The Economic Geography of Innovation

PRELIMINARY VERSION

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

Nicole Loumeau†
ETH Zürich

February 2018

Abstract: This paper outlines a multi-region quantitative model to assess the importance of country-level investment incentives towards innovation at the level of 5,633 micro-regions of different 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 cross-sectional unit of observation underlying the quantitative analysis are REGPAT regions, whose patenting output we measure and link to population as well as income statistics. 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 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. Particularly, regions with high amenities and a low degree of transport remoteness tend to benefit from such policy instruments.

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.

The authors gratefully acknowledge funding from the Swiss National Science Foundation (SNF) through project number CRSII 1_154446.

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 such innovations generated elsewhere. Economic policy has a number of instruments at hand which are aimed at stimulating innovation. Earlier research concerned with the effect of such incentives on innovations – which are commonly measured by patent filings or citations – and economic outcome focuses on reduced-form effects which largely abstract from general-equilibrium repercussions (see, among others, Warda, 2002; Westmore, 2013; Baumann et al., 2014; Boesenberg and Egger, 2016). The present paper formulates, estimates key parameters of, and calibrates a quantitative, multi-place general-equilibrium 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.

The model builds on earlier work by Allen and Arkolakis (2014) and Desmet et al. (2017) 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 last, 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. As compared to Allen and Arkolakis (2014) and Desmet et al. (2017), 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.

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 data on patent registrations from REGPAT and seven 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 abroad due to technology spillovers. Second, the long-run relocation effects due to a hypothetical abolishment of all R&D policy instruments are substantial and lead to a re-shifting of population towards high-density areas. Last, the quantitative analysis suggests that particularly regions with high amenities and a low degree of transport remoteness tend to benefit from such policy instruments. The former can be explained by the fact that regions with high amenity values are able to attract labor which is key not only for production but also innovation. The latter follows the arguments that well-connected regions can generate a high return on innovations to firms through their greater global market potential.

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 The Role of Innovation for 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. The first component is an endogenous decision on the level of innovation $\phi_{rt}(\omega)$. The second component is 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 its 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. Hence,

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

where $\gamma_1, \gamma_2 \in (0,1)$. 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 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. Contrary to Desmet et al. (2017), this paper considers an additional R&D-worker-specific productivity shifter $h_{rt} \geq 1$, which is region-specific and decreases the cost of innovation per unit of innovation produced.

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

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

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 $p(\omega)_{rt} = o_{rt}/z_{rt}(\omega)$, with

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

b_{rt} is the firms bid rent for land, which can be derived from the first order conditions as function of the per-unit costs of innovation $w_{rt}\phi_{rt}(\omega)^{\xi}h_{rt}^{-1}$, so that

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

Each firm considers their production unit costs as given, which is why o_{rt} is not product-specific.

2.2 The Role of Innovation for 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

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

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

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

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

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

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.

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 worker is $b_{rt}/\bar{L}_{rt}$. As we assume that workers can not write debt contracts with each other and there is perfect local competition, it follows that each worker consumes all his 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 at period t. As in Eaton and Kortum (2002), the share of consumption in region r of products produced in region s is determined by

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

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

2.4 Equilibrium in Each Period

Each equilibrium maximizes only current profits, as each period is self-contained and firms are not forward-looking. There is no mechanism to borrow from or lend to other workers.

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

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

where Ω is an importance parameter that is micro-founded as a Fréchet parameter of a location-specific preference shock as in Desmet et al. (2017). Overall, population mobility is restricted by the location-specific 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 congestions forces are greater than agglomeration forces. Hence,

$$\frac{\alpha}{\theta} + \frac{\gamma_1}{\xi} \leq \lambda + 1 - \mu + \Omega \tag{12}$$

A detailed proof of the uniqueness condition is presented in Appendix B.

2.5 Balanced Growth Path

In a balanced growth path (BGP) technology growth rates are constant, implying that $\frac{\tau_{rt+1}}{\tau_{rt}}$ is constant over time and space and $\frac{\tau_{st}}{\tau_{rt}}$ is constant over time. The firm's investment decision into innovation is constant, however it will be different across regions. Furthermore, we assume that the innovation-worker-productivity shifter h_{rt} is constant over time, even when the economy is only moving towards a BGP. This implies

$$\frac{\tau_{st}}{\tau_{rt}} = \left[\frac{\phi_{st}}{\phi_{rt}}\right]^{\frac{\theta\gamma_1}{1-\gamma_2}} = \left[\frac{\bar{L}_{st}h_{st}}{\bar{L}_{rt}h_{rt}}\right]^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}} \tag{13}$$

There exists a unique growth path if

$$\frac{\alpha}{\theta} + \frac{\gamma_1}{\xi} + \frac{\gamma_1}{[1-\gamma_2]\xi} \leq \lambda + 1 - \mu + \Omega. \tag{14}$$

This is a stronger equilibrium condition than (12), as it implies that with an additional dynamic agglomeration effect, $\frac{\gamma_1}{[1-\gamma_2]\xi}$, congestions forces still have to be greater than agglomeration forces for the equilibrium to be unique.[1] In the BGP aggregate welfare and real consumption grow according to

$$\frac{u_{rt+1}}{u_{rt}} = \frac{C_{rt+1}}{C_{rt}} = \left[\frac{1}{S}\right]^{\frac{1-\gamma_2}{\theta}} \left[\frac{\gamma_1/\nu}{\gamma_1+\mu\xi}\right]^{\frac{\theta\gamma_1}{\xi}} \left(\int_S (\bar{L}_s h_s)^{\frac{\theta\gamma_1}{[1-\gamma_2]\xi}} \;\; ds\right)^{\frac{1-\gamma_2}{\theta}} \tag{15}$$

where $a_{rt} = a_{rt+1}$ as the population density in each region is constant over time in the BGP.[2]

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 regions and region-specific land endowments which are given in the data, these are initial productivity levels in production for all regions and trade costs between all pairs of regions.[3] 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.

[1]A detailed derivation of this equilibrium condition is presented Appendix C.1.

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

[3]Initial productivity levels, τ_0, are obtained using the structure of the model and observed population and wage levels in year 2005. Technical details on the derivation of τ_0 are explained in Appendix A.

3.1 Delineation of Regions and Land Endowments

The delineation of regions used in our analysis is dictated by the definitions used in regional patent-statistical database (REGPAT) of the Organization of Economic Cooperation and Development (OECD).[4] In 2005, REGPAT distinguishes 5,633 regions across 213 countries around 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 delineated 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 most of the regions are in North America (United States, Canada and Mexico) and Europe, which together constitute 90 percent of the regions in the data. Using the regional code in the REGPAT database we can link the REGPAT regions with spatial information from two sources: (i) we employ the Geographical Information and Maps (GISCO) database from eurostat for spatial information on European countries (NUTS3 regions, 2010) and (ii) we use the Global Administrative Areas (GADM) spatial database on administrative boundaries for all other countries. Then, we extract the land mass for each region using ArcGIS software after excluding water sheds within the boundaries of a region. We 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. (2017) and using passing-through parameters from Allen and Arkolakis (2014);[5] (ii) a correspondence of

[4]The REGPAT database links the Worldwide Statistical Patent Database (PATSTAT) from the European Patent Office (EPO) to more that 5,500 regions across the globe, utilizing the addresses of the applicants and inventors.

[5]We modify those costs by symmetrifying them (using the average for costs from A to B and B to A) and by assuming that intra-cell transport costs are (essentially) zero as is customary in Ricardian work (see Eaton and

these transport costs to the level of REGPAT regions by 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 ASJP measure which we use here.

3.3 Estimating Amenity-function Parameters

Taking logs of $a_{rt} = \bar{a}_{rt}\bar{L}_{rt}^{-\lambda}$ in equation (7) obtains

$$\log(a_{rt}) = -\lambda \log \bar{L}_{rt} + const. + \varepsilon_{rt}^{a} \tag{16}$$

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_{rt}^{a}, 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 (16). Therefore, we estimate (16) by two-stage least squares (2SLS) for the baseline year 2005, instrumenting $\bar{L}_{rt}$ with a region-specific area-weighted remoteness index, $R_r = weight_r^{area}\left(\frac{1}{S}\sum_S \zeta_{rs}\right)$, which does not depend on individual location decisions. 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 (16), we use the structure of the model, substitute the indirect utility (8) into (10) and solve for a_{rt} as in equation (23) that is derived in Appendix A.

– Table 2 about here –

Table 2 reports the estimation results from estimating (16), with the congestion parameter estimated at a value of $\lambda = 0.60$. Furthermore, the table reports first-order and second-order

Kortum, 2002; Donaldson, 2017).

moments of $\widehat{a}_{rt}$. As described above, the region-specific exogenous amenity attribute is defined as $\widehat{a}_{rt} \equiv exp(\widehat{const.} + \widehat{\varepsilon}^{a}_{rt})$. In the general-equilibrium analysis, $\widehat{a}_{rt}$ is kept constant for all time periods.

3.4 Technology and Productivity-evolution Parameters

Table 1 summarizes the assumed values of the technology parameters $\{\alpha, \theta, \mu\}$ and the productivity-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 the relationship in equation (15). Taking logs and discretizing (15) obtains

$$\begin{aligned} \log(u_{rt+1}) - \log(u_{rt}) =& \log(y_{rt+1}) - \log(y_{rt}) \\ =& \frac{(1-\gamma_2)}{\theta}\log(\eta) + \frac{\gamma_1}{\xi}\log(\Psi) + \frac{1-\gamma_2}{\theta}\log(\sum_r \bar{L}^*_{rt}) \end{aligned} \tag{17}$$

where $\Psi \equiv \frac{\gamma_1/\nu}{\gamma_1+\mu\xi}$, $S = 5,633$, and $\bar{L}^*_{rt} \equiv \left[\bar{L}_{rt} h_{rt}\right]^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}}$. Note that equation (17) depends on both population density ($\bar{L}_{rt}$) as well as on real-income data (y_{rt+1}, y_{rt}). Either type of data is available at the 1° by 1° resolution from the G-Econ 4.0 Research Project at Yale University, which we aggregate to the required (non-gridded) regional level as described in Appendix D.2.[6] We use the mentioned data for $t \in \{1990, 1995, 2000\}$ and $t+1 \in \{1995, 2000, 2005\}$ and approximate the log difference between years $t+1$ and t by a five-year interval. Moreover, we determine $\{\gamma_1, \gamma_2\}$ by pooling the data for the mentioned three year tuples $\{t, t+1\}$.

For the estimation of $\{\gamma_1, \gamma_2\}$, note that $\gamma_1, \gamma_2 \in (0, 1)$ and that h_{rt} and, hence, $\bar{L}^*_{rt}$ itself depends on γ_1. However, using the notion of a bivariate bounded parameter space, we can search for the optimal values of $\{\gamma_1, \gamma_2\}$ by doing a grid search on the bivariate unit interval, employing the respective value of $h_{rt}(\gamma_1)$. It will become clear in the subsequent subsection how h_{rt} depends on γ_1. Adopting this procedure, we obtain the estimates $\hat{\gamma}_1 = 0.273$ and $\hat{\gamma}_2 = 0.9897$ listed in Table 1.

[6] Notice that we do have population data from two sources, namely SEDAC and the G-Econ 4.0 Research Project. Whereas SEDAC provides gridded population data with an output solution of 30 arc-seconds (approx. 1km at the equator), the G-Econ project provides the same data on an aggregated 1° by 1° resolution. For consistency reasons, we inform the estimating equation based on (17) by population and real-income data from the G-Econ 4.0 Research Project. For all other applications, we use population data from SEDAC directly, to avoid measurement error from aggregation.

3.5 Estimation of h_{rt} (preliminary)

We structurally estimate h_{rt} using (6) and knowing (5). To link the available data to our model, we assume $\phi_{rt}^{\xi} = Patents_{rt}^{\tilde{\xi}}$, so that

$$\phi_{rt}^{\xi} = Patents_{rt}^{\tilde{\xi}} = \frac{\gamma_1}{\xi\nu[\mu + \gamma_1/\xi]} \bar{L}_{rt} h_{rt} \tag{18}$$

where $Patents_{rt}$ are registered patents per unit of land in region r at year t from REGPAT.[7] Here, $\bar{L}_{rt}$ is the population density in region r at year t, where population levels are taken from SEDAC and aggregated to the regional level. We parametrize h_{rt} as

$$h_{rt} = \exp(\mathbf{D}_{rt}\beta + |lat_r|\mathbf{D}_{rt}\gamma) \tag{19}$$

where $\mathbf{D}_{rt}$ describes a vector of binary R&D-policy indicators from Boesenberg and Egger (2016). The indicators in $\mathbf{D}_{rt}$ are country-year-specific and each region r is associated with a single country. The indicators include a binary indicator variable for partial exemptions of returns on R&D investments, also known as patent boxes ($Dpatentbox_{rt}$), R&D investment related grants from the government ($Dgrants_{rt}$), tax credit on R&D investments ($Dtaxcredit_{rt}$), tax holidays for firms with R&D investment ($Dtaxholiday_{rt}$), super deductions of more than 100% for R&D investments from profits ($Dsuperd_{rt}$), any form of deductions of R&D investments from profits other than super deductions ($Ddeduc_{rt}$), and a lower effective average tax rate on R&D investments as compared to non-R&D-related investments ($Deatrrd_{rt}$). $Deatrrd_{rt}$ is coded as a nonlinear combination of the other front-end R&D-policy instruments (i.e., the instruments except $Dpatentbox_{rt}$), as it takes on the maximum value for each rt across all the respective instruments. Hence, whenever there is any front-end instrument in place, $Deatrrd_{rt}$ is unity. 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 ($|lat_r|\mathbf{D}_{rt}$) for two reasons. First, it allows us to account for differences on how productively a region in terms of its distance to the equator can use an adopted R&D policy instrument. 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 policy instruments across regions.

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

[7] Coelli et al. (2016) provide evidence that patenting can be understood as innovation as they find an increasing citation rate of registered patents over time.

specific tax revenues generated from such investments, it is not possible to validate a structural form of the associated nonlinear relationship. From this perspective, it appears customary to resort to a reduced-form nexus between the instruments and patenting and consider treatment effects of the instruments based on this reduced-form nexus by embedding it in the structure of the general equilibrium model.

Reformulating a stochastic version of (18) in terms of an exponential-family-model specification, we arrive at

$$Patents_{rt} = \exp(\beta_0 + \frac{1}{\tilde{\xi}} \log \tilde{\bar{L}}_{rt} + \frac{1}{\tilde{\xi}} \log h_{rt} + \tilde{\varepsilon}_{rt}) \tag{20}$$

where $\tilde{\bar{L}}_{rt} = \gamma_1/\xi\nu[\mu + \gamma_1/\xi]\bar{L}_{rt}$ and $\tilde{\varepsilon}_{rt}$ is the error term. As the location decision of individuals may be endogenous to innovation potential of the region, we instrument $\tilde{\bar{L}}_{rt}$ with the unweighted average trade costs through a control function approach. After substituting $\log(h_{rt})$ with $(\mathbf{D}_{rt}\beta + |lat_{rt}|\mathbf{D}_{rt}\gamma)$ according to (19), we estimate (20) to obtain the parameter estimates $\{\widehat{\beta}_0, \widehat{1/\tilde{\xi}}, \widehat{\beta}, \widehat{\gamma}\}$ using a cross section of the data for the baseline year $t = 2005$ and negative binomial regression.[8] Given the parameter estimates we obtain an estimate of h_{rt} for each region r in 2005 as

$$\widehat{h}_{rt} = \exp(\mathbf{D}_{rt}\widehat{\beta} + |lat_r|\mathbf{D}_{rt}\widehat{\gamma}) \tag{21}$$

and $\widehat{h}_{rt}$ is kept constant for each period in the general-equilibrium analysis.

In Table 3, we summarize all variables which inform this procedure. The table is organized in three vertical blocks: the one at the top provides information on patents; the one in the center summarizes moments of the land and population distribution as well as $\log(\tilde{\bar{L}}_{rt})$ which combines the two; and the block at the bottom summarizes the elements of $\mathbf{D}_{rt}$ as well as $|lat_r|$ used in $|lat_r|\mathbf{D}_{rt}$, underlying the parametrization of h_{rt}.

– Table 3 about here –

The figures on patent registrations at the top of Table 3 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 for the average year in 2000-2010. While for the average year zero-registration events are scarcer than for a single year such as 2005, the average value of patents

[8] The negative binomial model accommodates the present over-dispersion of the data, whereby the variance in patents increases with the conditional mean.

is smaller than in the center of the time interval due to the surge of patenting. Moreover, the respective figures suggest that inventions are more dispersed than applications (i.e., applications are more concentrated). This pattern is shown in higher first as well as second moments of patent applications. The latter shows also in a higher frequency of zeros in the applications data than the inventions data, which is not obvious from the table.

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_{rt}. The respective indicators – all of which are measured in 2005 – suggest that almost all regions r are exposed to some type of R&D-policy instrument, as their effective tax rate on R&D-related profits is lower than the one on non-R&D-related profits ($Deatrrd_r$). Moreover, more than two-thirds of the regions operate under a regime with tax credits ($Dtaxcredit_r$). Other R&D-policy instruments are used much less frequently (by fewer countries or by countries with not very fine-grained regions). For example, a grants system is applied in only about eight percent of the regions, super deductions in only about five percent of the regions, and deductions, tax holidays and patent boxes 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 4.

– Table 4 about here –

In Table 4, we report on marginal effects of the covariates in (20) for four specifications. The ones numbered (1)-(2) pertain to inventor-based patents per region the average year in 2000-2010 while the ones numbered (3)-(4) pertain to applicant-based ones.[9] In columns (1) and (3) we set $\gamma = 0$ (interaction terms with $|lat_r|$ are excluded) while we abandon this restriction in columns (2) and (4). Apart from marginal effects of $\tilde{\tilde{L}}_{rt}$ as well as individual elements in $\mathbf{D}_{rt}$ we report the model fit through the correlation of the data with the model prediction on $\log(h_{rt})$ as well as the number of observations (regions) used.

The results suggest that more densely populated regions (i.e., the ones with higher values of $\tilde{\tilde{L}}_{rt}$) display higher counts of patent inventions as well as applications. Tax holidays ($Dtaxholidays_r$) and grants ($Dgrants_r$) tend to raise patent counts, while patent boxes ($Dpatentbox_r$; a back-end incentive which primarily promotes the ownership but not the invention of patents) reduce patent registrations. A reduction of the effective average tax rate on R&D-investment-based relative to

[9]We choose patent registrations for the average year in 2000-2010 as a dependent variable in all specifications reported in Table 4 to account for the time it takes to get a patent filing granted.

other investments benefits patent registrations ($Deatrrd_r$; in particular, for inventions).[10] Also regular deductions ($Ddeduc_r$) of R&D investments from profits display a positive effect on patent registrations. Overall, the explanatory power of the model is relatively high, as can be seen from the correlation coefficient between $\log(Patents_{rt})$ and $\log(\widehat{Patents}_{rt})$ in the table.[11] In what follows, we will use the specification in column (2) as the preferred model, since the explanatory power is relatively high and there is variation in the marginal effect of policy instruments across regions.

– Figure 2 about here –

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

For the estimation of the R&D worker-specific productivity shifter, h_{rt}, we assumed $\phi_{rt}^{\xi} = Patents_{rt}^{\tilde{\xi}}$ in order to link the model to our available data. Given all model parameters presented in Table 1, we can simulate the model according to the five-step-procedure that is described in Appendix A.

– Figure 3 about here –

Figure 3 shows that our model predictions for ϕ_{rt} match well the observed inventor-based patent registrations $Patents_{rt} > 0$ in each region for the year 2010. We report a correlation coefficient for observed and predicted innovation output of 0.864.

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 are abandoned. Effectively, that means in the counterfactual we set the R&D worker-specific productivity shifter equal to one in all regions r, i.e., $h_{rt}^c = 1$, $\forall r \in S$. We

[10] Notice that $Deatrrd_{rt}$ is coded as a nonlinear combination of the other front-end R&D-policy instruments (i.e., the instruments except $Dpatentbox_{rt}$), as it takes on the maximum value for each rt across all the respective instruments. Hence, whenever there is any front-end instrument in place, $Deatrrd_{rt}$ is unity. This shows also in the large number of regions for which $Deatrrd_{rt}$ is unity. However, one consequence of this design is that there is some collinearity between $Deatrrd_{rt}$ and the other front-end R&D-policy instruments, which reduces the degree to which their differential impact on patenting can be estimated.

[11] As discussed in the footnote of Table 4, the overdispersion parameter involved in the negative binomial model is statistically significant in all estimated models. Hence, the data reject an analysis by the poisson estimator.

split the analysis in two parts. First, we investigate how economic outcomes react in response to abandoning all incentives towards innovation and distinguish between policy-adopting and 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.

4.1 Economic Outcomes and R&D-policy Instruments

In Figure 4 we display the variation in long-run (period-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 4 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 (period-100) long-run changes are quite substantial: some regions gain about 10 percent in population while others loose more than 20 percent due to the hypothetical policy change in the long run; the effects on overall productivity are detrimental throughout and even larger than the population changes; also the welfare changes are negative throughout and almost as large to the productivity changes. Note that the distribution of log changes in population levels does not integrate to one. Given the logarithmic transformation of the displayed population change, this implies that few largely populated places gain and many less populated places loose from the hypothetical abolishment of all R&D policy indicators. Lastly, Table 6 presents moments of the real GDP growth rate in the short- (T=10), medium- (T=50), and log-run (T=100). The table shows that regions converge towards a model-induced balanced growth path, as the dispersion of growth rates decreases with time.

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

In Figure 5, we focus on the welfare changes as in the third panel of Figure 4 and plot them against the size of the direct treatment changes – i.e., the change in h_{rt} induced by abolishing all R&D policy instruments together, which is nothing else than $-\log \widehat{h}_{rt}$. 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 red color 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 5.

– Figures 5 and 6 about here –

In Figure 6, we shed further light on potentially important mediators of the general-equilibrium treatment effect on welfare changes. While Figure 5 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 6, 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, Asia and Oceania (incl. 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. The right panel in Figure 6 reveals a negative relationship between remoteness and the welfare change (i.e., more remote regions loose less on welfare from the global abolishment of R&D-promoting policies) for regions in North America, Oceania (incl. Australia), and also Africa, while for regions on other continents this relationship tends to be positive.

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 in terms of registered patent inventions (and, alternatively, applications) 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 a region in terms of its distance to the equator 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 instruments benefit from their use abroad due to technology spillovers. However, adopting countries and regions benefit significantly more strongly. Also, the long-run relocation effects are substantial: some regions gain about 10 percent in population while others loose more than 20 percent due to a hypothetical abolishment of all R&D policy instruments. This is mainly due a loss in competition for workers from otherwise less attractive regions, which could compete for mobile workers through the use of R&D policy instruments.

Furthermore, the quantitative analysis suggests that particularly regions with high amenities and a low degree of transport remoteness tend to benefit from such policy instruments, as they are able to attract labor which is key not only for production but also innovation, and as they can generate a high return on innovations to firms through their greater global market potential. This result is especially true for regions in North America and Oceania, whereas the effect is less predominant in Europe or Asia.

References

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

Baumann, M., B. Knoll, and N. Riedel (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.

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

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

Coelli F., A. Moxnes and K. H. Ulltveit-Moe (2016): "Better, Faster, Stronger: Global Innovation and Trade Liberalization," NBER Working Papers 22647, National Bureau of Economic Research, Inc.

Desmet, K., D. K. Nagy, and E. Rossi-Hansberg (2017): "The Geography of Development", *Journal of Political Economy* (forthcoming).

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. (2017): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure", *American Economic Review (forthcoming).*

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

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

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

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. (2015): "Commuting, Migration and Local Employment Elasticities", *NBER Working Paper #21706.*

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

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

Warda, J. (2002): "Measuring the value of R&D tax treatment in OECD countries", *OECD Science, Technology and Industry Review, Special Issue on New Science and Technology Indicators*, 27, OECD, Paris.

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.

Table 1: Calibration Overview

1. Land Endowments		
G_r	Extract land mass for each region.	Arc GIS Software
	(G_r is normalized by $\frac{1}{S}\sum_{r=1}^{S} G_r$)	
2. Preferences		
$\sigma = 4$	Elasticity of substitution.	Bernard et al. (2003)
$\lambda = 0.60$	Relation between amenities and population.	Own estimation, section 3.3
$\bar{a}_{rt}$	Exogenous amenity attribute.	Own estimation, section 3.3
$\Omega = 0.5$	Elasticity of migration flows w.r.t. income.	Monte et al. (2015)
3. 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)
$\gamma_1 = 0.273$	Elasticity of tomorrow's productivity	Own estimation, section 3.4
	w.r.t. today's innovation.	
4. Evolution of productivity		
$\gamma_2 = 0.9897$	Elasticity of tomorrow's productivity	Own estimation, section 3.4
	w.r.t. today's productivity.	
$\xi = 125$	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. (2017)
5. Transport Costs		
	Based on Allen and Arkolakis (2014) and Fast Marching Algorithm.	
6. Other Trade Costs		
$tariffs_{rs}$	Weighted applied import tariffs for manufactures	World Development Indicator (WDI)
$\rho = 0.078$	Elasticity of trade w.r.t linguistic proximity.	Melitz and Toubal (2014)
7. Productivity-shifter for R&D Workers		
h_{rt}	Estimation using binary R&D policy indicators	Own estimation, section 3.5.
	$\widehat{h}_{rt} = \exp(\mathbf{D}_{rt}\widehat{\beta} + \lvert lat_r \rvert \mathbf{D}_{rt}\widehat{\gamma})$	

Table 2: Amenity Parameter Estimation

Regressor	Parameter	Coeff. (Std. err.)	**Moments of** $\widehat{a}_r \equiv exp(\widehat{const.} + \widehat{\varepsilon}^a_r)$				
First Stage: Dep. Var. $\log(\bar{L}_{rt})$			Mean			Std. Dev.	
$log(R_r)$	ρ_1	-0.473*** (0.014)	63,314			373,515	
Second Stage: Dep. Var.		$\log(a_{rt})$	5%	10%	50%	90%	95%
$\widehat{\log(\bar{L}_r)}$	$-\lambda$	-0.600*** (0.033)	203.4	461.9	6,563.9	81,232.7	165,393.5
#obs 5,633							

Table 3: Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.
Patents per norm. unit of land				
$Patents_{rt}$ (inv) 2005	800.98	5,366.5	0	205,728.4
$Patents_{rt}$ (inv) avg 2000-2010	775.97	5104.1	0	196,549.3
$Patents_{rt}$ (app) 2005	1,154.44	13,781.7	0	617,404.1
$Patents_{rt}$ (app) avg 2000-2010	1,124.27	13,563.3	0	610,453.4
G_r	1	11.83	1.9e-04	624.27
$\bar{L}_{rt}G_r$	1,108,491	7,637,692	5	2.17e+08
$\log(\tilde{\bar{L}}_{rt})$	9.89	2.01	-0.428	16.44
R&D-policy indicators				
Dtaxcredit_{rt}	0.715	0.452	0	1
Dsuperd_{rt}	0.053	0.224	0	1
Dtaxholiday_{rt}	0.023	0.151	0	1
Dgrants_{rt}	0.081	0.273	0	1
Dpatentbox_{rt}	0.022	0.147	0	1
Ddeduc_{rt}	0.029	0.169	0	1
Deatrrd_{rt}	0.982	0.131	0	1
$\lvert lat_r \rvert$	40.205	9.583	0.2	74.728

Notes: $Patents_{rt}(inv)$ refers to a regional denomination of patents according to the residence of inventors (inv). $Patents_{rt}(app)$ refers to a regional denomination of patents according to the residence of applicants (app).

Table 4: Estimation Results h_{rt} (Marginal Effects)

	(1)	(2)	(3)	(4)
	$Patents_{rt}$ (inv)	$Patents_{rt}$ (inv)	$Patents_{rt}$ (app)	$Patents_{rt}$ (app)
	avg 2000-2010	avg 2000-2010	avg 2000-2010	avg 2000-2010
$\log(\tilde{\tilde{L}}_{rt})$	1.254***	1.527***	1.197**	1.557***
	(0.245)	(0.257)	(0.548)	(0.481)
Dtaxcredit_{rt}	-0.084	0.218	0.312	0.777
	(0.601)	(0.593)	(0.828)	(0.758)
Dsuperd_{rt}	0.156	-0.148	0.096	-1.650**
	(0.772)	(0.865)	(0.963)	(0.700)
Dtaxholiday_{rt}	2.529**	0.987	3.167***	2.023**
	(1.198)	(0.857)	(1.200)	(1.148)
Dgrants_{rt}	1.336***	1.594	1.474**	1.944
	(0.394)	(2.017)	(0.577)	(2.259)
Dpatentbox_{rt}	-2.102*	-1.514**	-3.087***	-3.122***
	(1.168)	(0.754)	(0.747)	(0.711)
Ddeduc_{rt}	0.111	0.032	1.449*	0.928
	(0.333)	(0.312)	(0.871)	(0.902)
Deatrrd_{rt}	1.798**	2.206***	0.075	0.614
	(0.763)	(0.800)	(0.951)	(0.729)
# obs	5,633	5,633	5,633	5,633
$\log(\tilde{\tilde{L}}_{rt})$ instr.	YES	YES	YES	YES
$\lvert lat_r \rvert \mathbf{D}_{rt}$	NO	YES	NO	YES
overall fit	0.42	0.51	0.37	0.30

Notes: Robust and county-level-clustered standard errors in parentheses. Across all estimated models the dispersion parameter for the negative binominal model against the poisson model is positive and statistically significant at the 1 percent level.

Table 5: R&D Policy Instruments in 2005

R&D Policy Instrument	Description	Countries (in 2005)
Dtaxcredit_{rt}	Tax credits on R&D investments	Austria, Canada, China, France, Ireland, Japan, Mexico, Netherlands, Norway, Portugal, South Korea, Spain, Taiwan, US, Venezuela.
Dtaxholiday_{rt}	Tax holidays for firms with R&D investments.	France, Malaysia, Singapore, Switzerland.
Dgrants_{rt}	R&D investment related grants from the government.	Germany, Hungary, Ireland, Israel.
Dpatentbox_{rt}	(Partial) exemption of returns on R&D investments.	France, Hungary.
Ddeduc_{rt}	Any form of deductions on R&D investments.	Australia, Belgium, Ireland, Japan, South Korea.
Dsuperd_{rt}	Super deductions of more than 100% for R&D investments.	Australia, China, Czech Republic, Hungary, India, Malaysia, Malta, Puerto Rico, Singapore, UK.
Deatrrd_{rt}	Effective average tax rate is lower on returns on R&D investments than on other investments.	114 of 213 countries in the data.

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 6: Moments of Real GDP Growth

Period	Min	Max	Mean	Std
T=10	0.010	0.050	0.033	0.003
T=50	0.015	0.043	0.032	0.002
T=100	0.02	0.038	0.031	0.001

Figure 1: REGPAT Regions

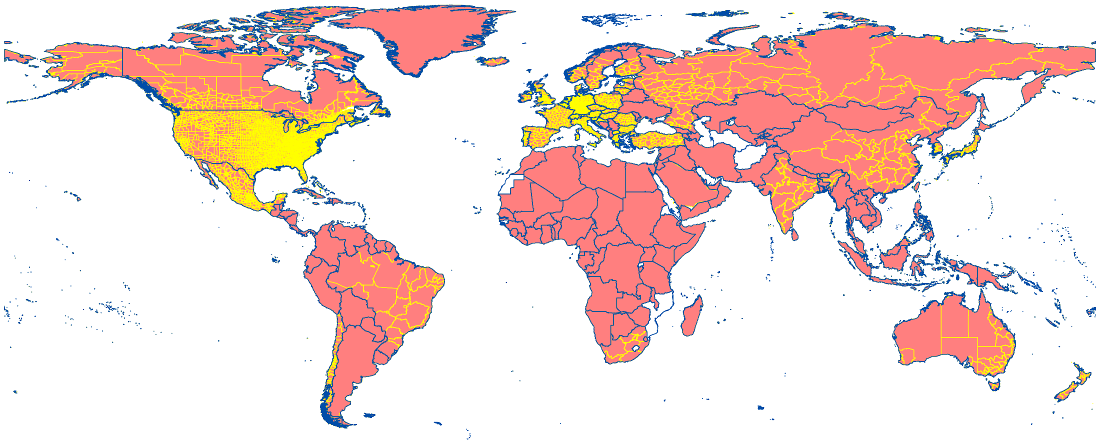


Figure 2: Kernel Density $\hat{h}_{rt}$

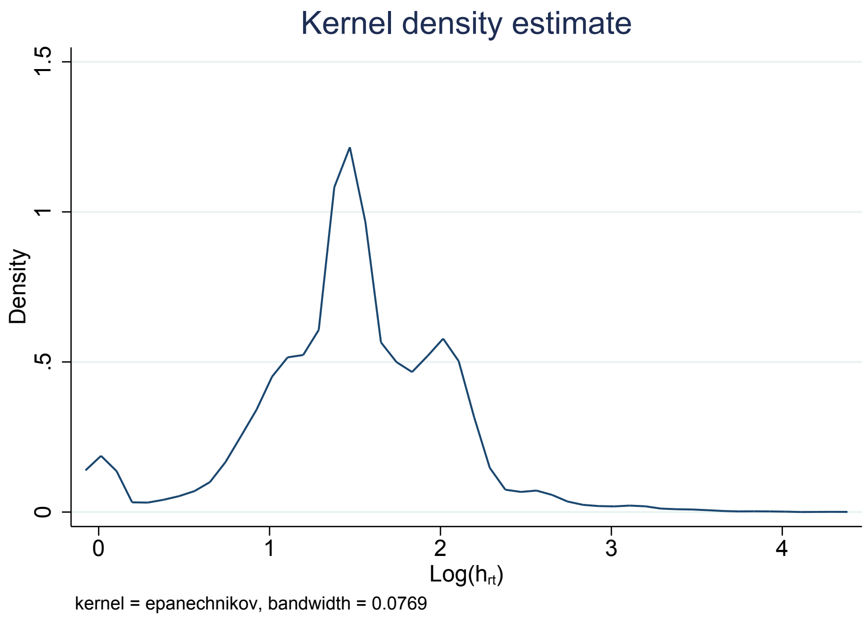


Figure 3: Correlation Innovation and Patents in 2010

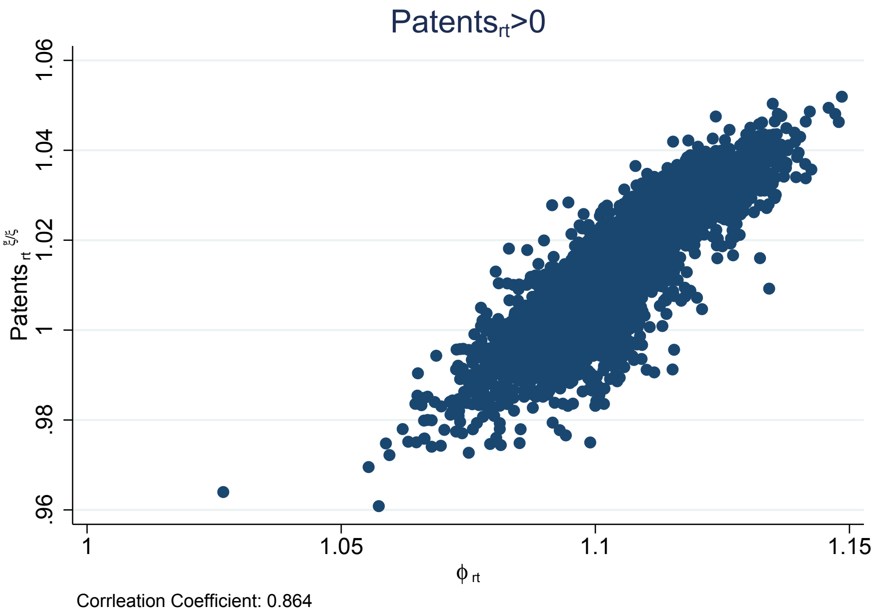


Figure 4: Density Estimates of Counterfactual Changes, T=100

(Population, Productivity, Welfare)

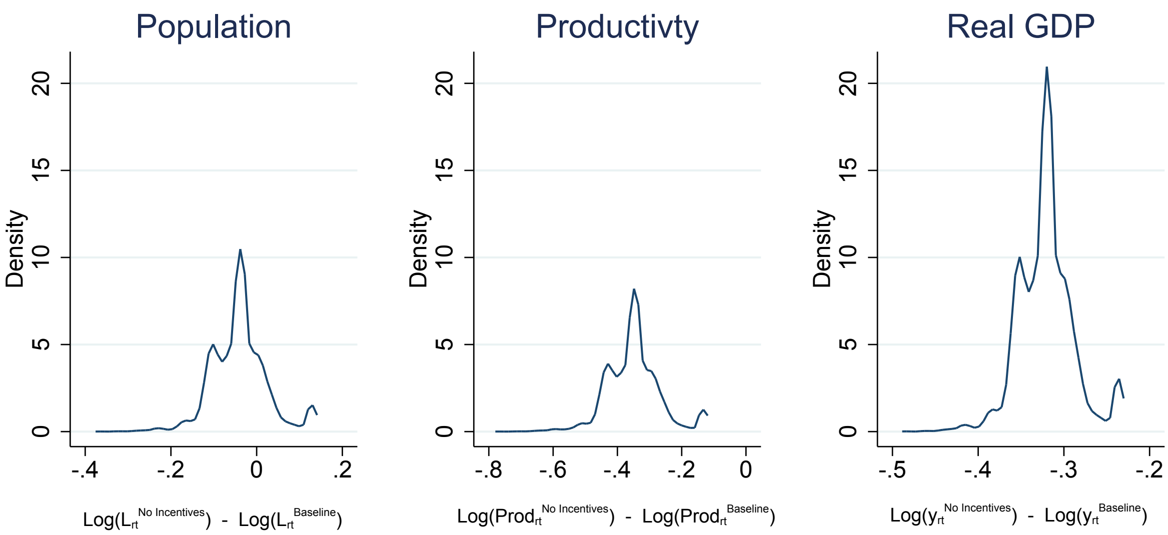


Figure 5: Welfare Change at T=100 and Changes in h_{rt}

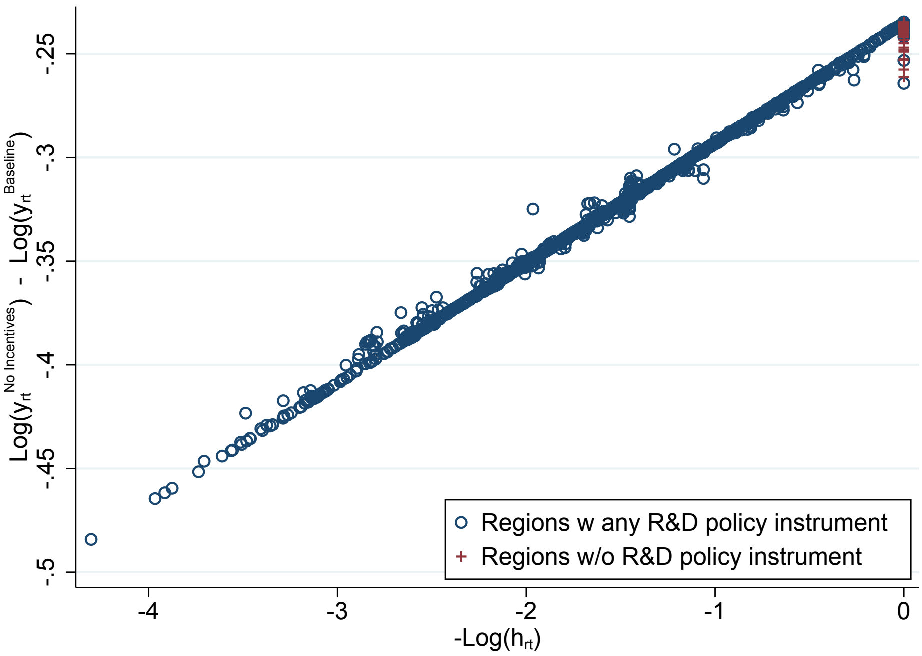
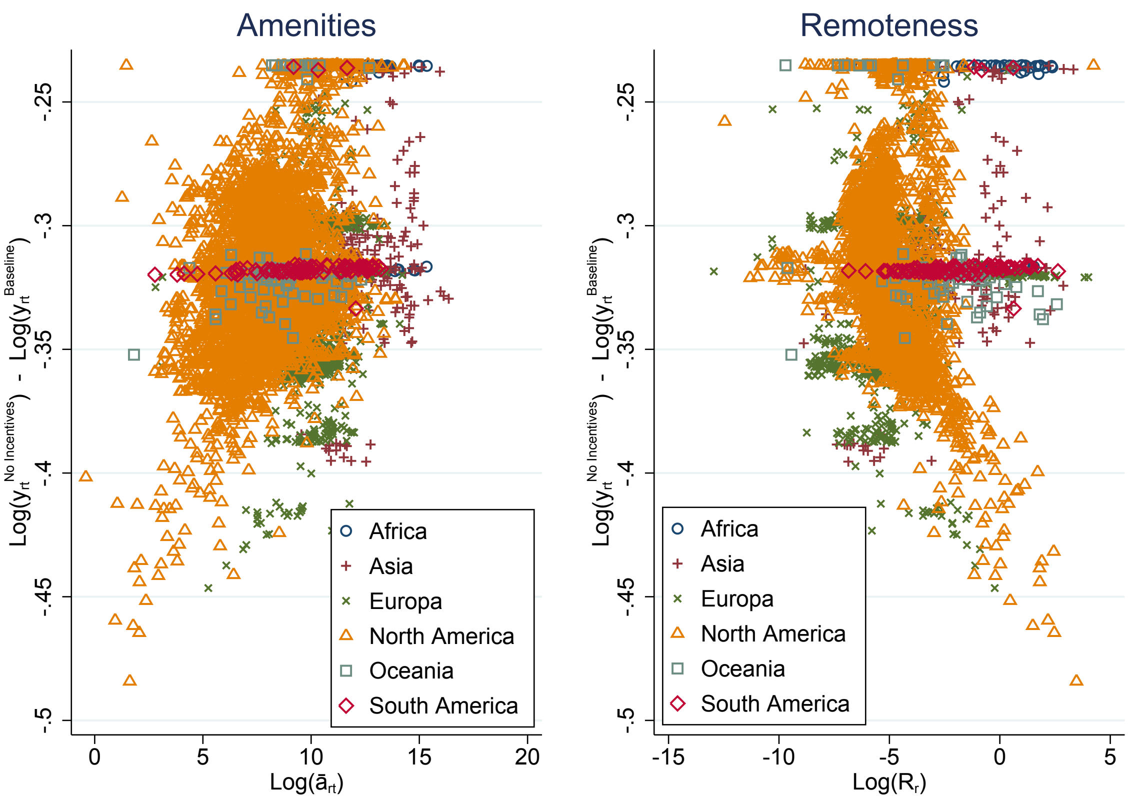


Figure 6: Welfare Change at T=100 and Amenity/Remoteness Levels (by continents)



Appendix

A Simulation of the Model

The simulation of the model proceeds as follows.

First, we identify the initial efficiency distribution τ_0. 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 τ_{rt}. Then, after defining

$$\Pi_{kt} \equiv \bar{L}_{kt}^{\iota} G_k w_{kt}^{-\theta} h_{kt}^{-\theta\gamma_1/\xi} \tau_{kt} \zeta_{ks}^{-\theta} \quad \text{and} \quad \iota \equiv \alpha - (1 - \mu - \gamma_1/\xi)\theta$$

$$\tau_{rt} = \frac{\bar{L}_{rt}^{1-\iota} G_r w_{rt}^{1+\theta} h_{rt}^{-\theta\gamma_1/\xi}}{\int\limits_S w_{st} \bar{L}_{st} G_s \zeta_{rs}^{-\theta} \left[\int_S \Pi_{kt} dk \right]^{-1} ds} \tag{22}$$

Now, we numerically solve for τ_0 by applying an iterative procedure[12] and using observed levels of population densities, $\bar{L}_0$ and wages, w_0 for the benchmark year 2005. We use population levels from SEDAC and wage levels 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 $\bar{L}_{rt}$.

Second, we identify the initial distribution of amenities a_0. 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}. Hence,

$$a_{rt} = \left(\frac{\bar{L}_{rt} G_r}{\bar{L}} \right)^{\Omega} \frac{1}{w_{rt}} \left[\int_S (a_{kt} w_{kt})^{1/\Omega} \left(\int_S \Pi_{st} ds \right)^{1/\Omega\theta} dk \right]^{\Omega} \left[\int_S \Pi_{kt} dk \right]^{-1/\theta} \tag{23}$$

Again, we apply an iterative procedure to solve for the initial amenity distribution a_0 in 2005 using observed population densities and wages. With a_{rt} we estimate the exogenous region-specific amenity-shock $\bar{a}_r t$ as described in section 3.3.

Third, having identified the initial efficiency distribution τ_0 and the region-specific amenity shock $\bar{a}_r$, we can solve for the equilibrium wage and population density levels using (10) and

[12] We apply a standard contraction mapping procedure as it is described in Appendix B.7 in Desmet et al. (2017).

(11) in an iterative procedure. We manipulate (10) and (11) as described in section 2.4, then the equilibrium equations are

$$w_{rt} = \left[\frac{\tau_{rt}\bar{L}_{rt}^{\iota-1}h_{rt}^{\theta\gamma_1/\xi}}{G_r} \int_S \frac{w_{st}\bar{L}_{st}G_s\zeta_{rs}^{-\theta}}{\int_S \Pi_{kt}dk} ds \right]^{1/(1+\theta)} \tag{24}$$

$$\bar{L}_{rt} = \left(\frac{\bar{L}}{G_r}\right)^{\frac{\Omega}{\Omega+\lambda}} (\bar{a}_r w_{rt})^{\frac{1}{\Omega+\lambda}} \frac{\left[\int_S \Pi_{kt}dk\right]^{\frac{1}{\theta(\Omega+\lambda)}}}{\left[\int_S \left(\bar{a}_k w_{kt}\bar{L}_{kt}^{-\lambda}\right)^{1/\Omega}\left(\int_S \Pi_{st}ds\right)^{1/\Omega\theta} dk\right]^{\frac{\Omega}{\theta(\Omega+\lambda)}}} \tag{25}$$

Fourth, once we determined w and $\bar{L}$ for a given period t, we can solve for social welfare in every region, u_{rt}, using (8) and for innovation levels, ϕ_{rt}, using (6).

Lastly, we update the efficiency level, τ, and rerun the simulation for every given period t.

B Equilibrium: Existence and Uniqueness

The uniqueness condition in (12) can be derived along the lines of Desmet et al. (2017) (see 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,

$$P_{rt} = \kappa_0 \left[\int_S \tau_{st}\bar{L}_{st}^{\alpha-(1-\mu-\gamma_1/\xi)} w_{st}^{-\theta}\zeta_{rs}^{-\theta}h_{st}^{\theta\gamma_1/\xi}ds\right]^{-\frac{1}{\theta}} \tag{26}$$

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 (26) into (8) gives

$$\left[\frac{\bar{a}_r}{u_{rt}}\right]^{-\theta} \bar{L}_{rt}^{\theta\lambda}w_{rt}^{-\theta} = \kappa_1 \int_S \tau_{st}\bar{L}_{st}^{\alpha-(1-\mu-\gamma_1/\xi)\theta} w_{st}^{-\theta}\zeta_{rs}^{-\theta}h_{st}^{\theta\gamma_1/\xi}ds \tag{27}$$

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

$$w_{rt}G_r\bar{L}_{rt} = \bar{p}^{-\theta}\int_S T_{rt}[o_{rt}\zeta_{sr}]^{-\theta}P_{st}^{-\theta}w_{st}G_s\bar{L}_{st}ds \tag{28}$$

Substituting unit costs (3) and $T_{rt} = \tau_{rt}\bar{L}_{rt}^{\alpha}$ into the previous equation yields

$$\tau_{rt}^{-1}w_{rt}^{1+\theta}G_r h_{rt}^{-\frac{\theta\gamma_1}{\xi}}\bar{L}_{rt}^{1-(\alpha-(1-\mu-\gamma_1/\xi)\theta)} = \kappa_1\int_S\left[\frac{\bar{a}_s}{u_{st}}\right]^{\theta}\zeta_{sr}^{-\theta}w_{st}^{1+\theta}G_s\bar{L}_{st}^{1-\lambda\theta}ds \tag{29}$$

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 (27) and (29):

$$\bar{f}_r = \frac{\tau_{rt}^{-1} w_{rt}^{1+\theta} G_r h_{rt}^{-\frac{\theta\gamma_1}{\xi}} \bar{L}_{rt}^{1-(\alpha-(1-\mu-\gamma_1/\xi)\theta)}}{\left[\frac{\bar{a}_r}{u_{rt}}\right]^{-\theta} \bar{L}_{rt}^{\theta\lambda} w_{rt}^{-\theta}} \tag{30}$$

Equivalently, $\bar{f}_r$ also equals the RHS's of (27) and (29) that is

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

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

$$\bar{f}_r = \frac{\int_S \bar{f}_s^{-\lambda} \bar{\bar{f}}_s r \quad ds}{\int_S \bar{f}_s^{-(1+\lambda)} \bar{\bar{f}}_{sr} \quad ds} \tag{32}$$

where

$$\bar{\bar{f}}_{sr} = \left[\frac{\bar{a}_s}{u_{st}}\right]^{\theta(1+\lambda)} \tau_{st}^{-\lambda} G_s^{1+\lambda} \, \zeta_{sr}^{-\theta} h_{st}^{-\lambda\frac{\theta\gamma_1}{\xi}} w_{st}^{1+\theta+(1+2\theta)\lambda} \tag{33}$$

Rewrite (32) as

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

Then, changing the notation to

$$\bar{g}_r = \bar{f}_r^{-\lambda} \quad \text{and} \quad \bar{\bar{g}}_r = \bar{f}_r^{-(1+\lambda)} \tag{35}$$

and rewrite both as follows

$$\bar{g}_r = \int_S \bar{\bar{\bar{f}}}_r \bar{\bar{f}}_{sr} \bar{g}_s \ ds \quad \text{and} \quad \bar{\bar{g}}_r = \int_S \bar{\bar{\bar{f}}}_r \bar{\bar{f}}_{sr} \bar{\bar{g}}_s \ ds \tag{36}$$

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

$$x_r = \int_S K_{rs} \ x_s \ ds. \tag{37}$$

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 (37). Hence,

$$\bar{g}_r = \varpi \bar{\bar{g}}_r \quad \Rightarrow \bar{f}_r^{-\lambda} = \varpi \bar{f}_r^{-(1+\lambda)} \quad \Rightarrow \bar{f}_r = \varpi \tag{38}$$

where ϖ is a constant. Therefore, we have

$$\frac{\tau_{rt}^{-1} w_{rt}^{1+\theta} G_r h_{rt}^{-\frac{\theta\gamma_1}{\xi}} \bar{L}_{rt}^{1-(\alpha-(1-\mu-\gamma_1/\xi)\theta)}}{\left[\frac{\bar{a}_r}{u_{rt}}\right]^{-\theta} \bar{L}_{rt}^{\theta\lambda} w_{rt}^{-\theta}} = \varpi \tag{39}$$

and solving for w_{rt} gives

$$w_{rt} = \bar{w} \left[\frac{\bar{a}_r}{u_{rt}}\right]^{-\frac{\theta}{1+2\theta}} \tau_{rt}^{\frac{1}{1+2\theta}} G_r^{-\frac{1}{1+2\theta}} \bar{L}_{rt}^{\frac{\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta}{1+2\theta}} h_{rt}^{\frac{\theta\gamma_1/\xi}{1+2\theta}} \tag{40}$$

where $\bar{w} = \varpi^{\frac{1}{1+2\theta}}$. Substituting (40) into (27) gives

$$\begin{aligned} &\left[\frac{\bar{a}_r}{u_{rt}}\right]^{-\frac{\theta(1+\theta)}{1+2\theta}} \tau_{rt}^{-\frac{\theta}{1+2\theta}} G_r^{\frac{\theta}{1+2\theta}} \bar{L}_{rt}^{\lambda\theta-\frac{\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]} h_{rt}^{-\frac{\theta(\theta\gamma_1/\xi)}{1+2\theta}} \\ &\qquad = \kappa_1 \int_S \left[\frac{\bar{a}_s}{u_{st}}\right]^{\frac{\theta^2}{1+2\theta}} \tau_{st}^{\frac{1+\theta}{1+2\theta}} G_s^{\frac{\theta}{1+2\theta}} \zeta_{rs}^{-\theta} \bar{L}_{st}^{1-\lambda\theta+\frac{1+\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]} h_{st}^{\frac{(1+\theta)(\theta\gamma_1/\xi)}{1+2\theta}} \, ds \end{aligned} \tag{41}$$

Inserting (10) into (41) gives

$$\begin{aligned} &\bar{B}_{rt} \ \hat{u}_{rt}^{\frac{1}{\Omega}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]+\frac{\theta(1+\theta)}{1+2\theta}} \\ &\qquad = \kappa_1 \int_S \hat{u}_{st}^{\frac{1}{\Omega}\left[1-\lambda\theta+\frac{1+\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta\right]\right]-\frac{\theta^2}{1+2\theta}} \bar{\bar{B}}_{st} \zeta_{rs}^{-\theta} \ ds \end{aligned} \tag{42}$$

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

$$\hat{u}_{rt} = u_{rt} \left[\frac{\bar{L}}{\int_S u_{kt}^{1/\Omega}}\right]^{\Omega\left[1-\frac{\theta}{\frac{1}{\Omega}\left[\left[\lambda+(1-\mu)-\frac{\gamma_1}{\xi}\right]\theta-\alpha\right]+\theta}\right]} \tag{43}$$

Rewrite (42) as

$$\bar{B}_r f_r^{\tilde{\gamma}_1} = \kappa_1 \int_S \bar{\bar{B}}_s \zeta_{rs}^{-\theta} f_s^{\tilde{\gamma}_2} \ ds \tag{44}$$

and apply Theorem 2.19 in Zabreyko et al. (1975), then the solution $f_{(\cdot)}$ to equation (44) 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

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

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

$$\frac{\tau_{rt+1}}{\tau_{rt}} = \phi_{rt}^{\theta\gamma_1}\left[\frac{1}{S}\int_S \frac{\tau_{st}}{\tau_{rt}} ds\right]^{1-\gamma_2} \tag{45}$$

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,

$$\underbrace{\frac{\frac{\tau_{rt+1}}{\tau_{rt}}}{\frac{\tau_{st+1}}{\tau_{st}}}}_{=1} = \left[\frac{\tau_{st}}{\tau_{rt}}\right]^{1-\gamma_2}\left[\frac{\phi_{rt}}{\phi_{st}}\right]^{\theta\gamma_1}\underbrace{\left[\frac{\int_S \tau_{st} ds}{\int_S \tau_{rt} dr}\right]^{1-\gamma_2}}_{=1} \quad \Rightarrow \quad \frac{\tau_{st}}{\tau_{rt}} = \left[\frac{\phi_{st}}{\phi_{rt}}\right]^{\frac{\theta\gamma_1}{1-\gamma_2}} = \left[\frac{\bar{L}_s h_s}{\bar{L}_r h_r}\right]^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}} \tag{46}$$

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

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

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

$$\int_S G_s \bar{L}_s ds = \bar{L} = \tau_{rt}^{-\frac{(1-\gamma_2)\xi}{\theta\gamma_1}} \bar{L}_r h_r \int_S G_s \tau_{st}^{\frac{(1-\gamma_2)\xi}{\theta\gamma_1}} h_s^{-1} ds \quad \Rightarrow \quad \tau_{rt} = \tilde{\kappa}_t (h_r \bar{L}_r)^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}} \tag{47}$$

where $\tilde{\kappa}_t$ depends on time but not on location. Take the last equation and substitute it into (41) such that

$$\left[\frac{\bar{a}_r}{u_{rt}}\right]^{-\frac{\theta(1+\theta)}{1+2\theta}} G_r^{\frac{\theta}{1+2\theta}} \bar{L}_r^{\lambda\theta - \frac{\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]} h_r^{-\frac{\theta(\theta\gamma_1/\xi)}{1+2\theta}(1+\frac{1}{1-\gamma_2})}$$

$$= \kappa_1 \tilde{\kappa}_t \int_S \left[\frac{\bar{a}_s}{u_{st}}\right]^{\frac{\theta^2}{1+2\theta}} G_s^{\frac{\theta}{1+2\theta}} \zeta_{rs}^{-\theta} \bar{L}_s^{1 - \lambda\theta + \frac{1+\theta}{1+2\theta}\left[\alpha - 1 + \left[\lambda + \frac{\gamma_1}{\xi} - [1-\mu]\right]\theta + \frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]} h_s^{\frac{(1+\theta)(\theta\gamma_1/\xi)}{1+2\theta}(1+\frac{1}{1-\gamma_2})} ds$$

$$(48)$$

Inserting (10) in (48) and rearranging conveniently, yields

$$\begin{aligned}
&\bar{D}_r \; \hat{\hat{u}}_{rt}^{\frac{1}{\Omega}\left[\lambda\theta-\frac{\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta+\frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]\right]+\frac{\theta(1+\theta)}{1+2\theta}} \\
&= \kappa_1\tilde{\kappa}_t \int_S \hat{\hat{u}}_{st}^{\frac{1}{\Omega}\left[\lambda\theta+\frac{1+\theta}{1+2\theta}\left[\alpha-1+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta+\frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]\right]-\frac{\theta^2}{1+2\theta}} \; \bar{\bar{D}}_s \; \zeta_{rs}^{-\theta} \; ds
\end{aligned} \tag{49}$$

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

$$\hat{\hat{u}}_{rt} = u_{rt} \left[\frac{\bar{L}}{\int_S u_{kt}^{1/\Omega}}\right]^{\Omega\left[1-\frac{\theta}{\frac{1}{\Omega}\left[\left[\lambda+(1-\mu)-\frac{\gamma_1}{\xi}\right]\theta-\alpha-\frac{\theta\gamma_1}{(1-\gamma_2)\xi}\right]+\theta}\right]} \tag{50}$$

Analogously to the existence and uniqueness proof in section B, we can rewrite (49) as

$$\bar{D}_r g_r^{\tilde{\bar{\gamma}}_1} = \kappa_1\tilde{\kappa}_t \int_S \bar{\bar{D}}_s \zeta_{rs}^{-\theta} g_s^{\tilde{\bar{\gamma}}_2} \; ds. \tag{51}$$

According to Theorem 2.19 in Zabreyko et al. (1975) $g_{(\cdot)}$ is a solution to the system of equations in (51) that is unique if $\left|\frac{\tilde{\bar{\gamma}}_2}{\tilde{\bar{\gamma}}_1}\right| \leq 1$. 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 (14) 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 (40) as follows

$$\tau_{rt} = \bar{w}^{-(1+2\theta)} \left[\frac{\bar{a}_r}{u_{rt}}\right]^{\theta} w_r^{1+2\theta} G_r \bar{L}_r^{\frac{1-\alpha+\left[\lambda+\frac{\gamma_1}{\xi}-[1-\mu]\right]\theta}{1+2\theta}} h_r^{-\frac{\theta\gamma_1}{\xi}} \tag{52}$$

Substituting the previous equation into (47) and solving for u_{rt} gives

$$u_{rt} = \tilde{\kappa}_t^{\frac{1}{\theta}} E_r \tag{53}$$

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

$$\frac{u_{rt+1}}{u_{rt}} = \left(\frac{\tilde{\kappa}_{t+1}}{\tilde{\kappa}_t}\right)^{\frac{1}{\theta}} = \left(\frac{\tau_{rt+1}}{\tau_{rt}}\right)^{\frac{1}{\theta}} \tag{54}$$

where the last equality follows from (47). From (45) and (46) we know

$$\frac{\tau_{rt+1}}{\tau_{rt}} = \phi_{rt}^{\theta\gamma_1}\left[\frac{1}{S}\int_S \frac{\tau_{st}}{\tau_{rt}} ds\right]^{1-\gamma_2} = \left(\frac{\gamma_1/\nu}{\gamma_1+\mu\xi}\bar{L}_r h_r\right)^{\frac{\theta\gamma_1}{\xi}} \left[\frac{1}{S}\int_S \left(\frac{\bar{L}_s h_s}{\bar{L}_r h_r}\right)^{\frac{\theta\gamma_1}{(1-\gamma_2)\xi}} ds\right]^{1-\gamma_2} \tag{55}$$

Rearranging the previous equation and substituting it into (54) 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.

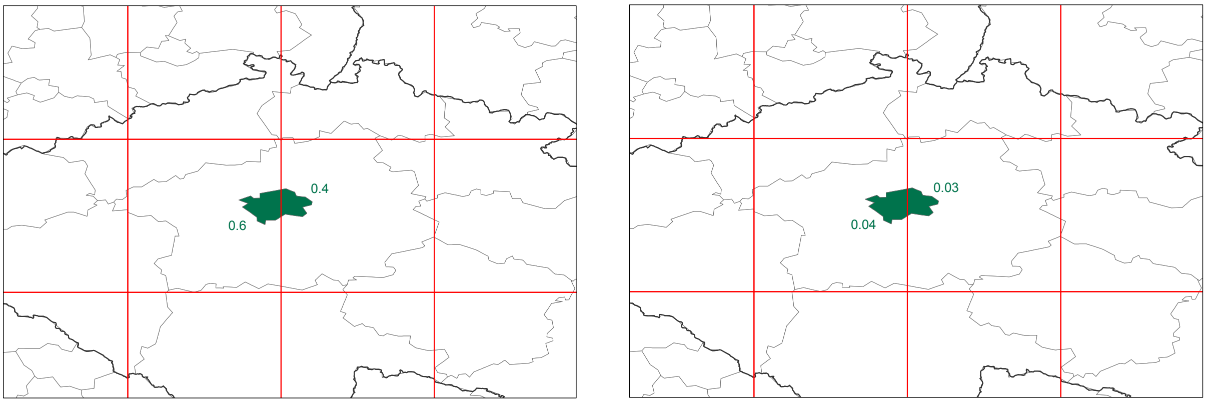
D.1 Population Data from SEDAC

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 7 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 7: 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. (2017). 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 7). 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

$$T_{rxr} = W'_{nxr} T_{nxn} W_{nxr} \tag{56}$$

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