A Non-Parametric Estimation of Productivity with Idiosyncratic and Aggregate Shocks: The Role of Research and Development (RD) and Corporate Tax

Bournakis, Ioannis and Tsionas, Mike G.

SKEMA Business School, Lancaster University

21 July 2023

A Non-Parametric Estimation of Productivity with Idiosyncratic and Aggregate Shocks: The Role of Research and Development (R&D) and Corporate Tax

Abstract

We developed a non-parametric technique to measure Total Factor Productivity (TFP). Our paper has two major novelties in estimating the production function. First, we propose a productivity modelling with both idiosyncratic firm factors and aggregate shocks within the same framework. Second, we apply Bayesian Markov Chain Monte Carlo (MCMC) estimation techniques to overcome restrictions associated with monotonicity between productivity and variable inputs and moment conditions in identifying input parameters. We implemented our methodology in a group of 4286 manufacturing firms from France, Germany, Italy, and the United Kingdom (2001-2014). The results show that: (i) aggregate shocks matter for firm TFP evolution. The global financial crisis of 2008 caused severe adverse effects on TFP albeit short in duration; (ii) there is substantial heterogeneity across countries in the way firms react to changes in R&D and taxation. German and U.K. firms are more sensitive to fiscal changes than R&D, while Italian firms are the opposite. R&D and taxation effects are symmetrical for French firms; (iii) the U.K. productivity handicap continued for years after the financial crisis; (iv) industrial clusters promote knowledge diffusion among German and Italian firms.

Keywords: Total Factor Productivity (TFP), Control Function, Non-parametric Bayesian Estimation, Markov Chain Monte Carlo(MCMC), Research and Development (R&D), Taxation, European firms

JEL Classification: C11, D24, Q55, H21

Data Availability

The data underlying this article cannot be shared publicly due to confidentiality imposed by Bruegel Organization, which is the data provider. Nonetheless, the data will be shared with the editor and journal subscribers on reasonable request to the corresponding author.

1 Introduction

The task of estimating a firm's production function is perhaps one of the most challenging in applied industrial organization and econometrics. The main challenge is to identify input choices conditional on productivity shocks, which despite being known to the firm manager, remain invisible to the researcher. This identification problem is well noted in Marschak and Andrews (1944) and labelled in the literature as "transmission bias" (Gandhi et al., 2020). More importantly, the simultaneity between unobserved shocks and input selection makes ordinary least squares (OLS) an inappropriate estimator for deriving the labour and capital coefficients of the production function. To mitigate "transmission bias" from the correlation of unobserved productivity shocks and input selection, the literature has considered various approaches. One candidate for resolving the problem is the use of instruments for identifying input selection with exogenously determined variables (i.e. prices). The practical implementation of this approach stumbles on the lack of valid instruments, which eventually causes bias similar to OLS (Griliches and Mairesee, 1998).[1] The use of index numbers (Caves et al, 1982) has often been a good alternative for researchers seeking to measure productivity assuming flexible forms of underlying production functions. Despite the attractive properties of index numbers, problems arise when output and input markets are imperfect (Van Beveren, 2010)).[2]

More recently, the literature mitigates "transmission bias" by employing control functions, which include demand of either investment (Olley and Pakes, 1996, OP hereafter) or intermediate materials (Ackerberg et al.2015, ACF hereafter; Levinsohn and Petrin, 2003, LP hereafter; Petrin, Wooldridge, 2009; Petrin and Sivadasan, 2013; Hu et al., 2020) to approximate unobserved productivity shocks. These semi-parametric approaches assume that investment and (or) materials are monotone in productivity movements. Nonetheless, if monotonicity does not hold the proxy function cannot be inverted, thus the productivity function becomes undefinable. Violating monotonicity condition [3] remains an empirical challenge with scholars exploring alternative ways to estimate production input coefficients and measure productivity. This is the starting point of our paper that

[1]Despite the increased availability of firm-level data over the last twenty-five years, unit price data are still scarce. It is mainly for this reason that instrumental variable (IV) estimators are not recommended for estimating production functions.

[2]If capital is not always fully utilized and (or) there are labour adjustment costs due to imperfect competition, input returns are not equal to the revenue shares observed in the data. To find out more about the advantages and disadvantages of purely parametric methods of firm-level productivity estimation, see Van Beveren (2010).

[3]Monotonicity holds when a productivity shock raises marginal revenue, and the profit-maximizing firm increases output up to the point that the marginal revenue product equals prices. These are assumptions 4 and 5 in ACF and imply that as a productivity shock raises the marginal product of capital increases and firms invest more. This is not the case if: (i) capital is subject to adjustment costs, (ii) there are labour market frictions and (iii) producers exert market power. If one of the three conditions holds, prices rise without output increasing; thus, productivity is not monotone in input use. Studies that apply LP and OP rarely test explicitly for monotonicity, rather assume that materials or investment are continuously increasing in productivity.

develops an estimation framework with two novelties. First, in our modelling of productivity we do not rely on monotonicity or any other specific control function for identifying productivity shocks, instead, we assume that productivity is a function of all variable inputs (i.e. labour, materials and services) of the production function. Second, the endogenous process that governs productivity consists of two components, an aggregate shock that is common to all firms at a given period and an idiosyncratic component that is firm-specific. Furthermore, these productivity components are augmented with two covariates that are expected to be important productivity determinants. Building upon the propositions of the innovation and growth literature (Chamley, 2001; Gentry and Hubbard, 2004; Gordon and Lee, 2006), productivity function includes Research and Development (R&D) and corporate taxes (fiscal environment). Our proposed methodology is a more realistic set-up of the productivity parameter, while our estimation technique mitigates some of the big challenges encountered in the literature of production function estimations.

Like semi-parametric techniques (OP, LP, ACF), our approach assumes a Markovian AR(1) process for productivity. Nonetheless, we follow a different modelling and estimation methodology. To start with, we do not consider productivity as a element of the demand of observables (materials or investment), instead, the functional forms of productivity is identified using the methods of sieves (Gallant and Nychka, 1987; Olley and Pakes, 1993;Gandhi et al.,2020). The sieves method is a non-parametric procedure that approximates the unknown functional forms of the productivity equation by using polynomial orders. Another key novelty in the modelling of the productivity parameter is the augmentation with factors that affect physical efficiency (De Loecker and Goldberg, 2013). R&D investment takes the form of process innovation, which enhances firm productivity (Griffith et al. 2004; O'Mahony and Vecchi 2009; Hall and Rosenberg, 2010). Changes in the corporate tax rate captures changes in the user cost of capital, which can result in adverse effects in investment decisions(Jacob, 2021; Romero-Jordán et al. 2020). Less investment can potentially slow down technological progress as the latter is primarily embodied in the purchase of new capital assets. Previous studies also suggest the inclusion of exogenous firms characteristics in the evolution of productivity [4],nonetheless conditions in the fiscal environment has not been one of them and second within our estimation framework, these characteristics are allowed to affect both the aggregate and the idiosyncratic component of productivity.[5]We use a Bayesian estimator with Markov Chain Monte Carlo (MCMC) methods to estimate the coefficients of the production function. This estimator is also known as the Particle Filtering technique. We apply this novel estimation into a groups of 4286 manufacturing firms from four European countries over 2001-2014. Comparing our results with estimates of productivity from a modified "canonical" ACF approach, we found that the latter

[4]Doraszelski and Jaumandreu (2013) include R&D in the endogenous productivity process, while De Loecker (2013) searches for learning by exporting gains by augmenting the productivity function with the export status of the firm.

[5]R&D and corporate tax are not the only factors that potentially drive productivity but they are among the most relevant in an advanced economy context and fully observable in the current data set. Changes in the trade policy regime can also improve productivity through intensified competition but this scenario is less relevant in the present context.

tends to overestimate productivity growth during this period. We attribute this difference to the absence of a systematic treatment of adverse macro effects in ACF that are apparent in the period under study (i.e., the 2008 global financial crisis) and are expected to drive productivity.

Despite the significance of aggregate shocks in governing firm TFP, they are found to be short in duration than productivity changes originated from firm idiosyncratic factors. Firm productivity is highly heterogeneous across countries in response to changes in R&D and taxation. German and UK firms are more sensitive to fiscal changes than R&D but the opposite is true for Italian firms. Another key remark from our results is the productivity stagnation of U.K. firms. The slowdown started before the 2008 global financial crisis but U.K. firms have not recovered since then. Other findings include the importance of industrial clusters in promoting knowledge diffusion among German and Italian firms that boost overall productivity. The rest of the paper is organized as follows: section two sets up the model. To test how our Bayesian estimates perform relative to results from "canonical" models, we replicate ACF estimates from a model that augments productivity with firm-specific shifters and aggregate shocks. Section two compares TFP growth rates between ACF and Bayesian techniques. The ACF estimates are close approximations of our modeling framework, except that the former rely on monotonicity assumptions and Generalized Methods of Moment (GMM) techniques. Section three shows the estimation steps of the new model. Section four discusses the data and empirical results in section four and section five concludes the paper.

2 Estimation Framework

2.1 Model Set-up

The objective is to estimate the following production function, which is specified in a generic form as:

$$y_{it} = f(x_{it}, k_{it}; \beta_i) + \omega_{it} + v_{it}, \tag{1}$$

where y_{it} is log output; $x_{it} = [x_{it,1}, ..., x_{it,3}]'$ are logs of variable inputs: $x_{it} \in \mathbb{R}$ including labour, intermediate materials and services[6]; $k_{it} \in \mathbb{R}$ is log of capital, the quasi-fixed input. Parameters $\beta_i \in \mathcal{B} \subseteq \mathbb{R}^{d_\beta}$ are to be estimated. Subscript $i \in \mathbb{I} = \{1, ..., n\}$ indexes firms and subscript $t \in \mathbb{T} = \{1, ..., T\}$ indexes time. Parameter ω_{it} stands for productivity and v_{it} is an error term with standard statistical properties $N(0, \sigma^2)$. The specific form of $f(\cdot)$ is a Cobb-Douglas function to conform to ACF, although any other functional forms can be also accommodated without further implications.

[6]Materials and services can be further decomposed into domestic and imported. In the estimations, we use the aggregate values.

2.2 Control Function Approach with Modifications

Before showing our estimation framework, we outline how "canonical" semi-parametric models use control function approaches to estimate production function (1). Without losing generality, OP assumes capital (k_{it}) as a state variable determined in period $t-1$ and function of $k_{it} = \kappa(k_{it-1}, i_{it-1})$. In OP and LP set-up, labour is a free variable that allows to change in period t. In the first stage of OP, OLS is applied to estimate $y_{it} = \beta_l l_{it} + \phi(k_{it}, i_{it}) + \epsilon_{it}$ and retrieve $\hat{\beta}_l$ and $\hat{\phi}(k_{it}, i_{it})$. The control function in OP is investment $i_{it} = \mathcal{I}(k_{it}, l_{it}, \omega_{it})$ and the coefficient of capital, β_k is recovered in the second stage by regressing non-parametrically y_{it} on $\hat{\beta}_l + \hat{\phi}(k_{it}, i_{it})$ assuming that: $E[\epsilon_{it}|i_{it}] = E[y_{it} - \beta_l - \phi(k_{it}, i_{it})|i] = 0$. Analogously, LP specifies materials, $m_{it} = \mathcal{M}(k_{it}, l_{it}, \omega_{it})$, as control function.

The second stage in OP, LP and ACF specifies productivity as a Markovian process: $\omega_{it} = g(\omega_{it-1}) + \xi^{\omega}_{it}$, where ξ^{ω}_{it} is an idiosyncratic error term uncorrelated with ω_{it-1}. In OP and LP, ω_{it} is expected to be strictly monotonous to $\mathcal{I}(\cdot)$ and $\mathcal{M}(\cdot)$, respectively. ACF also choose $\mathcal{M}(\cdot)$ as a proxy function but treats labour predetermined in period $t-s$, with $0 < s < 1$ and before selection of materials. This implies that labour also depends on unobserved productivity. In ACF, ω_{it}, β_l and β_k are recovered in the second stage. The original formulation of semi-parametric techniques has two limitations,(i) ω_{it} is only defined as function of observables,(ii) omitted variables bias in the endogenous process of ω_{it}. We discuss these limitations and suggest a modified ACF approach that can be later used as a means fo comparison for the TFP growth obtained from the MCMC Bayes estimator.

First, control function approaches assume that ω_{it} can only be defined as a function of observables, for example, $\omega_{it} = \mathcal{M}^{-1}(k_{it}, l_{it}, m_{it})$. This requires strictly monotonicity of either $\mathcal{I}(\cdot)$ or $\mathcal{M}(\cdot)$ in ω_{it} and matrix inversion. Second, the ξ^{ω}_{it} idiosyncratic term in the AR(1) process of ω_{it} is assumed to be uncorrelated with ω_{it-1} but this potentially causes omitted variables bias (De Loecker and Goldberg, 2013; Doraszelski and Jaumandreu, 2013; Iyoha, 2020) by treating systematic productivity drivers as pure noise. Our novel modelling of ω_{it} allows R&D and taxation to enter the vectors of predetermined covariates. In an ACF notation, this augmentation means that: $\omega_{it} = g(\omega_{it-1}, \Upsilon_{it-1}) + \epsilon_{it}$, where Υ_{it-1} is $1 \times \vartheta$ vector of predetermined firm-specific characteristics such as R&D and tax rate.[7] Table 3 replicates ACF estimates from a Cobb-Douglass production function where $g(\cdot)$ is augmented with term Υ_{it-1}, containing R&D, tax rate and year dummies for 2007 and 2008. Dummy variables capture the effects of the global financial crisis in firm productivity. This modified ACF (MACF, hereafter) specification with firm-specific covariates and time fixed effects in the evolution of ω_{it} is conceptually a close variant to our model that follows.

[7] We assume these firm activities to be exogenous (or predetermined) to productivity and allow them to enter the function in lags. Considering contemporaneous values exacerbates endogeneity bias as these activities are expected to be correlated with ξ^{ω}_{it}.

2.3 A Novel Modelling of Firm Productivity

We now present the key features of our specification in which ω_{it} also follows an AR(1), while it is a function of both firm specific and aggregate shocks. Parameter ω_{it} in (1) is composed by:

$$\omega_{it} = \lambda_i \Omega_t + \Phi_\omega(z_{it}; \gamma_i) + \tilde{\omega}_{it} + \xi_{it}^{\omega}, \forall i \in \mathbb{I}, t \in \mathbb{T}, \tag{2}$$

where Ω_t refers to an aggregate productivity shock with common factor loadings λ_i; $\tilde{\omega}_{it}$ represents the idiosyncratic productivity (firm-specific) component; Z_t and z_{it} are $d_Z \times 1$ and $d_z \times 1$ vectors of industry-wide and firm specific covariates that explain movements in Ω_t and $\tilde{\omega}_{it}$, respectively. Parameters γ_i are defined as $\gamma_i \in \Gamma \subseteq \mathbb{R}^{d_\gamma}$. The idiosyncratic component of productivity, $\tilde{\omega}_{it}$ follows an AR(1) process [8]:

$$\tilde{\omega}_{it} = \varrho_i \tilde{\omega}_{i,t-1} + \xi_{it}^{\tilde{\omega}} \, \forall i \in \mathbb{I}, t \in \mathbb{T}, \tag{3}$$

Similarly, the aggregate productivity shock Ω_t evolves as:

$$\Omega_t = \rho \Omega_{t-1} + \Phi_\Omega(Z_t; \lambda_i) + \xi_t^{\Omega}, \ \xi_t^{\Omega} \sim i.i.d\, N(0, \sigma_{\xi^\Omega}^2) \, \forall t \in \mathbb{T}, \tag{4}$$

Within the current context the covariates in Z_t and z_{it} include R&D and tax rates,[9] while Φ_ω and Φ_Ω are degree three polynomials. To account for non-linearities in the way aggregate shocks impact productivity, we modify the benchmark specification (2) as follows:

$$\omega_{it} = \lambda_i \Omega_t + \tfrac{1}{2}\psi_i \Omega_t^2 + \Phi_\omega(z_{it}; \gamma_i) + \tilde{\omega}_{it} + \xi_{it}^{\omega}, \forall i \in \mathbb{I}, t \in \mathbb{T}, \tag{5}$$

Differentiating (5) with respect to Ω_t, we obtain $\lambda_i + \psi_i \Omega_t$. Accordingly, positive values of λ_i and γ_i indicate that $\tilde{\omega}_{it}$ co-moves with aggregate productivity shocks Ω_t and firms are likely to be "unicorns". In this set-up, macro shocks create "unicorns", which are identified as firms with exceptional productivity performance. Finally, (5) is augmented with an error term ξ_{it}, which is defined as:

$$\xi_{it}^{\omega} \sim \mathcal{N}(0, \sigma_{\xi^\omega}^2). \tag{6}$$

To sum up, the formulation of ω_{it} in equations (2) to (6) is the first one in the productivity literature -to the best of our knowledge that accommodates with the same framework both firm-specific z_{it} and industry-wide Z_t factors in the evolution of ω_{it}. Using this approach, we are better able to capture the true realizations of firm productivity, which are often influenced not only by

[8]The autoregressive process of ω_{it} is also referred to the literature as the law of motion of productivity.

[9]As mentioned previously, z_{it} can include other variables such as prices of input and output, as well as other relevant productivity shifters. Our framework allows for such modifications subject to data availability in each application. We do not include input or output prices due to lack of data. We are restricted to R&D and taxes due to their importance as productivity drivers.

decisions within firms, but also by macroeconomic factors such as cyclical effects and policy changes.

2.4 Bayesian Markov Chain Monte Carlo (MCMC) Estimation

This section defines the other parameters of the production function and shows the estimation procedure of the production function. In estimating (1), we assume that firms minimize costs with regard to inputs and first order conditions (f.o.c) are specified as [10]:

$$x_{it} = \Phi(k_{it}, \omega_{it}, z_{it}; \gamma_i) + V_{it} \, \forall i \in \mathbb{I}, \, t \in \mathbb{T}, \tag{7}$$

where $z_{it} \in \mathbb{R}^{d_z}$ includes input prices and firm-specific productivity drivers. Output y_{it} also enters the f.o.c through the production function alone. For k f.o.c's in (7), we assume $\Phi_k(\cdot)$ is a low order polynomial. The vector field is defined as: $\Phi : \mathbb{R}^k \times \mathbb{R} \times \mathbb{R}^{d_z} \to \mathbb{R}^k$ with V_{it}: being the scalar vector error term. The stochastic specification of the error term is defined as:

$$\mathbf{v}_{it} \equiv [v_{it}, V_{it}, \xi_{it}^{\omega}, \xi_{it}^{\tilde{\omega}}]' \sim \text{i.i.d}\, \mathcal{N}(\mathbf{0}, \Sigma). \tag{8}$$

Below we discuss the assumptions about parameters in (1) and their associated priors. Notably, our coefficients are firm-specific to account for unobserved firm heterogeneity, which is expected to prevalent in the panel structure of the data. Below, we list the priors:

1. In the case of β_i that imposes monotonicity at certain points, set B includes such restrictions. Other than that
$$p(\beta) \propto 1_{\mathcal{B}}(\beta), \tag{9}$$
where $1_{\mathcal{S}}(x) = 1$ if $x \in \mathcal{S}$ and zero otherwise.

2. The factor loadings λ_i in Ω_t and the autoregressive parameter ϱ_i follow:
$$\lambda_i \sim \text{i.i.d}\, \mathcal{N}(0, \sigma_\lambda^2), \varrho_i \sim \text{i.i.d}\, \mathcal{N}(0, \sigma_\varrho^2). \tag{10}$$

3. Parameter γ_i follows random coefficient specifications:
$$\gamma_i \sim \mathcal{N}(\bar{\gamma}, \Sigma_\gamma) \forall i \in \mathbb{I}. \tag{11}$$

$$\bar{\gamma} \sim \mathcal{N}(\mathbf{0}, h^2\mathbf{I}), \bar{\gamma}, \gamma_\Omega \sim \mathcal{N}(\mathbf{0}, h^2\mathbf{I}), \tag{12}$$

[10] Our estimation approach is sufficiently general to cover all different cases of optimization in f.o.c. This implies that the formulation does not change if we assume that firms are profit maximizers with regard to output.

where h follows the same prior as in prior (14) below. Moreover, Σ and Σ_γ have the inverted Wishart prior:

$$p(\Sigma) \propto |\Sigma|^{-(\bar{n}+dim(\Sigma)+1)/2} \exp\left\{-\tfrac{1}{2}\text{tr}\bar{A}\Sigma^{-1}\right\}, \tag{13}$$

where "tr" is the operator that takes the trace (sum of diagonal elements) of a matrix, and $\bar{A} = \bar{q}\mathbf{I}$ (Zellner, 1971, p. 395).

4. For scale parameters, say σ^2, like σ_λ^2 and σ_ϱ^2:

$$\tfrac{\bar{q}}{\sigma^2} \sim \chi^2_{\bar{N}},\ \bar{q}, \bar{N} \geq 0 \tag{14}$$

see Zellner (1971, p. 371) where $\bar{q}, \bar{N} \geq 0$ are hyperparameters. We take $\bar{n} = 1$ and $\bar{q} = 10^{-4}$ so that the prior is proper but diffuse.

One of the novelties in estimating (1) is that we do not need to define the inverse function of materials or any other variable input. Also, we do not make any other strong assumption about the functional form of ω_{it}, instead, we estimate the system (7) by approximating the unknown functional forms using a method of sieves (Gallant and Nychka, 1987; Olley and Pakes, 1993; Gandhi et al., 2020). The implementation is straightforward and enables inference on functional forms through standard two-stage parametric results (Hahn et al., 2018):

$$\begin{aligned} \Pi_2(\omega_{it}, \boldsymbol{k}_{it}; \gamma_{\Pi_1}) &= \textstyle\sum_{\iota_1=0}^{Q} \cdots \sum_{\iota_{d_k}=0}^{Q} \sum_{\iota_\omega=0}^{Q} k_{it,1}^{\iota_1} \ldots k_{it,d_k}^{\iota_{d_k}} \omega_{it}^{\iota_\omega} \gamma_{\Pi_1,\iota_1,\ldots,\iota_{d_k},\iota_\omega}, \\ &\vdots \\ \Pi_K(\omega_{it}, \boldsymbol{k}_{it}; \gamma_{\Pi_k}) &= \textstyle\sum_{\iota_1=0}^{Q} \cdots \sum_{\iota_{d_k}=0}^{Q} \sum_{\iota_\omega=0}^{Q} k_{it,1}^{\iota_1} \ldots k_{it,d_k}^{\iota_{d_k}} \omega_{it}^{\iota_\omega} \gamma_{\Pi_{d_k},\iota_1,\ldots,\iota_{d_k},\iota_\omega}, \end{aligned} \tag{15}$$

where $\gamma_{\Pi_1,\iota_1,\ldots,\iota_{d_k},\iota_\omega}, \ldots, \gamma_{\Pi_K,\iota_1,\ldots,\iota_{d_k},\iota_\omega}$ are unknown parameters

$$\boldsymbol{\gamma}_\Pi = [\boldsymbol{\gamma}'_{\Pi_2}, \ldots, \boldsymbol{\gamma}'_{\Pi_k}]'$$

The system of equations in (15) is semi-parametric polynomial model. We assume common polynomial orders (Q) for simplicity with the optimal value of Q at 3 using the marginal likelihood criterion which is standard particle filtering in Sequential Monte Carlo (Andrieu et al., 2010).

As the posterior is non-standard (see Appendix A.1) we use Markov Chain Monte Carlo (MCMC) to estimate (1) and coefficients in $\tilde{\omega}_{it}$ and Ω_t. Specifically, we apply the Riemannian manifold Hamiltonian Monte Carlo of Girolami and Calderhead (2011) and its refinement in Durmus et al.(2017). This method is chosen because numerical integration allows latent unobservable variables like productivity to be integrated out of the posterior. MCMC uses first- and second-order derivative information from the log posterior (computed numerically), which is an improved process relative to other Metropolis-Hastings algorithms. The algorithm provides a set of draws $\left\{\theta^{(s)}, s = 1, \ldots, S\right\}$ that converge (as $S \to \infty$) to the distribution whose density is given by the posterior. To integrate out productivity, which is a dynamic latent variable we use the Creal and Tsay (2015) Sequential

Monte Carlo algorithm. Appendix A.3 provides more details.

3 Data Sources and Variables

The data used to estimate (1) are drawn from EFIGE (European Firms in a Global Economy: internal policies for external competitiveness) and cover balance sheet information of 4286 manufacturing firms (with 10 employees and above) for four EU countries, France, Germany, Italy and U.K. over the period 2001-2014. This survey contains 150 questions that cover six broad areas of firms' economic activity: organisational structure, labour force, innovation, internationalization, market structure, and finance (Altomonte and Aquilante, 2012; Altomonte et al., 2013). For estimating (1), output (y) is defined as the operating revenue (OPRE), capital stock (k) is the book value of fixed assets (FIAS), labour (l) is the number of employees(EMPL), materials(MATE) include purchases of intermediate inputs domestically and abroad, and services (SERV) include purchases of intangible intermediate inputs domestically and abroad. We remove firms with missing observations, negative values for fixed assets, operating revenue and cost of materials to perform the log estimations of (1)and (5).[11] We deflate nominal values of operating revenue, material and services costs, and fixed assets using a 2-digit NACE industry[12] production price index (2005=100) from Eurostat and Office of National Statistics (ONS). For the specification of productivity ω_{it}, equation (5), we define R&D intensity as expenditure to total turnover and tax rate as the ratio of tax payments to gross profit.[13] Due to the insufficient number of industries in some countries, we estimate (7) by country. Table 1 shows summary statistics of the variables used.

[11]Firms with missing observations or negative values in key variables are non-viable entities and behave as outliers. These firms do not invest in any assets (either tangible or intangible) and report negative income values. Therefore, the economic structure of these firms is not of interest to our analysis as they are kept financially active by external support. Data cleaning is necessary to ensure outliers do not drive our results.

[12]See Appendix B for the full list of Nace Rev.2 digit industries

[13]We have defined R&D intensity as R&D expenditure per worker and tax rate as the ratio of tax payments to earnings before interest and tax (ebit). There are no qualitative differences to the estimates shown in Tables 2, 3 and 4 after using alternative definitions for R&D and tax rate. These estimations are available from the authors upon request.

Table 1: Summary Statistics

Variable	Mean	SD	Min	Max
y_{it}	8.05	1.28	-2.32	15.24
k_{it}	6.88	1.81	-1.79	15.66
l_{it}	3.57	1.13	0	11.55
materials (m_{it})	7.48	1.98	-3.87	23.15
$services_{it}$	7.11	1.97	1.02	13.08
R&D (R_{it})	47.5	46.8	0	100
tax rate (tax_{it})	28.23	4.43	19.47	37.06

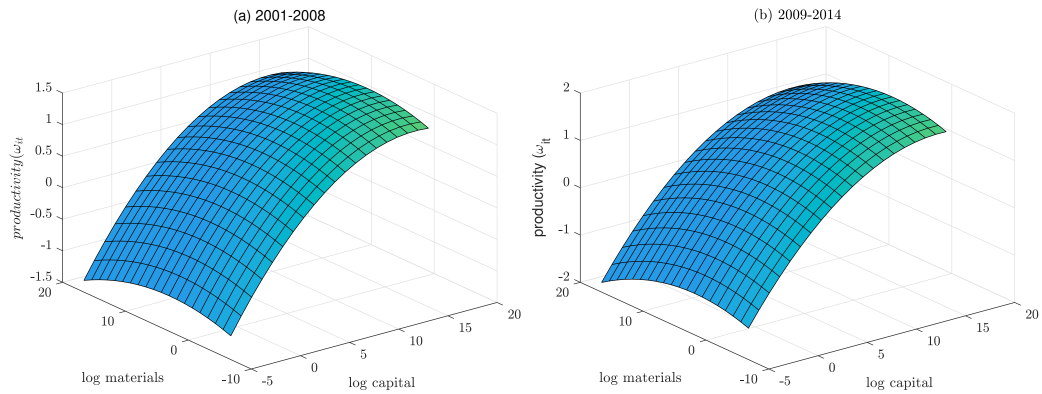
Notes: l, m and *services* are variable inputs; k is a state variable. Materials (m_{it}) and $services_{it}$ are the sum of domestic and imported purchases. The data set is an unbalanced panel of 57,629 observations.

4 Empirical results

4.1 Comparing Results Between Estimation Methods

We are particularly interested to assess the differences in TFP growth rates ($\Delta\omega_{it}$) between methods(MCMC vs.MACF). We first test whether ω_{it} is monotonous with respect to inputs by demonstrating plots in Figure 1 for the sub-periods, 2001-2008 and 2009-2014. We split the sample to capture possible discontinuity effects between productivity and input use associated with major global economic events (i.e., the sub-prime crisis in the USA).The graphs from both sub-periods confirm that demand of variable inputs increases *ceteris paribus* after a positive productivity shock.

Figure 1: Productivity as a function of capital and materials



Notes: The figures present the plots of productivity versus logs of capital and material for the sub-period 2001-2008 (panel (a)) and sub-period 2009-2014 (panel (b)). Analogous plots are drawn for labour and they are available from the authors upon request.

In the MACF estimation, ω_{it} is specified as: $\omega_{it} = \rho\omega_{it-1} + \theta_R R_{it} + \theta_{tax} tax_{it} + \theta_{2007} D_{2007} + \theta_{2008} D_{2008}$, where R stands for R&D, tax stands for tax rate, and D_{2007} and D_{2008} are year dummies for 2007 and 2008. We show average growth rates from three sub-periods to identify different patterns in TFP during periods of global economic turmoils. Comparing the average $\Delta\omega_{it}$ from the two approaches in Table 2, considerable differences are revealed. The first remark is that the MACF estimation tends to underestimate TFP growth in two sub-periods but this is not true for all countries. $\Delta\omega_{it}$ from MACF is high for Germany, while the opposite is true for Italy and the U.K. For the entire sample period, the MCMC method estimates the average $\Delta\omega_{it}$ at 0.11%, whereas MACF estimates it at 0.58%.

Table 2 shows that a more systematic inclusion of aggregate productivity shocks in the modelling of ω_{it} in MCMC corrects the estimate of TFP growth obtained by the "canonical" MACF estimation method. It is not negligible that $\Delta\omega_{it}$ from MACF is biased upward by 23 percentage points. We have more confidence in the MCMC approach and ω_{it} formulation in (2) than in MACF. This is because the former approach is more appropriate for accommodating aggregate shocks that are evident in the period studied.

Table 2: Average $\Delta\omega_{it}$ in % Using Different Estimations Techniques

country	Bayes-MCMC	MACF	Bayes-MCMC	MACF	Bayes-MCMC	MACF	Bayes-MCMC	MACF
FRA	0.50	0.31	-1.50	-0.31	1.20	0.8	0.07	0.26
GER	1.40	2.41	-1.30	3.46	1.40	8.2	0.50	4.68
ITA	1.10	0.53	-2.10	-0.85	1.00	-0.3	0.00	-0.21
UK	1.30	-1.44	-1.20	-2.32	-0.50	-3.4	-0.13	-2.40

Notes: Averages are calculated from 4286 firms from France, Germany, Italy and U.K. for 2001-2014. Bayes-MCMC refers to a Bayesian Markov Chain Monte Carlo technique and MACF refers to a modified specification of the Ackerberg et al.(2015) model. See the text for further details. The following moment conditions are formed to retrieve coefficients of inputs in the second stage of MACF: $\begin{bmatrix} E(l_{it-1}\xi^{\omega}_{it}) = 0 \\ E(l^2_{it-1}\xi^{\omega}_{it}) = 0 \end{bmatrix}$, $\begin{bmatrix} E(k_{it}\xi^{\omega}_{it}) = 0 \\ E(k^2_{it}\xi^{\omega}_{it}) = 0 \\ E(k_{it-1}\xi^{\omega}_{it}) = 0 \\ E(k^2_{it-1}\xi^{\omega}_{it}) = 0 \end{bmatrix}$, $\begin{bmatrix} E(l_{it-1}k_{it}\xi^{\omega}_{it}) = 0 \\ E(l_{it-1}k^2_{it}\xi^{\omega}_{it}) = 0 \end{bmatrix}$.

4.2 Diagnostics for the Bayesian MCMC Estimation

Before presenting results from the Bayesian MCMC estimation, we first show some diagnostic tests about the best-fitting specification. Table 3 reports Bayes factors of various specs testing how models perform based on the covariates included in each productivity component ($\tilde{\omega}_{it}$ and Ω_t). This test guides us on whether to include these covariates in the firm idiosyncratic or the aggregate component. The selection criterion here is the specification with the largest Bayes factor.[14] We start with a specification that assumes an AR(1) process of $\tilde{\omega}_{it}$ and Ω_t without any other covariates, then we add R&D and tax rate gradually into $\tilde{\omega}_{it}$ and Ω_t. The best-performing model (**S**5) has a Bayes factor of 738,265 and includes both R&D and tax rates in $\tilde{\omega}_{it}$ and Ω_t. The specification with the second highest factor (**S**3) (272,835.44) includes R&D and tax rates in $\tilde{\omega}_{it}$ and only tax rates in Ω_t. The third most effective alternative (**S**8) excludes R&D from both $\tilde{\omega}_{it}$ and Ω_t and reduces the Bayes factor to 213,864. Note that a linear specification (**S6**) in the best-performing model of the reduced form also results in a dramatic drop in the Bayes factor (4.77) 10^{-47}, showing the existence of non-linearities, which point to a quadratic specification in (2). According to this diagnostic, R&D and tax rate best fit the data when they are modelled through both idiosyncratic productivity and aggregate shock components.

[14]Bayes factors are computed as ratios of marginal likelihoods and the latter are the standard output of Sequential Monte Carlo.

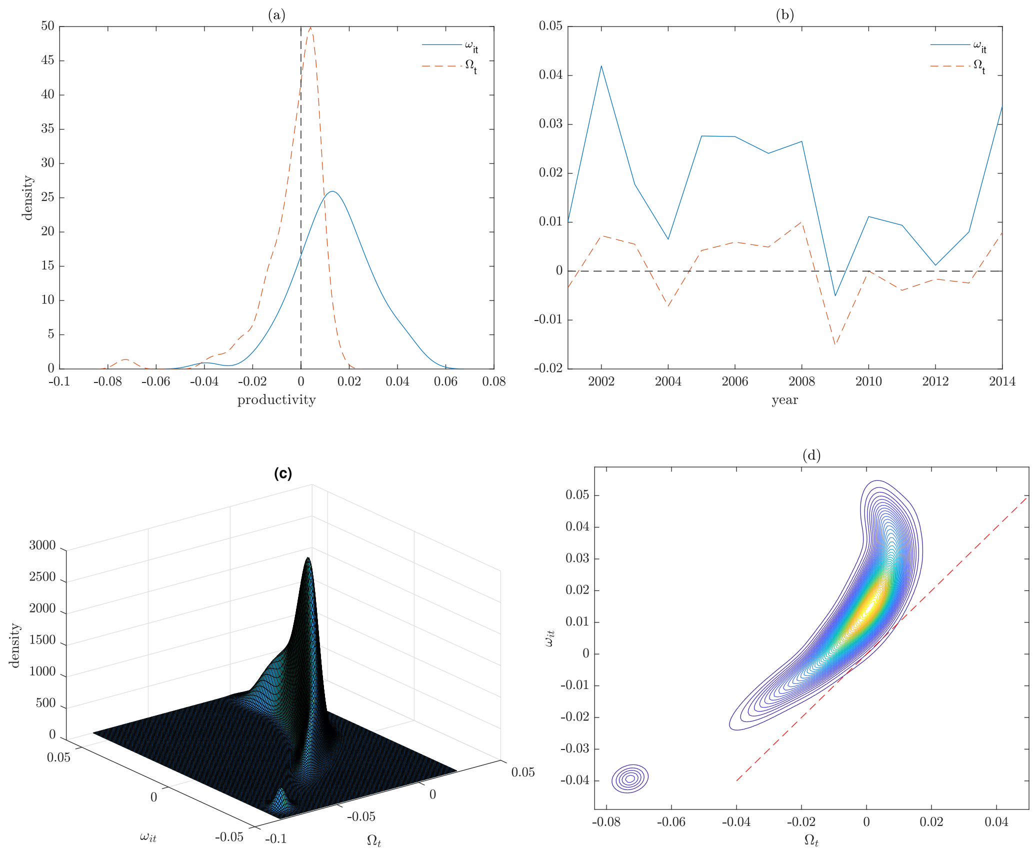
Table 3: Bayes Factors for Various Specifications

Hypothesis (Specification)	Bayes factor
S1: Both $\tilde{\omega}_{it}$ and Ω_t, no covariates	1,715
S2: R&D and tax rates in $\tilde{\omega}_{it}$, no covariates in Ω_{it}	125,825
S3: R&D and tax rates in $\tilde{\omega}_{it}$, tax rate in Ω_t	272,835
S4: R&D and tax rates in $\tilde{\omega}_{it}$, R&D in Ω_t	7,317
S5: R&D and tax rates in $\tilde{\omega}_{it}$, R&D and tax rates in Ω_t	738,265
S6: S5 with linear specification of Φ_ω and Φ_Ω	$4.77\ 10^{-47}$
S7: R&D in $\tilde{\omega}_{it}$ and Ω_t	1,578
S8: tax rate in $\tilde{\omega}_{it}$ and Ω_t	213,864

Notes: Bayes factors are against a model with $\tilde{\omega}_{it}$ without Ω_t and taxes or R&D, and in favor of the hypothesis specified in the first column. Notice that tax rates and R&D are firm-specific. Relative Bayes factors can be computed as well; for example, the Bayes factor for the model with no R&D, taxes in $\tilde{\omega}_{it}$ and Ω_t relative to the model with R&D and tax rates in $\tilde{\omega}_{it}$ and Ω_t is 213,864.15 / 1,578.39 = 135.5, approximately.

Panel (a) in Figure 2 presents the sample distributions of ω_{it} and the aggregate (industry-specific) productivity component Ω_t.The temporal evolution of these measures is presented in panel (b). Panels (c) and (d) show their joint bivariate sampling distributions. From panel (a) the sampling distribution of ω_{it} is shifted to the right relative to Ω_t. The fact is also evident in the contour plots reported in panel (d), which suggest that aggregate shocks cause severe firm productivity movements. It can be seen from panel (b) that the evolution of ω_{it} follows a very similar pattern to the aggregate component Ω_t, underlining the importance of incorporating macro-oriented shocks into firm productivity models. The trends in ω_{it} and the component Ω_t in panel (b) indicate a remarkable decline in 2008, which represents the effect of the subprime crisis initiated in the USA that led to a decrease in ω_{it} and Ω_t by nearly four and two percentage points, respectively.

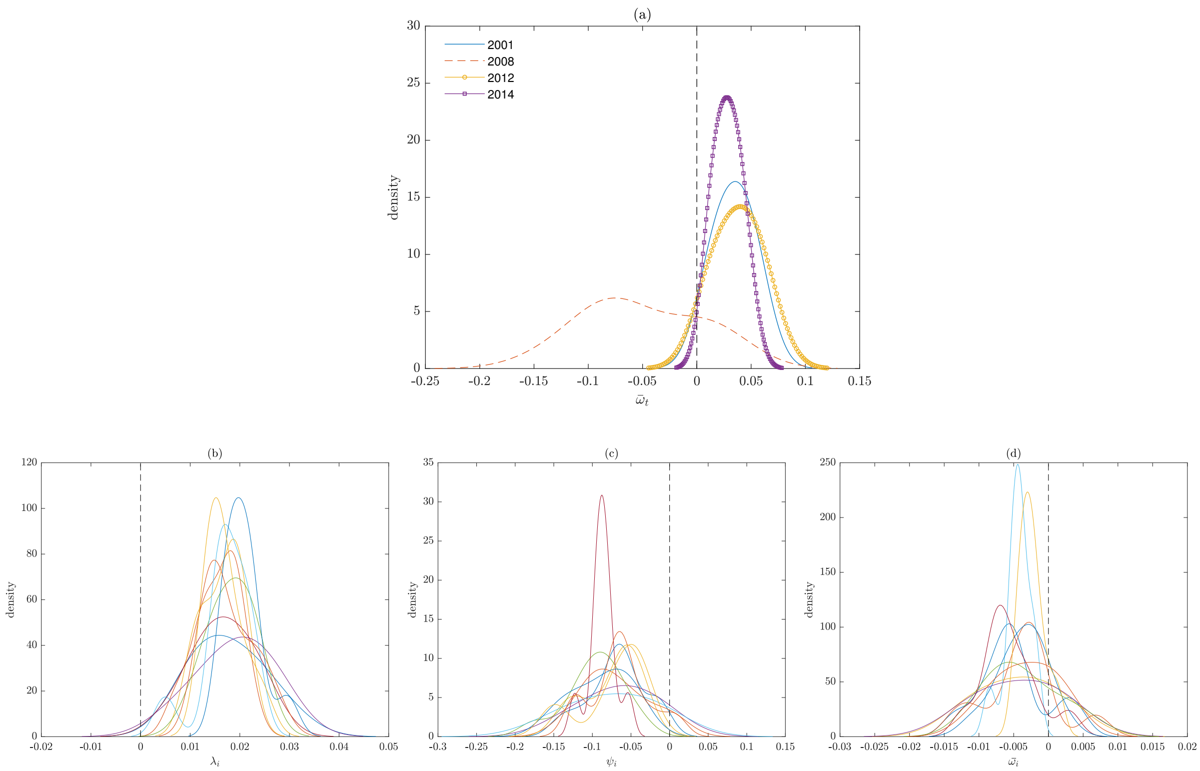
Figure 2: Aspects of Productivity ω_{it} and Aggregate Productivity Shocks Ω_t



Notes: Panel (a) reports sampling distributions of firm productivity (ω_{it}) and its industry-specific productivity component (Ω_{it}). Panels (b) and (c) show the bivariate density and its contours, respectively. The dotted line in panel (d) represents the 45° line.

The sample average of $\bar{\omega}_t = n^{-1}\sum_{i=1}^{n}\omega_{it}$ can be defined as: $\bar{\omega}_t = \bar{\lambda}+\bar{\psi}\Omega_t+\bar{\tilde{\omega}}_t$. Before displaying estimates, we show a graphical analysis of the results. We first show the sample distributions of average productivity ($\overline{\omega}_t$) in Figure 3 (panel a) for different periods. Density on the y-axis measures the probability density of ω_{it} against possible average values of the parameter that are denoted on the x-axis. Although the global financial crisis impacted substantially on firm-specific productivity, the distributions for 2001, 2008, and 2012 are practically the same implying that productivity in most countries has recovered in the aftermath of the global financial crisis. If $\bar{\omega}_t > 0$ then the area of the marginal posterior density of U_t to the right of zero represents the probability of "unicorns". Specifically, if $U_t \geq 1$, the productivity increase in "unicorns" is proportionally bigger than the aggregate productivity shock, while if $U_t \geq 0$ productivity co-moves with aggregate productivity shocks. The large variance of $\bar{\omega}_t$ in panel(a) for 2008 shows a negative aggregate productivity shock, which is associated with the US financial turbulence that also impacted other major economies and members of the G7. Given that the model is estimated by country and not by industry, panels (b), (c), and (d) in Figure 3 dig further into the results by reporting sample distributions of λ_i, ψ_i and $\overline{\omega}_i$ for ten randomly selected firms. Parameter λ_i is positive for most firms, ψ_i is negative and $\overline{\omega}_i$ exhibits a negative mode with negative values predominating.

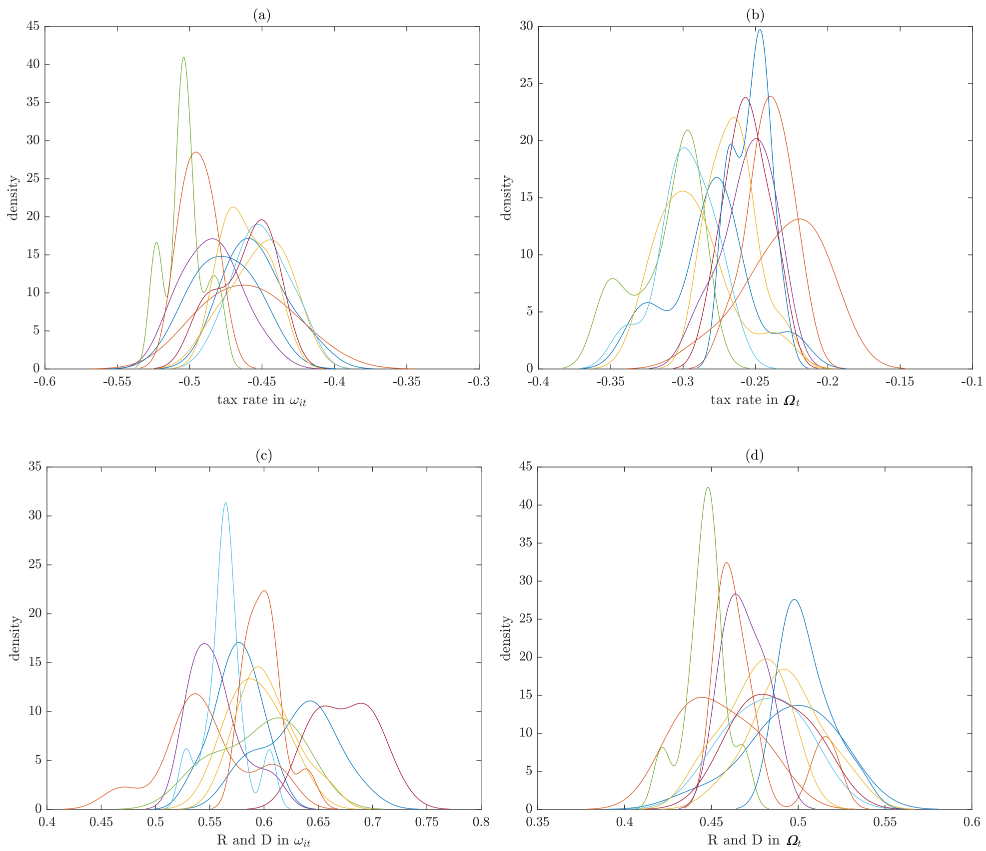
Figure 3: Marginal posterior densities of sample-averaged productivity ($\bar{\omega}_t$)



Notes: Panel (a) reports marginal posterior densities of $\bar{\omega}$ for different periods. Panels (b), (c) and (d), report marginal posterior densities of λ_i, ψ_i and $\bar{\omega}$ for ten randomly selected firms.

Figure 4 shows the probability density functions of R&D and tax rate (i.e., y-axis) separately for ω_{it} and Ω_t against sample mean values of these parameters (i.e., x-axis). The main message from these graphs is that the estimated effect of these variables confirms our theoretical predictions that R&D and tax rates impact productivity positively and negatively, respectively.

Figure 4: Sample Densities of Marginal Effects of R&D and Tax Rates andon Productivity



4.3 Bayesian MCMC Estimation Results

We now turn to the point estimates of (7), which show posterior moments of production inputs and other relevant parameters included in ω_{it}. Table 4 represents the average value the parameter is expected to take. We start with the generalized $\mathcal{R}^2$ at the bottom of the table, which measures the fit of the entire system consisting of the production function, the first-order conditions in the reduced form, and the productivity equations. The high $\mathcal{R}^2$ values indicate that the specifications used provide a very good fit for the data. The first four rows show posterior means of inputs (k_{it}, l_{it}, m_{it}, $services_{it}$) for the whole sample and for each country separately in the remaining columns. With the sum of input posterior means close to one, y_{it} exhibits constant returns to scale. The capital coefficient is high exceeding the standard threshold estimate of one-third, documented in aggregate production functions (Guerriero, 2019). As the current sample consists of large European enterprises excluding SMEs and self-employed firms, we expect higher levels of capital intensity in production.

Next, we look at posterior means and posterior standard deviations for parameters ρ, σ_ϵ, and σ_ξ.Estimates of ρ for the entire sample as well as for individual countries highlight persistence in productivity shocks with the mean value of the parameter close to 0.515 (posterior s.d. 0.016). The sample average of the idiosyncratic productivity term $\tilde{\omega}_{it}$ is marginally negative. This suggests that higher tax liabilities cancel out R&D gains. Parameters λ_i and ψ_i capture non-linear effects from aggregate shocks. Accordingly, the signs of these posterior mean estimates (column 1) imply an inverted U-shaped relationship between productivity and aggregate shocks. For the entire sample during 2001-2014, negative aggregate productivity shocks are stronger ($|\bar{\psi}| > \bar{\lambda}$).

In absolute terms, the effect of R&D in column (1) for the entire sample is more pronounced compared to tax rate. A change in a ω_{it} is more likely to result from a change in R&D intensity (0.586) rather than a change in the tax rate (-0.476). Nonetheless, this pattern varies across countries. Specifically, German and UK firms are more sensitive to tax changes than R&D. In France, the negative marginal effect of tax and the positive marginal effect of R&D are almost symmetrical. In Italy, productivity is more responsive to higher R&D than higher taxation. These effects indicate remarkable heterogeneity in the way firms capitalize on innovation gains and absorb additional tax burdens. Heterogeneity also exists in the non-linear effect of aggregate productivity shocks. The inverted U-shaped pattern is more evident in Italian firms than in any other counterpart in the sample. In a world with heterogeneous firms the effects of R&D and taxation on productivity can be better understood in conjunction with other characteristics. For example, firms that are internationally exposed are less affected by an increase in the statutory tax rate as they can outsource part of their production to jurisdictions with lower tax liabilities. Therefore, the tax burden is less relevant for investment decisions and productivity. Conversely, firms that serve solely the domestic market innovate less systematically, resulting in higher marginal productivity returns from innovation. A study of how firm heterogeneity impacts performance is beyond the scope of this paper but it will suffice to say that the existence of heterogeneity documented in the present estimates accords well with other strands of literature.[15]

The posterior marginal effects in the aggregate productivity component Ω_t suggest that common positive effects from R&D exceed losses from a common adverse taxation shock in all countries. This finding highlights the importance of R&D spillovers that enable individual firms to benefit from the innovation efforts of other peers in the industry [16]

[15] See Greenaway and Kneller (2007) for the role of heterogeneity in the micro-econometrics of trade literature.

[16] R&D spillovers at the industry level imply that social returns to R&D are larger than private returns. This is to say that R&D activity of firm i benefits the performance of firm i' thus aggregate industry productivity increases.

Table 4: Posterior Moments

	Sample	France	Germany	Italy	U.K.
	1	2	3	4	5
k_{it}	0.533	0.488	0.579	0.558	0.475
	(0.022)	(0.013)	(0.044)	(0.023)	(0.016)
l_{it}	0.120	0.138	0.159	0.185	0.166
	(0.017)	(0.022)	(0.017)	(0.030)	(0.029)
m_{it}	0.106	0.165	0.11	0.085	0.12
	(0.003)	(0.016)	(0.015)	(0.024)	(0.021)
$services_{it}$	0.07	0.1	0.11	0.11	0.13
	(0.015)	(0.007)	(0.005)	(0.012)	(0.01)
ρ	0.515	0.418	0.486	0.572	0.505
	(0.016)	(0.012)	(0.025)	(0.032)	(0.044)
σ_ϵ	0.023	0.019	0.031	0.025	0.022
	(0.0074)	(0.0045)	(0.0052)	(0.0030)	(0.0065)
σ_ξ	0.048	0.055	0.046	0.055	0.048
	(0.015)	(0.012)	(0.013)	(0.015)	(0.011)
$n^{-1}\sum_{i=1}^{n}\varrho_i$	0.272	0.210	0.198	0.316	0.302
	(0.0044)	(0.007)	(0.044)	(0.055)	(0.017)
$\bar{\tilde{\omega}}_t = n^{-1}\sum_{i=1}^{n}\tilde{\omega}_{it}$	–0.0044	0.0012	–0.0035	0.0014	–0.0010
	(0.0045)	(0.0039)	(0.0017)	(0.0026)	(0.0028)
$\bar{\lambda} = n^{-1}\sum_{i=1}^{n}\lambda_i$	0.017	0.012	0.0078	0.016	0.020
	(0.0052)	(0.0037)	(0.0044)	(0.0057)	(0.0092)
$\bar{\psi} = n^{-1}\sum_{i=1}^{n}\psi_i$	–0.071	0.035	0.007	–0.0095	0.037
	(0.043)	(0.056)	(0.046)	(0.032)	(0.024)
tax rate in ω_{it} $^{(*)}$	–0.476	–0.517	–0.672	–0.499	–0.713
	(0.018)	(0.013)	(0.022)	(0.026)	(0.035)
tax rate in Ω_t $^{(*)}$	–0.265	–0.217	–0.313	–0.287	–0.310
	(0.024)	(0.018)	(0.019)	(0.018)	(0.014)
R&D in ω_{it} $^{(*)}$	0.586	0.516	0.487	0.519	0.488
	(0.032)	(0.018)	(0.023)	(0.010)	(0.032)
R&D in Ω_t $^{(*)}$	0.473	0.376	0.518	0.488	0.475
	(0.017)	(0.017)	(0.023)	(0.016)	(0.022)
$\mathcal{R}^2$	0.985	0.972	0.975	0.988	0.965
	(0.0074)	(0.0052)	(0.0072)	(0.0084)	(0.0055)
observations	57,629	2,826	4,021	20,417	30,365

Notes: Reported are posterior means with posterior standard deviations in parentheses. $\mathcal{R}^2$ is the squared correlation coefficient between actual and fitted values average over MCMC draws. $^{(*)}$ means that these are marginal effects computed from flexible functions (low order polynomials).

Figure 5 shows sample densities of marginal effects of tax rates and R&D on productivity in individual countries. Although they are not symmetric, they are concentrated around negative and positive values respectively. The tax rate has a negative effect as expected both in ω_{it} and Ω_{it}. Based on the distributions displayed, the effects tend to be stronger in ω_{it} as shown with the posterior moments in Table 4. These findings indicate that changes in R&D and tax rates have more permanent effects on a firm's productivity as these changes represent long-term decisions made intentionally to enhance performance. Aggregate shocks despite being occasionally severe like the global financial crisis in 2008 tend to die out faster as firms can more easily absorb them. Firm heterogeneity also matters for absorbing aggregate shocks. As shown in Figure 5 firms from different countries have the flexibility to adjust the input mix in response to an external shock, while firm-specific productivity effects are structural and persistent. Although we do not provide a theoretical framework for explaining the mechanisms through which firms absorb macro shocks, Grisby (2022) has shown that in frictionless input markets firms tend to adjust promptly. In other words, aggregate shocks are more transitory [17], while firm-specific productivity changes are permanent and can explain more compositional changes in aggregate output over time (Franco and Philippon, 2007).

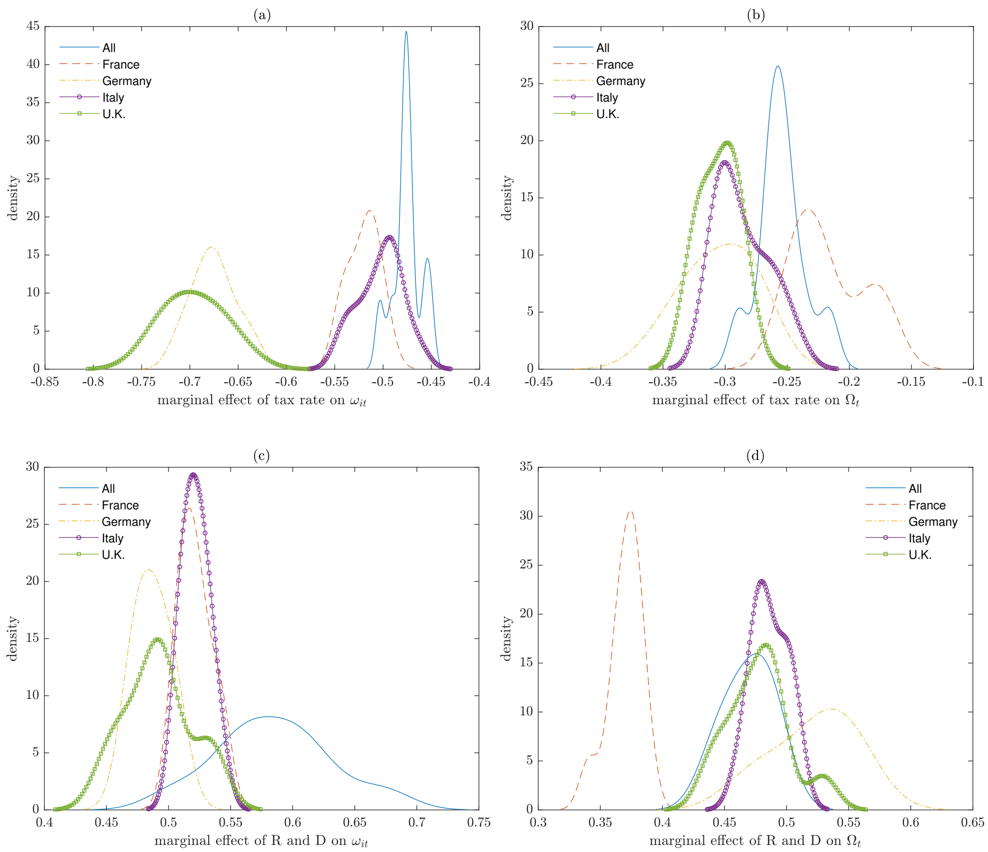
4.4 Country Specific Insights

Looking closer at individual country estimates reveals interesting insights. Figure 5, panels (a) and (b) show that UK firms have been affected more severely than their other counterparts by fiscal changes. This is a finding highly related to the aggregate UK productivity slowdown documented in the period under study. [18] Turning to R&D in the lower panel of Figure 5 ((c) and (d)), we observe that R&D spending tends to be very beneficial for productivity improvements in French and Italian firms while looking at the industry pattern, Germany does exceptionally well in transmitting aggregate productivity gains from the innovation activity of individual firms. The latter pattern suggests that intra-industry clusters accelerate technological diffusion within the same industry, which results in higher overall productivity. Italy also has industrial clusters that promote $\omega_[it]$ to a lesser extent.

[17] A transitory effect within the context of the present analysis is also likely to co-move with the business cycle (Franco and Philippon, 2007).

[18] Bournakis and Mallick (2021) have also shown, albeit with a different estimation framework that higher corporate taxes damage firms' ability to converge towards the productivity frontier. As in Bournakis and Mallick (2021) the same implication is drawn from the present results, higher tax liabilities increase the user cost of capital and decrease the amount of capital available for productivity-enhancing investment.

Figure 5: Marginal Effects of R&D and Tax Rates on Productivity By Country



Notes: Reported are marginal effects of tax rates (panels (a) and (b)) and R&D (panels (c) and (d)) on industry-specific and aggregate productivity by country.

In Figure 6 we compare the temporal evolution of productivity growth for the whole sample (panel (a)) and individual countries (panel (b) to (e)) with 95% Bayes probability interval bands. Productivity growth decelerated sharply between 2007 and 2008. The biggest productivity losses from the global financial crisis are experienced in Italy. Even so, Italy, France, and Germany recovered after 2008 and only the UK showed productivity stagnation throughout the sample period. The UK productivity slowdown (panel(e)) started before the USA subprime crisis and persisted for another ten years or so. Although the productivity backwardness of the UK economy is not a new concept, it has already been discussed, *inter alia*, in Harris and Moffat (2017) and Goodridge et al.(2018),[19] our approach highlights the complexity of the phenomenon emphasizing that poor

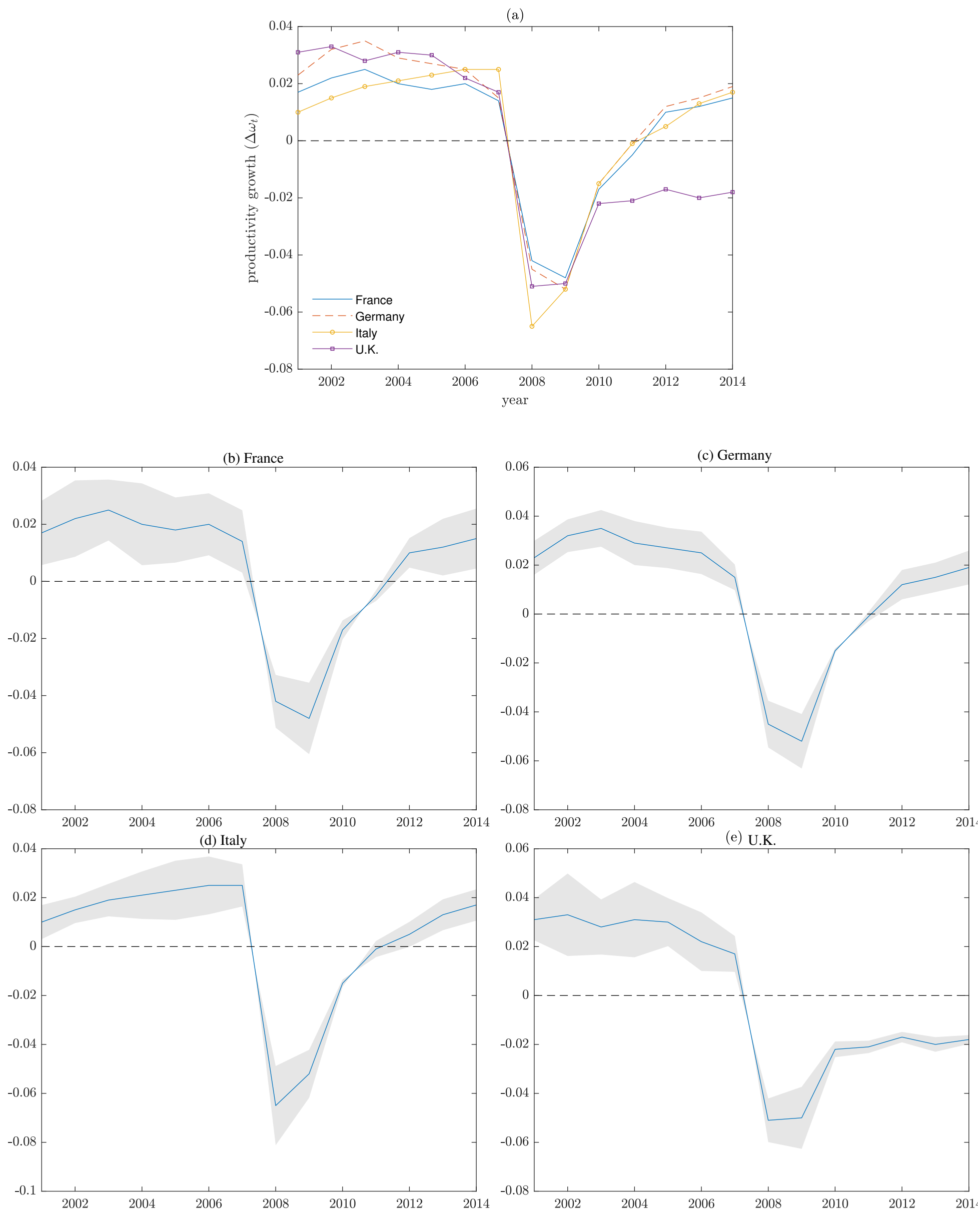
[19]These studies focus only on the UK and cover a wide spectrum of industries. Their estimation framework is based on growth accounting (Goodridge et al., 2018) and GMM estimation of TFP (Moffat and Harris, 2018). Harris and Moffat (2018) concluded that the UK productivity slowdown stems from small businesses and services. Goodridge et al.(2018) attribute the TFP slowdown to low labour productivity, poor performance in specific sectors (i.e. finance and oil) and business cycle effects. Although the present analysis differs in terms of methodology and industry coverage, we also document a substantial slowdown in

productivity in the UK has both aggregate and firm-specific aspects. While most European firms recovered immediately after 2008, UK firms remained outliers not because they did not invest enough in R&D [20] but mainly, as also documented in Table 4, due to high tax liabilities on individual firms. The new element that our study contributes to the UK productivity slowdown agenda is that higher tax burdens impact productivity through transitory and permanent channels. The transitory channel works through reduced incentives to invest, while the permanent channel works through reduced capital accumulation over the long term. Our results also show that changes in the fiscal environment can have complex effects on corporate performance and welfare.

UK productivity over the last twenty years. In addition, we identify that the UK falls behind in productivity relative to other EU economies because of aggregate shocks that might incorporate cyclical characteristics.

[20]Estimates in Table 4 show that R&D has been a key driver of UK productivity.

Figure 6: Productivity growth in France, Germany, Italy, and the U.K.

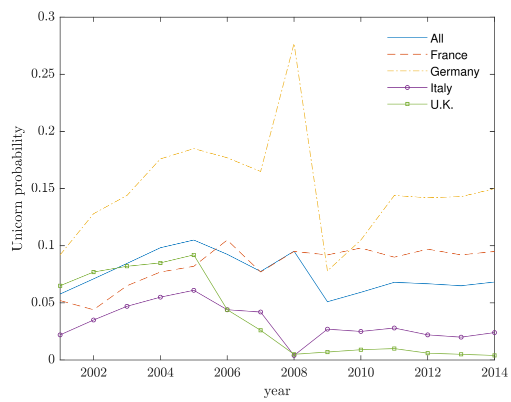


Notes: Shown in panel (a) are productivity growth rates by country, computed, for a given year, as $\Delta\omega_t = n^{-1}\sum_{i=1}^{n}(\omega_{it} - \omega_{i,t-1})$ where the ω_{it}s are averaged across MCMC draws. In panels (b), (c), (d), and (e) we provide the same information with 95% Highest Posterior Density intervals.

Figure 7 shows the probability of observing "unicorns" in individual countries, viz firms with exceptional productivity. As a result of the USA subprime crisis in 2008, the probability of identifying "unicorns" dropped across the entire sample (blue line). However, this probability pattern shows substantial cross-country variation. The largest reduction is found in Germany (from 0.25 to

0.1), followed by the UK and Italy (green and purple lines, respectively). During the 2008 financial crisis, the "unicorn" probability in France changed the least. It is clear from Figure 7 that German firms' productivity was hit severely in 2008 but improved immediately the following year. Figure 7 also captures the empirical regularity that the UK never recovered after the 2008 global financial crisis with productivity remaining unchanged for another six years (the flat part in the green line). Although the financial crisis affected German firms the most, they did not experience recursive productivity losses. German firms are characterized by high responsiveness to global shocks but also a strong capacity to absorb them. In other European firms the recovery tends to take longer. Overall, Figure 7 demonstrates the European productivity leadership of German manufacturing.

Figure 7: Unicorn posterior probability



Notes: Reported is the probability of observing "unicorns", in the sense that $P(\bar{\omega}_t > 0|\text{Data}) > 0$. The probability is displayed over time for different countries and the sample as a whole.

5 Concluding Remarks

The paper proposes a novel estimation framework for measuring firm-level TFP. Unlike conventional semi-parametric techniques, our approach overcomes identification bias without relying on the monotonicity between productivity and arbitrary input. First, our model captures the evolution of firm productivity by incorporating both firm-idiosyncratic factors and macro shocks. This modelling approach offers a more realistic set-up than earlier literature, as it recognizes that firm productivity is driven not only by firm-specific decisions but also subject to negative exogenous industry-wide shocks. In our framework, we consider R&D and corporate tax liabilities as the most prominent determinants that capture the effects of innovation efforts and fiscal environment conditions on productivity. Second, we apply a new Bayesian non-parametric technique that makes each variable input a function of the productivity parameter without relying on monotonicity between productivity and input use. We tested this novel approach on a set of 4286 firms from France, Germany, Italy and the UK over 14 years (2001-2014). Our results reveal some interesting findings regarding European firms' performance in a period of substantial global economic turbulence. The USA subprime crisis of 2008 and the financial downturn that followed negatively impacted productivity in all countries. This finding supports our choice to develop a framework that credits firms for adverse aggregate shocks. Although all countries suffered severe negative effects, they were short-lived, with the UK being the only country that has not recovered after many years.

Not surprisingly, our results suggest that R&D enhances firm productivity, while higher tax liabilities harm it. Nonetheless, the magnitude of these effects varies remarkably across countries. Productivity gains of French firms from R&D are almost symmetrical to high tax losses. Italian firms' R&D activity outweighs tax losses. German and UK companies experience productivity losses from taxation that are not offset by higher R&D intensity. Two generic policy implications can be drawn from the present results. First, firms need a stable fiscal environment for productivity to flourish. Any change in taxes can have detrimental performance effects, which cannot always be compensated for by proportionally higher innovation gains. This result contradicts the traditional view (Lucas, 1976) that changes in fiscal policy are neutral in determining supply-side variables like productivity, while it supports evidence (Slemrod, 1992) that marginal changes in fiscal policy are related to long-term economic outcomes. Our results suggest that tax policy fluctuations affect investors' incentives to undertake new projects. As the acquisition of new assets reflects technological progress, higher taxes can potentially decelerate productivity growth. Second, governments should amplify policies that offer tax exceptions to R&D firms. Social returns to innovation are high, leading to a diffusion of knowledge within the industry and an overall increase in productivity. Whenever a firm wants to innovate, incentives should be put in place to encourage it, since R&D activity can benefit the firm's peers in the long run.

Several country-specific observations are also drawn from the analysis. UK policymakers should reconsider corporate tax policy given the remarkable productivity laggardness and high sensitivity of firm productivity to tax liability changes. A decline in UK productivity had already occurred before the subprime crisis, followed by an extended period of stagnation. Our results show that the UK productivity handicap is not due to inadequate innovation efforts at the firm level. Instead, it highlights structural problems that tend to derive from the fiscal environment that does not favour investment in infrastructure and other productivity-enhancing sectors. As a first step, a policy recommendation is to extend fiscal incentives for investment projects other than R&D as these are also crucial for promoting technological progress. The significant contribution of R&D to aggregate productivity in the present study suggests that micro- to macro-level productivity gains can be achieved by clusters. R&D activity in German firms embodies social returns through knowledge spillovers that enhance aggregate productivity. Based on this and the rapid rebound of German firms after the financial crisis, we conclude that clustering is an effective channel for technological advancement among firms. Italian firms are also highly responsive in R&D investment and adhere to the German paradigm in benefiting from industrial clusters. French firms have lower social R&D investment returns than their counterparts in Germany and Italy. Implementing more targeted policy interventions may help French firms increase social R&D returns. These could aim to intensify ties between firms through increased innovation synergies and research collaborations.

References

[1] Ackerberg, D. A., K. Caves, & G. Frazer. (2015). Identification Properties of Recent Production Function Estimators. Econometrica 83 (6), 2411–2451.

[2] Andrieu, C., Doucet, A., Holenstein, R. (2010). Particle Markov chain Monte Carlo methods (with discussion). Journal of the Royal Statistical Society Series B 72 (2), 1–33.

[3] Arellano, M., & S. Bond (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. Review of Economic Studies 58 (2), 277–97.

[4] Blundell, R., & S. Bond (1998). Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. Journal of Econometrics 87 (1), 115–143.

[5] Blundell, R., & S. Bond (2000). GMM Estimation with Persistent Panel Data: An Application to Production Functions. Econometric Reviews 19 (3), 321–340.

[6] Bournakis, I. and Mallick, S., 2018. TFP estimation at firm level: The fiscal aspect of productivity convergence in the UK. Economic Modelling, 70, pp.579-590.

[7] Bournakis, I. and Mallick, S., 2021. Do corporate taxes harm economic performance? Explaining distortions in R&D-and export-intensive UK firms. Macroeconomic Dynamics, 25(1), pp.5-27.

[8] Chamley, C., 2001. Capital Income Taxation, Wealth Redistribution and Borrowing Constraints. Journal of Public Economics, 79, pp.55-69.

[9] Chernozhukov, V. and Hong, H. (2003). An MCMC approach to classical estimation. Journal of Econometrics 115, 293–346.

[10] Chopin, N., Singh, S.S., 2013. On the particle Gibbs sampler. Working paper, ENSAE. http://arxiv.org/abs/1304.1887.

[11] Creal, D.D., 2012. A survey of sequential Monte Carlo methods for economics and finance. Econometric Reviews 31 (3), 245–296.

[12] Creal, D., and R. Tsay (2015). High dimensional dynamic stochastic copula models. Journal of Econometrics 189 (2), 335–345.

[13] De Loecker, J. and Warzynski, F., 2012. Markups and firm-level export status. American economic review, 102(6), pp.2437-71.

[14] De Loecker, J. and Goldberg, P.K., 2013.Firm performance in a global market (No. w19308). National Bureau of Economic Research.

[15] De Loecker, J., 2013. Detecting learning by exporting. American Economic Journal: Microeconomics, 5(3), 1-21.

[16] Diewert, W. E., and T. J. Wales (1987). Flexible Functional Forms and Global Curvature Conditions. Econometrica 55, 43–68.

[17] Doraszelski, U., & J. Jaumandreu (2013). R&D and Productivity: Estimating Endogenous Productivity. Review of Economic Studies 80 (4), 1338–1383.

[18] Durmus, A., G. O. Roberts, G. Vilmart, and K. C. Zygalakis (2017). Fast Langevin based algorithm for MCMC in high dimensions. The Annals of Applied Probability 27 (4), 2195–2237.

[19] Franco, F. and Philippon, T., 2007. Firms and aggregate dynamics. The Review of Economics and Statistics, 89(4), pp.587-600.

[20] Gallant, A. R., R. Giacomini & G. Ragusa (2017). Bayesian estimation of state space models using moment conditions. Journal of Econometrics 201, 198-211.

[21] Gallant, A. R. and G. H. Golub (1984). Imposing Curvature Restrictions on Flexible Functional Forms. Journal of Econometrics, 26, 295–322.

[22] Gallant, A. R., and D. W. Nychka (1987). Semi-Nonparametric Maximum Likelihood Estimation. Econometrica 55 (2), 363–390.

[23] Gandhi, A., S. Navarro, & D. A. Rivers (2020). On the Identification of Gross Output Production Functions. Journal of Political Economy 128 (8), 2973–3016.

[24] Gentry, W., and G. Hubbard (2004). Success Taxes, Entrepreneurial Entry and Innovation. NBER Working Paper 10551.

[25] Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4* (J. M. Bernado, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Clarendon Press, Oxford, UK, 169–193.

[26] Girolami, M., and B. Calderhead (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. Journal of the Royal Statistical Society Series B, 73 (2), 123–214.

[27] Godsill, S.J., Doucet, A., West, M., 2004. Monte Carlo smoothing for nonlinear time series. Journal of the American Statistical Association 99 (465), 156–168.

[28] Goodridge, P., Haskel, J. and Wallis, G., 2018. Accounting for the UK productivity puzzle: a decomposition and predictions. Economica, 85(339), pp.581-605.

[29] Gordon, R. and Lee, Y., 2008. Interest Rates, Taxes and Corporate Financial Policies. National Tax Journal, 61(2), pp.159-160.

[30] Greenaway, D. and Kneller, R., 2007. Firm heterogeneity, exporting and foreign direct investment. The Economic Journal, 117(517), pp.F134-F161.

[31] Griffith, R., Redding, S. and Reenen, J.V., 2004. Mapping the two faces of R&D: Productivity growth in a panel of OECD industries. Review of economics and statistics, 86(4),pp.883-895.

[32] Grisby, J.R., 2022. Skill heterogeneity and aggregate labor market dynamics (No. w30052). National Bureau of Economic Research.

[33] Guerriero, M., 2019. The labor share of income around the world: Evidence from a panel dataset. *Labor Income Share in Asia*, Springer, pp.39-79.

[34] Hadjidoukas, P.E., P. Angelikopoulos , C. Voglis, D.G. Papageorgiou, I.E. Lagaris (2014). NDL-v2.0: A new version of the numerical differentiation library for parallel architectures. Computer Physics Communications 185 (7), 2217–2219.

[35] Hahn, J., Liao, Z. and Ridder, G., 2018. Nonparametric two-step sieve M estimation and inference. Econometric Theory, 34(6), pp.1281-1324.

[36] Hall BH, Rosenberg N, editors. Handbook of the Economics of Innovation. Elsevier; 2010 May 14.

[37] Harris, R. and Moffat, J. (2017). The UK productivity puzzle, 2008?2012: evidence using plant-level estimates of total factor productivity. Oxford Economic Papers, 69(3), pp.529-549.

[38] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57, 97–10.

[39] Hu, Y., H. Huang, & Y. Sasaki (2020). Estimating production functions with robustness against errors in the proxy variables. Journal of Econometrics 215, 375–398.

[40] Huljak, I., R. Martin, D. Moccero (2019). The cost-efficiency and productivity growth of euro area banks. ECB Report 2305.

[41] Iyoha, E., 2020. Estimating productivity in the presence of spillovers: Firm-level evidence from the us production network. Conference of the European Association for Research in Industrial Economics (EARIE) .

[42] Jacob, M. (2021). Dividend taxes, employment, and firm productivity. Journal of Corporate Finance, 69 , 102040.

[43] Kasahara, H., & J. Rodrigue. (2008). Does the Use of Imported Intermediates Increase Productivity? Plant-level Evidence. Journal of Development Economics 87 (1), 106–118.

[44] Lee, Y., A. Stoyanov, & N. Zubanov, 2019. Olley and Pakes-style Production Function Estimators with Firm Fixed Effects. Oxford Bulletin of Economics and Statistics 81 (1), 79–97.

[45] Levinsohn, J., & A. Petrin (2003). Estimating Production Functions Using Inputs to Control for Unobservables. Review of Economic Studies 70 (2), 317–342.

[46] Liu, J. S. (2008). *Monte Carlo strategies in scientific computing*. Springer Series in Statistics. Springer, New York.

[47] Lucas, R. E., 1967a. Adjustment costs and the theory of supply, Journal of Political Economy 75, 321–334.

[48] Lucas, R. E., 1967b. Optimal investment policy and the flexible accelerator, International Economic Review 8, 78–85.

[49] Lucas, R.E., 1976. Econometric Policy Evaluation: A Critique, w: K. Brunner, AH Meltzer (wyd.): The Phillips Curve and Labour Markets.

[50] Malikov, E., S. C. Kumbhakar, & M. G. Tsionas (2016). A Cost System Approach to the Stochastic Directional Technology Distance Function with Undesirable Outputs: The Case of U.S. Banks in 2001-2010. Journal of Applied Econometrics, Vol. 31, No. 7, 2016, pp. 1407–1429.

[51] Marschak, J., & W. H. Andrews, Jr. (1944). Random Simultaneous Equations and the Theory of Production. Econometrica, 12 (3/4), 143–205.

[52] Mundlak, Y. (1961). Empirical Production Function Free of Management Bias. Journal of Farm Economics, 43, 44–56.

[53] Norets, A. (2010). Approximation of conditional densities by smooth mixtures of regressions. The Annals of Statistics 38 (3), 1733–1766.

[54] Olley, G. S., & A. Pakes (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. Econometrica 64 (6), 1263–1297.

[55] O'Mahony, M. and Vecchi, M., 2009. R&D, knowledge spillovers and company productivity performance. Research Policy, 38(1), pp.35-44.

[56] Petrin, A., & J. Sivadasan (2013). Estimating Lost Output from Allocative Inefficiency, with an Application to Chile and Firing Costs. The Review of Economics and Statistics 95 (1), 286–301.

[57] Pindyck, R. S., 1982. Adjustment Costs, Uncertainty, and the Behavior of the Firm. American Economic Review, 72 (3), 415–427.

[58] Robert, C. P., and G. Casella (2004). *Monte Carlo statistical methods*. Springer Texts in Statistics. Springer-Verlag, New York, second edition.

[59] Romero-Jordán, D., Sanz-Labrador, I., and Sanz-Sanz, J. F. (2020). Is the corporation tax a barrier to productivity growth? Small Business Economics, 55 (1), 23-38.

[60] Slemrod, J., 1992. Do Taxes Matter? Lessons from the 1980's. The American Economic Review, 82(2), 250-256.

[61] Terrell, D. (1996). Imposing monotonicity and concavity restrictions in flexible functional forms, Journal of Applied Econometrics 11, 179–194.

[62] Voglis, C., P.E. Hadjidoukas, E. Lagaris, D.G. Papageorgiou (2009). A numerical differentiation library exploiting parallel architectures. Computer Physics Communications 180 (8), 1404–1415.

[63] White, H. (1989). Learning in artificial neural networks: A statistical perspective. Neural Computation, 1, 425–464.

[64] White, H. (1990). Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings, Neural Networks 3, 535–549.

[65] Whiteley, N., Sumeetpal, S., Godsill, S., 2010. Auxiliary Particle Implementation of Probability Hypothesis Density Filter. IEEE Transactions on Aerospace and Electronic Systems 46 (3), 1437–1454.

[66] Wooldridge, J. M. (2009). On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables. Economics Letters 104 (3):112–114.

[67] Zellner, A. (1971). An introduction to Bayesian inference in econometrics. Wiley, New York.

Appendix A

A.1 Posterior

Our system of equations consists of the production function in (1), the general f.o.c.'s in (7), the productivity specification in (3) and (4), and the priors in (9). Our posterior is as follows.

$$
\begin{gathered}
p(\theta, \Lambda, \Omega, \Sigma, \Sigma_\gamma, \Sigma_{\gamma^\omega}, \sigma_\lambda, \sigma_\varrho, \sigma_{\xi^\omega} | D) \propto |\Sigma|^{-(nT+\bar{n}+dim(\Sigma)+1)/2} \\
\exp\left\{ -\tfrac{1}{2}\bar{q}\mathrm{tr}\Sigma^{-1} - \tfrac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T} \mathbf{V}_{it}(\theta_i, \Lambda_{it}, \Omega_t)'\Sigma^{-1}\mathbf{V}_{it}(\theta_i, \Lambda_{it}, \Omega_t) \right\} \cdot \\
\prod_{i=1}^{n} p(\theta_i) \cdot p(\Theta) \cdot \\
(\sigma_{\xi^\Omega})^{-(T+\bar{n}+1)/2)} \exp\left\{ -\frac{1}{2\sigma^2_{\xi^\Omega}} \left[\bar{q} + \sum_{t=1}^{T}(\Omega_t - \rho\Omega_{t-1} - \Phi_\Omega(Z_t; \gamma_\Omega))^2 \right] \right\},
\end{gathered}
\tag{A.1}
$$

where D denotes the data, $\Lambda_t = (\omega_{it}, \tilde{\omega}_{it}, i \in \mathbb{I}, t \in \mathbb{T})$, $\Omega = (\Omega_t, t \in \mathbb{T})$, $\theta = (\theta_i, i \in \mathbb{I})$, $\theta_i = (\beta_i, \gamma_i, \gamma_{\omega,i}, \lambda_i, \psi_i, \varrho_i)$, the "structural" parameters are $\Theta = (\gamma_\Omega, \gamma_\Pi, \bar{\gamma}, \sigma_{\xi^\Omega})$, and

$$
\mathbf{V}_{it}(\theta, \Lambda_{it}) \equiv \begin{bmatrix} y_{it} - f(x_{it}, k_{it}; \beta_i) - \omega_{it} \\ x_{it} - \Phi(k_{it}, \omega_{it}, z_{it}; \gamma_i) \\ \omega_{it} - \lambda_i\Omega_t - \frac{1}{2}\psi_i\Omega_t^2 - \Phi_\omega(z_{it}; \gamma_{\omega,i}) - \tilde{\omega}_{it} \\ \tilde{\omega}_{it} - \varrho_i\tilde{\omega}_{i,t-1} \end{bmatrix}. \tag{A.2}
$$

Integrating with respect to $\Sigma, \Sigma_\gamma, \Sigma_{\gamma^\omega}, \sigma_\lambda, \sigma_\varrho, \sigma_{\xi^\Omega}$, we obtain (Zellner, 1971, p. 273)

$$
\begin{gathered}
p(\theta, \Lambda, \Omega | D) \propto \left\| \left\{ \bar{q}\mathbf{I} + \sum_{i=1}^{n}\sum_{t=1}^{T} \mathbf{V}_{it}(\theta_i, \Lambda_{it}, \Omega_t)\mathbf{V}_{it}(\theta_i, \Lambda_{it}, \Omega_t)' \right\} \right\|^{-(nT+\bar{n}+1)/2} \cdot \\
\left| \bar{q} + \sum_{i=1}^{n}(\gamma_i - \bar{\gamma})(\gamma_i - \bar{\gamma})' \right|^{-(n+\bar{n})/2)} \cdot \left| \bar{q} + \sum_{i=1}^{n}(\gamma_{\omega,i} - \bar{\gamma})(\gamma_{\omega,i} - \bar{\gamma})' \right|^{-(n+\bar{n})/2)} \cdot \\
\left(\sum_{t=1}^{T}(\Omega_t - \rho\Omega_{t-1} - \Phi_\Omega(Z_t; \gamma_\Omega)) \right)^{-(T+\bar{n})/2} \cdot \\
\prod_{i=1}^{n} p(\theta_i) \cdot p(\Theta).
\end{gathered}
\tag{A.3}
$$

We can draw the θ_is using simple and well-known Gibbs sampling steps using their posterior conditional distributions and the same is true for Θ (notice that for γ_Π we resort to fast MCMC as described below.) For the latent variables we use the Sequential Monte Carlo algorithm of Creal and Tsay (2015).

A.2 Fast MCMC

Regarding the Girolami and Calderhead (2011, GC) algorithm to update draws for $\boldsymbol{\theta}$, we use local information about both the gradient and the Hessian of the log-posterior conditional of $\boldsymbol{\theta}$ at the existing draw. A Metropolis test is again used for accepting the candidate so generated but the GC algorithm moves considerably faster relative to our naive scheme previously described. It has been found that the GC algorithm performs much better than a standard Metropolis-Hastings algorithm and autocorrelations are, more often than not, much smaller. Suppose $\mathscr{L}(\boldsymbol{\theta}) = \log p(\boldsymbol{\theta}|\boldsymbol{X})$ is used to denote for simplicity the log posterior of $\boldsymbol{\theta}$. Moreover, define:

$$\boldsymbol{G}(\boldsymbol{\theta}) = \text{est.cov} \tfrac{\partial}{\partial \boldsymbol{\theta}} \log p(\boldsymbol{X}|\boldsymbol{\theta}), \tag{A.4}$$

which is the empirical counterpart of:

$$\boldsymbol{G}_o(\boldsymbol{\theta}) = -\mathbb{E}_{\mathscr{Y}|\boldsymbol{\theta}} \tfrac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \log p(\boldsymbol{X}|\boldsymbol{\theta}). \tag{A.5}$$

The Langevin diffusion is given by the stochastic differential equation:

$$d\boldsymbol{\theta}(t) = \tfrac{1}{2} \tilde{\nabla}_{\boldsymbol{\theta}} \mathscr{L} \{\boldsymbol{\theta}(t)\} dt + d\mathbf{B}(t), \tag{A.6}$$

where

$$\tilde{\nabla}_{\boldsymbol{\theta}} \mathscr{L} \{\boldsymbol{\theta}(t)\} = -\boldsymbol{G}^{-1} \{\boldsymbol{\theta}(t)\} \cdot \tilde{\nabla}_{\boldsymbol{\theta}} \mathscr{L} \{\boldsymbol{\theta}(t)\}, \tag{A.7}$$

is the so called "natural gradient" of the Riemann manifold generated by the log posterior. The elements of the Brownian motion are

$$\begin{aligned} \boldsymbol{G}^{-1} \{\boldsymbol{\theta}(t)\} d\mathbf{B}_i(t) = & |\boldsymbol{G}\{\boldsymbol{\theta}(t)\}|^{-1/2} \sum_{j=1}^{K_\beta} \tfrac{\partial}{\partial \boldsymbol{\theta}} \left[\boldsymbol{G}^{-1} \{\boldsymbol{\theta}(t)\}_{ij} |\boldsymbol{G}\{\boldsymbol{\theta}(t)\}|^{1/2} \right] dt \\ & + \left[\sqrt{\boldsymbol{G}\{\boldsymbol{\theta}(t)\}} d\boldsymbol{B}(t) \right]_i . \end{aligned} \tag{A.8}$$

The discrete form of the stochastic differential equation provides a proposal as follows:

$$\begin{aligned} \tilde{\boldsymbol{\theta}}_i = & \boldsymbol{\theta}_i^o + \tfrac{\varepsilon^2}{2} \left\{ \boldsymbol{G}^{-1}(\boldsymbol{\theta}^o) \nabla_{\boldsymbol{\theta}} \mathscr{L}(\boldsymbol{\theta}^o) \right\}_i - \varepsilon^2 \sum_{j=1}^{K_\theta} \left\{ \boldsymbol{G}^{-1}(\boldsymbol{\theta}^o) \frac{\partial \boldsymbol{G}(\boldsymbol{\theta}^o)}{\partial \boldsymbol{\theta}_j} \boldsymbol{G}^{-1}(\boldsymbol{\theta}^o) \right\}_{ij} \\ & + \tfrac{\varepsilon^2}{2} \sum_{j=1}^{K_\theta} \left\{ \boldsymbol{G}^{-1}(\boldsymbol{\theta}^o) \right\}_{ij} \text{tr} \left\{ \boldsymbol{G}^{-1}(\boldsymbol{\theta}^o) \frac{\partial \boldsymbol{G}(\boldsymbol{\theta}^o)}{\partial \boldsymbol{\theta}_j} \right\} + \left\{ \varepsilon \sqrt{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)} \boldsymbol{\xi}^o \right\}_i \\ = & \boldsymbol{\mu}(\boldsymbol{\theta}^o, \varepsilon)_i + \left\{ \varepsilon \sqrt{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)} \boldsymbol{\xi}^o \right\}_i , \end{aligned} \tag{A.9}$$

where $\boldsymbol{\beta}^o$ is the current draw. We select ε so that, approximately, 20% of candidates are eventually accepted. The proposal density is:

$$\tilde{\boldsymbol{\theta}} | \boldsymbol{\theta}^o \sim \mathcal{N}_{K_\theta} \left(\tilde{\boldsymbol{\theta}}, \varepsilon^2 \boldsymbol{G}^{-1}(\boldsymbol{\theta}^o) \right). \tag{A.10}$$

Finally, convergence to the invariant distribution suggests using the Metropolis-Hastings probability:

$$\min\left\{1,\ \frac{p\left(\tilde{\boldsymbol{\theta}}|\cdot,\mathscr{Y}\right)q\left(\boldsymbol{\theta}^o|\tilde{\boldsymbol{\theta}}\right)}{p\left(\boldsymbol{\theta}^o|\cdot,\mathscr{Y}\right)q\left(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}^o\right)}\right\}. \tag{A.11}$$

We use a recent advance on the Metropolis Adjusted Langevin Algorithm (MALA) called fast MALA (fMALA), see Durmus et al. (2017). Suppose we have a parameter vector $\boldsymbol{\theta} \in \Re^d$, and we target $\pi(\boldsymbol{\theta})$ which represents the posterior, omitting the dependence on data to ease notation. We consider a Langevin diffusion defined by:

$$\mathrm{d}\boldsymbol{\theta}_t = \tfrac{1}{2}\boldsymbol{\Sigma}\cdot\nabla\ln\pi(\boldsymbol{\theta}_t) + \boldsymbol{\Sigma}^{1/2}\mathrm{d}\boldsymbol{W}_t, \tag{A.12}$$

where $\{\boldsymbol{W}_t, t \geq 0\}$ is a standard d-dimensional Brownian motion, and $\boldsymbol{\Sigma}$ is a given positive definite self-adjoint matrix. Under appropriate assumptions on π one can show that the dynamics generated by (A.12) are ergodic and result in $\pi(\boldsymbol{\theta})$ as the unique invariant distribution. A standard approach is to discretize (A.12) using a one step integrator, and sample using the averages over the numerical trajectories. This approach introduces a bias because the posterior does not coincide in general with the exact π. An alternative way of sampling from π exactly, i.e. that is not biased by discretizing (1), is by using the Metropolis-Hastings algorithm (Hastings, 1970). The idea is to construct a Markov chain $\{\boldsymbol{\theta}_j\}$, where at each step j, given $\boldsymbol{\theta}_j$, a new sample proposal $\boldsymbol{\theta}^c$ is generated from the Markov chain with transition kernel $q(\boldsymbol{\theta},\cdot)$. This proposal is then accepted ($\boldsymbol{\theta}_{j+1} = \boldsymbol{\theta}^c$) with probability $\alpha(\boldsymbol{\theta}_j, \boldsymbol{\theta}^c)$ and rejected ($\boldsymbol{\theta}_{j+1} = \boldsymbol{\theta}_j$) otherwise. If we have

$$\alpha(\boldsymbol{\theta},\boldsymbol{\theta}^c) = \min\left\{1,\ \frac{\pi(\boldsymbol{\theta}^c)q(\boldsymbol{\theta}^c,\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})q(\boldsymbol{\theta},\boldsymbol{\theta}^c)}\right\}, \tag{A.13}$$

then the resulting Markov chain $\{\theta_j\}$ is π-invariant and will, for large j generate samples from π under mild ergodicity assumptions (Liu, 2008, Robert and Casella, 2004). In general, a candidate is generated as:

$$\boldsymbol{\theta}^c = \boldsymbol{\mu}(\boldsymbol{\theta},h) + \boldsymbol{S}(\boldsymbol{\theta},h)\boldsymbol{\zeta}, \tag{A.14}$$

where $\boldsymbol{\zeta} \sim \mathcal{N}_d(\mathbf{0},\mathbf{I}_d)$. The specific fMALA proposal has

$$\boldsymbol{\mu}(\boldsymbol{\theta},h) = x + \tfrac{h}{2}f(\boldsymbol{\theta}) - \tfrac{h^2}{24}\nabla f(\boldsymbol{\theta})\cdot f(\boldsymbol{\theta}) + \{\boldsymbol{\Sigma} : \nabla^2 f(\boldsymbol{\theta})\}, \tag{A.15}$$

$$\boldsymbol{S}(\boldsymbol{\theta},h) = \left(h^{1/2}\mathbf{I}_d + \tfrac{h^{3/2}}{12}Df(\boldsymbol{\theta})\right)\boldsymbol{\Sigma}^{1/2}, \tag{A.16}$$

where $f(\boldsymbol{\theta}) \triangleq \boldsymbol{\Sigma}\cdot\nabla\ln\pi(\boldsymbol{\theta})$, $\nabla f(\boldsymbol{\theta})$ and $\nabla^2 f(\boldsymbol{\theta})$ are the $d\times d$ Jacobian and $d\times d^2$ Hessian of $f(\boldsymbol{\theta})$, respectively, and $\boldsymbol{\Sigma} = \boldsymbol{S}(\boldsymbol{\theta},h)$. Let $\nabla^2 f(\boldsymbol{\theta}) = [\boldsymbol{H}_1(\boldsymbol{\theta}),\ldots,\boldsymbol{H}_d(\boldsymbol{\theta})]$ where $[\boldsymbol{H}_i(\boldsymbol{\theta})]_{jk} = \frac{\partial f_i(\theta)}{\partial\theta_k\partial\theta_j}$. Then, $\{\boldsymbol{\Sigma} : \nabla^2 f(\boldsymbol{\theta})\}_i \triangleq tr\left[\boldsymbol{\Sigma}'\boldsymbol{H}_i(\boldsymbol{\theta})\right]$. The scaling constant has received detailed attention in Durmus et al. (2017) and it is related directly to the discretization of (A.12). Specifically, Durmus et al. (2017)

recommend $h = \varepsilon d^{-1/5}$ for some positive constant, ε. The optimal acceptance rate maximizing the first-order efficiency is very close to the limiting value of 0.704 predicted in Theorem 3.2 of Durmus et al. (2017). Therefore, one can calibrate the constant ε (during the burn-in phase) so that the acceptance rate is close to 0.70. This approach has been found to perform excellently once ε is and h are calibrated correctly during the burn-in phase. All derivatives are computed numerically[21] during the burn-in phase, and they are interpolated[22] in the main phase of the MCMC algorithm. This results in dramatic computational savings and, as a matter of fact, different chain can be run in parallel in computers with multiple nodes. We run ten different chains starting from randomly selected initial conditions and we compare the chains after 150,000 iterations with a burn-in phase consisting of 50,000 iterations. Our transition density $q(\theta, \theta^c)$ is a d-dimensional Student-t distribution with five degrees of freedom. We monitor convergence using the standard diagnostics of Geweke (1992).

A.3 Sequential Monte Carlo

Additionally, we use a recent advance in sequential Monte Carlo methods known as the particle Gibbs (PG) sampler, see Andrieu et al. (2010). The algorithm allows us to draw paths of the state variables in large blocks. Particle filtering is a simulation based algorithm that sequentially approximates continuous, marginal distributions using discrete distributions. This is performed by using a set of support points called ''particles'' and probability masses; see Creal (2012) for a review.

The PG sampler draws a single path of the latent or state variables from this discrete approximation. As the number of particles M goes to infinity, the PG sampler draws from the exact full conditional distribution. As mentioned in Creal and Tsay (2015, p. 339): "The PG sampler is a standard Gibbs sampler but defined on an extended probability space that includes all the random variables that are generated by a particle filter. Implementation of the PG sampler is different than a standard particle filter due to the ''conditional'' resampling algorithm used in the last step. Specifically, in order for draws from the particle filter to be a valid Markov transition kernel on the extended probability space, Andrieu et al. (2010) note that there must be positive probability of sampling the existing path of the state variables that were drawn at the previous iteration. The pre-existing path must survive the resampling steps of the particle filter. The conditional resampling step within the algorithm forces this path to be resampled at least once. We use the conditional

[21] We use the `Fortran77` subroutines in package `NDL` of Voglis et al. (2009). Specifically we use version 2.0 of Hadjidoukas et al. (2014), `https://data.mendeley.com/datasets/j2fhmszg85/1`, see also `http://cpc.cs.qub.ac.uk/summaries/AEDG_v1_0.html`

[22] We use the `Fortran` subrourines in `finterp` by Jacob Williams in:`https://github.com/jacobwilliams/finterp/blob/master/README.md`
Alternatively, we use for comparison:`RBF_INTERP_ND` in:
`https://people.sc.fsu.edu/~jburkardt/f_src/rbf_interp_nd/rbf_interp_nd.html`.
`RBF_INTERP_ND` is a `Fortan90` library by John Burkardt, which defines and evaluates radial basis function (RBF) interpolants to multidimensional data.

multinomial resampling algorithm from Andrieu et al. (2010), although other resampling algorithms exist, see Chopin and Singh (2013)." We follow Creal and Tsay (2015). Suppose the posterior is $p(\theta, \Lambda_{1:T}|\mathbf{y}_{1:T})$ where $\Lambda_{1:T}$ denotes the latent variables whose prior can be described by $p(\Lambda_t|\Lambda_{t-1}, \theta)$. In the PG sampler we can draw the structural parameters $\theta|\Lambda_{1:T}, \mathbf{y}_{1:T}$ as usual, from their posterior conditional distributions. This is important because, in this way, we can avoid mixture approximations or other Monte Carlo procedures that need considerable tuning and may not have good convergence properties. As such posterior conditional distributions we omit the details and focus on drawing the latent variables. Suppose we have $\Lambda_{1:T}^{(1)}$ from the previous iteration. The particle filtering procedure consists of two phases.

Phase I: Forward filtering (Andrieu et al., 2010).

- Draw a proposal $\Lambda_{i,t}^{(m)}$ from an importance density $q(\Lambda_{i,t}|\Lambda_{i,t-1}^{(m)}, \theta), m = 2, \ldots, M$.
- Compute the importance weights:

$$w_{i,t}^{(m)} = \frac{p(y_{i,t}; \Lambda_{i,t}^{(m)}, \theta)p(\Lambda_{i,t}^{(m)}|\Lambda_{i,t-1}^{(m)}, \theta)}{q(\Lambda_{i,t}|\Lambda_{i,t-1}^{(m)}, \theta)}, m = 1, \ldots, M. \tag{A.17}$$

- Normalize the weights: $\tilde{w}_{i,t}^{(m)} = \frac{w_{it}^{(m)}}{\sum_{m'=1}^{M} w_{it}^{(m')}}, m = 1, \ldots, M$.
- Resample the particles $\{\Lambda_{i,t}^{(m)}, m = 1, \ldots, M\}$ with probabilities $\{\tilde{w}_{i,t}^{(m)}, m = 1, \ldots, M\}$.

In the original PG sampler, the particles are stored for $t = 1, \ldots, T$ and a single trajectory is sampled using the probabilities from the last iteration. An improvement upon the original PG sampler was proposed by Whiteley et al. (2010), who suggested drawing the path of the latent variables from the particle approximation using the backwards sampling algorithm of Godsill et al. (2004). In the forwards pass, we store the normalized weights and particles and we draw a path of the latent variables as we detail below (the draws are from a discrete distribution).

Phase II: Backward filtering (Chopin and Singh, 2013, Godsill et al., 2004).

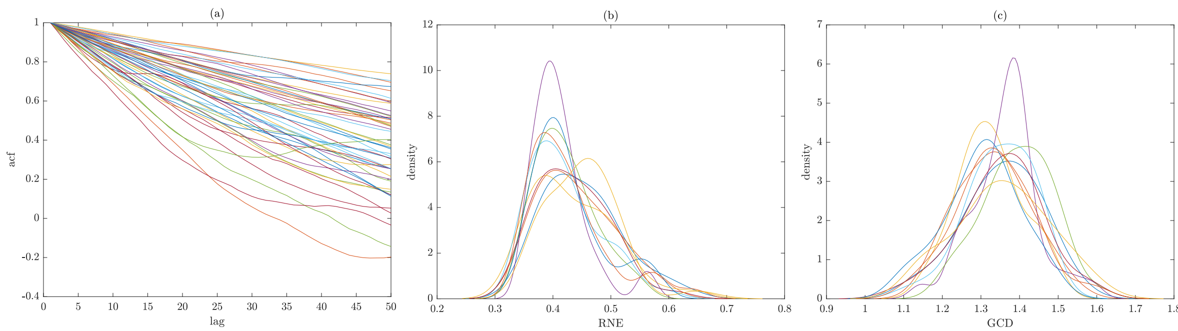
- At time $t = T$ draw a particle $\Lambda_{i,T}^{*} = \Lambda_{i,T}^{(m)}$.
- Compute the backward weights: $w_{t|T}^{(m)} \propto \tilde{w}_{t}^{(m)} p(\Lambda_{i,t+1}^{*}|\Lambda_{i,t}^{(m)}, \theta)$.
- Normalize the weights: $\tilde{w}_{t|T}^{(m)} = \frac{w_{t|T}^{(m)}}{\sum_{m'=1}^{M} w_{t|T}^{(m')}}, m = 1, \ldots, M$.
- Draw a particle $\Lambda_{i,t}^{*} = \Lambda_{i,t}^{(m)}$ with probability $\tilde{w}_{t|T}^{(m)}$.

Therefore, $\Lambda^*_{i,1:T} = \{\Lambda^*_{i1}, \ldots, \Lambda^*_{iT}\}$ is a draw from the full conditional distribution. The backwards step often results in dramatic improvements in computational efficiency. For example, Creal and Tsay (2015) find that $M = 100$ particles is enough. There remains the problem of selecting an importance density $q(\Lambda_{i,t}|\Lambda_{i,t-1}, \theta)$. We use an importance density implicitly defined by $\Lambda_{i,t} = a_{i,t} + \sum_{p=1}^{P} b_{i,t}\Lambda^p_{i,t-1} + h_{i,t}\xi_{i,t}$ where $\xi_{i,t}$ follows a standard (zero location and unit scale) Student-t distribution with $\nu = 5$ degrees of freedom. That is, we use polynomials in $\Lambda_{i,t-1}$ of order P. We select the parameters $a_{i,t}, b_{i,t}$ and $h_{i,t}$ during the burn-in phase (using $P = 1$ and $P = 2$) so that the weights $\{\tilde{w}^{(m)}_{i,t}, m = 1, \ldots, M\}$ and $\{\tilde{w}^{(m)}_{t|T}, m = 1, \ldots, M\}$ are approximately not too far from a uniform distribution. Chopin and Singh (2013) have analyzed the theoretical properties of the PG sampler, and proved that the sampler is uniformly ergodic. They also prove that the PG sampler with backwards sampling strictly dominates the original PG sampler in terms of asymptotic efficiency. Alternatively, when the dimension of the state vector is large, we can draw $\Lambda_{i,1:T}$, conditional on all other paths $\Lambda_{-i,1:T}$ that are not path i. Therefore, we can draw from the full conditional distribution $p(\Lambda_{i,1:T}|\Lambda_{-i,1:T}, \mathbf{y}_{1:T}, \theta)$.

A.4 MCMC performance

MCMC can be potentially misleading if the draws are highly autocorrelated. Of course, some autocorrelation is unavoidable but the empirical issue is whether it is so large as to prevent thorough exploration of the posterior implied by highly complex, deep ANNs. Autocorrelation can be detected using the autocorrelation functions (acf) corresponding to MCMC draws. Additionally, one can look into relative numerical efficiency (RNE, Geweke, 1992) which is equal to one in the case of i.i.d draws from the posterior. Finally, convergence can be tested using Geweke's (1992) z-diagnostic known as Geweke Convergence Diagnostic (GCD). As we have a large number of parameters in the reduced form, we report maximum autocorrelations across all parameters, minimum RNE and maximum GCDs in Figure **??**. Autocorrelation functions (acf's) are reported in panel (a), RNE densities in panel (b), and GCDs in panel (c). From this evidence, although there is autocorrelation, this does not affect MCMC convergence (panel (c)) and, certainly, autocorrelation is not destructively large.

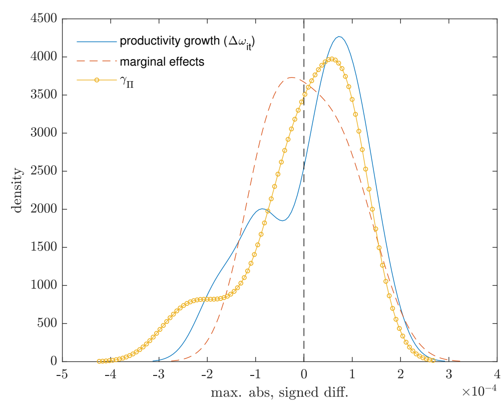
Figure A.1: MCMC performance



A.5 Posterior robustness

Bayesian analysis can be potentially sensitive to results in the prior. In this section, we examine posterior sensitivity of productivity growth and marginal effects under different, automatically generated priors. Specifically, we change $\bar{n}, \bar{q}$,by selecting them uniformly in the intervals (0.001, 100) a total of 10,000 times. $\bar{\gamma}$ and $\bar{\gamma}_{\Pi}$ are changed by drawing their elements uniformly in the interval $(-100, 100)$. In Figure A.2 we report (i) the maximum absolute signed changes in estimates (posterior means) of productivity growth, (ii) the maximum absolute signed changes in estimates (posterior means) of marginal effects of taxes and R&D in ω_{it} and Ω_t; the fundamental outputs of our model and finally, (iii) the maximum absolute signed changes in estimates (posterior means) of γ_{Π}, relative to the benchmark which is the posterior means corresponding to the default prior choice developed in the main text. From this evidence, posterior robustness is quite impressive and the effects of prior specification are dominated strongly by the data.

Figure A.2: Posterior robustness



Appendix B

B1:Two Digit Nace Industry Classification

Code	Description
10	Food products
11	Beverages
12	Tobacco products
13	Textiles
14	Wearing apparel
15	Leather and related products
16	Wood products
17	Paper and paper products
18	Printing and reproduction
19	Coke and refined petroleum
20	Chemicals
21	Basic pharmaceutical
22	Rubber and plastics
23	Other non-metallic
24	Basic metals
25	Fabricated metal products
26	Electronics
27	Electrical equipment
28	Machinery and equipment
29	Motor vehicles
30	Other transport equipment
31	Furniture
32	Other manufacturing
33	Repairs