The Economic Geography of Innovation

Peter H. Egger* — ETH Zürich — CEPR, CESifo, GEP, WIFO

Nicole Loumeau† — ETH Zürich

Februar 2019

Abstract: This paper outlines a quantitative global multi-region model to assess the importance of country-level investment incentives towards innovation at the level of 5,633 regions of heterogeneous size. While incentives vary across countries (and time), the responses are largely heterogeneous across regions within as well as across countries. The reason for this heterogeneity roots in average technology differences – in terms of the production of both output and innovation – as well as in the geography (location) and amenities across regions. The model and quantitative analysis take the tradability of output as well as the mobility of people across regions into account. In the counterfactual equilibrium analysis we focus on the effects of R&D-investment incentives on three key variables – place-specific employment, productivity, and welfare – in a scenario where investment incentives towards innovation are abandoned. We find that the use of policy instruments which are designed to stimulate private R&D are globally beneficial in terms of productivity and welfare. In particular, low-amenity, peripheral places, and ones where patenting is relatively less common than elsewhere benefit more strongly than others, which implies that the studied nation-wide investment incentives also work as place-based policies. According to the quantification, about one-tenth of the long-run growth rate of real GDP on the globe can be attributed to the use of R&D investment incentives as used in the year 2005 alone.

Keywords: Economic geography; Innovation; Trade; Labor mobility; Quantitative general equilibrium; Structural estimation.

JEL classification: C68; F13; F14; O31; R11.

*ETH Zürich, Department of Management, Technology, and Economics, Leonhardstrasse 21, 8092 Zurich, Switzerland; E-mail: egger@kof.ethz.ch; + 41 44 632 41 08.

†ETH Zürich, Department of Management, Technology, and Economics, Leonhardstrasse 21, 8092 Zurich, Switzerland; E-mail: nicole.loumeau@kof.ethz.ch; +41 44 632 33 48.

We gratefully acknowledge financial support by the Swiss National Science Foundation grant CRSII1_154446 "Inequality and Globalization: Demand versus Supply Forces and Local Outcomes". We thank Pol Antràs, Johnathan Eaton, Pablo Fajgelbaum, Ana Cecilia Fieler, Wolfgang Keller, Stephen Redding, Esteban Rossi-Hansberg, Felix Tintelnot, and Daniel Xu for comments and suggestions. We are also grateful to conference and seminar participants at Florence, Hongkong, Kaifeng, NBER, Sarnen, Tokio, Villars and Zurich for helpful comments.

1 Introduction

Technology and productivity are key drivers not only of production potential of places but also of the attractiveness for mobile factors to locate there and, hence, of demand potential and well-being. The technological capabilities of production factors located in a place are influenced to a major extent by local innovation and the capability of absorbing innovations generated elsewhere. Policy makers have a number of instruments at hand which are particularly aimed at stimulating innovation for exactly that reason. Earlier research concerned with the effect of innovation incentives – where innovations are commonly measured by patent filings and other patenting behavior – on economic outcomes focuses largely on reduced-form effects, which abstract from general-equilibrium repercussions. One related strand of reduced-form work focuses on the effect of R&D tax incentives on innovation (see De Jong and Verhoeven, 2007, for the Netherlands; Ernst and Spengel, 2011, for multiple EU countries; Westmore, 2013, for 19 OECD countries; Araclia and Botric, 2013, for Croatia; Baumann et al., 2014, for European countries; Czarnitzki et al., 2014, for Canada; Boesenberg and Egger, 2016, for 106 countries). Another strand of reduced-form work highlights the effect of R&D tax incentives on productivity (see Caiumi, 2011, for Italy; Hallépée and Garcia, 2012, for France; and Cappelen et al., 2007, for Norway). The shortcoming of reduced-form analysis is that quantitatively potentially important interdependencies of outcomes, markets, and places are ignored by assumption. Also, the heterogeneity of places in their response to even a homogeneous treatment of economic policy is beyond the reach of a reduced-form analysis, at least to the degree that is suggested by general-equilibrium theory.

The present paper adopts a structural approach, which permits accounting for direct and indirect (spillover plus general-equilibrium) effects of customary innovation-stimulating policies. For this purpose, it formulates, estimates key parameters of, and calibrates a quantitative, multi-place model of trade and factor mobility among places in order to assess the economic value of innovation incentives and their consequences for the location of supply and demand across places as well as for the well-being of consumers there. With this agenda, the paper particularly relates to three lines of work. The one on the social cost-benefit and aggregate analysis of individual tax incentives towards R&D (see Cornet, 2001, and Lokshin and Mohnen, 2012, for the Netherlands; Parsons and Phillips, 2007, for Canada; and Bloom et al., 2013, for the United States) which is based on reduced-form estimates but aims at accounting for effects on various outcomes. The structural aggregate (macro-economic) modeling approach in Atkeson and

Burstein (2018), which permits gauging global effects of innovation policies, abstracting from the multi-place structure of the world economy. And a host of studies with a focus on structural-quantitative, multi-region models with mobile goods and factors without a deeper consideration of innovation policies (see Allen and Arkolakis, 2014; Ahlfeldt et al., 2015; Donaldson and Hornbeck, 2016; Nagy, 2017; Allen and Donaldson, 2018; Monte et al., 2018; Caliendo et al., 2015 and 2018; Desmet et al., 2018; Donaldson, 2018; see Redding and Rossi-Hansberg, 2017, for an extensive review of that line of work).

The model we propose builds on Allen and Arkolakis (2014), Desmet et al. (2018), and Allen and Donaldson (2018) and describes a world in which each place is unique in terms of amenities, productivity, and geography. Firms have an incentive to innovate as it improves their productivity and competitiveness. However, the benefits from innovation which are exclusive to the firm are short-lived, and knowledge about any newly-invented technology becomes public after one period. The technology available to firms in a place evolves through an endogenous dynamic process. Innovation is produced under constant returns to scale, using research labor for each unit of innovation produced. In contrast to Allen and Arkolakis (2014) and Desmet et al. (2018), total factor productivity consists of a random and a chosen part through (optimal) investments in innovation. The parametrization and estimation of the endogenous productivity component as well as of the dynamic technology process are at the heart of the paper's interest. Firms benefit from R&D investment incentives in places, *ceteris paribus*, as they reduce the costs of generating innovations all else equal. Firms use patented as well as non-patented innovations in doing so.

Our analysis considers 5,633 places/regions in 213 countries around the globe, where the delineation of places follows the definition by the Organization of Economic Cooperation and Development (OECD) and their Regional Patent-statistical Database (REGPAT). For the estimation of the R&D-worker-specific productivity shifter, we use region-specific efficiency levels that are recovered from the model structure and five country-specific indicators on R&D investment incentives which are geared towards innovations from Boesenberg and Egger (2016).

In the counterfactual equilibrium analysis we focus on the effects on three key variables – place-specific employment, productivity, and welfare – in a scenario where investment incentives towards innovation are abandoned. There are three main take-aways from the analysis. First, the use of policy instruments which are designed to stimulate private R&D are globally beneficial in terms of productivity and welfare. In other words, also countries and their regions who do not use such instruments benefit from their use elsewhere due to technology spillovers. Second, the long-run relocation effects due to a hypothetical abolishment of R&D tax incentives are substantial and lead to a re-shifting of the population towards high-density areas (i.e., centrally-located ones with great exogenous amenities). Hence, transport accessibility and good exogenous amenities work as a quasi-insurance against adverse innovation policy shocks. Analogously, the quantitative analysis suggests that a nation-wide innovation policy works indirectly as a place-based policy, where low-amenity, peripheral regions benefit, *ceteris paribus*, relatively strongly than high-amenity, centrally located ones. Low-amenity, peripheral regions, *ceteris paribus*, gain relatively in international competitiveness from national R&D policies due to the cross-border mobility of labor.

Furthermore, the quantitative analysis suggests that about one-tenth of the long-run growth rate of real GDP on the globe can be attributed to the use of R&D policy instruments as used in the year 2005 alone. The findings also imply that only a relatively small fraction of that should be attributed to the stimulus on patenting, but the share of non-patented innovations triggered by such policy instruments is relatively large.

The remainder of the paper is organized as follows. Section 2 presents the model, states the equilibrium conditions for each period and defines the underlying assumptions for a unique balanced growth path to exist. Section 3 discusses the calibration of key model parameters, including a methodology to determine or estimate them. Section 4 presents the results of our counterfactual analysis. Section 5 concludes.

2 The Model

We consider a world where S is a set of regions r on a two-dimensional surface, i.e., $r \in S$. Region r has land density $G_r > 0$, where G_r is exogenously given and normalized by the average land density of all regions in the world. The world is inhabited by a measure $\bar{L}$ of workers, who are freely mobile between regions and endowed with one inelastically-supplied unit of labor each. Each region is unique in terms of geography, amenities and productivity.

2.1 Innovation and Production

In each region, firms produce product varieties ω, innovate, and trade subject to iceberg transport costs. A firm's production of ω per unit of land in the intensive form is defined as

$$q_{rt}(\omega) = \phi_{rt}(\omega)^{\gamma_1} \; z_{rt}(\omega) \; L_{rt}(\omega)^{\mu}, \quad \gamma_1, \mu \in (0,1]. \tag{1}$$

Output depends on production labor per unit of land, $L_{rt}(\omega)$, and the firm's total factor productivity, which is determined by two components: an endogenous innovation component, $\phi_{rt}(\omega)$, and an exogenous, product-specific productivity factor, $z_{rt}(\omega)$, which is drawn from a Fréchet distribution with location parameter $T_{rt} = \tau_{rt}\bar{L}_{rt}^{\alpha}$ and shape parameter θ, where $\alpha \geq 0$ and $\theta > 0$. Where in the productivity distribution a firm is located depends on the total workforce at region r in period t, $\bar{L}_{rt}$, and the region's level of efficiency, τ_{rt}.

The value of τ_{rt} is determined by an endogenous dynamic process, which depends on past investments into local innovations, and the capability of absorbing innovations that were generated elsewhere and now diffuse globally. Assuming a first-order autoregressive process about efficiency,[1] we postulate

$$\tau_{rt} = \phi_{rt-1}^{\gamma_1\theta} \quad \left[\int_S W_{rs}\tau_{st-1}ds\right]^{1-\gamma_2} \tau_{rt-1}^{\gamma_2}, \tag{2}$$

where $\gamma_1, \gamma_2 \in (0,1)$ and W_{rs} is an rs-specific technology diffusion weighting scalar. The value of γ_2 determines the strength of technological diffusion. The higher γ_2, the more a region benefits from own investments in technology. In return, low levels of γ_2 imply that the aggregate level of investment into technology in a region is relatively more important than local investments.

Firms have an incentive to invest into innovation as it improves their productivity in (1). This allows them to post a higher bid for the regionally fixed factor of production, land. However, due to a decreasing marginal product of labor, the innovation effort will be finite. The latter is guaranteed by the parameter configuration where land intensity is larger than the cost normalized innovation intensity in production, $[1-\mu] > \gamma_1/\xi$. Innovation, $\phi_{rt}(\omega)$, is produced under Cobb-Douglas technology and with constant returns to scale, such that a firm has to employ $\nu\phi_{rt}(\omega)^{\xi}h_{rt}^{-1}$ additional units of labor in order to innovate, where $h_{rt} \geq 1$ is a region-time-specific R&D-worker productivity shifter, which reduces the cost of innovation per unit of innovation produced. The latter will be key to the analysis here, as it captures the influence of R&D tax incentives.

Firms enjoy the benefit of their innovation for only one period. In the next period all

[1] Allowing for a longer memory in the process would be technically straightforward. E.g., Allen and Donaldson (2018) consider a second-order process. However, the available data for the present paper do not permit doing so, as the time series available for each region is extremely short, as will become clear below.

entrants to the market have the same access to technology. This simplifies the dynamic profit maximization to a sequence of static problems. After learning their productivity draw $z_{rt}(\omega)$, firms maximize their profits with choosing the level of employment and innovation through

$$\max_{L_{rt}(\omega),\phi_{rt}(\omega)} \quad p_{rt}(\omega)\ \phi_{rt}(\omega)^{\gamma_1}\ z_{rt}(\omega)\ L_{rt}(\omega)^{\mu} - w_{rt}[L_{rt}(\omega) + \phi_{rt}(\omega)^{\xi} h_{rt}^{-1}] - b_{rt},$$

where p_{rt} is the price a firm charges for a product that is sold in region r and period t. A firm's productivity affects prices without changing unit costs, o_{rt}, such that $p_{rt}(\omega) = o_{rt}/z_{rt}(\omega)$, with

$$o_{rt} \equiv \left[\frac{1}{\mu}\right]^{\mu} \left[\frac{\nu\xi}{\gamma_1}\right]^{1-\mu} \left[\frac{b_{rt}\gamma_1}{w_{rt}\nu(\xi(1-\mu)-\gamma_1)}\right]^{(1-\mu)-\frac{\gamma_1}{\xi}} h_{rt}^{-\frac{\gamma_1}{\xi}} w_{rt}. \tag{3}$$

Each firm considers their production unit costs as given, which is why o_{rt} is not product-specific. b_{rt} reflects the firms' bid rent for land, which can be derived from the first-order conditions as a function of the per-unit costs of innovation $w_{rt}\phi_{rt}(\omega)^{\xi}h_{rt}^{-1}$, so that

$$b_{rt} = \left[\frac{\xi(1-\mu)}{\gamma_1} - 1\right] \nu w_{rt}\phi_{rt}(\omega)^{\xi} h_{rt}^{-1}. \tag{4}$$

2.2 Innovation and Total Employment

Total employment in region r at period t is the sum of production workers, $L_{rt}(\omega)$, and innovation workers, $\nu\phi_{rt}(\omega)^{\xi}h_{rt}^{-1}$, so that

$$\bar{L}_{rt}(\omega) = L_{rt}(\omega) + \nu\phi_{rt}(\omega)^{\xi}h_{rt}^{-1} = L_{rt}(\omega)\left[1 + \frac{\gamma_1}{\mu\xi}\right], \tag{5}$$

where the last equality follows from the first-order relation between production labor and innovation labor,

$$\frac{\xi}{\gamma_1}\nu\phi_{rt}(\omega)^{\xi}h_{rt}^{-1} = \frac{L_{rt}(\omega)}{\mu} \quad \Rightarrow \quad \nu\phi_{rt}(\omega)^{\xi}h_{rt}^{-1} = \frac{\gamma_1}{\xi[\mu+\gamma_1/\xi]}\bar{L}_{rt}(\omega). \tag{6}$$

2.3 Utility and Consumption

When choosing residence in region r, a representative worker in period t derives utility from local amenities, a_{rt}, and from consuming a set of differentiated product varieties ω

with CES preferences according to

$$u_{rt} = a_{rt}C_{rt} = a_{rt}\left[\int_0^1 c_{rt}(\omega)^{\frac{\sigma-1}{\sigma}}d\omega\right]^{\frac{\sigma}{\sigma-1}} \quad \text{with} \;\; a_{rt} = \bar{a}_{rt}\;\bar{L}_{rt}^{-\lambda}, \tag{7}$$

where a_{rt} are amenities at r in t, with $\bar{a}_{rt}$ being an exogenous amenity attribute and $\lambda \geq 0$ being a congestion externalities parameter. C_{rt} is the real consumption bundle, and $\sigma \in (1, \infty)$ is the elasticity of substitution between products ω.

Consumer-workers earn income from work, w_{rt}, and from local ownership of land. Local land rents are uniformly distributed among all residents in a region, i.e., the land rent per resident is $b_{rt}/\bar{L}_{rt}$. As we assume that agents cannot write debt contracts with each other and there is perfect local competition, it follows that each consumer-worker spends all her income. Hence, the indirect utility is defined as

$$u_{rt} = a_{rt}y_{rt} = a_{rt}\frac{w_{rt} + b_{rt}/\bar{L}_{rt}}{P_{rt}}, \tag{8}$$

where $P_{rt} = \Gamma\left(\frac{1-\sigma}{\theta}+1\right)^{\frac{1}{1-\sigma}}\left[\int_S T_{kt}[o_{kt}\zeta_{ks}]^{-\theta}dk\right]^{-\frac{1}{\theta}}$ is the price index in region r and period t. As in Eaton and Kortum (2002), the share of consumption in region r of products produced in region s is determined by

$$\pi_{rst} = \frac{T_{rt}[o_{rt}\zeta_{rs}]^{-\theta}}{\int_S T_{kt}[o_{kt}\zeta_{ks}]^{-\theta}dk}, \;\; \forall r, s \in S, \tag{9}$$

where $\zeta_{rs} > 1$ denote the iceberg costs of transporting a product from r to s.

2.4 Equilibrium in Each Period

Profits and utility are maximized within each period, as neither firms nor consumers are forward-looking; see also Desmet et al. (2018) and Allen and Donaldson (2018).

The equilibrium population density will be evaluated as a measure of the location specific utility, u_{rt}, such that

$$\bar{L}_{rt} = \frac{\bar{L}}{G_r}\frac{{u_{rt}}^{1/\Omega}}{\int_S u_{kt}^{1/\Omega}dk}, \quad \text{with} \int_S G_r L_{rt}dr = \bar{L}, \tag{10}$$

where Ω is a Fréchet dispersion parameter of a location-specific preference shock as in Desmet et al. (2018).[2] Overall, population mobility is restricted by the location-specific

[2]Notice that location-time-specific utility, u_{rt}, and, more specifically, the amenity parameter, a_{rt}, is proportional to average migration costs in region r. In general, residence-region-specific migration costs are isomorphic to location-specific amenities. Hence, the population-share specification in (10) accounts

preference parameter (Ω), an amenity-reducing congestion parameter (λ) and the land-intensity in production ($1-\mu$).

Product-market clearing requires total revenues in region r to be equal to total expenditures on products from region r. Hence,

$$w_{rt} G_r \bar{L}_{rt} = \int_S \pi_{rst} w_{st} G_s \bar{L}_{st} \; ds \quad \forall r, s \in S, \tag{11}$$

where L_{rt} can be replaced with $\bar{L}_{rt}$ as production labor is proportional to total labor across all regions.

In equilibrium, population density in each region is determined by (10), replacing u_{rt} by the indirect utility in (8). The product-market clearing pins down wages, with substituting (4) into (3) and using this expression to replace it into the trade share (9), which in return can be substituted in (11).

An equilibrium exists and is unique if dispersion forces are greater than agglomeration forces. Hence,

$$\underbrace{\frac{\alpha}{\theta} + \frac{\gamma_1}{\xi}}_{\text{Static agglomeration forces}} \leq \underbrace{\lambda + 1 - \mu + \Omega}_{\text{Static dispersion forces}} . \tag{12}$$

A detailed proof of the uniqueness condition can be based on the insights from Allen and Arkolakis (2014), Desmet et al. (2018) and Allen and Donaldson (2018), and it is presented in Appendix B.

2.5 Balanced Growth Path

In a balanced growth path (BGP), technology growth rates are constant and identical across regions at constant fundamentals, implying that $\frac{\tau_{rt+1}}{\tau_{rt}}$ is constant over time and space, and $\frac{\tau_{st}}{\tau_{rt}}$ is constant over time. Firms' investment decisions into innovation are constant, but they differ across regions. In order for a BGP to materialize, we assume that the R&D-worker-specific productivity shifter is constant over time, $h_{rt} = h_r$, in the BGP as well as in the transition towards it. Rewriting the endogenous dynamic process in (2), the growth rate of τ_{rt} can then be expressed as

$$\frac{\tau_{rt+1}}{\tau_{rt}} = \phi_{rt}^{\theta\gamma_1} \left[\int_S \frac{W_{rs} \tau_{st}}{\tau_{rt}} ds \right]^{1-\gamma_2} . \tag{13}$$

---

for such costs.

That growth rate relative to region s' is

$$\frac{\frac{\tau_{rt+1}}{\tau_{rt}}}{\frac{\tau_{st+1}}{\tau_{st}}} = \left[\frac{\tau_{st}}{\tau_{rt}}\right]^{1-\gamma_2} \left[\frac{\phi_{rt}}{\phi_{st}}\right]^{\theta\gamma_1} \left[\frac{\int_S W_{rs}\tau_{st}ds}{\int_S W_{sr}\tau_{rt}dr}\right]^{1-\gamma_2}. \tag{14}$$

In a BGP, $\left(\frac{\tau_{rt+1}}{\tau_{rt}} / \frac{\tau_{st+1}}{\tau_{st}}\right) = 1$. Furthermore, for a BGP to exist, technological diffusion, which is governed by W_{rs}, needs to be uniform across space, implying that $[\int_S W_{rs}\tau_{st}ds / \int_S W_{sr}\tau_{rt}dr] = 1$, see Egger and Pfaffermayr (2006).[3] Desmet et al. (2018) propose to specify $W_{rs} = \left[\frac{1}{S}\right], \forall rs$, and we follow them in this regard. With the latter assumption, (14) reduces to

$$\frac{\tau_{rt}}{\tau_{st}} = \left[\frac{\phi_{rt}}{\phi_{st}}\right]^{\frac{\theta\gamma_1}{1-\gamma_2}} = \left[\frac{\bar{L}_{rt}h_r}{\bar{L}_{st}h_s}\right]^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}}. \tag{15}$$

Then, there exists a unique BGP of the system, if

$$\underbrace{\frac{\alpha}{\theta} + \frac{\gamma_1}{\xi}}_{\text{Static agglomeration forces}} + \underbrace{\frac{\gamma_1}{[1-\gamma_2]\xi}}_{\text{Dynamic agglomeration forces}} \leq \underbrace{\lambda + 1 - \mu + \Omega}_{\text{Static dispersion forces}}, \tag{16}$$

which is the same as in Desmet et al. (2018); see Appendix C.1 for a proof.

In the BGP, aggregate welfare and real consumption growth depends on the population density, the R&D-worker-specific productivity shifter and their distribution in space, according to

$$\frac{u_{rt+1}}{u_{rt}} = \frac{C_{rt+1}}{C_{rt}} = \left[\frac{1}{S}\right]^{\frac{1-\gamma_2}{\theta}} \left[\frac{\gamma_1/\nu}{\gamma_1 + \mu\xi}\right]^{\frac{\theta\gamma_1}{\xi}} \left(\int_S (\bar{L}_s h_s)^{\frac{\theta\gamma_1}{[1-\gamma_2]\xi}} \; ds\right)^{\frac{1-\gamma_2}{\theta}}, \tag{17}$$

where $a_{rt} = a_{rt+1}$, as the population density in each region is constant over time in the BGP.[4]

3 Calibration of Key Model Parameters

To compute the quantitative multi-region equilibrium for each time period from a given year to the steady state (long run), we need the parameters contained in the equations above and summarized in Table 1. Apart from parameters that are common to all

[3] To see this, consider the following thought experiment. Suppose each region r would receive the same time-invariant, common growth impulse. If $[\int_S W_{rs}\tau_{st}ds \neq \int_S W_{sr}\tau_{rt}dr]$, the same impulse would have region-specific consequences due to the importance of the regions' location in the spillover network. Then, the same impulse would be amplified (or moderated) to a heterogeneous degree, and regional growth would be heterogeneous in the BGP, as a result.

[4] A detailed derivation of the growth rate of aggregate welfare is presented in Appendix C.2.

regions and region-specific land endowments which are given in the data, these are initial efficiency levels in production and exogenous amenity levels for all regions, R&D-worker-specific productivity shifter as well as trade costs between all pairs of regions. Table 1 alludes to the sources of these parameters, some of which are collected from other work and some of which are derived (computed or estimated) here.

– Table 1 about here –

We organize the remainder of this section in subsections which pertain to important model blocks based on which estimating equations are formulated or key parameters can be backed out.

3.1 Delineation of Regions and Land Endowments

The delineation of regions used in our analysis is dictated by the definitions used in the Regional Patent-Statistical Database (REGPAT) of the Organization of Economic Cooperation and Development (OECD).[5] In 2005, REGPAT distinguishes 5,633 regions across 213 countries on the globe. The size of regions by land mass (somewhat less so by income or patenting) differs to a large extent. In some countries the granularity of regions is very fine, while it is coarse in others. In some cases, even a whole country is a region (e.g., in some African or Asian and South American countries). This pattern is related to the intensity of patenting in a country: economies with more patents tend to be organized in a more fine-grained fashion, while the ones with less patenting tend to be more coarsely captured. Figure 1 shows a world map of all regions that indicates all countries in the sample with a red color and countries not part of the sample with a white color. In the figure, country borders are drawn in blue and regional borders in yellow. Whenever region and country borders coincide, the yellow region borders are not visible.

– Figure 1 about here –

The map shows that REGPAT region are relatively small (and numerous) in North America (United States, Canada, and Mexico) and Europe. We can link the REGPAT regions with spatial information from two sources: (i) the Geographical Information and Maps (GISCO) database from Eurostat for spatial information on European countries (NUTS3 regions, 2010), and (ii) the Global Administrative Areas (GADM) spatial

[5]The REGPAT database links the Worldwide Statistical Patent Database (PATSTAT) from the European Patent Office (EPO) to 5,633 regions across the globe, utilizing the addresses of the applicants and inventors.

database on administrative boundaries for all other countries. We extract the land mass for each region using ArcGIS software after excluding water sheds within the boundaries of a region and normalize the region-specific land mass by the average landmass, $\frac{1}{S}\sum_{r=1}^{S} G_r$.

3.2 Trade-cost-function Parameters

In constructing trade costs, we employ three ingredients: (i) fast-marching-algorithm-based transportation costs between pairs of 1° grid cells along the lines of Desmet et al. (2018), using passing-through parameters from Allen and Arkolakis (2014);[6] (ii) a correspondence of these transport costs to the level of REGPAT regions by weighted averaging them within regions as explained in Appendix D.3; (iii) the consideration of discontinuities in trade costs at national borders due to tariffs and linguistic proximity. Tariffs and common language are among the most important factors which are used in parameterizing the international trade-cost function beyond mere transportation costs. We follow the customary approach to specify the trade-elasticity-scaled trade costs as a product of their scaled ad-valorem ingredients – here a transport-cost factor, a tariff factor, and a language factor. We specify the tariff factor between regions r and s as $(1 + tariff_{rs})^{-\theta}$, where $tariff_{rs}$ is the weighted applied import tariff on manufactures in 2005 (which differs between most-favored-nation partners and customs-union or free-trade-area members). To acknowledge the language factor in trade costs we follow Melitz and Toubal (2014) and use $\exp(\rho \times proxling_{rs})$, where $proxling_{rs} \in [0, 1]$ is the linguistic proximity and $\rho = 0.078$ is the corresponding parameter estimate favored in Melitz and Toubal (2014, p.357, Table 3, column 6) on their Automated Similarity Judgment Program (ASJP) measure, which we use here.

3.3 Initial Efficiency Distribution

Simulating the model requires knowledge on the spatial distribution of the initial efficiency (τ_{rt}) in the benchmark year 2005. We use the product-market clearing condition in (11) to rewrite τ_{rt} as a function of observables, replacing the R&D-worker-specific productivity shifter, which is unknown at this point, by the BGP relationship,

[6] We modify those costs by symmetrifying them (using the average for costs from r to s and s to r) and by assuming that intra-cell transport costs are (essentially) zero as is customary in quantitative Ricardian work (see Eaton and Kortum, 2002; Donaldson, 2017).

$\tau_{rt} \propto \left(\bar{L}_r h_r\right)^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}}$. Hence, we can express a scaled version of τ_{rt} as follows

$$\tau_{rt}^{(2-\gamma_2)} = \frac{\bar{L}_{rt}^{1-\iota_1} G_r w_{rt}^{1+\theta}}{\int\limits_S w_{st}\bar{L}_{st} G_s \zeta_{rs}^{-\theta}\left[\int_S \tau_{kt}^{(2-\gamma_2)} \bar{L}_{rt}^{\iota_1} \zeta_{rk}^{-\theta} w_{kt}^{-\theta} dk\right]^{-1} ds}, \tag{18}$$

where $\iota_1 \equiv \alpha - (1-\mu)\theta$. We numerically solve for $\tau_{rt}^{2-\gamma_2}$ by applying a standard contraction mapping procedure and using observed levels of population densities, $\bar{L}_{rt}$, and wages, w_{rt}, for the benchmark year 2005. Population levels are from SEDAC and wage levels from the G-Econ Project, which are both aggregated to the regional level as described in Appendices D.1 and D.2, respectively. Technical details on the derivation of (18) are presented in Appendix A.

3.4 Estimation of the Productivity Shifter for R&D workers

To estimate the R&D-worker-specific productivity shifter governing the BGP and the transition towards it, h_r, we use (15) along with the derived (scaled) initial efficiency distribution from the previous section.[7] Taking logs of the BGP relationship and multiplying both sides with the exponent $(2-\gamma_2)$ obtains

$$\log\left[\tau_{rt}^{2-\gamma_2}/\overline{\tau_{st}^{2-\gamma_2}}\right] = \frac{\theta\gamma_1(2-\gamma_2)}{(1-\gamma_2)\xi}\log\left[\bar{L}_r/\overline{\bar{L}_s}\right] + \frac{\theta\gamma_1(2-\gamma_2)}{(1-\gamma_2)\xi}\log\left[h_r/\overline{h_s}\right]. \tag{19}$$

We parameterize h_r as

$$h_r = \exp(\mathbf{D}_r\beta + |lat_r|\mathbf{D}_r\gamma + \delta|lat_r|), \tag{20}$$

where $\mathbf{D}_r$ describes a vector of binary R&D-policy indicators from Boesenberg and Egger (2016) which are measured in the same year as h_r (here, 2005). The indicators in $\mathbf{D}_r$ are

[7] Using the BGP relationship to estimate the R&D-worker specific productivity shifter implicitly assumes that all regions in the sample are growing at the steady-state rate already in 2005. Notice that also Desmet et al. (2018, pp. 927 and 929) have to assume that the data are characterized by a BGP in order to determine the relative importance of technology inertia and diffusion parameters. We have to make this assumption also in order to calibrate the productivity levels across regions in the benchmark year and to link it to the R&D-policy variables. For robustness regarding the latter, we ran the analysis only for OECD member countries, including Singapore, and found that the parameter estimates in estimating (19) are very similar when running the regression for the mentioned sub-sample relative to the full data-set (see Table 7 in the Appendix). However, we should admit that solving for τ_{rt} in Subsection 3.3 inevitably requires assuming all places to grow at the BGP rate. Otherwise, the market-clearing condition for goods for all places depends on both $\{\tau_{rt}, h_{rt}\}$ for every $\{rt\}$. Hence, the mentioned robustness analysis should be taken with a grain of salt. We also reran the analysis for all non-OECD member countries (excluding Singapore). In this subsample of 434 regions, only two out of five R&D-policy instruments are used, and the most important instrument in this sample are tax holidays. We present the table summarizing the corresponding results for OECD countries plus Singapore on the one hand (5,199 regions) and the other 434 regions as Table 7 in the Appendix.

country-specific and pertain to all regions in a country. Specifically, $\mathbf{D}_r$ includes a binary indicator variable for partial exemptions of returns on R&D investments, also known as patent boxes ($Dpatentbox_r$), R&D investment related grants from the government which act akin to subsidies ($Dgrants_r$), tax credit on R&D investments ($Dtaxcredit_r$), tax holidays for firms with R&D investment ($Dtaxholiday_r$), and any form of deductions of R&D investments from profits other than super deductions ($Ddeduc_r$). In any case, a binary indicator is set to unity, if the respective kind of provision is in place in the year 2005 and zero else. We use binary indicators for R&D instruments for one specific reason: combining these instruments into specific rates requires information about the detailed investment structure of firms in each region and time. These data are not available globally.

Additionally, we include an interaction term of each binary R&D policy indicator with the absolute value of the latitude of the region's centroid ($|lat_r|\mathbf{D}_r$) for two reasons. First, it allows us to account for differences on how productively a region can use an adopted R&D policy instrument depending on its distance to the equator. Notable contributions that have highlighted a relation between a firm's ability to adopt new technologies and its distance to the equator are Theil and Chen (1995) and Hall and Jones (1997), among others. And, second, it adds variation in the marginal effect of country-level policy instruments across regions. Clearly, the absolute value of the latitude is a better representation for between- rather than within-country variations when considering the ability to adopt new technologies. However, there is also evidence for a within-country variation as a number of economies in our sample display a clear north-south divide in economic activity, e.g., Italy, France or the US.[8] The specification which takes interactions of the national binary R&D policy environment with latitudes into account allows for general latitude-related patterns of the unobserved investment structure of firms, which affects the bite of the R&D-policy environment for R&D-cost reductions.

We refrain from explicitly modeling any budgetary effects of the considered R&D policy instruments for the following reason. The employed instruments affect the marginal tax rate on returns generated from R&D in a highly nonlinear way. However, as countries do not report specific tax revenues generated from such investments, it is not possible to validate a structural form of the associated nonlinear relationship. From this perspective,

[8]We tried to estimate (19) using a region-specific remoteness index being interacted with each binary R&D-policy indicator. The results are comparable to the ones presented here, but remoteness interactions are less statistically significant than latitude interactions.

it appears customary to resort to a reduced-form nexus between the instruments and innovation and consider treatment effects of the instruments based on this reduced-form nexus by embedding it in the structure of the general equilibrium model.[9]

As the location decision of individuals is endogenous to the productivity potential of a region, we instrument $\log(\bar{L}_r)$ in the year 2005 with a region-specific remoteness index in logs, $\log(R_r) = \log(areashare_r) + \log\left(\frac{1}{S}\sum_{s=1}^{S}\zeta_{rs}\right)$.[10] After substituting $\log(h_r)$ with $(\mathbf{D}_r\beta + |lat_r|\mathbf{D}_r\gamma + \delta|lat_r|)$ according to (20), we estimate (19) with two-stage least squares (2SLS) to obtain the parameter estimates $\{\widehat{\frac{\theta\gamma_1(2-\gamma_2)}{(1-\gamma_2)\xi}}, \widehat{\beta}, \widehat{\gamma}, \widehat{\delta}\}$ based on data for the baseline year $t = 2005$.

In Table 2, we summarize all variables which inform this procedure. The table is organized in three vertical blocks: the one at the top summarizes moments of the scaled initial efficiency, the land and population distribution as well as $\log(\bar{L}_r)$, which combines the two and the remoteness index; the one in the center summarizes the elements of $\mathbf{D}_r$ as well as $|lat_r|$ used in $|lat_r|\mathbf{D}_r$, underlying the parametrization of h_r; and the block at the bottom provides information on registered patents per unit of land in region r from REGPAT, which will be further discussed in the decomposition exercise in Section 3.6.

– Table 2 about here –

While the information about the population and land data may be interesting to some readers, we suppress a discussion here for the sake of brevity and rather focus on the R&D-policy instruments used in the parametrization of h_r. The respective indicators suggest that more than two-thirds of the regions operated under a regime with tax credits ($Dtaxcredit_r$), while other R&D-policy instruments were used much less frequently (by fewer countries or by countries with not very fine-grained regions) in 2005. For example, a grants system was applied in only about eight percent of the regions, and deductions, tax holidays, and patent boxes were used in only about two to three percent of the regions.

The parameter estimates and some other statistics based on the aforementioned procedure and data are summarized in Table 3. There, we report on marginal effects of the covariates in (19) for three specifications. The first column presents the ordinary

[9]Modelling tax revenue effects more explicitly with the R&D tax instruments at hand would require detailed information in the structure of a region's capital stock (the ratio of buildings versus machinery in that stock and its financing, etc.), see Egger and Loretz (2010). As such information is not available for the regions at hand, we resort to the parsimonious approach adopted here.

[10]Notice that trade frictions are among the few exogenous parameters in the model. Hence, they are natural candidate instruments for endogenous variables in the model. Any (highly nonlinear) reduced form of the model would involve trade costs as a determinant of every one of the endogenous variables in the model.

least squares (OLS) results, while columns (2) and (3) show the results from the 2SLS regression, which takes the potential endogeneity of an individual location decision into account. The second and third columns differ, as the specification in column (3) controls for continent fixed effects, while the one in column (2) does not.[11] Apart from marginal effects of $\log(\bar{L}_r)$ as well as individual elements in $\mathbf{D}_r$ and the absolute value of the latitude we report the overall model fit through the correlation of the data with the model prediction as well as the number of observations (regions) used for estimation.[12] As key variables of interest are measured at the country level, all standard errors and test statistics are robust to clustering at the country level.

– Table 3 about here –

We document in the upper block of Table 3 that the proposed instrument is highly relevant. The OLS and second-stage 2SLS results suggest that more densely populated regions (i.e., ones with higher values of $\log(\bar{L}_r)$) have higher efficiency values, as predicted by the model. Comparing the parameter on $\log(\bar{L}_r)$ in column (1) to those in column (2) and (3), it becomes evident that accounting for endogeneity not only reduces the importance of population density on efficiency levels but also reveals significant effects of country-specific R&D policy instruments on regional efficiency. The latter is concealed by the bias of the OLS estimates. In particular, tax holidays ($Dtaxholidays_r$) and grants ($Dgrants_r$) tend to raise efficiency according to columns (2) and (3), while patent boxes ($Dpatentbox_r$; a back-end incentive which primarily promotes the ownership but not the invention of patents) reduce efficiency levels.[13] Also regular deductions ($Ddeduc_r$) of R&D investments from profits display a positive effect on efficiency levels. The explanatory power of the model is relatively high, as can be seen from the overall fit measured by the correlation coefficient between the data and the model prediction as reported at the bottom of the table. Overall, these results document that, as postulated and hypothesized, a favorable so-called front-end R&D-policy environment indeed appears to have cost-reducing effects on innovation and productivity – which is the very intention of the associated policies – and, hence, boosts productivity as intended in a way which is measurable at the regional level.

[11] Continent fixed effects inter alia capture the heterogeneity in the granularity of regions as classified in REGPAT. Moreover, they capture a heterogeneity at the macro-regional level in terms of the desirability of patenting among innovative firms.

[12] Clearly, as the elements in $\mathbf{D}_r$ are binary, what we report is the average effect of an indicatoir being unity versus zero for the considered R&D tax-policy instruments.

[13] Patent box is the only policy instrument in our analysis for which the invention does not need to have taken place at the same location as where the tax incentive would be enjoyed. This is why the point estimate is likely to differ in sign compared to other instruments.

In what follows, we will use the specification in column (3) as the preferred model, since its explanatory power is relatively highest among the two 2SLS models, and the parameters on R&D policy instruments are all statistically relevant predictors of regional efficiency levels. Given the parameter estimates, we obtain an estimate of h_r for each region r in 2005 and the transition towards as well as the BGP as

$$\widehat{h}_r = \exp(\mathbf{D}_r\widehat{\beta} + |lat_r|\mathbf{D}_r\widehat{\gamma} + \widehat{\delta}|lat_r|). \tag{21}$$

The R&D-policy instruments included in $\mathbf{D}_r$ jointly contribute to a sizable variation of $\log(\widehat{h}_r)$ in the data. We illustrate the latter by way of a kernel density plot in Figure 2.

– Figure 2 about here –

3.5 Technology and Efficiency-evolution Parameters

Table 1 summarizes the assumed values of the technology parameters $\{\alpha, \theta, \mu\}$ and the efficiency-evolution parameters $\{\xi, \nu\}$ which we take from others' work. Here, we focus on the two remaining parameters $\{\gamma_1, \gamma_2\}$ which are elemental but for which existing estimates are not available given the adopted model structure. Specifically, the BGP implies that welfare grows according to (17). Taking logs and expressing (17) for a finite number of regions obtains

$$\begin{aligned}\log(u_{rt+1}) - \log(u_{rt}) =& \log(y_{rt+1}) - \log(y_{rt}) \\ =& \frac{(1-\gamma_2)}{\theta}\log(\frac{1}{S}) + \frac{\gamma_1}{\xi}\log(\Psi) + \frac{1-\gamma_2}{\theta}\log(\sum_{s=1}^{S}\left(\bar{L}_s h_s\right)^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}}),\end{aligned} \tag{22}$$

where $\Psi \equiv \frac{\gamma_1/\nu}{\gamma_1+\mu\xi}$ and $S = 5,633$. Note that equation (22) depends on both population density ($\bar{L}_{rt}$) and on real-income (y_{rt+1}, y_{rt}). Either type of data is available at the $1^\circ \times 1^\circ$ resolution from the G-Econ 4.0 Research Project at Yale University. However, as the estimation is informed by parameter values established in the estimation of Section 3.4, we employ the population data from SEDAC for consistency.[14]

For identification of the parameters it is useful to see that the left-hand side of (22) is indexed by t, whereas none of the parameters and variables on the right-hand side

[14] Whereas SEDAC provides gridded population data with an output solution of 30 arc-seconds (approx. 1 km at the equator), the G-Econ project provides the same data on an aggregated $1^\circ \times 1^\circ$ resolution. We use population data from SEDAC directly to avoid measurement error from aggregation. However, we reran the analysis with population data from the G-Econ Project as a robustness check, and the parameter estimates do not change significantly, when doing so.

is. Moreover, γ_1 can be expressed as a function of γ_2 (and vice versa), and all the other parameters are known at this point. Hence, for a single year, γ_2 could be exactly solved for. For identification we pool the mentioned data for $t \in \{1990, 1995, 2000\}$ and $t+5 \in \{1995, 2000, 2005\}$ and approximate the log difference between years $t+1$ and t by the average annual change within any five-year interval. We use the estimated parameter $\widehat{\frac{\theta\gamma_1(2-\gamma_2)}{(1-\gamma_2)\xi}}$ of Section 3.4 and rearrange all parameters dependent on γ_1 in (22) to express them as a function of γ_2. Noting that $\gamma_1, \gamma_2 \in (0,1)$, we can search for the optimal value of γ_2 by doing a grid search on the unit interval with an objective function that minimizes the sum of squared residuals between the left-hand side and the right-hand side of (22) for the mentioned three year tuples $\{t, t+1\}$ together. Adopting this procedure obtains the grid-search estimates $\hat{\gamma}_2 = 0.979$ and the implied $\hat{\gamma}_1 = 0.234$ as listed in Table 1.

3.6 Patented vs. Non-patented Innovations

Patenting is often used as a measure of innovation (see e.g., Griliches, 1990; and Nagaoka et al., 2010). However, not all innovations are patented. In fact, non-patented innovations appear much more common than patented ones on average (see more details on this in the discussion below). The model structure allows us to obtain a measure of the overall innovation level for each region, ϕ_{rt}, and data on patenting permit attributing it to patented innovations versus (residual) non-patented ones.

For this, we use data on patent registrations, assuming a Cobb-Douglas relationship: $\phi_{rt} = (\phi_{rt}^{Patent})^{\alpha_{1r}} (\phi_{rt}^{Rest})^{1-\alpha_{1r}}$. Taking logs we obtain

$$\log(\phi_{rt}) = \alpha_{1r} \log(\phi_{rt}^{Patent}) + (1 - \alpha_{1r}) \log(\phi_{rt}^{Rest}), \tag{23}$$

where $\phi_{rt} \propto \tau_{rt}^{\frac{1-\gamma_2}{\theta\gamma_1}}$ according to the BGP relationship in (15), ϕ_{rt}^{Patent} is a measure of patent registrations from REGPAT, ϕ_{rt}^{Rest} is a measure of non-patented (unobserved) innovations, and $\alpha_{1r} \in (0,1)$ is a region-specific Cobb-Douglas weight. Table 2 reports figures on patent registrations at the bottom, which are expressed in normalized units of land, G_r. The two lines at the top of the respective block pertain to a regional denomination of patents according to the residence of inventors (*inv*), whereas the two lines at the bottom of the respective block pertain to a regional denomination of patents according to the residence of applicants (*app*). For each concept, we report the average normalized patent registration counts for 2005 as well as the patent stock counts from 1995 to 2005. The respective figures suggest that inventions are more dispersed than applications (i.e., applications are more concentrated). This pattern shows in higher first and second moments of patent applications as well as in a higher frequency of zeros across regions in the applications data than the inventions data, which is not obvious from the table.

It is useful to introduce a parametrization of $\alpha_{1r}\log(\phi_{rt}^{Patent})$ in order to gauge the relative importance of observable patented innovations and unobservable non-patented ones. In particular, we parameterize $\alpha_{1r}\log(\phi_{rt}^{Patent})$ as a weighted average of the log of the normalized patent stock in a region ($\log(Patentstock_{rt})$) and an interaction term thereof with the log normalized land mass ($\log(G_r)$). The reason for an inclusion of the latter is that REGPAT regions tend to be larger in areas of the globe where patenting is relatively rare, and the mentioned interaction term captures this pattern. Then, using inventor-based patent data, the suggested parametrization reads

$$\alpha_{1r}\log(\phi_{rt}^{Patent(inv)}) \equiv \alpha_2\log(Patentstock_{rt}^{(inv)}) + \alpha_{3r}\log(Patentstock_{rt}^{(inv)}) \times \log(G_r). \tag{24}$$

Based on this, we can replace $\alpha_{1r}\log(\phi_{rt}^{Patent})$ in (23) by the expression on the right-hand side of (24) and obtain $(1-\alpha_{1r})\phi_{rt}^{Rest}$ as a residual, in order to yield region-specific shares for patent-related innovations as $\hat{\alpha}_{1r} = \hat{\alpha}_2 + \hat{\alpha}_{3r}\log(G_r)$.[15] According to the data and estimates, (inventor-based) patent-related innovation stocks explain about 42 percent of the variation in $\log(\phi_{rt})$ (in terms of the R^2), and their Cobb-Douglas share $\hat{\alpha}_{1r}$ ranges from 0.005 to 0.014, with an average of 0.009 and a standard deviation of 0.0008. Hence, the cost share of patented innovations in the generation of all innovations is with about one percent on average relatively small, and it does not vary too starkly in the data.

One may assume that this low cost share is driven by regions in which the patent law is such that the patent stock would over-represent inventive activities and, hence, bias our estimates. Nagaoka et al. (2010) mention this problem by reference to the Japanese patent law. Including a binary indicator for Japan in the analysis, however, does not reveal any significant effect, which we take as evidence that the institutional differences do not seem to play a significant role for our results (when conditioning on the included factors determining endogenous innovation, ϕ_{rt}).

In fact, the notion of a relatively low cost share of patented innovations in all innovations squares with earlier evidence. For instance, Bloom and Van Reenen (2000) find a relatively low elasticity of total-factor productivity with respect to patents of about 0.03. Moreover, the evidence in Danguy et al. (2009) suggests that an increase in R&D

[15] The results are similar when estimating (23) with any other measure of patent registrations that is listed in Table 2.

expenditures raises patents at an elasticity of only 0.12. Moser (2013) documents that, using historical exhibition data, the share of inventors who chose to patent their innovations varied between 5 and 20 percent across industries. Sierotowicz (2015) finds that the average number of patents per million euros of R&D expenditures in leading European Union countries varied between 0.03 in Spain and 0.26 in Germany. Nagaoka et al. (2010) summarize the reasons for why innovations may not be patented. Clearly, in the proposed model a micro foundation of the choice of patenting is absent, and firms are characterized as to rely on both patented and non-patented innovations for technological reasons.[16]

In Figure 3, we display the relationship between calibrated log overall innovative productivity in the benchmark year 2005 ($\log(\phi)_r$) and the estimated region-specific importance weight of patented innovations therein ($\widehat{\alpha}_{1r}$). Interestingly, this relationship is negative, though weakly so. This means that in larger regions -– where patents are, on average, relatively rare and $\widehat{\alpha}_{1r}$ is relatively high — the overall productivity is relatively lower than on average, in spite of the higher weight of the (fewer) patents. Hence, larger regions enjoy on average a lesser degree of non-patented innovations. However, we should acknowledge that the R^2 underlying the linear relationship in Figure 3 is as low as 0.04.

It is worth mentioning that (23) postulates a relationship between $\log(\phi_{rt})$ and $\log(\phi_{rt}^{Patent})$ which does not vary too starkly around 0.009. Using the estimates $\{\widehat{\alpha}_2, \widehat{\alpha}_{3r}\}$ and data on $\log(Patentstock_{rt}^{(inv)})$ and, alternatively, $\log(Patentstock_{rt}^{(app)})$ as well as $\log(G_r)$, we can plot $\log(\phi_{rt})$ against the estimates $\log(\widehat{\phi}_{rt}^{Patent})$. Figure 4 does so by way of scatter plots using binned data, where we group regions into 20 equally-sized bins and compute averages within bins for inventor-based (left panel) and applicant-based patents (right panel). The result is a non-parametric visualization of the conditional expectation function, and the figure suggests that the data support relatively well a low variability of the log-linear relationship between $\log(\phi_{rt})$ and $\log(\widehat{\phi}_{rt}^{Patent})$, as expected.

3.7 Estimating Amenity-function Parameters

Before we can simulate the model and do counterfactual analyses, we need to estimate the amenity-function parameters. We postulate and expect overall amenities to decrease with population density as described in equation (7). Taking logs of $a_{rt} = \bar{a}_{rt}\bar{L}_{rt}^{-\lambda}$ obtains

$$\log(a_{rt}) = -\lambda \log(\bar{L}_{rt}) + const. + \varepsilon_{rt}^{a}, \tag{25}$$

[16] This is consistent with a notion of patented innovations to be technologically different from non-patented ones.

where $\log(\bar{a}_{rt})$ is specified as a common constant (*const.*, which measures the average of $\log(\bar{a}_{rt})$ across all regions) plus a deviation from it (ε^a_{rt}, i.e., a disturbance term). Clearly, as population density $\bar{L}_{rt}$ depends on people's location choice in the model which itself depends on a_{rt}, it should be treated as endogenous in estimating the region-specific exogenous amenity parameter $\bar{a}_{rt}$ and the congestion parameter λ based on (25). Therefore, we estimate (25) by two-stage least squares (2SLS) for the baseline year 2005, instrumenting $\bar{L}_{rt}$ with a region-specific area-weighted remoteness index, $R_r = weight^{area}_r \left(\frac{1}{S}\sum_{s=1}^{S}\zeta_{rs}\right)$, which does not depend on individual location decisions (see Footnote 10 for a reasoning regarding this instrumentation strategy). In order to measure $\bar{L}_{rt}$ we use gridded population data from the Socioeconomic Data and Application Center (SEDAC) which we aggregate to the required (non-gridded) regional level. Technical details on this aggregation are described in Appendix D.1.

To construct the dependent variable based on a_{rt} in (25), we use the structure of the model, substitute the indirect utility (8) into (10) and solve for a_{rt} as in equation (27) in Appendix A.

– Table 4 about here –

Table 4 reports the estimation results from estimating (25), with the congestion parameter estimated at a value of $\widehat{\lambda} = 0.596$. Furthermore, the table reports first-order and second-order moments of $\widehat{\bar{a}}_{rt}$. As described above, the region-specific exogenous amenity attribute is defined as $\widehat{\bar{a}}_{rt} \equiv exp(\widehat{const.} + \widehat{\varepsilon}^a_{rt})$. In the general-equilibrium analysis, $\widehat{\bar{a}}_{rt}$ is kept constant at its level of the year $t = 2005$ for all subsequent time periods.

4 Counterfactual Analysis

In the counterfactual equilibrium analysis we focus on the effects on three key variables – place-specific employment, productivity, and welfare – in a scenario where investment incentives towards innovation – except for patent boxes – are abandoned. Effectively, this means that in the counterfactual analysis the R&D-worker-specific productivity shifter equals $h^c_r = \exp(\hat{\beta}_{PB} DPatentbox_r + \hat{\gamma}_{PB}|lat_r|DPatentbox_r + \hat{\delta}|lat_r|)$, $\forall r \in S$. We split the analysis in three parts. First, we investigate how economic outcomes react in response to abandoning incentives towards innovation and distinguish between regions in policy-adopting vs. policy-non-adopting countries. Table 5 lists all policy-adopting countries for each instrument in the year 2005, according to Boesenberg and Egger (2016). The second part of the analysis concentrates on the role of the treatment-size, exogenous amenities, and remoteness for welfare responses. Lastly, we investigate the role of the patented innovation weight for innovation responses.

4.1 Economic Outcomes and R&D-policy Instruments

In Figure 5 we display the variation in long-run ($T = 100$) counterfactual changes in important economic outcomes across all regions in the data. These three outcomes are log population levels ($\log(\bar{L}_{rt}G_r)$), log (overall) productivity levels of the Fréchet location parameter ($\theta^{-1}\log(\tau_{rt}\bar{L}_{rt}^{\alpha})$), and log welfare levels of the representative household as expressed in real GDP ($\log(y_{rt}) = \log(u_{rt}/a_{rt})$).

– Figure 5 about here –

The three panels in the figure suggest that all three economic aggregates are reduced on average when abolishing the considered R&D-policy instruments. However, a non-trivial mass of regions gains population – mainly due to a loss in competition for workers from otherwise less attractive regions that could compete for mobile workers through the use of R&D-policy instruments on the benchmark BGP. The ($T = 100$) long-run changes are quite substantial: some regions gain about eight percent in population while others lose more than 30 percent due to the hypothetical policy change in the long run. Note that the distribution of log changes in population levels does not integrate to one. Given the logarithmic transformation of the displayed population change, the hypothetical abolishment of R&D-policy indicators implies that workers move to high-density places (large agglomerations) away from (previously competitive) low-density places. Hence, country-wide R&D tax incentives increase the competitiveness of less attractive low-density places in comparison to similar low-density places abroad, where these policy indicators are not adopted. Accordingly, these nationally adopted instruments indirectly work as place-based policies in an international context for two reasons. First, they raise the attractiveness of low-density (peripheral and low-amenity) places relative to high-density places at a national level, where the policy is adopted, and, second, they raise the attractiveness of low-density places in policy-adopting countries relative to such places in non-adopting economies.

The effects of abandoning R&D-policy instruments on overall productivity are detrimental throughout and even larger than on population changes; also the welfare changes are negative throughout and almost as large as the productivity changes. The fact that welfare and productivity changes are negative throughout the distribution implies that also countries and their regions which do not use such instruments benefit from their use elsewhere due to technology spillovers.

– Table 6 about here –

Table 6 presents moments of the real GDP growth rate in the short- (T=10), medium- (T=50), and long-run (T=100). The table shows that regions converge towards a model-induced balanced growth path, as the dispersion of growth rates declines with time. The lower panel of the table presents counterfactual-minus-benchmark growth rate differences in percentage points. The corresponding panel suggests that one-tenth of the average long-run real GDP growth can be attributed to the R&D policy instruments alone.

4.2 The Role of Treatment Size, Remoteness, and Amenities for Welfare Responses

In Figure 6, we focus on the welfare changes as in the third panel of Figure 5 and plot them against the size of the direct treatment changes – i.e., the change in h_r induced by abolishing R&D-policy instruments – and differentiate between all possible combinations of R&D-policy instruments that were in place in 2005. That figure suggests that the relationship between the treatment change and the associated change in utility is almost linear. Hence, the direct (or partial) effect entails a strong signal for the long-run response. There are indirect effects, which are most obvious for the non-adopting regions in 2005 (about one-percent of the regions displayed in blue circles in the upper-right corner of the figure). The indirect, technology-spillover plus general-equilibrium effects on the other regions materialize inter alia as deviations of the data points from the latent linear relationship in Figure 6.

– Figures 6 and 7 about here –

In Figure 7, we shed further light on potentially important mediators of the general-equilibrium treatment effect on welfare changes. While Figure 6 alluded to the nexus between the treatment signal and the welfare response, we focus on the role of exogenous amenities in 2005 ($\log(\bar{a}_{rt})$; in the left panel) and a region's remoteness ($\log(R_r)$; in the right panel).

In the two panels of Figure 7, we use different color to plot the relationships for different continents. Interestingly, the left panel reveals a positive relationship between amenities and the welfare change for regions in North America, Europe and Oceania (including Australia). Hence, a better endowment with good amenities provides for a better insurance against adverse effects from the global abolishment of R&D-stimulating policies. That relationship is still positive but weaker for regions in South America, while it is negative for regions in Africa and Asia. The right panel in Figure 7 reveals a negative relationship between remoteness and the welfare change (i.e., more remote regions lose less on welfare from the global abolishment of R&D-promoting policies) for regions in North America, Oceania (including Australia), and also South America, while for regions on other continents this relationship tends to be positive.

In summary, greater (exogenous) amenities and a higher degree of centrality of a place in the transport network provides for a better quasi-insurance against adverse effects from weak R&D-policy institutions, on average. Moreover, an R&D policy at the national level with a homogeneous direct effect of treatment of all innovations across the places there has indirect place-based effects which are *ceteris paribus* stronger for more peripheral places with less attractive amenities.

4.3 The Role of the Patented Innovation Weight for Innovation Responses

In Subsection 3.6 we discussed the relative importance of patented and non-patented innovations for overall innovation in the data. In the model, the weight of patented in all innovations is α_{1r}; see equation (23). The respective parameter is indexed by region, because the importance of patenting depends on the land mass of a region, according to equation (24). The latter was introduced to capture the fact that the delineation of regional borders in the REGPAT database was done according (with region size being inversely related) to the frequency of patenting. However, the overall role of innovation in a region is not a simple function of land mass only but also depends on other fundamentals (such as amenities, market access, etc.).

– Figure 8 about here –

In this subsection, we shed light on the nexus between the effects of general R&D incentives as studied here on outcome depending on the relative importance of patenting in a region as captured by α_{1r}. In Figure 8 we plot the log change in overall (patented plus non-patented) innovation as induced by the counterfactual change in R&D tax instruments against α_{1r}. There could be a pattern in this relationship, if the land mass of the regions were related to the latitude (as the effectiveness of R&D incentives may vary with the latitude of a region) or to the actual use of instruments (e.g., through the more intensive use of the instruments by countries where patenting is common and, hence, the average land mass of a region is small). The figure suggests that the relationship between the counterfactual-to-benchmark change in $\log(\phi_r)$ and α_{1r} is weak: recall that the R^2 of a linear regression of $\log(\phi_r)$ on α_{1r} was 0.04, and the one of a linear regression of $\log(\phi_r^c) - \log(\phi_r)$ on α_{1r} is 0.02 when considering the change after $T = 100$ periods. However, the slope of the regression line for the change is positive. Hence, larger regions (i.e., ones with a lower patent count on average in the outset which are also the ones where the overall innovation level $\log(\phi_r)$ was low and α_{1r} was high) are the ones which gain more in overall innovation than on average. It turns out that this relationship is mainly driven by changes in Asia and not on other continents.

5 Conclusion

This paper outlines a multi-regional model of innovation, production, trade, and factor mobility with a dynamic technology diffusion process. The key parameters of the model are estimated and the model is otherwise calibrated to 5,633 REGPAT regions. One of the main goals of the paper is to provide a quantitative account of the consequences and the value of innovation for regional and national economies as well as the global economy. Since nationally implemented policy instruments towards firm-level R&D are particularly important, we put emphasis on quantifying the role of such incentives. We document that, in spite of their national inception, these instruments affect regions between but also within adopting and non-adopting countries heterogeneously. The degree of heterogeneity depends on the extent of the treatment – how many and which instruments are used and how productively (in terms of its absorptive capacity) a region can use them. Moreover, the degree of heterogeneity depends on other fundamentals such as a region's integration in the national and international transport network as well as its attractiveness for the location of mobile labor in terms of the available amenities.

One important insight is that the use of policy instruments which are designed to stimulate private R&D are globally beneficial in terms of productivity and welfare. In other words, also countries and their regions who do not use such incentives benefit from their use abroad due to technology spillovers. Also, the long-run relocation effects from a hypothetical abolishment of R&D investment incentives are substantial and lead to a re-shifting of the population towards high-density areas. This is mainly due to a loss in competition for workers from otherwise less attractive regions, which could gain in international competitiveness for mobile factors through the use of R&D policy instruments.

In line with the previous result, the quantitative analysis suggests that particularly low-amenity, peripheral places – and, on average, ones where the patenting of innovations is less common than elsewhere – benefit relatively more strongly from R&D investment incentives than others. The latter implies that these instruments work as place-based policies. This result is especially true for regions in North America and Oceania, whereas the effect is less predominant in Europe or Asia. Overall, R&D-policy instruments affect endogenous innovations primarily through non-patented innovations, as the estimated range of weights of patented innovations in all innovations is relatively small around the globe.

References

Ahlfeldt, G., Redding, S., Sturm, D. and Wolf, N. (2015): "The Economics of Density: Evidence From the Berlin Wall", *Econometrica*, vol. 83, 2127-2189.

Allen, T. and Arkolakis, C. (2014): "Trade and the Topography of the Spatial Economy", *The Quarterly Journal of Economics*, 129(3), 1085-1140.

Allen, T. and Donaldson, D. (2018): "Geography and Path Dependence", unpublished manuscript.

Aralica, Z. and Botric, V. (2013): "Evaluation of Research and Development Tax Incentive Schemes in Croatia", *Ekonomska istrazivanja*, 26(3), 63-80.

Atkeson, K. and Burstein, A. T. (2018): "Aggregate Implications of Innovation Policy", forthcoming *Journal of Political Economy.*

Baumann, M., Knoll, B. and Riedel, N. (2014): "The global effects of R&D tax incentives: evidence from micro-data", Unpublished manuscript.

Bernard, A. B., Eaton, J., Jensen, J.B., and Kortum, S. (2003): "Plants and Productivity in International Trade", *American Economic Review*, 93(4), 1268-1290.

Bloom, N., Schankerman, M. and Van Reenen, J. (2013): "Identifying technology spillovers and product market rivalry", *Econometrica*, 81(4), 1347-1393.

Bloom, N. and Van Reenen, J. (2000): "Patents, productivity and market value: evidence from a panel of UK firms," IFS Working Papers W00/21, Institute for Fiscal Studies.

Boesenberg, S. and Egger, P. H. (2016): "R&D tax incentives and the emergence and trade of ideas", *Economic Policy*, 89, 32.

Caiumi, A. (2011), "The Evaluation of the Effectiveness of Tax Expenditures - A Novel Approach: An Application to the Regional Tax Incentives for Business Investments in Italy", OECD Taxation Working Papers, No. 5, OECD Publishing, Paris.

Caliendo, L., Dvorkin, M. and Parro, F. (2015): "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock", NBER Working Paper No. 21149.

Caliendo, L., Parro, F., Rossi-Hansberg, E. and Sarte, P.-D. (2018): "The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy", forthcoming *Review of Economic Studies.*

Cappelen A., Raknerud, A. and Rybalka, M. (2007): "Effekter av SkatteFUNN pa foretakenes produktivitet og Ionnsomhet", *Okonomiske analyser*, 2007(6), 1-54.

Carlino, G., Chatterjee, S., and Hunt R. (2007): "Urban Density and the Rate of Invention", *Journal of Urban Economics*, 61, 389-419.

Cornet, M. (2001): "De maatschappelijke kosten en baten van technologiesubsidies zoals de WBSO", CPB document (no. 008).

Czarnitzki D. and Lopes-Bento, C. (2014): "Innovation Subsidies: Does the Funding Source Matter for Innovation Intensity and Performance? Empirical Evidence from Germany," *Industry and Innovation, Taylor & Francis Journals*, 21(5), 380-409.

Danguy, J., de Rassenfosse, G. and van Pottelsberghe de la Potterie, B. (2009): "The R&D-patent relationship: An industry perspective", *EIB Papers*, ISSN 0257-7755, European Investment Bank (EIB), Luxembourg, Vol. 14, Iss. 1, pp. 170-195.

De Jong, J. P. J. and Verhoeven, W. H. J. (2007): "Evaluatie WBSO 2001-2005: effecten, doelgroepbereik en uitvoering", Ministerie van Economische Zaken, Netherlands.

Desmet, K., Nagy, D. K. and Rossi-Hansberg, E. (2018): "The Geography of Development", *Journal of Political Economy* 126(3), 903–983.

Desmet, K. and Rappaport, J. (2015): "The Settlement of the United States, 1800 to 2000: The Long Transition Towards Gibrat's Law", *Journal of Urban Economics*, 98, 50-68 .

Desmet, K. and Rossi-Hansberg, E. (2015): "On the Spatial Economic Impact of Global Warming", *Journal of Urban Economics*, 88, 16-37.

Donaldson, D. (2018): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure", *American Economic Review*, 108(4-5), 899-934.

Donaldson, D. and Hornbeck, R. (2016): "Railroads and American Economic Growth: a Market Access Approach", *Quarterly Journal of Economics*, 131(2), 799-858.

Eaton, J. and Kortum, S. (2002): "Technology, Geography, and Trade", *Econometrica*, 70(5), 1741-1779.

Egger, P.H. and Loretz, S. (2010): "Homogeneous Profit Tax Effects for Heterogeneous Firms?," *The World Economy*, Wiley Blackwell, 33(8), 1023-1041.

Egger, P.H. and Pfaffermayr, M. (2006): "Spatial Convergence", *Papers in Regional Science*, 85(2), 199-215.

Ernst, C. and Spengel, C. (2011): "Taxation, R&D Tax Incentives and Patent Application in Europe", ZEW Discussion Paper No. 11-024, Mannheim.

Greenwood, J., Hercowitz, Z., and Krusell, P. (1997): "Long-Run Implications of Investment-Specific Technological Change", *American Economic Review*, 87(3), 342-362.

Griliches, Z. (1990): "Patent Statistics as Economic Indicators: A Survey", *Journal of Economic Literature*, 28(4), 1661-1707.

Hall, R. E. and Jones, C. I. (1997): "Levels of Economic Activity Across Countries.", *American Economic Review*, 87 (2): 173-177.

Hallépée S. and Garcia, A. H. (2012): "Évaluation du dispositif JEI", DGCIS/ Ministère du Redressement Productif, Paris, France.

Lokshin, B. and Mohnen, P. (2012): "How effective are level-based R&D tax credits? Evidence from the Netherlands", *Applied Economics*, 44(12), 1527-1538.

Melitz, J. and Toubal, F. (2014): "Native Language, Spoken Language, Translation and Trade", *Journal of International Economics*, 93(2), 351-363.

Monte, F., Redding, S. and Rossi-Hansberg, E. (2018): "Commuting, Migration and Local Employment Elasticities", American Economic Review, 108(12), 3855-3890.

Moser, P. (2013): "Patents and Innovation: Evidence from Economic History", *Journal of Economic Perspectives*, 27(1), 23-44.

Nagaoka, S., Motohashi, K. and Goto, A. (2010): "Patent Statistics as an Innovation Indicator", Chapter 25 in *Handbook of the Economics of Innovation*, vol. 2, 1083-1127, Elsevier.

Nagy, D. K. (2017): "City location and economic development", unpublished manuscript.

Parsons, M. and Phillips, N. (2007): "An evaluation of the federal tax credit for scientific research and experimental development", Department of Finance, Economic and Fiscal Policy Branch.

Redding, S. and Rossi-Hansberg, E. (2017): "Quantitative Spatial Economics", *Annual Review of Economics*, vol. 9, 21-58.

Simonovska, I. and Waugh, M.E. (2014): "Trade Models, Trade Elasticities, and the Gains from Trade", Unpublished manuscript.

Sierotowicz, T. (2015): “Patent activity as an effect of the research and development of the business enterprise sectors in the countries of the European Union”, *Journal of International Studies*, 8(2), 101-113.

Theil, H. and Chen, D. (1995): "The Equatorial Grand Canyon.", *De Economist*, 143(3): 317-327.

Westmore, B. (2013): "R&D patenting and growth", *OECD Economics Department Working Papers 1047*, OECD, Paris.

Zabreyko, P., Koshelev, A., Krasnoselskii, M., Mikhlin, S., Rakovshchik, L., and Stetsenko, V. (1975): "Integral Equations: A Reference Text", Noordho International Publishing. Leyden.

Parameter common to all regions		
1. Preferences		
$\sigma = 4$	Elasticity of substitution.	Bernard et al. (2003)
$\lambda = 0.596$	Relation between amenities and population.	Own estimation, Section 3.7
$\Omega = 0.5$	Elasticity of migration flows w.r.t. income.	Monte et al. (2018)
2. Technology		
$\alpha = 0.06$	Elasticity of productivity w.r.t. pop. density.	Carlino, Chatterjee and Hunt (2007)
$\theta = 6.5$	Trade elasticity and dispersion of productivity.	Eaton and Kortum (2002), Simonovska and Waugh (2014)
$\mu = 0.8$	Labor share in production (non-land share).	Greenwood et al. (1997); Desmet and Rappaport (2015)
$\gamma_1 = 0.234$	Elasticity of tomorrow's productivity w.r.t. today's innovation.	Own estimation, Section 3.5
3. Evolution of Productivity		
$\gamma_2 = 0.979$	Elasticity of tomorrow's productivity w.r.t. today's productivity.	Own estimation, Section 3.5
$\xi = 125$	Elasticity of innovation costs w.r.t. innovation.	Desmet and Rossi-Hansberg (2015)
$\nu = 0.15$	Intercept parameter in innovation cost function.	Desmet et al. (2018)
Region-specific parameter		
1. Land Endowments		
G_r	Extract land mass for each region. (G_r is normalized by $\frac{1}{S}\sum_{r=1}^{S} G_r$)	Arc GIS Software
2. Initial Efficiency in 2005		
τ_{rt}	Initial efficiency distribution.	Own estimation, Section 3.3
2. Amenities in 2005		
a_{rt}	Initial amenity distribution.	Own estimation, Section 3.7
$\bar{a}_{rt}$	Exogenous amenity attribute.	Own estimation, Section 3.7
3. Productivity-shifter for R&D workers in 2005		
h_r	Estimation using binary R&D policy indicators $h_r = \exp(\mathbf{D}_r\widehat{\beta} + \lvert lat_r\rvert \mathbf{D}_r\widehat{\gamma} + \hat{\delta}\lvert lat_r\rvert))$	Own estimation, Section 3.4.
4. Transport Costs		
ζ_{rs}	Based on Allen and Arkolakis (2014) and Fast Marching Algorithm.	
5. Other Trade Costs		
$tariffs_{rs}$	Weighted applied import tariffs for manufactures	World Development Indicator (WDI)

Table 1: Calibration Overview

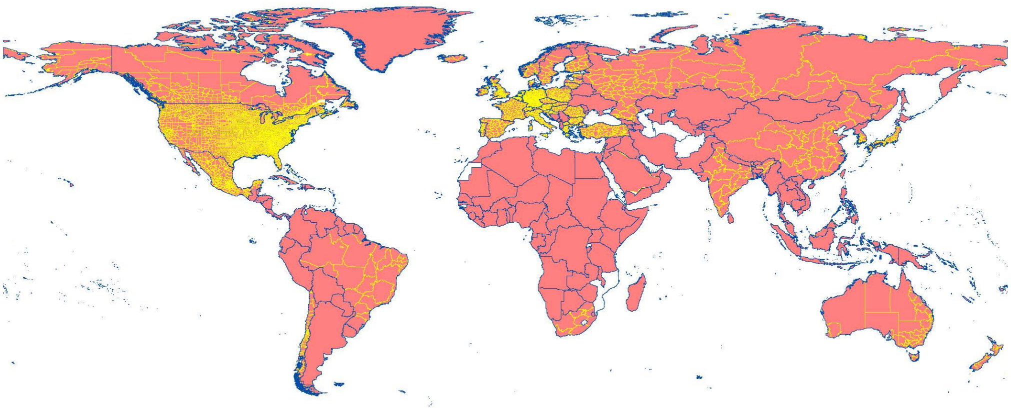

Figure 1: REGPAT Regions

Variable	Mean	Std. Dev.	Min.	Max.
Scaled initial efficiency ($\tau_{rt}^{2-\gamma_2}$)	11.487	413.097	5.43e-09	30,142
Population ($\bar{L}_{rt}G_r$)	1,108,491	7,637,692	5	2.17e+08
Normalized land (G_r)	1	11.83	1.9e-04	624.27
$\log(\bar{L}_{rt})$	9.89	2.01	-0.428	16.44
Remoteness (R_r)	0.171	1.668	0	69.754
R&D-policy indicators				
Dtaxcredit$_r$	0.715	0.452	0	1
Dtaxholiday$_r$	0.023	0.151	0	1
Dgrants$_r$	0.081	0.273	0	1
Dpatentbox$_r$	0.022	0.147	0	1
Ddeduc$_r$	0.029	0.169	0	1
Absolute latitude ($\|lat_r\|$)	40.205	9.583	0.2	74.728
Patents per norm. unit of land				
$Patents_{rt}^{(inv)}$ 2005	1,278.1	8,648.6	0	297,026.4
$Patentstock_{rt}^{(inv)}$ 1995-2005	10,366.8	67,655.1	0	2,474,476.5
$Patents_{rt}^{(app)}$ 2005	1,749.2	20,186.2	0	832,164.6
$Patentstock_{rt}^{(app)}$ 1995-2005	9,330.5	105,772.2	0	4,488,536.5

Notes: $Patents_{rt}^{(inv)}$ and $Patentstock_{rt}^{(inv)}$ refer to a regional denomination of patents in 2005 and patent stocks from 1995-2005, respectively, according to the residence of inventors (inv). $Patents_{rt}^{(app)}$ and $Patentstock_{rt}^{(app)}$ refer to a regional denomination of patents in 2005 and patent stocks from 1995-2005, respectively, according to the residence of applicants (app).

Table 2: Summary Statistics (2005)

$\log(\tau_r^{2-\gamma_2})$	(1) OLS	(2) 2SLS	(3) 2SLS
First Stage		$\log(\bar{L}_r)$	$\log(\bar{L}_r)$
$\log(R_r)$		-0.576***	-0.688***
		(0.102)	(0.074)
Second Stage			
$\widehat{log(\bar{L}_r)}$	1.154***	0.620***	0.593***
	(0.104)	(0.096)	(0.087)
Dtaxcredit$_r$	0.216	-0.464	0.915**
	(0.342)	(0.313)	(0.368)
Dtaxholiday$_r$	0.873	1.931***	1.299**
	(0.738)	(0.598)	(0.628)
Dgrants$_r$	0.602	1.552**	1.838***
	(0.645)	(0.614)	(0.525)
Dpatentbox$_r$	-0.301	-0.572	-1.679**
	(0.715)	(0.569)	(0.650)
Ddeduc$_r$	0.763	1.300***	0.715**
	(0.502)	(0.352)	(0.410)
$\|lat_r\|$	0.076***	0.039***	0.017*
	(0.008)	(0.011)	(0.009)
continent FE	NO	NO	YES
# obs	5,633	5,633	5,633
Corr. coeff. $\{\log(\tau_r^{2-\gamma_2}); \widehat{\log(\tau_r^{2-\gamma_2})}\}$	0.734	0.707	0.708

Notes: Robust and country-level clustered std. errors in parentheses.

Table 3: Estimation Results (Marginal Effects)

Regressor	Parameter	Coeff. (Std. err.)	**Moments of** $\widehat{a}_r \equiv exp(\widehat{const.} + \widehat{\varepsilon}_r^a)$				
First Stage: Dep. Var. $\log(\bar{L}_r)$			Mean			Std. Dev.	
$log(R_r)$	ρ_1	-0.473*** (0.014)	60,107			352,390	
Second Stage: Dep. Var.		$\log(a_{rt})$	5%	10%	50%	90%	95%
$\log(\widehat{\bar{L}_r})$	$-\lambda$	-0.596*** (0.033)	194.8	440.0	6,272.7	77,787.9	158,529.7
#obs 5,633							

Table 4: Amenity Parameter Estimation Results

R&D Policy Instrument	Description	Adopting Countries (in 2005)
Dtaxcredit_r	Tax credits on R&D investments.	Austria, Canada, China, France, Ireland, Japan, Mexico, Netherlands, Norway, Portugal, South Korea, Spain, Taiwan, US, Venezuela.
Dtaxholiday_r	Tax holidays for firms with R&D investments.	France, Malaysia, Singapore, Switzerland.
Dgrants_r	R&D investment related grants from the government.	Germany, Hungary, Ireland, Israel.
Dpatentbox_r	(Partial) exemption of returns on R&D investments.	France, Hungary.
Ddeduc_r	Deductions on R&D investments other than super deductions.	Australia, Belgium, Ireland, Japan, South Korea.

France incl. Guadeloupe, French Guiana, Martinique, Reunion; Netherlands incl. Bonaire; US incl. American Samoa, US Minor Outlying Islands; Australia incl. Cocos Islands; UK incl. Falkland Islands, Gibraltar, Montserrat, Pitcarn, St. Helena.

Table 5: R&D Policy Instruments in 2005

Period	Min	Max	Mean	Std
Baseline in %				
T=10	0.9	6.5	3.6	0.53
T=50	1.6	4.1	2.8	0.24
T=100	2.1	3.1	2.6	0.09
Counterfactual in %				
T=10	0.8	6.5	3.4	0.52
T=50	1.4	3.9	2.5	0.24
T=100	1.9	2.8	2.3	0.09
Counterfactual-Baseline in %pts				
T=10	-0.94	-0.00	-0.23	0.10
T=50	-0.58	-0.16	-0.26	0.05
T=100	-0.38	-0.22	-0.26	0.02

Table 6: Moments of Real GDP Growth

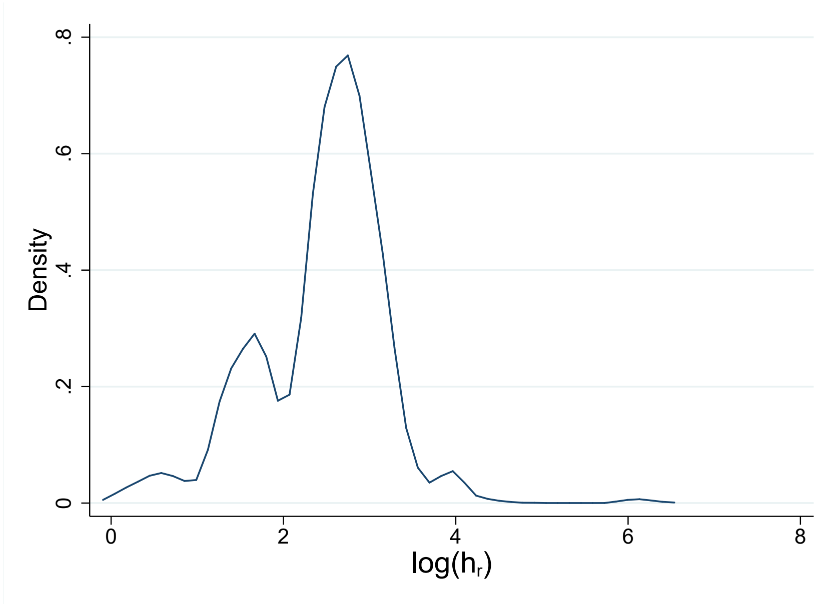

Figure 2: Kernel density of the estimated log R&D-worker-specific productivity shifter

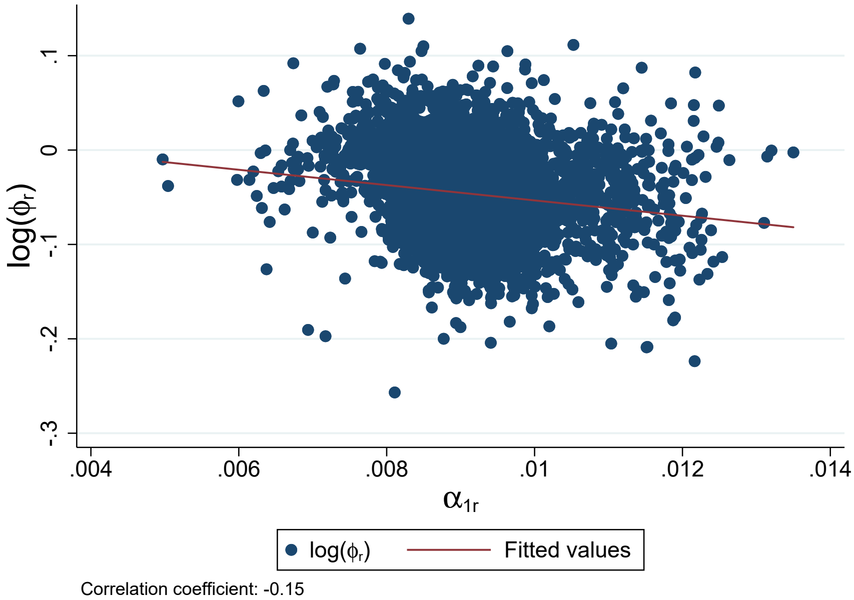

Figure 3: Relationship between log overall innovative productivity and estimated region-specific importance weight of patented innovations (2005)

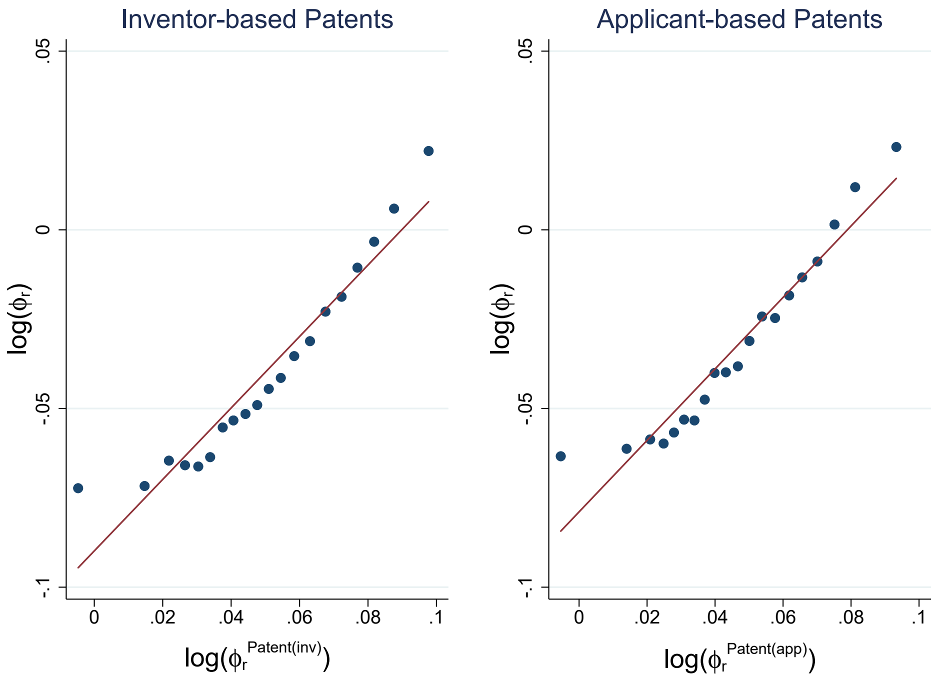

Figure 4: Binned scatterplots of the relationship between log overall innovative productivity and log patented innovations (2005)

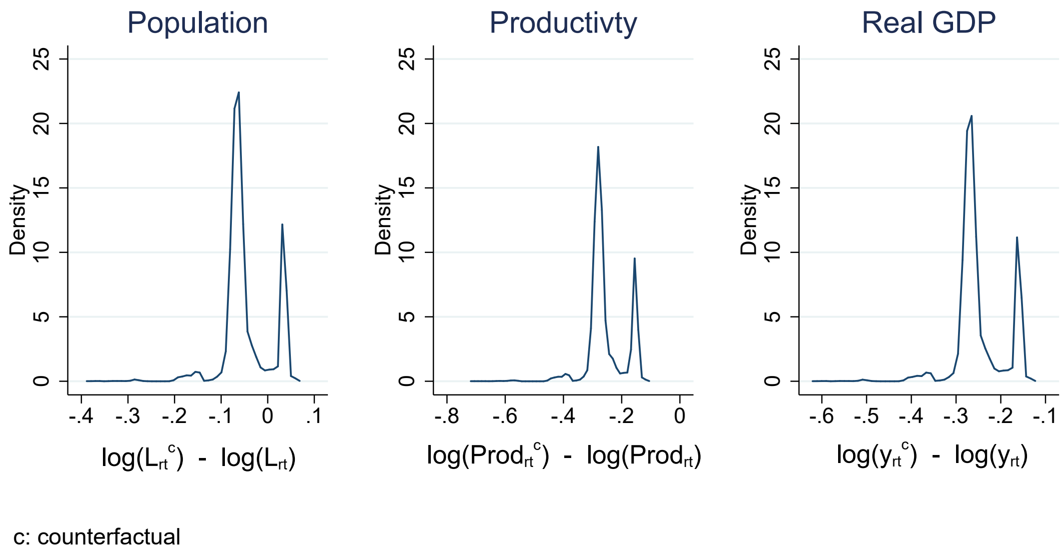

Figure 5: Density estimates of counterfactual changes, T=100 (Population, Productivity, Welfare)

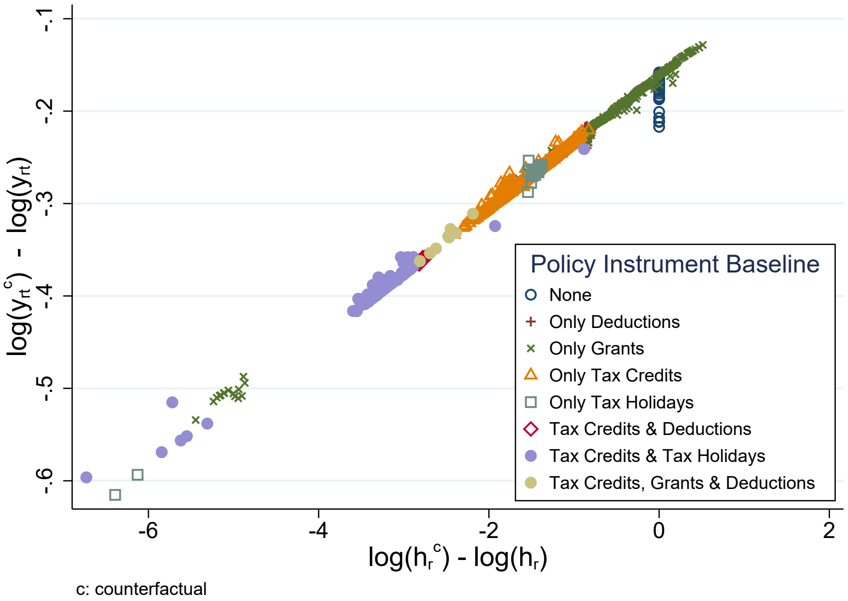

Figure 6: WELFARE CHANGE AT T=100 AND CHANGES IN h_r

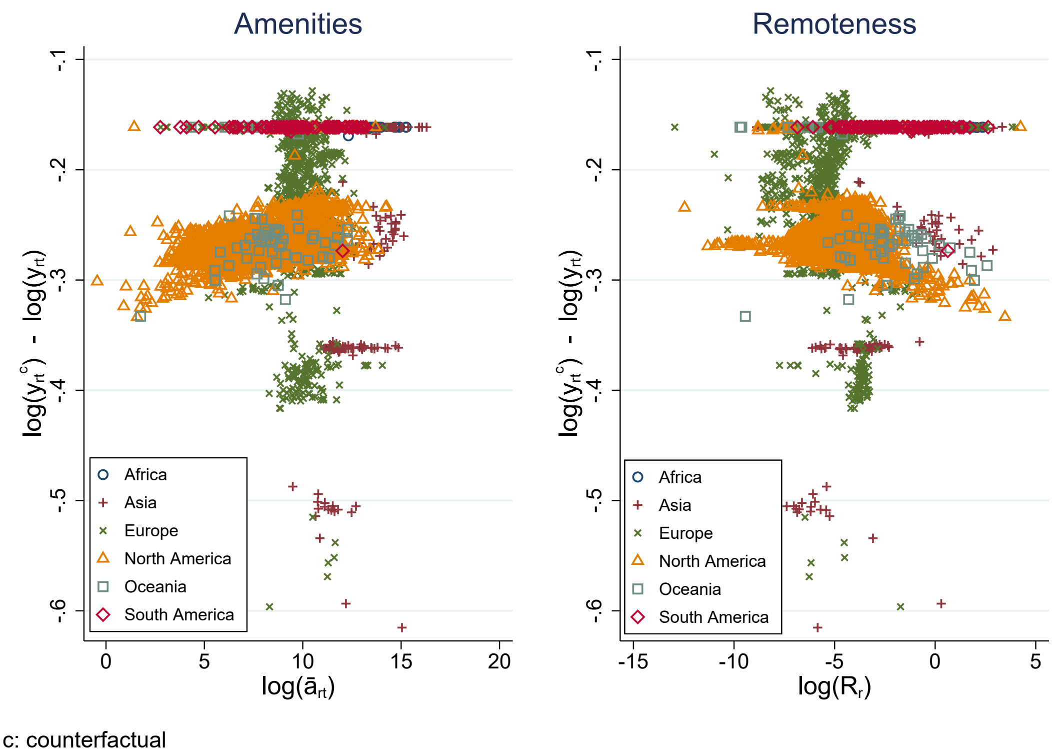

Figure 7: WELFARE CHANGE AT T=100 AND AMENITY/REMOTENESS LEVELS (BY CONTINENTS)

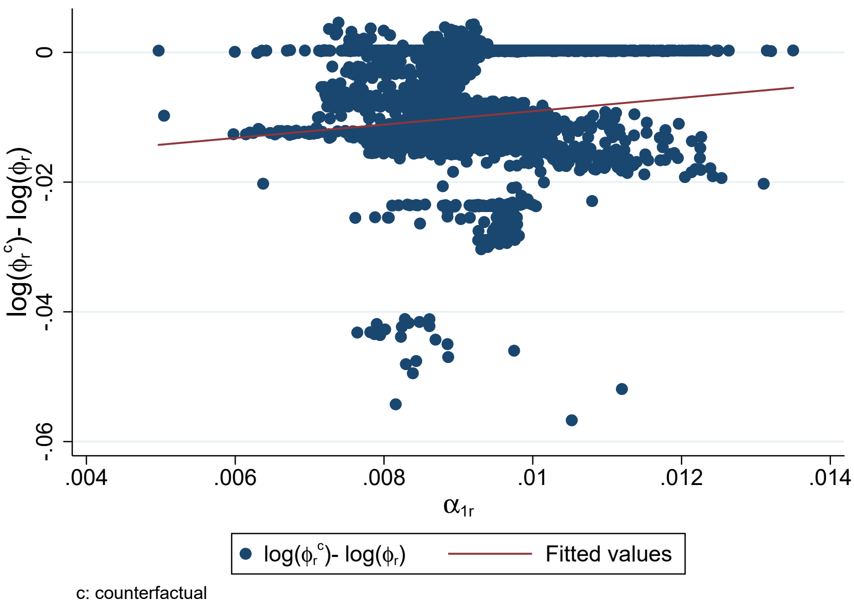

Figure 8: Relationship between long-term log changes in overall innovation and estimated region-specific importance weight of patented innovations (T=100)

Appendix

$\log(\tau_r^{2-\gamma_2})$	OECD Countries (with Singapore) (1) OLS	(2) 2SLS	(3) 2SLS	Non-OECD Countries (without Singapore) (4) OLS	(5) 2SLS	(6) 2SLS
First Stage		$\log(\bar{L}_r)$	$\log(\bar{L}_r)$		$\log(\bar{L}_r)$	$\log(\bar{L}_r)$
$\log(R_r)$		-0.793***	-0.814***		-0.442***	-0.517***
		(0.074)	(0.065)		(0.070)	(0.063)
Second Stage						
$\widehat{log(\bar{L}_r)}$	1.200***	0.620***	0.606***	0.608**	0.350	0.283
	(0.088)	(0.063)	(0.087)	(0.274)	(0.434)	(0.389)
Dtaxcredit$_r$	0.090	-0.650	1.150**	-0.080	0.265	-1.103
	(0.421)	(0.421)	(0.448)	(0.541)	(0.669)	(0.894)
Dtaxholiday$_r$	-0.262	1.098***	0.418*	8.843***	8.752***	8.626***
	(0.441)	(0.320)	(0.201)	(0.648)	(0.576)	(0.830)
Dgrants$_r$	0.265	1.293**	1.787***	(*omitted*)	(*omitted*)	(*omitted*)
	(0.706)	(0.636)	(0.489)			
Dpatentbox$_r$	0.710*	0.225	-1.238***	(*omitted*)	(*omitted*)	(*omitted*)
	(0.381)	(0.308)	(0.419)			
Ddeduc$_r$	0.564	1.176***	0.660*	(*omitted*)	(*omitted*)	(*omitted*)
	(0.514)	(0.392)	(0.324)			
$\|lat_r\|$	0.079***	0.039***	0.014**	0.033	0.023	0.060*
	(0.009)	(0.011)	(0.007)	(0.021)	(0.020)	(0.033)
continent FE	NO	NO	YES	NO	NO	YES
# obs	5,199	5,199	5,199	434	434	434
Corr. coeff. $\{\log(\tau_r^{2-\gamma_2}); \widehat{\log(\tau_r^{2-\gamma_2})}\}$	0.736	0.704	0.709	0.549	0.417	0.120

Notes: Robust and country-level clustered std. errors in parentheses. In columns (4)-(6) the binary indicators Dgrants$_r$, Dpatentbox$_r$ and Ddeduc$_r$ are omitted because none of these policy instruments was in place in any of the non-OECD countries in 2005.

Table 7: Robustness Estimation Results (Marginal Effects): Subsamples

A Initial Efficiency and Amenity Distribution

To identify the initial efficiency distribution, we need to derive an expression for τ_{rt}, using the model structure. To do so, we replace unit costs (3) into the bilateral trade share in (9), plug it into the product-market clearing (11) and solve for a scaled τ_{rt}. At this point, we do not have any information on the R&D-worker-specific productivity shifter h_r. However, we can use the BGP relationship, $\tau_{rt} \propto \left(\bar{L}_r h_r\right)^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}}$, and replace h_r as a function of population density and efficiency levels. Then,

$$\tau_{rt}^{(2-\gamma_2)} = \frac{\bar{L}_{rt}^{1-\iota_1} G_r w_{rt}^{1+\theta}}{\int\limits_S w_{st}\bar{L}_{st} G_s \zeta_{rs}^{-\theta} \left[\int_S \tau_{kt}^{(2-\gamma_2)} \bar{L}_{rt}^{\iota_1} \zeta_{rk}^{-\theta} w_{kt}^{-\theta} dk\right]^{-1} ds}, \tag{26}$$

where $\iota_1 \equiv \alpha - (1-\mu)\theta$. Now, we numerically solve for the scaled τ_{rt} by applying a standard contraction mapping procedure as it is described in Appendix B.7 in Desmet et al. (2018), and using observed levels of population densities, $\bar{L}_{rt}$ and wages, w_{rt} for the benchmark year 2005. Population levels come from SEDAC and wage levels come from the G-Econ Project, which are aggregated to the regional level as described in D.1 and D.2, respectively. Note that $\bar{L}_{rt}$ represents population density, hence, population levels are divided by normalized land G_r to obtain $\bar{L}_{rt}$.

After learning h_r and parameters values γ_1 and γ_2 as described in Section 3.4 and 3.5, respectively, we identify the initial distribution of amenities, a_{rt} in the year 2005. To do so, we replace the unit costs (3) in the price index and plug the price index into the indirect utility function in (8). Then we replace the utility in (10) and solve for amenities, a_{rt}. Then, after defining

$$\Pi_{st} \equiv \bar{L}_{st}^{\iota_1} G_s w_{st}^{-\theta} h_s^{-\theta\gamma_1/\xi} \tau_{st} \zeta_{rs}^{-\theta},$$

$$a_{rt} = \left(\frac{\bar{L}_{rt} G_r}{\bar{L}}\right)^{\Omega} \frac{1}{w_{rt}} \left[\int_S (a_{kt} w_{kt})^{1/\Omega} \left(\int_S \Pi_{st} ds\right)^{1/\Omega\theta} dk\right]^{\Omega} \left[\int_S \Pi_{kt} dk\right]^{-1/\theta}. \quad (27)$$

Again, we apply an iterative procedure to solve for the initial amenity distribution a_{rt} using observed population densities and wages. With a_{rt} we estimate the exogenous region-specific amenity-shock $\bar{a}_{rt}$ as described in Section 3.7.

B Equilibrium: Existence and Uniqueness

The uniqueness condition in (12) can be derived along the lines of Desmet et al. (2018) (see their Section B.3). We can manipulate the system of equations that defines an equilibrium as follows. For the first set of equations, we substitute (4) into (3) and replace that expression in the price index. Then,

$$P_{rt} = \kappa_0 \left[\int_S \tau_{st} \bar{L}_{st}^{\alpha-(1-\mu-\gamma_1/\xi)} w_{st}^{-\theta} \zeta_{rs}^{-\theta} h_{st}^{\theta\gamma_1/\xi} ds\right]^{-\frac{1}{\theta}}, \quad (28)$$

where $\kappa_0 = \bar{p}\left(\frac{1}{\mu}\right)^{\mu} \left(\frac{\xi\nu}{\gamma_1}\right)^{\gamma_1/\xi} \left(\frac{\xi\mu+\gamma_1}{\xi}\right)^{-(1-\mu-\gamma_1/\xi)}$ and $\bar{p} = \Gamma\left(\frac{1-\sigma}{\theta}+1\right)^{\frac{1}{1-\sigma}}$. Substituting (28) into (8) gives

$$\left[\frac{\bar{a}_r}{u_{rt}}\right]^{-\theta} \bar{L}_{rt}^{\theta\lambda} w_{rt}^{-\theta} = \kappa_1 \int_S \tau_{st} \bar{L}_{st}^{\alpha-(1-\mu-\gamma_1/\xi)\theta} w_{st}^{-\theta} \zeta_{rs}^{-\theta} h_{st}^{\theta\gamma_1/\xi} ds, \quad (29)$$

where $\kappa_1 = \left(\kappa_0 \frac{\mu\xi+\gamma_1}{\xi}\right)^{-\theta}$. For the second set of equations, we insert (9) and the price index into the product-market clearing (11) so that

$$w_{rt} G_r \bar{L}_{rt} = \bar{p}^{-\theta} \int_S T_{rt} [o_{rt} \zeta_{sr}]^{-\theta} P_{st}^{\theta} w_{st} G_s \bar{L}_{st} ds. \tag{30}$$

Substituting unit costs (3) and $T_{rt} = \tau_{rt} \bar{L}_{rt}^{\alpha}$, as well as replacing the price index with the indirect utility in the previous equation yields

$$\tau_{rt}^{-1} w_{rt}^{1+\theta} G_r h_{rt}^{-\frac{\theta\gamma_1}{\xi}} \bar{L}_{rt}^{1-(\alpha-(1-\mu-\gamma_1/\xi)\theta)} = \kappa_1 \int_S \left[\frac{\bar{a}_s}{u_{st}}\right]^{\theta} \zeta_{sr}^{-\theta} w_{st}^{1+\theta} G_s \bar{L}_{st}^{1-\lambda\theta} ds. \tag{31}$$

Assuming symmetric trade costs, we follow the proof of Theorem 2 in Allen and Arkolakis (2014), which is based on Theorem 2.19 in Zabreyko et al. (1975). Let us introduce the following function $\bar{f}_r$, which is the ratio of LHS's of (29) and (31):

$$\bar{f}_r = \frac{\tau_{rt}^{-1} w_{rt}^{1+\theta} G_r h_{rt}^{-\frac{\theta\gamma_1}{\xi}} \bar{L}_{rt}^{1-(\alpha-(1-\mu-\gamma_1/\xi)\theta)}}{\left[\frac{\bar{a}_r}{u_{rt}}\right]^{-\theta} \bar{L}_{rt}^{\theta\lambda} w_{rt}^{-\theta}}. \tag{32}$$

Equivalently, $\bar{f}_r$ also equals the RHS's of (29) and (31) that is

$$\bar{f}_r = \frac{\int_S \left[\frac{\bar{a}_s}{u_{st}}\right]^{\theta} \zeta_{sr}^{-\theta} w_{st}^{1+\theta} G_s \bar{L}_{st}^{1-\lambda\theta} ds}{\int_S \tau_{st} \bar{L}_{st}^{\alpha-(1-\mu-\gamma_1/\xi)\theta} w_{st}^{-\theta} \zeta_{rs}^{-\theta} h_{st}^{\theta\gamma_1/\xi} ds}. \tag{33}$$

Applying symmetric trade costs, $\zeta_{rs} = \zeta_{rs}$, we can rewrite $\bar{f}_r$ as follows

$$\bar{f}_r = \frac{\int_S \bar{f}_s^{-\lambda} \bar{\bar{f}}_{sr} \; ds}{\int_S \bar{f}_s^{-(1+\lambda)} \bar{\bar{f}}_{sr} \; ds}, \tag{34}$$

where

$$\bar{\bar{f}}_{sr} = \left[\frac{\bar{a}_s}{u_{st}}\right]^{\theta(1+\lambda)} \tau_{st}^{-\lambda} G_s^{1+\lambda} \, \zeta_{sr}^{-\theta} h_{st}^{-\lambda\frac{\theta\gamma_1}{\xi}} w_{st}^{1+\theta+(1+2\theta)\lambda} \bar{L}_{st}^{1-\lambda\theta-\lambda[\alpha-1+(\lambda+\frac{\gamma_1}{\xi}-(1-\mu))\theta]}. \tag{35}$$

Rewrite (34) as

$$\bar{\bar{\bar{f}}}_r = \frac{\bar{f}_r^{-\lambda}}{\int_S \bar{f}_s^{-\lambda} \bar{\bar{f}}_{sr} \; ds} = \frac{\bar{f}_r^{-(1+\lambda)}}{\int_S \bar{f}_s^{-(1+\lambda)} \bar{\bar{f}}_{sr} \; ds}. \tag{36}$$

Then, changing the notation to

$$\bar{g}_r = \bar{f}_r^{-\lambda} \quad \text{and} \quad \bar{\bar{g}}_r = \bar{f}_r^{-(1+\lambda)}, \tag{37}$$

and rewrite both as follows

$$\bar{g}_r = \int_S \bar{\bar{\bar{f}}}_r \bar{\bar{f}}_{sr} \bar{g}_s \; ds \quad \text{and} \quad \bar{\bar{g}}_r = \int_S \bar{\bar{\bar{f}}}_r \bar{\bar{f}}_{sr} \bar{\bar{g}}_s \; ds. \tag{38}$$

Define $\bar{\bar{\bar{f}}}_r \bar{\bar{f}}_{sr}$ as kernel K_{sr}. Hence, $\bar{g}_r$ and $\bar{\bar{g}}_r$ are both solutions to the integral equation

$$x_r = \int_S K_{rs} \; x_s \; ds. \tag{39}$$

We have to ensure that K_{sr} is (i) non-negative, (ii) measurable and (iii) square-integrable. Non-negativity holds as $\bar{\bar{f}}$ and $\bar{\bar{\bar{f}}}$ are non-negative. Measurability holds because it can be shown that $\bar{\bar{f}}$ and $\bar{\bar{\bar{f}}}$ are approximately continuous everywhere. Square-integrablility holds as long as population at any given location is bounded from below and above. The former is true because by construction population cannot shrink to zero unless nominal wages are zero or amenities are infinitely high. The latter is true because population at any given location cannot exceed the level of world population $\bar{L}$.

Given the properties of K_{sr}, Theorem 2.19 in Zabreyko et al. (1975) guarantees that there exists a unique (to scale) strictly positive function that satisfies the system of equations in (39). Hence,

$$\bar{g}_r = \varpi \bar{\bar{g}}_r \quad \Rightarrow \bar{f}_r^{-\lambda} = \varpi \bar{f}_r^{-(1+\lambda)} \quad \Rightarrow \bar{f}_r = \varpi, \tag{40}$$

where ϖ is a constant. Therefore, we have

$$\frac{\tau_{rt}^{-1} w_{rt}^{1+\theta} G_r h_{rt}^{-\frac{\theta\gamma_1}{\xi}} \bar{L}_{rt}^{1-(\alpha-(1-\mu-\gamma_1/\xi)\theta)}}{\left[\frac{\bar{a}_r}{u_{rt}}\right]^{-\theta} \bar{L}_{rt}^{\theta\lambda} w_{rt}^{-\theta}} = \varpi, \tag{41}$$

and solving for w_{rt} gives

$$w_{rt} = \bar{w} \left[\frac{\bar{a}_r}{u_{rt}}\right]^{-\frac{\theta}{1+2\theta}} \tau_{rt}^{\frac{1}{1+2\theta}} G_r^{-\frac{1}{1+2\theta}} \bar{L}_{rt}^{\frac{\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta}{1+2\theta}} h_{rt}^{\frac{\theta\gamma_1/\xi}{1+2\theta}}, \tag{42}$$

where $\bar{w} = \varpi^{\frac{1}{1+2\theta}}$. Substituting (42) into (29) gives

$$\begin{aligned}
&\left[\frac{\bar{a}_r}{u_{rt}}\right]^{-\frac{\theta(1+\theta)}{1+2\theta}} \tau_{rt}^{-\frac{\theta}{1+2\theta}} G_r^{\frac{\theta}{1+2\theta}} \bar{L}_{rt}^{\lambda\theta-\frac{\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]} h_{rt}^{-\frac{\theta(\theta\gamma_1/\xi)}{1+2\theta}} \\
&\qquad = \kappa_1 \int_S \left[\frac{\bar{a}_s}{u_{st}}\right]^{\frac{\theta^2}{1+2\theta}} \tau_{st}^{\frac{1+\theta}{1+2\theta}} G_s^{\frac{\theta}{1+2\theta}} \zeta_{rs}^{-\theta} \bar{L}_{st}^{1-\lambda\theta+\frac{1+\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]} h_{st}^{\frac{(1+\theta)(\theta\gamma_1/\xi)}{1+2\theta}} \; ds.
\end{aligned} \tag{43}$$

Inserting (10) into (43) gives

$$\begin{aligned}\bar{B}_{rt}\ \hat{u}_{rt}^{\frac{1}{\Omega}\left[\lambda\theta-\frac{\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]\right]+\frac{\theta(1+\theta)}{1+2\theta}} \\ =\kappa_1\int_S \hat{u}_{st}^{\frac{1}{\Omega}\left[1-\lambda\theta+\frac{1+\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]\right]-\frac{\theta^2}{1+2\theta}}\bar{\bar{B}}_{st}\zeta_{rs}^{-\theta}\ ds,\end{aligned} \tag{44}$$

where

$$\bar{B}_{rt}=\bar{a}_r^{-\frac{\theta(1+\theta)}{1+2\theta}}\tau_{rt}^{-\frac{\theta}{1+2\theta}}G_r^{\frac{\theta}{1+2\theta}[\alpha+[\lambda+\gamma_1/\xi-(1-\mu)]\theta]-\lambda\theta}h_{rt}^{-\frac{\theta(\theta\gamma_1/\xi)}{1+2\theta}},$$

and

$$\bar{\bar{B}}_{st}=\bar{a}_s^{\frac{\theta^2}{1+2\theta}}\tau_{st}^{\frac{1+\theta}{1+2\theta}}G_s^{\frac{\theta}{1+2\theta}-1+\lambda\theta-\frac{1+\theta}{1+2\theta}[\alpha-1+[\lambda+\gamma_1/\xi-(1-\mu)]\theta]}h_{st}^{\frac{(1+\theta)(\theta\gamma_1/\xi)}{1+2\theta}},$$

and

$$\hat{u}_{rt}=u_{rt}\left[\frac{\bar{L}}{\int_S u_{kt}^{1/\Omega}dk}\right]^{\Omega\left[1-\frac{\theta}{\frac{1}{\Omega}\left[\left[\lambda+(1-\mu)-\frac{\gamma_1}{\xi}\right]\theta-\alpha\right]+\theta}\right]}. \tag{45}$$

Rewrite (44) as

$$\bar{B}_r f_r^{\tilde{\gamma}_1}=\kappa_1\int_S \bar{\bar{B}}_s\zeta_{rs}^{-\theta}f_s^{\tilde{\gamma}_2}\ ds, \tag{46}$$

and apply Theorem 2.19 in Zabreyko et al. (1975), then the solution $f_{(\cdot)}$ to equation (46) exists and is unique if (a) the function $\kappa_1\bar{B}_r^{-1}\bar{\bar{B}}_s\zeta_{rs}^{-\theta}$ is strictly positive and continuous, and (b) $\left|\frac{\tilde{\gamma}_2}{\tilde{\gamma}_1}\right|\leq 1$. The latter implies

$$\frac{\frac{1}{\Omega}\left[1-\lambda\theta+\frac{1+\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]\right]-\frac{\theta^2}{1+2\theta}}{\frac{1}{\Omega}\left[\lambda\theta-\frac{\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]\right]+\frac{\theta(1+\theta)}{1+2\theta}}\leq 1,$$

which after some simplification can be written as the uniqueness condition (12) as stated in Section 2.4

$$\frac{\alpha}{\theta}+\frac{\gamma_1}{\xi}\leq\lambda+1-\mu+\Omega.$$

C Balanced Growth Path: Derivation

C.1 Uniqueness and Existence Condition in the BGP

Efficiency evolves according to a endogenous dynamic process in (2) and, hence, the growth rate of τ_{rt} is given by

$$\frac{\tau_{rt+1}}{\tau_{rt}}=\phi_{rt}^{\theta\gamma_1}\left[\int_S\frac{W_{rs}\tau_{st}}{\tau_{rt}}ds\right]^{1-\gamma_2}, \tag{47}$$

where we define $W_{rs} \equiv 1/S, \forall rs$ as described in Section 2.5. Divide both sides by the corresponding equation for region s, and rearrange, knowing that $\frac{\tau_{rt+1}}{\tau_{rt}}$ is constant over time and space and $\frac{\tau_{st}}{\tau_{rt}}$ is constant over time. Hence,

$$\underbrace{\frac{\frac{\tau_{rt+1}}{\tau_{rt}}}{\frac{\tau_{st+1}}{\tau_{st}}}}_{=1} = \left[\frac{\tau_{st}}{\tau_{rt}}\right]^{1-\gamma_2} \left[\frac{\phi_{rt}}{\phi_{st}}\right]^{\theta\gamma_1} \underbrace{\left[\frac{\int_S \tau_{st} ds}{\int_S \tau_{rt} dr}\right]^{1-\gamma_2}}_{=1} \quad \Rightarrow \quad \frac{\tau_{st}}{\tau_{rt}} = \left[\frac{\phi_{st}}{\phi_{rt}}\right]^{\frac{\theta\gamma_1}{1-\gamma_2}} = \left[\frac{\bar{L}_s h_s}{\bar{L}_r h_r}\right]^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}}, \tag{48}$$

where the last equality follows from (6). We drop the time subscript to demonstrate that population density remains constant in the BGP. Rewrite the last equation as

$$\bar{L}_s = \left[\frac{\tau_{st}}{\tau_{rt}}\right]^{\frac{(1-\gamma_2)\xi}{\theta\gamma_1}} \bar{L}_r \frac{h_r}{h_s},$$

and integrate both sides over s and apply the labor market clearing condition, $\int_S G_s \bar{L}_{st} ds = \bar{L}$ such that

$$\int_S G_s \bar{L}_s ds = \bar{L} = \tau_{rt}^{-\frac{(1-\gamma_2)\xi}{\theta\gamma_1}} \bar{L}_r h_r \int_S G_s \tau_{st}^{\frac{(1-\gamma_2)\xi}{\theta\gamma_1}} h_s^{-1} ds \quad \Rightarrow \quad \tau_{rt} = \tilde{\kappa}_t (h_r \bar{L}_r)^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}}, \tag{49}$$

where $\tilde{\kappa}_t$ depends on time but not on location. Take the last equation and substitute it into (43) such that

$$\begin{aligned} &\left[\frac{\bar{a}_r}{u_{rt}}\right]^{-\frac{\theta(1+\theta)}{1+2\theta}} G_r^{\frac{\theta}{1+2\theta}} \bar{L}_r^{\lambda\theta - \frac{\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]} h_r^{-\frac{\theta(\theta\gamma_1/\xi)}{1+2\theta}(1+\frac{1}{1-\gamma_2})} \\ &= \kappa_1 \tilde{\kappa}_t \int_S \left[\frac{\bar{a}_s}{u_{st}}\right]^{\frac{\theta^2}{1+2\theta}} G_s^{\frac{\theta}{1+2\theta}} \zeta_{rs}^{-\theta} \bar{L}_s^{1-\lambda\theta + \frac{1+\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]} h_s^{\frac{(1+\theta)(\theta\gamma_1/\xi)}{1+2\theta}(1+\frac{1}{1-\gamma_2})} ds. \end{aligned} \tag{50}$$

Inserting (10) in (50) and rearranging conveniently, yields

$$\begin{aligned} &\bar{D}_r \; \hat{\hat{u}}_{rt}^{\frac{1}{\Omega}\left[\lambda\theta - \frac{\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]\right] + \frac{\theta(1+\theta)}{1+2\theta}} \\ &= \kappa_1 \tilde{\kappa}_t \int_S \hat{\hat{u}}_{st}^{\frac{1}{\Omega}\left[\lambda\theta + \frac{1+\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]\right] - \frac{\theta^2}{1+2\theta}} \; \bar{\bar{D}}_s \; \zeta_{rs}^{-\theta} \; ds, \end{aligned} \tag{51}$$

where

$$\bar{D}_r = \bar{a}_r^{-\frac{\theta(1+\theta)}{1+2\theta}} G_r^{\frac{\theta}{1+2\theta}[\alpha + [\lambda + \frac{\gamma_1}{\xi} - [1-\mu]]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}]} h_r^{-\frac{\theta(\theta\gamma_1/\xi)}{1+2\theta}(1+\frac{1}{1-\gamma_2})},$$

and

$$\bar{\bar{D}}_s = \bar{a}_s^{\frac{\theta^2}{1+2\theta}} G_s^{\frac{\theta}{1+2\theta} - 1 + \lambda\theta - \frac{1+\theta}{1+2\theta}[\alpha - 1 + [\lambda + \frac{\gamma_1}{\xi} - [1-\mu]]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}]} h_s^{\frac{(1+\theta)(\theta\gamma_1/\xi)}{1+2\theta}(1+\frac{1}{1-\gamma_2})},$$

are exogenously given, and

$$\hat{\hat{u}}_{rt} = u_{rt} \left[\frac{\bar{L}}{\int_S u_{kt}^{1/\Omega} dk} \right]^{\Omega \left[1 - \frac{\theta}{\frac{1}{\Omega}\left[\left[\lambda + (1-\mu) - \frac{\gamma_1}{\xi}\right]\theta - \alpha - \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right] + \theta} \right]}. \tag{52}$$

Analogously to the existence and uniqueness proof in Section B, we can rewrite (51) as

$$\bar{D}_r g_r^{\tilde{\tilde{\gamma}}_1} = \kappa_1 \tilde{\kappa}_t \int_S \bar{\bar{D}}_s \zeta_{rs}^{-\theta} g_s^{\tilde{\tilde{\gamma}}_2} \, ds. \tag{53}$$

According to Theorem 2.19 in Zabreyko et al. (1975) $g_{(\cdot)}$ is a solution to the system of equations in (53) that is unique if $\left| \frac{\tilde{\tilde{\gamma}}_2}{\tilde{\tilde{\gamma}}_1} \right| \leq 1$. This condition implies

$$\frac{\frac{1}{\Omega}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right] + \frac{\theta(1+\theta)}{1+2\theta}}{\frac{1}{\Omega}\left[1 - \lambda\theta + \frac{1+\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]\right] - \frac{\theta^2}{1+2\theta}} \leq 1,$$

from which, after some rearrangement, we get the uniqueness condition in the balanced growth path (16) as stated in Section 2.5

$$\frac{\alpha}{\theta} + \frac{\gamma_1}{\xi} + \frac{\gamma_1}{[1-\gamma_2]\xi} \leq \lambda + 1 - \mu + \Omega.$$

C.2 Growth Rate of Aggregate Welfare

To derive the growth rate of aggregate welfare, rewrite (42) as follows

$$\tau_{rt} = \bar{w}^{-(1+2\theta)} \left[\frac{\bar{a}_r}{u_{rt}} \right]^{\theta} w_r^{1+2\theta} G_r \bar{L}_r^{\frac{1-\alpha+\left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta}{1+2\theta}} h_r^{-\frac{\theta\gamma_1}{\xi}}. \tag{54}$$

Substituting the previous equation into (49) and solving for u_{rt} gives

$$u_{rt} = \tilde{\kappa}_t^{\frac{1}{\theta}} E_r, \tag{55}$$

where E_r is only dependent on the location and not on time. Hence,

$$\frac{u_{rt+1}}{u_{rt}} = \left(\frac{\tilde{\kappa}_{t+1}}{\tilde{\kappa}_t} \right)^{\frac{1}{\theta}} = \left(\frac{\tau_{rt+1}}{\tau_{rt}} \right)^{\frac{1}{\theta}}, \tag{56}$$

where the last equality follows from (49). From (47) and (48) we know

$$\frac{\tau_{rt+1}}{\tau_{rt}} = \phi_{rt}^{\theta\gamma_1} \left[\frac{1}{S} \int_S \frac{\tau_{st}}{\tau_{rt}} ds \right]^{1-\gamma_2} = \left(\frac{\gamma_1/\nu}{\gamma_1 + \mu\xi} \bar{L}_r h_r \right)^{\frac{\theta\gamma_1}{\xi}} \left[\frac{1}{S} \int_S \left(\frac{\bar{L}_s h_s}{\bar{L}_r h_r} \right)^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}} ds \right]^{1-\gamma_2} . \tag{57}$$

Rearranging the previous equation and substituting it into (56) gives

$$\frac{u_{rt+1}}{u_{rt}} = \left[\frac{1}{S} \right]^{\frac{1-\gamma_2}{\theta}} \left[\frac{\gamma_1/\nu}{\gamma_1 + \mu\xi} \right]^{\frac{\theta\gamma_1}{\xi}} \left(\int_S (\bar{L}_s h_s)^{\frac{\theta\gamma_1}{[1-\gamma_2]\xi}} \; ds \right)^{\frac{1-\gamma_2}{\theta}} .$$

D Data Aggregation

Our unit of interest are REGPAT regions. We use gridded data with different resolution for which we need an aggregation strategy to the regional level. Hereafter, we discuss the aggregation strategy for each data source separately.

D.1 Population Data from SEDAC

The Socioeconomic Data and Application Center (SEDAC) provides gridded population data with an output resolution of 30 arc-seconds (approximately 1 km at the equator). As the size of each grid cell is smaller than the smallest region in our data, we simply sum up the population count over all grid cells falling withing the regional border.

D.2 Population and GDP from G-Econ Project

The Geographically based Economic Data (G-Econ) project at Yale University provides SEDAC gridded population data aggregated to the 1° by 1° resolution (approximately 100km by 100km at the equator), which is about the same size as second level political entities in most countries. Besides population data, the G-Econ project offers gridded GDP data (gross cell product at purchasing power parity (PPP)) at the 1° by 1° resolution. We assign population and GDP values to each region through an area-weighted average aggregation. Figure 9 illustrates how the area-weights are assigned in the case of GDP data (left panel) and population count data (right panel). In both panels, the green area is the region of Prague, which falls into two different grid cells (bordered in red). Therefore, the GDP value of Prague is equal to six-tenth of the left grid cell plus four-tenth of the right grid cell. In the case of population count data, we construct the area-weight as the part of Prague that falls into the grid cell relative to the overall area of the grid cell. Hence, the population count of Prague is four-hundreds of the left grid cell pull three-hundreds of the right grid cell.

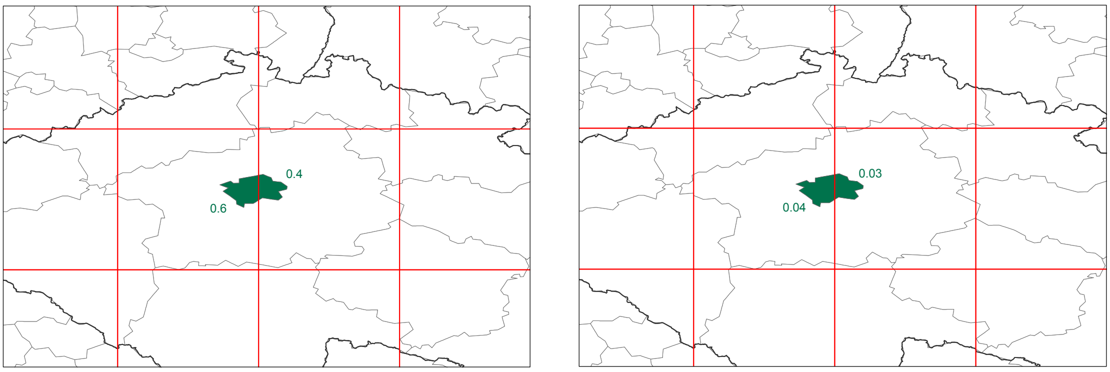

Figure 9: Aggregation for data with one degree resolution

D.3 Fast-marching-algorithm-based Transportation Costs

We derive the fast-marching-algorithm-based transportation costs between pairs of 1° grid cells along the lines of Desmet et al. (2018). To find a correspondence of these transportation costs to the level of REGPAT regions, we employ an area-weighted average assignment. The area-weights are constructed as the share of regional area falling into a grid cell relative to the total regional area (see left panel of Figure 9). Our averaging procedure can be best explained using matrix notation. Let W_{nx1} be the vector of area-weights for n sub-regions, where a sub-region refers to an intersection between a REGPAT region and a one-degree grid cell area. Furthermore, we define the fast-marching transportation costs matrix as T_{nxn}, which is blown up from the number of one-degree grid cells to the number of n sub-regions, using information on sub-region intersections with one-degree grid cells from ArcGIS. Lastly, we need a correspondence of sub-region to the final set of REGPAT regions r and define a selector matrix S_{nxr} using ArcGIS, where r is equal to 5,633. Then the regional transportation costs T_{rxr} can be obtained as follows

$$T_{rxr} = W'_{nxr} T_{nxn} W_{nxr}, \tag{58}$$

where $W_{mxr} = (W_{mx1} \iota'_{mx1}) \circ S_{mxr}$.