mathbf{P}} = \int \psi(x) d\mathbf{P}(x). \tag{D.10}$$

The function $\psi(x)$ is called the influence function (Hines et al., 2022; Ichimura and Newey, 2022; Fisher and Kennedy, 2021; Van der Vaart, 1998). The sample analogue of (D.10) is $\frac{1}{n}\sum_i \psi(x_i)$ which converges to $\widetilde{\phi}_{\mathbf{P}}$. Therefore, the asymptotic properties of estimator ϕ can be studied by studying the influence function since

$$\phi(\mathbb{P}_n) - \phi(\mathbf{P}) = \int \psi(x) d\mathbf{P}(x) + o(\|\mathbb{P}_n - \mathbf{P}\|) = \frac{1}{n}\sum_{i=1}^{n} \psi(x_i) + o_p(1) \tag{D.11}$$

More specifically, to study the asymptotic properties of ϕ, we need to study the average $\frac{1}{n}\sum_{i=1}^{n} \psi(x_i)$, and asymptotic theory for averages is straightforward. The asymptotic variance of $\sqrt{n}[\phi(\mathbb{P}_n) - \phi(\mathbf{P})]$ is the asymptotic variance of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi(x_i)$, which can be estimated by $\widehat{AsyVar} = \frac{1}{n}\sum_{i=1}^{n} \psi(x_i)\psi'(x_i)$. The standard errors are $\sqrt{\frac{\text{diag}(\widehat{AsyVar})}{n}}$.

Calculating the influence function (the chain rule for the influence function). Formulas for the influence function of the estimators GMM, SMD, MLE, and quantiles are available (see Newey and McFadden, 1994; Tsiatis, 2006). We can construct the influence function of S from the influence function of the mean, which is $x - E(x)$. If a statistic is a function of a mean, then the influence function can be derived using the chain rule. To see that let $\theta = h(\phi(\mathbb{P}_n))$. Then $\theta'_{\mathbf{P}} = h'(\phi(\mathbf{P}))\phi'_{\mathbf{P}} = h'(\phi(\mathbf{P}))\int \psi(x)d\mathbf{P}(x) = \int h'(\phi(\mathbf{P}))\psi(x)d\mathbf{P}(x)$. Therefore the influence function ϕ_θ of θ is $h'(\phi(\mathbf{P}))\psi(x)$. Thus, the influence function of S can be derived by successive applications of the chain rule.

Approximating the influence function. The influence function at each point x can also be calculated as the value of the Hadamard derivative at the direction $\epsilon(\delta_x - \mathbf{P})$, where δ_x represents the Dirac delta: a probability measure with all the mass concentrated at x. This directional derivative is the Gateaux derivative, which is guaranteed to exist since ϕ

is Hadamard differentiable and is defined as

$$\psi(x) = \lim_{\epsilon \downarrow 0} \frac{\phi(\mathbf{P} + \epsilon[\delta_x - \mathbf{P}]) - \phi(\mathbf{P})}{\epsilon}, \quad \epsilon \in (0,1) \tag{D.12}$$

$$\psi(x) = \lim_{\epsilon \downarrow 0} \frac{\phi([1-\epsilon]\mathbf{P} + \epsilon\delta_x) - \phi(\mathbf{P})}{\epsilon}, \quad \epsilon \in (0,1) \tag{D.13}$$

The Gateaux derivative in this direction resembles the partial derivative in $\mathbb{R}^N$ (see Kennedy, 2016; Ichimura and Newey, 2022; Kennedy, 2022). The advantage of the Gateaux derivative at x is that it is a limit to a difference between two numbers which can be approximated numerically for a wide class of statistics.

Approximating the influence function of the mean. To estimate the asymptotic distribution of the mean μ, we need the sample analogue of $\psi_\mu(x)$ which for the mean is $\psi_{\mu_n}(x_i) = x_i - (1/n)\sum_i(x_i)$. It turns out that the influence function $\psi_{\mu_n}(x_i)$ at an observation x_i is equivalent to the Gateaux derivative in expression (D.13) evaluated at the empirical measure $\mathbb{P}_n$ instead of the unknown true measure $\mathbf{P}$. To see this, note that

$$\psi_{\mu_n}(x_i) = \lim_{\epsilon\downarrow 0} \frac{\mu(\mathbb{P}_n + \epsilon[\delta_{x_i} - \mathbb{P}_n]) - \mu(\mathbb{P}_n)}{\epsilon}, \quad \epsilon \in (0,1] \tag{D.14}$$

$$= \lim_{\epsilon\downarrow 0} \frac{\mu([1-\epsilon]\mathbb{P}_n + \epsilon\delta_{x_i}) - \mu(\mathbb{P}_n)}{\epsilon} \tag{D.15}$$

$$= \lim_{\epsilon\downarrow 0} \frac{\sum_{j=1}^{n} \left[(1-\epsilon)\frac{1}{n}x_j + \epsilon x_i \mathbb{1}\{j=i\}\right] - \sum_{j=1}^{n} \frac{1}{n}x_j}{\epsilon} \tag{D.16}$$

$$= \lim_{\epsilon\downarrow 0} \frac{\epsilon x_i + (1-\epsilon)\sum_{j=1}^{n} \frac{1}{n}x_j - \sum_{j=1}^{n} \frac{1}{n}x_j}{\epsilon} \tag{D.17}$$

$$= \lim_{\epsilon\downarrow 0} \frac{\epsilon x_i - \epsilon\sum_{j=1}^{n} P_j x_j}{\epsilon} = \tag{D.18}$$

$$\frac{\epsilon(x_i - \sum_{j=1}^{n} \frac{1}{n}x_j)}{\epsilon} = x_i - \sum_{i=1}^{n} \frac{1}{n}x_j, \ \epsilon \in (0,1] \tag{D.19}$$

$$\Rightarrow \mu(\mathbb{P}_n + \epsilon[\delta_{x_i} - \mathbb{P}_n]) - \mu(\mathbb{P}_n) = \epsilon\psi_{\mu_n}(x_i), \ \forall\epsilon \in [0,1] \tag{D.20}$$

Approximating the influence function of functions of the mean. We can derive the influence function of $h(\mu)$ using the chain rule $\psi_{h(\mu)}(x) = h'(\mu)\psi_{\mu}(x)$. And its sample analogue is $\psi_{h(\mu_n)}(x_i) = h'(\mu_n)\psi_{\mu_n}(x_i)$. By the definition of the derivative h' we have

$$h'(\mu_n)\psi_{\mu_n}(x_i) = \lim_{t\to 0} \frac{h(\mu_n + t) - h(\mu_n)}{t} \tag{D.21}$$

For every t close to 0, and $\psi_{\mu_n}(x_i) \neq 0$ there exists an $\epsilon \in (0,1)$ such that $|t| = \epsilon|\psi_{\mu_n}(x_i)|$. Since $\psi_{\mu_n}(x_i)$ is finite, as $\epsilon \downarrow 0$, $t \to 0$. Therefore,

$$\lim_{\epsilon\downarrow 0} \frac{h(\mu_n + \epsilon\psi_{\mu_n}(x_i)) - h(\mu_n)}{\epsilon\psi_{\mu_n}(x_i)} = h'(\mu_n), \quad \epsilon \in (0,1], \ \psi_{\mu_n}(x_i) \neq 0 \tag{D.22}$$

$$\Rightarrow \lim_{\epsilon\downarrow 0} \frac{h(\mu_n + \epsilon\psi_{\mu_n}(x_i)) - h(\mu_n)}{\epsilon\psi_{\mu_n}(x_i)}\psi_{\mu_n}(x_i) = h'(\mu_n)\psi_{\mu_n}(x_i) = \psi_{h(\mu_n)}(x_i) \tag{D.23}$$

Therefore,

$$\psi_{h(\mu_n)}(x_i) = \lim_{\epsilon\downarrow 0} \frac{h(\mu_n + \epsilon\psi_{\mu_n}(x_i)) - h(\mu_n)}{\epsilon\psi_{\mu_n}(x_i)}\psi_{\mu_n}(x_i), \quad \epsilon \in (0,1],\ \psi_{\mu_n}(x_i) \neq 0 \tag{D.24}$$

$$= \lim_{\epsilon\downarrow 0} \frac{h(\mu_n + \epsilon\psi_{\mu_n}(x_i)) - h(\mu_n)}{\epsilon} \tag{D.25}$$

$$= \lim_{\epsilon\downarrow 0} \frac{h(\mu(\mathbb{P}_n) + \epsilon\psi_{\mu_n}(x_i)) - h(\mu(\mathbb{P}_n))}{\epsilon} \tag{D.26}$$

$$\stackrel{(D.20)}{=} \lim_{\epsilon\downarrow 0} \frac{h(\mu(\mathbb{P}_n) + \mu(\mathbb{P}_n + \epsilon[\delta_{x_i} - \mathbb{P}_n]) - \mu(\mathbb{P}_n)) - h(\mu(\mathbb{P}_n))n)}{\epsilon} \tag{D.27}$$

$$= \lim_{\epsilon\downarrow 0} \frac{h(\mu(\mathbb{P}_n + \epsilon[\delta_{x_i} - \mathbb{P}_n])) - h(\mu(\mathbb{P}_n))}{\epsilon} \tag{D.28}$$

$$= \lim_{\epsilon\downarrow 0} \frac{h(\mu([1-\epsilon]\mathbb{P}_n + \epsilon\delta_{x_i})) - h(\mu(\mathbb{P}_n))}{\epsilon}, \quad \epsilon \in (0,1],\ \psi_{\mu_n}(x_i) \neq 0 \tag{D.29}$$

If $\psi_{\mu_n}(x_i) = 0$, then $\psi_{\mu_n}(x_i)\lim_{\epsilon\downarrow 0}\frac{h(\mu([1-\epsilon]\mathbb{P}_n+\epsilon\delta_{x_i}))-h(\mu(\mathbb{P}_n))}{\epsilon} = 0 = h'(\mu_n)\psi_{\mu_n}(x_i) = \psi_{h(\mu_n)}(x_i), \epsilon \in (0,1]$, since $h'(\mu_n)$ is finite and $\mu[1-\epsilon]\mathbb{P}_n + \epsilon\delta_{x_i}) \to \mathbb{P}_n$ and $\mu(\mathbb{P}_n)$ is continuous with respect to $\mathbb{P}_n$ as a linear function. Therefore

$$\psi_{h(\mu_n)}(x_i) = \lim_{\epsilon\downarrow 0} \frac{h(\mu([1-\epsilon]\mathbb{P}_n + \epsilon\delta_{x_i})) - h(\mu(\mathbb{P}_n))}{\epsilon}, \quad \epsilon \in (0,1] \tag{D.30}$$

Equation (D.30) implies that the influence function of $h(\mu(\mathbb{P}_n))$ is the Gateaux derivative at the empirical measure $\mathbb{P}_n$ in the direction of δ_x. This representation means that calculating the sample analog of the influence function does not require any analytical derivations but can be approximated numerically using finite differences for small enough spacing ϵ. Successive approximations of the chain rule imply that the sample analog influence function of any estimator that is a smooth function of averages (it has continuous partial derivatives) can be approximated using equation (D.30). Functions of averages is a large class of estimators that includes misallocation statistics like S and also moments, variance, standard deviation, and correlation between two random variables.

Note that (D.30) is an estimator of the influence function. The proof is constructive and doesn't require any appeal to empirical process theory. Developing general estimators of influence functions requires tools from the theory of empirical processes because the influence function is an infinite dimensional parameter, but such a general approach is beyond the scope of this paper.

The influence function of yearly averages (time fixed effects) One way is to estimate the yearly means using a regression with time dummy variables and then obtain the influence function using the GMM formulas. The time dummy coefficients are parameters, and then we can use the standard von Mises expansion. The influence function derived from the GMM formula is

$$n(X_t'X_t)^{-1}X_t\left(x_{it^*}-\sum_{i,t^*}\frac{1}{N_t}X_{it^*}\mathbb{1}\{t^*=t\}\right) \tag{D.31}$$

$$=nN_t^{-1}\mathbb{1}\{t^*=t\}\left(x_{it^*}-\sum_{i,t^*}\frac{1}{N_t}X_{it^*}\mathbb{1}\{t^*=t\}\right) \tag{D.32}$$

Where, X_t is the dummy column for t.

Another approach is to calculate means from subsamples. Let θ_t denote the estimate of the expectation of X, $\theta_t = (1/N_t)\sum_i X_{it}$. The influence function of θ_t is

$$\psi_{\theta_t}(x_{i\tilde{t}}) = \begin{array}{rl} \frac{n}{N_t}\left(x_{i\tilde{t}}-(1/N_t)\sum_i X_i\right) & \text{if } t^*=t \\ 0 & \text{if } \tilde{t}\neq t \end{array} \tag{D.33}$$

In therms of probabilities, the yearly average estimator is

$$\theta_t(\mathbb{P}_n)=\frac{\sum_{i,t^*}p_{it^*}X_{it^*}\mathbb{1}\{t^*=t\}}{\sum_{i,t^*}p_{it^*}\mathbb{1}\{t^*=t\}} \tag{D.34}$$

Therefore, for $t^*\neq t$

$$\frac{\theta_t([1-\epsilon]\mathbb{P}_n + \epsilon\delta_{x_{it^*}}) - \theta_t(\mathbb{P}_n)}{\epsilon} = \frac{\theta_t(\mathbb{P}_n) - \theta_t(\mathbb{P}_n)}{\epsilon} = 0, \quad t^* \neq t \tag{D.35}$$

For $t^* = t$

$$\begin{aligned}
&\frac{\theta_t([1-\epsilon]\mathbb{P}_n + \epsilon\delta_{x_{it^*}}) - \theta_t(\mathbb{P}_n)}{\epsilon} = \\
&\left(\frac{\epsilon x_{it^*} + (1-\epsilon)\sum_{i,t^*} p_{it^*} X_{it^*} \mathbb{1}\{t^* = t\}}{\epsilon + (1-\epsilon)\sum_{i,t^*} p_{it^*} \mathbb{1}\{t^* = t\}} - \frac{\sum_{i,t^*} p_{it^*} X_{it^*} \mathbb{1}\{t^* = t\}}{\sum_{i,t^*} p_{it^*} \mathbb{1}\{t^* = t\}}\right) / \epsilon = \\
&\left(\frac{\epsilon x_{it^*} + (1-\epsilon)\sum_{i,t^*} \frac{1}{n} X_{it^*} \mathbb{1}\{t^* = t\}}{\epsilon + (1-\epsilon)\sum_{i,t^*} \frac{1}{n} \mathbb{1}\{t^* = t\}} - \frac{\sum_{i,t^*} \frac{1}{n} X_{it^*} \mathbb{1}\{t^* = t\}}{\sum_{i,t^*} \frac{1}{n} \mathbb{1}\{t^* = t\}}\right) / \epsilon \\
&= \left(\frac{\epsilon x_{it^*} + (1-\epsilon)\sum_{i,t^*} \frac{1}{n} X_{it^*} \mathbb{1}\{t^* = t\}}{\epsilon + (1-\epsilon)\frac{N_t}{n}} - \frac{\sum_{i,t^*} \frac{1}{n} X_{it^*} \mathbb{1}\{t^* = t\}}{\frac{N_t}{n}}\right) / \epsilon \\
&= \left(\frac{\frac{N_t}{n}\epsilon x_{it^*} + \frac{N_t}{n}(1-\epsilon)\sum_{i,t^*} \frac{1}{n} X_{it^*} \mathbb{1}\{t^* = t\} - \left(\epsilon + (1-\epsilon)\frac{N_t}{n}\right)\sum_{i,t^*} \frac{1}{n} X_{it^*} \mathbb{1}\{t^* = t\}}{\left(\epsilon + (1-\epsilon)\frac{N_t}{n}\right)\frac{N_t}{n}}\right) / \epsilon \\
&= \left(\frac{\frac{N_t}{n}\epsilon x_{it^*} - \epsilon\sum_{i,t^*} \frac{1}{n} X_{it^*} \mathbb{1}\{t^* = t\}}{\left(\epsilon + (1-\epsilon)\frac{N_t}{n}\right)\frac{N_t}{n}}\right) / \epsilon \\
&= \left(\frac{\epsilon x_{it^*} - \epsilon\sum_{i,t^*} \frac{1}{N_t} X_{it^*} \mathbb{1}\{t^* = t\}}{\left(\epsilon + (1-\epsilon)\frac{N_t}{n}\right)}\right) / \epsilon \\
&= \frac{x_{it^*} - \sum_{i,t^*} \frac{1}{N_t} X_{it^*} \mathbb{1}\{t^* = t\}}{\left(\epsilon + (1-\epsilon)\frac{N_t}{n}\right)}
\end{aligned}$$

$$\Rightarrow \lim_{\epsilon \downarrow 0} \frac{x_{it^*} - \sum_{i,t^*} \frac{1}{N_t} X_{it^*} \mathbb{1}\{t^* = t\}}{\left(\epsilon + (1-\epsilon)\frac{N_t}{n}\right)} = \frac{n}{N_t}\left(x_{it^*} - \sum_{i,t^*} \frac{1}{N_t} X_{it^*} \mathbb{1}\{t^* = t\}\right) \tag{D.36}$$

The limit in (D.36) is the same formula as the fixed effects regression formula in (D.32). Therefore, the influence function of yearly means (or any subsample mean) can be approximated numerically by finite differences using the empirical measure $\mathbb{P}_n$.

Comparison with resampling methods. An alternative approach to calculating the standard errors of estimator S_n is using resampling methods such as subsampling or bootstrapping. Note that resampling methods are another approach to perturb the empirical measure $\mathbb{P}_n$: bootstrap resamples with replacement, so an observation sampled twice has an empirical probability $2/N$ instead of $1/N$ under $\mathbb{P}_n$; subsampling decreases the sample size so the sampled observations have empirical probability $1/B > 1/N$, with B being the subsample size. For a formal analysis of the bootstrap as a perturbation method see Hong and Li (2020).

An advantage of the resampling approach versus the analytical influence function approach is its simplicity: devise a resampling algorithm and re-estimate the model parameters on different samples. The numerical approximation of the influence function makes it as simple as resampling without giving up precision (see Section D.1.3 for an assessment of the precision of the numerical influence function in this dataset).

The hybrid influence function approach. The advantage of the influence function approach is that it permits using the numerical approximation approach for the estimation steps that are easy to calculate and the analytical approach for the computationally intensive steps. For example, if an estimation step involves estimating a dynamic model using SMD (such as Bazdresch et al., 2018) and another a complicated function of means such as the misallocation statistic S_n, then the influence function for the dynamic parameters can be calculated using analytical formulas, while the influence function for S_n can be calculated using the approximate method and then combine the two influence functions to conduct inference on counterfactuals. In contrast, the bootstrap would require re-estimating the dynamic model, which is computationally inefficient.

D.1.2 Additional influence function formulas

Quantiles. Let $p \in (0,1)$ and the p quantile estimator $q_p = F^{-1}(p)$, where F is the cumulative distribution function. The influence function at an observation x_i is (see Van der Vaart, 1998, corollary 21.5):

$$\psi_{q_p}(x_i) = -\frac{\mathbb{1}\{x_i \leq \widehat{q}_p\} - p}{\widehat{f}(\widehat{q}_p)}, \tag{D.37}$$

where $\widehat{f}(\widehat{q}_p) > 0$ is the kernel estimate of the probability density function at $\widehat{q}_p$. I estimate the density using the Epanechnikov kernel with bandwidth equal to $h = \frac{0.9 * min\{\sqrt{\hat{\sigma}^2}, IQR/1.349\}}{N^{1/5}}$ which is the default in STATA. The median used to estimate α is a special case of the quantile estimator for $p = 0.5$.

M-estimators, Z-estimators, GMM A large class of estimators that includes the linear and non-linear regression take the following form:

$$\frac{1}{n}\sum_{i=1}^{n} m(x_i, \theta_{K\times 1}) = 0_{M\times 1}. \tag{D.38}$$

These estimator are often called Z-estimators (from zero; see Van der Vaart, 1998; Kosorok, 2008), or M-estimators (Tsiatis, 2006). Many extremum estimators, such as the MLE, the GMM, Simulate Minimum Distance, also belong to this class through their first-order optimality conditions. For example, the OLS estimator (such as the within fixed-effects estimator used for the estimation of γ), can be recast as

$$\frac{1}{n}\sum_{i=1}^{n} (y_i - x_i\beta_{K\times 1})\epsilon_i = 0_{K\times 1}. \tag{D.39}$$

In the just-identified case, where $M = K$, the influence function of the estimator is (see Tsiatis, 2006, ch. 3.2).

$$\psi(x_i,\theta)_{1\times K} = -\left(\frac{1}{n}\sum_i \frac{\partial m(x_i,\theta)}{\partial\theta}_{K\times K}\right)^{-1} m(x_i,\theta)_{K\times 1} \equiv -(G_{K\times K})^{-1} m(x_i,\theta)_{K\times 1} \quad \text{(D.40)}$$

In over-identified GMM models, a weighting matrix Ξ is necessary, and these estimators have the following influence function (see the chapter GMM and Minimum Distance estimation in Wooldridge, 2002, ch. 14):

$$\psi(x_i,\theta)_{K\times 1} = -(G'\Xi_{M\times M}G_{M\times K})^{-1}G'\Xi m(x_i,\theta)_{K\times 1} \quad \text{(D.41)}$$

If the model is just identified then $M = K$ and $\Xi = I$ the influence function becomes

$$\begin{aligned}
\psi(x_i,\theta) &= -\ (G'IG_{K\times K})^{-1}G'Im(x_i,\theta)_{K\times 1} \\
&= -\ (G'G_{K\times K})^{-1}G'm(x_i,\theta)_{K\times 1} \\
&\overset{G \text{ invertible as a square matrix}}{=} G^{-1}G'^{-1}G'm(x_i,\theta)_{K\times 1} \\
&= G^{-1}m(x_i,\theta)_{K\times 1}
\end{aligned}$$

which is identical to expression (D.40). If Ξ equals the optimal weighting matrix which is

$$W = \left[\frac{1}{n}\sum_i m(x_i,\theta)_{K\times 1}m(x_i,\theta)'\right]^{-1} \quad \text{(D.42)}$$

Then the influence function is:

$$\psi(x_i,\theta) = -(G'W_{M\times M}G_{M\times K})^{-1}G'Wm(x_i,\theta)_{K\times 1}$$

which matches the formula for the GMM in (Newey and McFadden, 1994, equation 3.6).

Maximum Likelihood Estimators (MLE). As discussed in (Tsiatis, 2006, ch. 3.2), the MLE, under regularity conditions, is a special case of a z-estimator, where the moment condition is the score S_θ of the log likelihood $l(x_i, \theta)$: $m(x_i, \theta) = S_\theta(x_i, \theta) = \frac{\partial l(x_i,\theta)}{\partial \theta}$. The influence function is

$$\psi(x_i, \theta) = -\left(\frac{1}{n}\frac{\partial^2 l(x_i, \theta)}{\partial \theta \partial \theta'}\right)^{-1} S_\theta(x_i, \theta)$$

, where the first term is the inverse of the expectation of the Hessian, which is the information matrix.

D.1.3 Equivalence among the different approaches to inference

This section compares three approaches to inference: the analytical influence function, the approximate influence function, and the bootstrap. More specifically, I focus on inference on the three counterfactuals of interest in Table 2 in the text.

I first compare the standard errors of each counterfactual of interest calculated using the analytical influence function and the bootstrap. The analytical influence function uses formulas for the influence function of the median estimator, the fixed-effects regression, and the means in the TFP expression to calculate the standard error of the counterfactual quantities, containing variation from all estimation stages. It is the approach used to calculate the standard errors in section 5. The first row of Table D.2 duplicates the standard errors of Table 2 in the text for the sake of completeness, while the second row of Table D.2 reports standard errors from 1,000 bootstrap samples, each drawn from the original sample with replacement. The standard errors calculated using the bootstrap align with the standard errors from the influence function approach.

Panels B and C of Table D.2 compare the analytical to the numerical influence function approach. The numerical approach applies only to the last stage of the estimation, which

Table D.2: Analytical influence, numerical influence, and the bootstrap

Approach to inference	Inframarginal	Marginal	Total
Panel A.	Standard error of each quantity above		
	All stage variation (including from α, γ)		
Influence function formula Ψ	0.242	0.315	1.385
Bootstrap (1,000 samples)	0.256	0.323	1.372
Panel B.	Last stage variation only (not from α, γ)		
Influence function formula Ψ	0.2371	0.3151	1.3761
Numerical influence $\widetilde{\Psi}(\epsilon = 10^{-9})$	0.2371	0.3151	1.3761
Numerical influence $\widetilde{\Psi}(\epsilon = 10^{-8})$	0.2371	0.3151	1.3761
Numerical influence $\widetilde{\Psi}(\epsilon = 10^{-7})$	0.2370	0.3151	1.3760
Numerical influence $\widetilde{\Psi}(\epsilon = 10^{-6})$	0.2368	0.3148	1.3747
Numerical influence $\widetilde{\Psi}(\epsilon = 10^{-5})$	0.2344	0.3118	1.3622
Panel C.	Accuracy of the approximate influence vector $\widetilde{\Psi}$		
$\text{MSE}[\Psi - \widetilde{\Psi}(\epsilon = 10^{-9})]$	$7*10^{-9}$	$6*10^{-9}$	$2*10^{-7}$
$\text{MSE}[\Psi - \widetilde{\Psi}(\epsilon = 10^{-8})]$	$6*10^{-7}$	$4*10^{-7}$	$1*10^{-5}$
$\text{MSE}[\Psi - \widetilde{\Psi}(\epsilon = 10^{-7})]$	$4*10^{-5}$	$6*10^{-5}$	$1*10^{-3}$
$\text{MSE}[\Psi - \widetilde{\Psi}(\epsilon = 10^{-6})]$	$4*10^{-3}$	$6*10^{-3}$	$1*10^{-1}$
$\text{MSE}[\Psi - \widetilde{\Psi}(\epsilon = 10^{-5})]$	$4*10^{-1}$	$6*10^{-1}$	$1*10^{+1}$

This table presents the assessment of the precision of three approaches (analytical influence function, numerical influence function, and the bootstrap) to calculate standard errors for the quantities of Table 2 in the text. MSE refers to the distance measure of the mean squared error, which is used to evaluate the degree of similarity between the analytical and the numerical influence vector. Standard errors are heteroskedasticity robust. Computations are carried out using the Julia programming language (Bezanson et al., 2017).

is a function of means, so Panel B reports standard errors that take into account variation only from the last stage of the estimation: the influence function includes only the first term in expansion (11), ignoring variation from the estimation of α and γ. The first row of panel B reports the standard errors from the analytical influence function. As expected, the standard error from the last stage is smaller than the standard error from the variance of all the stages (first row of panel A). Still, the difference is negligible, indicating that the bulk of the variation comes from the last stage.

The rest of the rows of panel B report standard errors from the last stage using the numerical approximation to the influence function. Since the numerical approach is a finite difference, it depends on the size of the spacing ϵ, and I report results for different magnitudes of spacing. Note that for spacing 10^{-7} or less, the standard errors from the numerical influence function are identical to the analytical influence function to three digits. For spacings 10^{-6} or 10^{-5}, the standard error's accuracy deteriorates but negligibly.

Panel C compares the vector of the analytical to the numerical influence function using the distance measure of mean squared error (MSE). Each vector has a length equal to the sample size, and the two influence vectors are very close for spacing 10^{-7} or less. Also note that a decrease in the spacing of one order of magnitude results in two orders of magnitude decrease in the MSE, as expected, since the mean squared error converges at rate ϵ^2. The main takeaway of Table D.2 is that the influence function approach to inference is precise and admits a numerical approximation that is easy to implement without giving up precision.

Bibliography

Bazdresch, S., R. J. Kahn, and T. M. Whited (2018). Estimating and testing dynamic corporate finance models. *The Review of Financial Studies 31*(1), 322–361.

Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah (2017). Julia: A fresh approach to numerical computing. *SIAM Review 59*(1), 65–98.

David, J. M. and V. Venkateswaran (2019). The sources of capital misallocation. *American Economic Review 109*(7), 2531–2567.

Fakos, A. (2023). Subsidies with misallocation. *SSRN working paper No. 3483259*.

Fernholz, L. T. (1983). *von Mises Calculus for Statistical Functionals*, Volume 19. Springer-Verlag.

Fisher, A. and E. H. Kennedy (2021). Visually communicating and teaching intuition for influence functions. *The American Statistician 75*(2), 162–172.

Hines, O., O. Dukes, K. Diaz-Ordaz, and S. Vansteelandt (2022). Demystifying statistical learning based on efficient influence functions. *The American Statistician*, 1–13.

Hong, H. and J. Li (2020). The Numerical Bootstrap. *The Annals of Statistics 48*(1), 397–412.

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

Ichimura, H. and W. K. Newey (2022). The influence function of semiparametric estimators. *Quantitative Economics 13*(1), 29–61.

Kennedy, E. H. (2016). Semiparametric theory and empirical processes in causal inference. In H. He, P. Wu, and D.-G. D. Chen (Eds.), *Statistical Causal Inferences and Their Applications in Public Health Research*, pp. 141–167. Springer International Publishing.

Kennedy, E. H. (2022). Semiparametric doubly robust targeted double machine learning: A review. *arXiv preprint arXiv:2203.06469*.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*, Volume 61. Springer.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engel and D. L. McFadden (Eds.), *Handbook of econometrics*, Volume 4, pp. 2111–2245. Elsevier.

Robins, J., L. Li, E. Tchetgen, and A. W. van der Vaart (2009). Quadratic semiparametric von mises calculus. *Metrika 69*, 227–247.

Tsiatis, A. A. (2006). *Semiparametric theory and missing data*. Springer New York.

van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

Van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge university press.

Whited, T. M. and J. Zhao (2021). The misallocation of finance. *The Journal of Finance 76*(5), 2359–2407.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT press.