Demographic and Geographic Determinants of Regional Physician Supply

(preliminary version)

Michael Kuhn* Carsten Ochsen†

June 5, 2009

Abstract

We provide a theoretical framework which allows us to analyse how physician supply at regional level depends on demographic (population size, age structure, mobility and fertility rates) and geographic determinants. Using regional data for Germany, we then examine econometrically what determines physician's local choice. We find that physician supply is related negatively to the population share of 60plus within rural areas and positively to population growth (due to migration or fertility). Spatial differences in terms of market tightness seem not to be important for the local choice. According to our results and due to the unfavorable age-structure of physicians we conclude that inequality in the regional distribution of physicians will increase in the future.

Keywords: age structure, physician supply, regional population ageing, regional migration, panel data, spatial model.

JEL classification: I11, J44, J10, R23, C33, C31

*Vienna Institute of Demography; Correspondence: Michael Kuhn,Vienna Institute of Demography, Wohllebengasse 12-14, A 1040 Vienna, Austria, e-mail: michael.kuhn@oeaw.ac.at

†University of Rostock; Correspondence: Carsten Ochsen, Department of Economics, University of Rostock, Ulmenstrasse 69, 18057 Rostock, Germany, e-mail: carsten.ochsen@uni-rostock.de

1 Introduction

Since the second half of the nineties it has been increasingly reported on a declining physician supply related to the number of inhabitants at regional level in Germany. Initially, this was limited to general physicians, but more and more this applies also to other physician specialisations. Estimates of physician density at regional level in the coming years underline that this observation in rural areas follows a trend.[1] On the other hand, however, has the number of physicians between 1990 and 2005 increased by about 30 % at the aggregate level. The aim of this study is to analyse the causes of the increasingly unequal distribution of physicians and, in particular, the role of demographic change in regional populations.

Population ageing is widely expected to come with an increased per capita demand for (ambulatory) physician services. Everything else equal regions with high population shares of old persons should then be particularly attractive on economic grounds for the location of physician practices and should therefore exhibit high physician to population ratios. However, cursory evidence suggests that this may not be the case. In Germany it appears that many rural regions with a high share of the elderly population are in particular danger of being under-doctored. Ageing may therefore lead to a widening gap in some regions between a high need for health care and a low supply that may well warrant policy intervention. Furthermore, there are concerns about growing regional disparities in the provision of health care. Generally, however, too little is known yet about how demographic, economic and spatial features of a region interact as determinants of physician supply.

This study uses panel data at regional level to examine the relationship between regional physician supply and its demographic and geographic determinants. Given that physicians' location choices are mostly long-term decisions, the intertemporal aspect implies that not only the current population structure will matter but also the expected population. In as far as they constitute measures of population change, we would therefore expect net population growth (births – deaths) and net migration to constitute significant determinants for physician supply besides the population and its age structure in of themselves.

Furthermore, we adopt the hypothesis that demographic and geographic characteristics do not determine in isolation the attractiveness of a region from a physician's perspective but interact in a particular way. More specifically, we posit that the high demand for services by elderly patients may be less attractive for a physician if it has to be served within a rural context. Long travel times and poor availability of public transport deter frail elderly patients from attending the physician's practice, implying a high demand

[1] See, for example, Andersen and Mühlbacher (2004) and Kopetsch (2004).

for home visits as compared to an urban context. Furthermore, due to the long travel times the provision of home visits is more costly for the physician in rural areas. Thus, a high demand from old populations may be served at a relatively low cost within an urban region but only at a high cost within a rural context. Given that the reimbursement system does not fully account for these differences, one should expect a positive correlation between age-structure and a measure of urbanity in its effect on physician supply.

Finally, we would expect that spatial interactions in terms of physician density are important. More specifically, our hypothesis is that the propensity to locate in a particular region is positively related to the physician density in neighbouring districts. This is because high densities of physicians in neighbouring regions are likely to be associated with strong competition. In turn, this would imply a greater relative attractiveness for the region under consideration.

Using annual data for 439 regions at the district level in Germany for the period 1995 to 2004 we examine if the share of 60+ in the population affects the physician density negatively in rural regions. We control for several regional characteristics which, according to the literature, are relevant for the geographical distribution of physician practices. In a first step we use a fixed effects panel estimator to analyze the effects of a change in the age structure of the population, population flows, and population shrinking. In addition, we apply a spatial dynamic panel data estimator, to find out if spatial interactions in physician density are of importance.

It is common to approximate physician location decisions with the physician density (i.e. physicians per capita) at the macro level. However, if ageing at the regional level is intensified by interregional migration, this measure could comprise a measurement error, e.g., because younger and older people exhibit differences in regional mobility. In fact it is observed empirically that younger people leave rural areas more often than older people. The number of habitants in these regions falls and, hence, the physician density increases for a given number of physicians. At the same time the share of older people increases and is therefore positively correlated with the physician density. Given that physicians' location choices are made for the long term, physician density comprises a measurement error that is expected to be positive. Therefore, we employ the number of physicians as dependent variable not subject to such a bias.

According to the results, physician supply is negatively related to the population share of 60plus within rural areas, while it is not significantly related to the share within urban regions. The effects of the population size are significant positive, which is consistent with theory. Hence, population ageing and shrinking affects physician geographic decisions as expected. Spatial interactions measured in contiguity effects are significantly related to the number of physicians in the local district. However, the effect is smaller in magnitude than the population effects. Since the speed of population ageing is much faster in rural regions, policy should provide incentives to move into these regions not only on equity grounds but also because a worsening provision of basic goods and services, such as health care, may reduce even further the development prospects (e.g. by way of job creation) for such regions. In addition, many rural regions in East Germany have a rural contiguity neighbor, which means that the benefit from the spatial effect disappears here.

2 Literature

Hingstman & Boon (1989) analyse the regional dispersion of primary health care professionals (GPs, dentists, physiotherapists, midwives and pharmacists) in the Netherlands. Using cross-sectional data, they examine the effect of local earning opportunities (average income, share of elderly, birth rate, population density) and local amenities (share of green belt, distance to nearest training centre) on the aggregated location choices of practicioners as measured by densities (i.e. number of practitioner per capita) within a region. Within a simple OLS regression, they identify a significant (at 5 percent) negative effect of the share of elderly on the densities of all practitioners (except midwives). The intuition for this turns on the features of the payment system. GPs and pharmacists are reimbursed a capitation for each publicly insured patient. This turns the elderly into relatively unprofitable patients as the relatively high treatment intensities have to be financed out of a fixed budget per patient. In contrast, dentists are reimbursed by fee for service. While this renders profitable the treatment of patients with high demand, for dentistry it is the young rather than the hold who require intensive treatment.

Kraft & v.d. Schulenburg (1986) employ cross sectional data from Switzerland to estimate jointly various measures of treatment intensity and the physician density. The share of old patients (55+) has a positive but insignificant impact on physician density.

Kopetsch (2007) uses cross sectional data (yr 2000 reimbursement data from KBV combined with German regional statistics INKAR) to estimate various measures of treatment intensity each combined with physician density within a 2SLS framework. Here, the share of population 50+ is used in the density equation and generally has a significant effect for most types of physicians, with the exception of psychotherapists.

Using cross-sectional data from German regional statistics (INKAR 2002), Juerges (2007) also obtains a positive impact of the population share 65+ on regional physician densities.

[TO BE COMPLETED]

3 Theoretical Framework

In this section we develop a theoretical framework as a basis for organising and interpreting our empirical findings. We consider a set of I regions, indexed by $i \in \{1, 2, ...I\}$. Define k_i as a set of regional characteristics. We will in some interpretations understand k_i to be a (continuous) index of (increasing) 'rurality'. The demographic structure, of each region is described by the size of the local population ℓ_i, the share of the elderly population (age 60 and over) $\overline{\lambda}_i \in [0, 1]$, and the share of the young population (below age 20) $\underline{\lambda}_i \in [0, 1]$. The remaining population (aged 20-59) is, thus, captured by the the share $1 - \underline{\lambda}_i - \overline{\lambda}_i$, obviously implying the additional constraint $\overline{\lambda}_i + \underline{\lambda}_i \in [0, 1]$.

Let n_i denote the number of physicians who practice in region i. Physicians draw their demand from the local population (only) and, facing a remuneration system parametrised by τ_i, they earn an income $Y\left(n_i, \ell_i, \overline{\lambda}_i, \underline{\lambda}_i, k_i, \tau_i\right)$.[2] Let $U(k_i)$ denote the physician's sub-utility from residing within a region characterised by k_i.[3] Assuming additive separability of income, we can then write

$$V\left(n_i, \ell_i, \overline{\lambda}_i, \underline{\lambda}_i, k_i, \tau_i\right) = Y\left(n_i, \ell_i, \overline{\lambda}_i, \underline{\lambda}_i, k_i, \tau_i\right) + U(k_i)$$

for a physician's overall utility from practicing in region i.

3.1 Determinants of current physician income

A representative physician's income is given by $Y = \sum_j \left(\tau^j - c^j\right) q^j$, where we suppress the regional index i for convenience. Here, τ^j denotes a fee per unit of service rendered to a patient belong to age group j, c^j denotes the cost per unit of service, and q^j denotes the demand for services from age-grup j. In the following we use $j = 1$ to denote the young age-group (below age 20), $j = 2$ to denote the middle age-group (aged 20-59) and $j = 3$ the group of elderly (aged 60 and over).

[2] Note that we disregard (from the physician's perspective) the age distribution within the middle age-group 20-59. This does not appear unreasonable because, with the exception of pregnancy and birth-giving, patients within this group experience relatively few health problems, whereas children and old patients exhibit significantly higher demand for physician services **(check refs: e.g. Pohlmeier and Ulrich 1995, Duscheiko et al. 2002, Dormont et al. 2006, Juerges 2007)**.

[3] More generally, a physician's residential preferences $U\left(k_i, \ell_i, \overline{\lambda}_i, \underline{\lambda}_i\right)$ may also include the demographic make-up of a region. E.g. the physician (as indeed any other individual) may have a preference for intermediate population densities - and thus against living in over-crowded or under-populated regions. Similarly young physicians may prefer to live in an environment which is reasonably 'young' and thus express a preference against a high $\overline{\lambda}_i$. For the purpose of this paper we ignore this by assuming that the effects of demographic structure on the physician's utility are negligible relative to the effects on income.

The cost per unit service $c^j = \gamma(j,k)\,x^j$ increases in the quality/intensity of the service, x^j, and varies with the patient's age j and the character of the region, k, according to the function $\gamma(j,k)$. Generally it may be assumed that the provision of services to the elderly and to the young may be relatively more resource intensive, as their treatment requires more time and as higher levels of morbidity render it necessary more frequently to provide a service by way of a home visit. Furthermore, the cost-weights $\gamma(j,k)$ are likely to vary with the regional structure k. For instance, the servicing of sick patients may require the physician to make a home visit. Due to higher levels of morbidity amongst the elderly young this is particularly true for age-group $j = 3$ and to a lesser extent for the young age group $j = 1$. Here, it is plausible that for rural regions with their longer travel times (i) home visits have to be carried out more frequently (as some of the share of the sick patients who are willing and able to travel decreases with travel time/distance) and (ii) that the cost of carrying out a home visit is higher for the physician (precisely for the longer travel times). Thus, we assume[4]

$$\begin{aligned} \gamma(2,k) &\leq \min\{\gamma(3,k), \gamma(1,k)\} \\ \gamma_k(2,k) &\leq \min\{\gamma_k(3,k), \gamma_k(1,k)\}. \end{aligned}$$

Demand $q^j = \phi(j,k)\,x^j\widetilde{\ell^j}$ from patient group j is given as the product of demand per patient $\phi(j,k)\,x^j$ and number of patients $\widetilde{\ell^j}$ within age group j. The demand per patient $\phi(j,k)\,x^j$ increases in the quality/intensity of medical services, x^j, offered to this group. Furthermore, it is assumed to vary with age and with the character of the region according to the function $\phi(j,k)$. The U-shaped age-profile of the demand for physician services would suggest that the intermediate age-group experiences the lowest levels of demand $\phi(2,k) < \min\{\phi(1,k), \phi(3,k)\}$. As rural regions generally involve longer travel times, this is likely to imply that patients are less prone to visit the physician and are potentially less responsive to the quality of services. It does not seem unreasonable to assume that due to their lack of mobility this effect is more pronounced for young and old patients. Thus, we assume

$$\begin{aligned} \phi(2,k) &\leq \min\{\phi(1,k), \phi(3,k)\} \\ \max\{\phi_k(1,k), \phi_k(3,k)\} &\leq \phi_k(2,k) \leq 0. \end{aligned}$$

The physician chooses the vector of quality levels $\left\{x^1, x^2, x^3\right\}$ so as to maximise income Y. From the set of first-order conditions $\frac{\partial Y}{\partial x^j} = \left(\tau^j - c^j\right)\phi(j,k)\,\widetilde{\ell^j} - \gamma(j,k)\,q^j = 0$ for $j = 1,2,3$ we obtain the set of optimal quality levels $\left\{x^{1*}, x^{2*}, x^{3*}\right\}$, where $x^{j*} = \frac{\tau^j}{2\gamma(j,k)}$.[5] We can thus determine (optimised)

[4] Note that we do not place an assumption on the sign of $\gamma_k(j,k)$. We only assume that rurality shifts relative costs towards age groups 1 and 3 as opposed to age-group 2.

[5] It is straightforward to verify that the second-order conditions are satisfied.

practice income as $Y = \sum_j \left(\tau^j\right)^2 \frac{\phi(j,k)}{4\gamma(j,k)} \widetilde{\ell^j}$ from which we readily obtain

$$\frac{dY}{d\widetilde{\ell^j}} = \left(\tau^j\right)^2 \frac{\phi(j,k)}{4\gamma(j,k)} > 0, \tag{1}$$

As physicians choose service provision in a way that guarantees a positive mark-up per patient this implies that a greater number patients, $\widetilde{\ell^j}$, wtihin any age-group always contributes towards a higher income. Recalling that for a representative physician the age-structure of the patient population reflects the age-structure of the regional population we can write $\widetilde{\ell^j} = \lambda^j \widetilde{\ell}$, where $\lambda^1 = \underline{\lambda}$, $\lambda^2 = 1 - \underline{\lambda} - \overline{\lambda}$ and $\lambda^3 = \overline{\lambda}$ and where $\widetilde{\ell} := \frac{\ell}{n}$ gives the total number of patients per practice. As we are considering a representative physician, it is reasonable to assume that patients are distributed equally across the n practices.[6] We can thus write physician income $Y\left(n, \ell, \overline{\lambda}, \underline{\lambda}, k, \tau\right)$ as a function of the current demographic structure, regional make-up and the payment system $\tau = \left\{\tau^1, \tau^2, \tau^3\right\}$. In particular, we obtain

the derivatives

$$\begin{aligned}
\frac{dY}{d\widetilde{\ell}} &= \sum_j \frac{dY}{d\widetilde{\ell^j}} \lambda^j = \sum_j \left(\tau^j\right)^2 \frac{\phi(j,k)}{4\gamma(j,k)} \lambda^j > 0, &(2)\\
\frac{dY}{d\ell} &= \frac{1}{n} \frac{dY}{d\widetilde{\ell}} > 0, \qquad \frac{dY}{dn} = -\frac{\widetilde{\ell}}{n} \frac{dY}{d\widetilde{\ell}} < 0, &(3)
\end{aligned}$$

implying that (i) for a given number of physicians a larger population leads to a higher income, whereas (ii) increases in the number of rival physicians trigger a reduction in income.[7] Furthermore, we obtain

$$\begin{aligned}
\frac{dY}{d\overline{\lambda}} &= \left(\frac{dY}{d\widetilde{\ell^3}} - \frac{dY}{d\widetilde{\ell^2}}\right) \widetilde{\ell} = \left(\frac{\left(\tau^3\right)^2 \phi(3,k)}{4\gamma(3,k)} - \frac{\left(\tau^2\right)^2 \phi(2,k)}{4\gamma(2,k)}\right) \widetilde{\ell} &(4)\\
\frac{dY}{d\underline{\lambda}} &= \left(\frac{dY}{d\widetilde{\ell^1}} - \frac{dY}{d\widetilde{\ell^2}}\right) \widetilde{\ell} = \left(\frac{\left(\tau^1\right)^2 \phi(1,k)}{4\gamma(1,k)} - \frac{\left(\tau^2\right)^2 \phi(2,k)}{4\gamma(2,k)}\right) \widetilde{\ell} &(5)
\end{aligned}$$

Whether or not a larger share of older patients, $\overline{\lambda}$, contributes to a higher practice income then depends on the income per patient $\frac{\tau^3 \phi(3,k)}{4\gamma(3,k)}$ gained for

[6] This would also be the equilibrium outcome of a model in which homogeneous physicians compete for patients.

[7] Note that in our model physicians do not compete for patients and thus behave like monopolists. A growing number of rival physicians thus only leads to a reduction in the number of patients per practice. In a model of oligopolistic physician competition (see e.g. Gravelle 1999, Nuscheler 2003) a greater number of rivals would additionally lead to a reduction in mark-up(s) per patient due to stronger competition. This would only strengthen the negative relationship between income and number of practices.

this group relative to the group of middle-aged patients. We see that relative profitability is thus determined by three factors: (i) differences in unit demand, (ii) differences in marginal cost per patient, and (iii) differences in reimbursement rates. A number of cases are possible depending on the reimbursement system and on the regional structure. Generally, it is plausible that unit demand by older patients exceeds that of younger patients, i.e. $\phi(3,k) > \phi(2,k)$, but that older patients are more costly to treat, i.e . $\gamma(3,k) > \gamma(2,k)$. In this case it would follow that old patients are more profitable in a system for which the fees are age-adjusted in such a way that $\tau^j = \pi\gamma(j,k)$ induces the same level of service quality for all age-groups: $\frac{dY}{d\overline{\lambda}} = \pi\left(\frac{\gamma(3,k)\phi(3,k)}{4} - \frac{\gamma(2,k)\phi(2,k)}{4}\right)\widetilde{\ell} > 0$. In contrast, a uniform fee $\tau^j = \tau$ would generate the same mark-up, $\tau^j - \gamma(j,k)\,x^{j*} = \frac{\tau^j}{2} = \frac{\tau}{2}$, per patient so that $\frac{dY}{d\overline{\lambda}} = (\tau)^2\left(\frac{\phi(3,k)}{4\gamma(3,k)} - \frac{\phi(2,k)}{4\gamma(2,k)}\right)\widetilde{\ell} > 0 \Leftrightarrow \frac{\phi(3,k)}{\phi(2,k)} > \frac{\gamma(3,k)}{\gamma(2,k)}$. Here, older patients are more profitable if and only if the relative increase in their unit demand more than compensates the relative increase in unit cost. Similar arguments apply to the effects of changes in the share of young patients, $\underline{\lambda}$.

In our empirical analysis we examine whether and how the role of age-structure for the supply of physicians (and implicitly for the physician's income) depends on the degree of rurality. Assuming that reimbursement rates do not vary across regions,[8] the cross effect between age-structure and rurality is described by the cross derivatives

$$\frac{d^2Y(\cdot)}{dkd\overline{\lambda}} = \left(\frac{dY}{dkd\widetilde{\ell}^3} - \frac{dY}{dkd\widetilde{\ell}^2}\right)\widetilde{\ell} = \left[\frac{dY}{d\widetilde{\ell}^3}\left(\varepsilon_\phi(3,k) - \varepsilon_\gamma(3,k)\right) - \frac{dY}{d\widetilde{\ell}^2}\left(\varepsilon_\phi(2,k) - \varepsilon_\gamma(2,k)\right)\right]\frac{\widetilde{\ell}}{k} \quad (6)$$

$$\frac{d^2Y(\cdot)}{dkd\underline{\lambda}} = \left(\frac{dY}{dkd\widetilde{\ell}^1} - \frac{dY}{dkd\widetilde{\ell}^2}\right)\widetilde{\ell} = \left[\frac{dY}{d\widetilde{\ell}^1}\left(\varepsilon_\phi(1,k) - \varepsilon_\gamma(1,k)\right) - \frac{dY}{d\widetilde{\ell}^2}\left(\varepsilon_\phi(2,k) - \varepsilon_\gamma(2,k)\right)\right]\frac{\widetilde{\ell}}{k} \quad (7)$$

where $\varepsilon_\phi(j,k) := \frac{\phi_k(j,k)k}{\phi(j,k)} < 0$ and $\varepsilon_\gamma(j,k) := \frac{\gamma_k(j,k)k}{\gamma(j,k)}$ are the age-specific elasticities with respect to k of the unit-demand and cost-parameter, respectively. The following is then readily verified.

Proposition 1 *Consider a region (or set of regions) of degree of rurality* $\widehat{k}^{j'2} \in [k_{\min}, k_{\max}]$, $j' = 1,3$ *for which* $\frac{dY}{d\widetilde{\ell}^{j'}} = \frac{dY}{d\widetilde{\ell}^2}$. *If* $\varepsilon_\phi(j',k) - \varepsilon_\gamma(j',k) < \varepsilon_\phi(2,k) - \varepsilon_\gamma(2,k)$ *then* $\frac{d^2Y(\cdot)}{dkd\overline{\lambda}}\big|_{k=\widehat{k}^{32}} < 0$ *and* $\frac{d^2Y(\cdot)}{dkd\underline{\lambda}}\big|_{k=\widehat{k}^{12}} < 0$, *implying that* $\frac{dY(\cdot)}{d\overline{\lambda}} < 0 \Leftrightarrow k > \widehat{k}^{32}$ *and* $\frac{dY(\cdot)}{d\underline{\lambda}} < 0 \Leftrightarrow k > \widehat{k}^{12}$.

[8] Indeed, in Germany the fees are fixed at the level of states (Bundesländer), each state comprising a number of regions (Kreise) in our terminology (Busse & Riesberg 2004). Thus, while there may be variation in fee levels across states there is no within-state variation of fees. In our estimation we control for state-level and time fixed effects, which should pick up the difference and variation in the fees.

The intuition for this Proposition is the following. If it is true that for the intermediate age-group ($j = 2$) the cost of provision and unit demand are least responsive to rurality (as measured by the sum of elasticities) then we may observe that old and young patients are profitable relative to the intermediate age-group only within the least rural regions. In this case the impact of the share of oldest and (youngest) patients on physician income changes in the degree of rurality from positive within urban areas to negative within rural areas. According to our previous arguments regarding cost of provision and unit demand for services such a finding does not lack intuition. Although we cannot measure the elasticities, it corresponds well with our empirical results in section 4.3. In an estimation where long-run effects are picked out by a lagged dependent variable, we are able to identify negative short term effects of age-structure on the number of physicians. As the latter are positively correlated with physician income, this implies negative cross-effects on income.

3.2 Determinants of Life-Time Income

Generally, the opening of a physician practice can be viewed as a long-term decision. Indeed, there is empirical evidence pointing to a very low geographical mobility of practitioners once they have settled within a particular region (Kopetsch and Munz 2007). In this case, it is plausible to assume that over their working lives physicians experience demographic change in terms of a changing size and age-structure of their patient population. A (young) physician who considers opening practice in a region i would then seek to make predictions not only about the current demand (depending on current population size and age-structure) but also about future demand. In a simple way this can be captured by considering a set-up, where physicians practice for two-periods with distinct population structure. Thus, for any two subsequent periods t and $t+1$ we can write a physician's expected utility as the sum of her utilities within the two periods

$$W_i = V\left(n_{it}, \ell_{it}, \overline{\lambda}_{it}, \underline{\lambda}_{it}, k_{it}, \tau_{it}\right) + \delta V\left(n_{it+1}, \ell_{it+1}, \overline{\lambda}_{it+1}, \underline{\lambda}_{it+1}, k_{it+1}, \tau_{it+1}\right), \quad (8)$$

where the second period utility is discounted by a factor δ. In the following we are interested in particular in the development of the population from which the physician recruits her patients.

Assume that individuals (including physisicans) live for four periods à 20 years and work for the middle two periods. Due to data restrictions we have to bunch in our empirical analysis the middle two age-groups 20-39 and 40-59 into a single group 20-59. The population structure and its development for the three age groups 0-19, 20-59 and 60-80 (in our data 60+) can thus be described by the following (unbalanced) OLG model comprising of three age groups $a = 1, 2, 3$. Suppressing the regional index i for convenience, we denote by ℓ_t^a the size of age-group a at time t. The total population at time t is then given by $\ell_t = \sum \ell_t^a$. Furthermore, define

$$\overline{\lambda}_t := \frac{\ell_t^3}{\ell_t}; \qquad \underline{\lambda}_t := \frac{\ell_t^1}{\ell_t}$$

as the population share of the oldest and youngest population, respectively. Assuming that all demographic events are counted at the end of each period, i.e. at the point of transition from t to $t+1$, and assuming that mortality arises (at rate 1) only for members of age-group 3, we can write the population dynamics as

$$\begin{aligned} \ell_{t+1}^3 &= \rho_t \ell_t^2 + \omega_t M_t \\ \ell_{t+1}^2 &= (1-\rho_t)\,\ell_t^2 + \ell_t^1 + (1-\omega_t)\,M_t \\ \ell_{t+1}^1 &= \mu_t \ell_t^2; \end{aligned}$$

where $\rho_t \in [0,1]$ denotes the share of individuals within age-group 2 who are aged 40-59 and therefore promoted to age group 3; where $\omega_t M_t$, with $\omega_t M_t \geq -\rho_t \ell_t^2$, and $(1-\omega_t)\,M_t$, with $(1-\omega_t)\,M_t \geq -\ell_t^1$, denote net migration at the end and at the beginning of working life, respectively; where M_t denotes total net migration; and where $\mu_t \geq 0$ denotes the fertility rate as applied to members of age-group 2 for which the majority of births occurs.[9] Substituting $\ell_t^2 = \left(1 - \overline{\lambda}_t - \underline{\lambda}_t\right)\ell_t$, $\ell_t^1 = \underline{\lambda}_t \ell_t$ and $M_t = m_t \ell_t$ we can express the population in period $t+1$ in terms of the demographic make-up $\left\{\ell_t, \overline{\lambda}_t, \underline{\lambda}_t, \mu_t, m_t\right\}$ in period t

$$\ell_{t+1}^3 = \left[\rho_t\left(1 - \overline{\lambda}_t - \underline{\lambda}_t\right) + \omega_t m_t\right]\ell_t \tag{9}$$

$$\ell_{t+1}^2 = \left[(1-\rho_t)\left(1 - \overline{\lambda}_t - \underline{\lambda}_t\right) + \underline{\lambda}_t + (1-\omega_t)\,m_t\right]\ell_t \tag{10}$$

$$\ell_{t+1}^1 = \mu_t\left(1 - \overline{\lambda}_t - \underline{\lambda}_t\right)\ell_t. \tag{11}$$

Population size and age-structure in period $t+1$ can thus be expressed

[9] Note that the migration streams at the beginning and end of working life may differ in their direction. For instance, a rural region that is attractive as residential area but offers poor employment prospects may be characterised by $(1-\omega_t)\,M_t < 0 < \omega_t M_t$. The converse may apply to agglomerations which are unattractive for residence but offer good employment opportunities.

as functions of the demographic make-up in period t

$$\begin{aligned}
\ell_{t+1}\left(\ell_t,\overline{\lambda}_t,\underline{\lambda}_t,\rho_t,\mu_t,m_t\right) &= \left[(1+\mu_t)\left(1-\overline{\lambda}_t-\underline{\lambda}_t\right)+\underline{\lambda}_t+m_t\right]\ell_t, \quad (12)\\
\overline{\lambda}_{t+1}\left(\ell_t,\overline{\lambda}_t,\underline{\lambda}_t,\rho_t,\mu_t,m_t\right) &= \frac{\left[\rho_t\left(1-\overline{\lambda}_t-\underline{\lambda}_t\right)+\omega_t m_t\right]}{(1+\mu_t)\left(1-\overline{\lambda}_t-\underline{\lambda}_t\right)+\underline{\lambda}_t+m_t},\\
\underline{\lambda}_{t+1}\left(\ell_t,\overline{\lambda}_t,\underline{\lambda}_t,\rho_t,\mu_t,m_t\right) &= \frac{\mu_t\left(1-\overline{\lambda}_t-\underline{\lambda}_t\right)}{(1+\mu_t)\left(1-\overline{\lambda}_t-\underline{\lambda}_t\right)+\underline{\lambda}_t+m_t},
\end{aligned}$$

which in turn allows us to express a physician's income in period $t+1$ as a function of the demographic make-up in period t,

$$Y\left(n_{t+1},\ell_{t+1}(\cdot),\overline{\lambda}_{t+1}(\cdot),\underline{\lambda}_{t+1}(\cdot),k_{t+1},\tau_{t+1}\right)=\widehat{Y}\left(n_{t+1},\ell_t,\overline{\lambda}_t,\underline{\lambda}_t,\mu_t,m_t,k_{t+1},\tau_{t+1}\right)=:Y_{t+1}$$

The following Lemma summarises the impact of the current demographic make-up on a physician's future income Y_{t+1}.

Lemma 2 *Physician income in period $t+1$ responds to the current demographic make-up in the following way*

$$\begin{aligned}
\frac{dY_{t+1}}{d\ell_t} &= \frac{(1+\mu_t)\left(1-\overline{\lambda}_t-\underline{\lambda}_t\right)+\underline{\lambda}_t+m_t}{n_{t+1}}\frac{dY}{d\widetilde{\ell}}>0,\\
\frac{dY_{t+1}}{d\overline{\lambda}_t} &= -\frac{\ell_t}{n_{t+1}}\left[\rho_t\frac{dY}{d\widetilde{\ell}^3}+(1-\rho_t)\frac{dY}{d\widetilde{\ell}^2}+\mu_t\frac{dY}{d\widetilde{\ell}^1}\right]<0,\\
\frac{dY_{t+1}}{d\underline{\lambda}_t} &= -\frac{\ell_t}{n_{t+1}}\left[\rho_t\left(\frac{dY}{d\widetilde{\ell}^3}-\frac{dY}{d\widetilde{\ell}^2}\right)+\mu_t\frac{dY}{d\widetilde{\ell}^1}\right]\\
&= -\frac{\ell_t}{n_{t+1}}\left[\frac{\rho_t}{\ell_t}\frac{dY}{d\overline{\lambda}}+\mu_t\frac{dY}{d\widetilde{\ell}^1}\right],\\
\frac{dY_{t+1}}{dm_t} &= \frac{\ell_t}{n_{t+1}}\left[\omega_t\frac{dY}{d\widetilde{\ell}^3}+(1-\omega_t)\frac{dY}{d\widetilde{\ell}^2}\right],\\
\frac{dY_{t+1}}{d\mu_t} &= \frac{\ell_t}{n_{t+1}}\left(1-\overline{\lambda}_t-\underline{\lambda}_t\right)\frac{dY}{d\widetilde{\ell}^1}>0,
\end{aligned}$$

where $\frac{dY}{d\widetilde{\ell}}>0$ and $\frac{dY}{d\widetilde{\ell}^j}>0$, as defined in (2) and (1) are evaluated at $t+1$, respectively.

Proof: The derivatives follow immediately when inserting into $\frac{dY_{t+1}}{d\ell_t}=\frac{1}{n_{t+1}}\frac{dY_{t+1}}{d\widetilde{\ell}}\frac{d\ell_{t+1}}{d\ell_t}$ and $\frac{dY_{t+1}}{dz_t}=\frac{1}{n_{t+1}}\sum_j\frac{dY_{t+1}}{d\widetilde{\ell}^j}\frac{d\ell^j_{t+1}}{dz_t}$ for $z_t\in\left\{\overline{\lambda}_t,\underline{\lambda}_t,\mu_t,m_t\right\}$

the derivatives

$$\begin{aligned}
\frac{d\ell_{t+1}}{d\ell_t} &= (1+\mu_t)\left(1-\overline{\lambda}_t-\underline{\lambda}_t\right)+\underline{\lambda}_t+m_t>0 \\
\frac{d\ell^3_{t+1}}{d\overline{\lambda}_t} &= -\rho_t\ell_t<0, \quad \frac{d\ell^2_{t+1}}{d\overline{\lambda}_t}=-\left(1-\rho_t\right)\ell_t<0, \quad \frac{d\ell^1_{t+1}}{d\overline{\lambda}_t}=-\mu_t\ell_t<0 \\
\frac{d\ell^3_{t+1}}{d\underline{\lambda}_t} &= -\rho_t\ell_t<0, \quad \frac{d\ell^2_{t+1}}{d\underline{\lambda}_t}=\rho_t\ell_t>0, \quad \frac{d\ell^1_{t+1}}{d\underline{\lambda}_t}=-\mu_t\ell_t<0 \\
\frac{d\ell^3_{t+1}}{dm_t} &= \omega_t\ell_t, \quad \frac{d\ell^2_{t+1}}{dm_t}=\left(1-\omega_t\right)\ell_t, \quad \frac{d\ell^1_{t+1}}{dm_t}=0 \\
\frac{d\ell^3_{t+1}}{d\mu_t} &= \frac{d\ell^2_{t+1}}{d\mu_t}=0, \quad \frac{d\ell^1_{t+1}}{d\mu_t}=\left(1-\overline{\lambda}_t-\underline{\lambda}_t\right)\ell_t>0
\end{aligned}$$

as obtained from (12) and (9)-(11).■

Unsurprisingly, the size of the population in $t+1$, as indeed the size of all individual age-groups, increases in the size of the current population, ℓ_t. The greater practice size thus implied always raises a physician's (future) income. In contrast, the current share of the elderly, $\overline{\lambda}_t$, has an unambiguously negative impact on the size of all future age-groups. This is a consequence of the fact that for a given $\underline{\lambda}_t$, a greater $\overline{\lambda}_t$ implies an unambiguously lower size of the intermediate age group ℓ^2_t, which is the group out of which all future age-groups are recruited according to the rates ρ_t, $1-\rho_t$, and μ_t, respectively. Consequently, the practice income expected for period $t+1$ decreases in the current share of the elderly. The effect of the current share of the young, $\underline{\lambda}_t$, on future practice income is ambiguous. It tends to decrease the size of the next-period old and young age-groups, ℓ^3_{t+1} and ℓ^1_{t+1}, respectively, but it increases the size of the intermediate age-group, ℓ^2_{t+1}. The impact on practice income thus depends on the relative profitability of these age-groups and, on the birth rate and on the rate ρ_t (i.e. the share of older individuals within age-group 2). A negative effect arises if treating old patients is not less profitable from the physician's point of view than treating middle-aged patients, such that $\frac{dY}{d\overline{\lambda}} \geq 0$. Recall from Proposition 1 that this may be true, in particular, within urban regions.

The impact of the rate-of net migration on the size of the future age-groups ℓ^3_{t+1} and ℓ^2_{t+1} depends on the parameter ω_t. If the migration streams at the beginning and end of working life take on the same direction (positive or negative) then we have $\omega_t \in [0,1]$ and thus both age-groups increase in the aggreagte rate of net-migration, m_t. The effect of migration becomes ambiguous, however, if the migration streams at the beginning and end of working life are opposed, as we then have $\omega_t \notin [0,1]$. Our data suggests that migration (in or out) takes place predominantly at the beginning of working life, implying that $|\omega_t|$ is small. For this case, we obtain $\frac{dY_{t+1}}{dm_t} > 0$.

The effect of the fertility rate on future income is unambigously positive, as higher fertility implies an increase in the size of the young age-group for constant levels of the other age-groups.

The cross effects between rurality and age-structure

$$\begin{aligned}\frac{d^2Y_{t+1}}{dkd\overline{\lambda}_t} &= -\frac{\ell_t}{n_{t+1}}\left[\rho_t\frac{dY}{dkd\widetilde{\ell}^3}+(1-\rho_t)\frac{dY}{dkd\widetilde{\ell}^2}+\mu_t\frac{dY}{dkd\widetilde{\ell}^1}\right],\\ \frac{d^2Y_{t+1}}{dkd\underline{\lambda}_t} &= -\frac{\ell_t}{n_{t+1}}\left[\rho_t\left(\frac{dY}{dkd\widetilde{\ell}^3}-\frac{dY}{dkd\widetilde{\ell}^2}\right)+\mu_t\frac{dY}{dkd\widetilde{\ell}^1}\right]\end{aligned}$$

are composits of the impact of rurality on the marginal profitability of the various patient groups, $\frac{dY_{t+1}}{dkd\widetilde{\ell}^j}$, these cross derivatives being ambiguous in their own right. If rurality lowers the marginal profitability of all patient groups, i.e. if $\frac{dY_{t+1}}{dkd\widetilde{\ell}^j} < 0$ for all j, then $\frac{d^2Y_{t+1}}{dkd\overline{\lambda}_t} > 0$ is always true and $\frac{d^2Y_{t+1}}{dkd\underline{\lambda}_t} > 0$ is true if $\frac{d^2Y}{dkd\overline{\lambda}_t} = \frac{dY_{t+1}}{dkd\widetilde{\ell}^3} - \frac{dY_{t+1}}{dkd\widetilde{\ell}^2} < 0$. However, this is not granted and ultimately these effects should be subject to empirical investigation.

We can now derive the full impact of the period t demographic make-up $\left\{\ell_t, \overline{\lambda}_t, \underline{\lambda}_t, \rho_t, \mu_t, m_t\right\}$ on a physician's (expected) utility (8), as transmitted through changes in present and future income

$$\frac{dW}{d\ell_t} = \frac{\overset{+}{dY_t}}{d\ell_t} + \delta\frac{\overset{+}{dY_{t+1}}}{d\ell_t} > 0, \tag{13}$$

$$\frac{dW}{d\overline{\lambda}_t} = \frac{\overset{\pm}{dY_t}}{d\overline{\lambda}_t} + \delta\frac{\overset{-}{dY_{t+1}}}{d\overline{\lambda}_t}, \qquad \frac{dW}{d\underline{\lambda}_t} = \frac{\overset{\pm}{dY_t}}{d\underline{\lambda}_t} + \delta\frac{\overset{\pm}{dY_{t+1}}}{d\underline{\lambda}_t}, \tag{14}$$

$$\frac{dW}{dm_t} = \delta\frac{\overset{+}{dY_{t+1}}}{dm_t} > 0 \quad \text{if } |\omega_t| \text{ small}, \qquad \frac{dW}{d\mu_t} = \delta\frac{\overset{+}{dY_{t+1}}}{d\mu_t} > 0. \tag{15}$$

We see from this exercise that the current size and age-structure of the population $\left\{\ell_t, \overline{\lambda}_t, \underline{\lambda}_t\right\}$ affects a physician's discounted income both in the current and future period. Future income is affected by the current demographic make-up, as it allows predictions about the size and structure of the future patient population. This latter argument applies also to the demographic rates $\{\mu_t, m_t\}$, which by assumption only affect the size and structure of the future population. A larger current population always increases the physician's life time income by increasing the size of the patient population both in the current period t and the future period $t+1$. Population shrinking therefore leads to an unambiguous decline in the physician's current and expected income and consequently in expected utility. A similar clear-cut argument applies to the rate of fertility and the rate of net migration given that net migration is determined mostly amongst the young, i.e. given that $|\omega_t|$ small. By leading to the expectation of a greater future population both fertility and migration lead to the expectation of a higher income in the second period of the physician's working life. The impact of the age-shares $\{\overline{\lambda}_t, \underline{\lambda}_t\}$ involves a potential trade-off between current income and future income. While a higher current share of the elderly always leads to the prediction of a lower future income, the effect on current income is indeterminate. If the conditions in Proposition 1 are satisfied, we have $\frac{dY_t}{d\overline{\lambda}_t} < 0$ for high levels of rurality, k, which would imply an unambiguously negative effect within rural regions. However, even where the treatment of elderly patients is relatively profitable for low levels of rurality, i.e. even if $\frac{dY_t}{d\overline{\lambda}_t} > 0$, this is offset by the negative impact of the share of the elderly on future income. Hence, within urban areas the effect of a higher share of the elderly on a physician's life time income is ambiguous. The impact on life-time income of the youth share, $\underline{\lambda}_t$, is ambiguous for all types of regions. If $\frac{dY}{d\overline{\lambda}} \geq 0$, as may be the case for urban areas, we have a negative impact of the youth share on future income $\frac{dY_{t+1}}{d\underline{\lambda}_t} < 0$. However, according to Proposition 1, urban areas are also more prone to exhibit $\frac{dY_t}{d\underline{\lambda}_t} > 0$, leaving the aggregate effect indeterminate. For rural areas this trade-off is reversed when the birth rate μ_t is low. Then, we obtain $sgn\frac{dY_{t+1}}{d\underline{\lambda}_t} = -sgn\frac{dY}{d\overline{\lambda}} > 0$, while at the same time $\frac{dY_t}{d\underline{\lambda}_t} < 0$ holds true. Thus, while theory helps us to identify the sources of the various trade-offs related to the age-distribution and its effect both on present and future income, it also yields the plausible yet unspectacular result that the effect of demographic change $\{d\ell_t < 0, d\overline{\lambda}_t > 0, d\underline{\lambda}_t < 0, d\mu_t < 0\}$ on a representative physician's lifetime income must necessarily be ambiguous. In fact, it will be part of our empirical exercise to shed some further light on the actual relationships.

3.3 Entry Equilibrium

In order to keep things simple, we assume that physicians are homogeneous in their preferences, their practice technology, and their outside utility $\overline{W} \geq 0$.[10] Recalling $W_{it}(n_{it}, En_{it+1})$, as defined in equation (8), as a physician's expected utility when settling in region i at time t, and assuming that physicians are free in their location choice, entry into region i takes place as long as $W_{it}(n_{it}, En_{it+1}) \geq \max\{W_{i't}(n_{i't}, En_{i't+1}), \overline{W}\}$, for any other region $i' \neq i$. Here, En_{it+1} is the expected number of physisians for the second period of practice. Ignoring the integer issue, we can then describe an entry equilibrium for period t by the set of conditions $W_{it}(n_{it}^*, En_{it+1}) = \overline{W}$ $\forall i$.[11] Defining $\widehat{Y}_{it} := \sum_j \left(\tau_{it}^j - \gamma(j,k)\, x_{it}^{j*}\right) \phi(j,k)\, x_{it}^{j*} \lambda_{it}^j \ell_{it}$ (as the total

[10] The model is readily extended to the case where physicians can be ranked according to their outside utility $\overline{W}(n)$, where $\overline{W}_n(n) \geq 0$.

[11] As $\frac{\partial W_{it}(n_{it}^*, \cdot)}{\partial n} = \frac{\partial Y_{it}(n_{it}^*, \cdot)}{\partial n} < 0$ such an equilibrium is stable.

income from physician services within region i in period t) we can write $W_{it}(n_{it}, En_{it+1}) = \frac{\widehat{Y}_{it}}{n_{it}} + U(k_{it}) + \delta\left(\frac{\widehat{Y}_{it+1}}{En_{it+1}} + U(k_{it+1})\right)$ and, thus obtain

$$\frac{\widehat{Y}_{it}}{n_{it}^*} + \delta\frac{\widehat{Y}_{it+1}}{En_{it+1}} = \overline{W} - (U(k_{it}) + \delta U(k_{it+1})) =: \widehat{W}_{it}$$

as condition for an entry equilibrium. According to this condition physicians enter up to the point at which their (discounted) life-time income equals their outside utility net of their discounted residential utility. In order to simplify the following analysis, we assume without much loss of generality $k_{it} = k_{i+t} = k_i$ and, thus, $\widehat{W}_{it} = \widehat{W}_i$ for all t. We then obtain

$$n_{it}^* = \frac{\widehat{Y}_{it}En_{it+1}}{\widehat{W}_i En_{it+1} - \delta\widehat{Y}_{it+1}}.$$

as the equilibrium number of physicians in region i and period t. As is readily checked it increases in pesent and future income $\widehat{Y}_{it}$ and $\widehat{Y}_{it+1}$, respectively, and decreases in the number of physicians expected for period $t+1$ and in the outside utility (net of residential utility). Denoting by $\widehat{n}_{it}$ the number of young physicians who enter region i in a period t and assuming that all physicians continue practice within the same region for the second period, we have $n_{it} = \widehat{n}_{it} + \widehat{n}_{it-1}$ and similarly $n_{it+t} = \widehat{n}_{it} + \widehat{n}_{it+1}$. We can then describe two (extreme) patterns of entry equilibrium as follows.[12]

Proposition 2 *(i) Full generational overlap: Assume that* $\frac{\widehat{Y}_{it}}{\widehat{n}_{it-1}} + \delta\frac{\widehat{Y}_{it+1}}{E\widehat{n}_{it+1}} > \widehat{W}_i$ *holds for all* $t \leq T$, *with* $T = \infty$. *An entry equilibrium for period* t *is then given by*

$$n_{it}^* = \frac{(1+\delta)\widehat{Y}_{it}}{\widehat{W}_i}.$$

(ii) No generational overlap: Assume that $\left\{\frac{\widehat{Y}_{it-1}}{\widehat{n}_{it-2}} + \delta\frac{\widehat{Y}_{it}}{E\widehat{n}_{it}}, \frac{\widehat{Y}_{it+1}}{\widehat{n}_{it}} + \delta\frac{\widehat{Y}_{it+2}}{E\widehat{n}_{it+2}}\right\} < \widehat{W}_i$ *holds. An entry equilibrium for period* t *and period* $t+1$ *is then given by*

$$n_{it}^* = n_{it+1}^* = \frac{\widehat{Y}_{it} + \delta\widehat{Y}_{it+1}}{\widehat{W}_i}.$$

[12] There exist other equilibrium patterns, depending on the period T at which zero entry takes place. As we show in the appendix, for zero entry at period T (equivalent to a finite number of periods T), the equilibrium number of firms in period t is given by, $n_{it}^* = \widehat{Y}_{it}n_{iT}\left[\frac{1-(-\delta)^{T-t}1}{1+\delta}\widehat{W}_i n_{iT} + (-\delta)^{T-t}\widehat{Y}_{iT}\right]^{-1}$, where n_{iT} is the number of firms which is active in period T. As by assumption no entry takes place in period T, we have $n_{iT} = \widehat{n}_{it-1}$.

Proof: In the following we drop the regional index i. To prove (i) consider a sequence of periods $\widehat{t} = t, t+1, ..., T-1$, with $T \geq t+1$. The assumption $\frac{\widehat{Y}_t}{\widehat{n}_{t-1}} + \delta \frac{\widehat{Y}_{t+1}}{E\widehat{n}_{t+1}} > \widehat{W}$ implies that there is some entry taking place in all periods, i.e. $\widehat{n}_{\widehat{t}} > 0$ for all $\widehat{t}$.In this case, we can write $n^*_{\widehat{t}} = \frac{\widehat{Y}_{\widehat{t}} n^*_{\widehat{t}+1}}{\widehat{W} n^*_{\widehat{t}+1} - \delta \widehat{Y}_{t+1}}$. Starting with $n^*_{T-1} = \frac{\widehat{Y}_{T-1} En_T}{\widehat{W} En_T - \delta \widehat{Y}_T}$ and substituting recursively, we obtain $n^*_t = \prod_{\widehat{t}=t}^{T-1} \widehat{Y}_{\widehat{t}} En_T \left[\frac{(-\delta)^{T-t}-1}{(-\delta)-1} \widehat{W} \prod_{\widehat{t}=t+1}^{T-1} \widehat{Y}_{\widehat{t}} En_T + (-\delta)^{T-t} \prod_{\widehat{t}=t+1}^{T} \widehat{Y}_{\widehat{t}} \right]^{-1} =$

$\widehat{Y}_t En_T \left[\frac{1-(-\delta)^{T-t}1}{1+\delta} \widehat{W} En_T + (-\delta)^{T-t} \widehat{Y}_T \right]^{-1}$. Assuming $T = \infty$, we then obtain $(-\delta)^T = 0$ and, therefore, $n^*_t = \frac{(1+\delta)\widehat{Y}_t}{\widehat{W}}$. To prove part (ii), we note that $\frac{\widehat{Y}_{t-1}}{\widehat{n}_{t-2}} + \delta \frac{\widehat{Y}_t}{E\widehat{n}_t} < \widehat{W}$ and $\frac{\widehat{Y}_{t+1}}{\widehat{n}_t} + \delta \frac{\widehat{Y}_{t+2}}{E\widehat{n}_{t+2}} < \widehat{W}$ imply that no entry takes place in periods $t-1$ and $t+1$, respectively, i.e. that $\widehat{n}_{t-1} = \widehat{n}_{t+1} = 0$. In this case, we obtain $En_{t+1} = n^*_t$ and, thus, $n^*_t = \frac{\widehat{Y}_i n^*_t}{\widehat{W}_i n^*_t - \delta \widehat{Y}_{t+1}}$, which solves to $n^*_t = n^*_{t+1} = \frac{\widehat{Y}_i + \delta \widehat{Y}_{t+1}}{\widehat{W}}$.■

In an equilibrium with full generational overlap entry is profitable for at least some physicians of every generation. There are then two generations of physicians in practice within any given period. The total number of physicians practicing in region i at period t then depends only on the potential income $\widehat{Y}_{it}$ for this same period but not on the income for the next period (nor on the income of the past period). Thus, for variations in variable $x_{it} \in \{\ell_{it}, \overline{\lambda}_{it}, \underline{\lambda}_{it}, \mu_{it}, m_{it}\}$ we obtain

$$\frac{dn^*_{it}}{dx_{it}} = \frac{1+\delta}{\widehat{W}_i} \frac{d\widehat{Y}_{it}}{dx_{it}} = \frac{(1+\delta)\, n^*_{it}}{\widehat{W}_i} \frac{dY_{it}}{dx_{it}},$$

with the derivatives $\frac{dY_{it}}{d\ell_{it}}, \frac{dY_{it}}{d\ell_{it}}, \frac{dY_{it}}{d\ell_{it}}$ as described by (3)-(5) and with $\frac{\partial Y_{it}}{\partial \mu_{it}} = \frac{\partial Y_{it}}{\partial m_{it}} = 0$. Thus, for an equilibrium with full generational overlap expectations over the future population have no bearing on the current number of physicians.[13]

However, it is by no means guaranteed that physicians of every generation enter at all times. As one alternative, entry may take place only every second period, in case of which there is no generational overlap. Such a

[13] The relationship $\widehat{n}_{it} + \widehat{n}_{it+1} = \frac{(1+\delta)\widehat{Y}_{it+1}}{\widehat{W}_i}$ suggests, of course, that entry $\widehat{n}_{it}$ in period t increases in the income expected for period $t+1$. However, the tendency towards higher entry in period t is balanced out by less entry in period $t-1$, implying that the total number of physicians in period t, $n^*_{it} = \widehat{n}_{it} + \widehat{n}_{it-1}$is invariant to income in the subsequent period $\widehat{Y}_{it+1}$ or previous period $\widehat{Y}_{it-1}$.

situation may arise for a period $t+1$, say, when a large number of entrants from the previous generation, $\widehat{n}_{it} >> 0$, crowd the market in period $t+1$, leading to low levels of current income Y_{it+1}, and if at the same time there is an expectation that future entry $E\widehat{n}_{t+2} >> 0$ depresses income Y_{it+2} in the second period of practice. In such a case, $\frac{\widehat{Y}_{it+1}}{\widehat{n}_{it}} + \delta \frac{\widehat{Y}_{it+2}}{E\widehat{n}_{it+2}} < \widehat{W}_i$ and entry becomes unrpofitable for members of generation $t+1$, implying that $\widehat{n}_{it+1} = 0$.

If the same applies to generation $t-1$, then a pattern of entry arises where only every second generation is setting up practice: $\widehat{n}_{it-2} = n^*_{it-2} = n^*_{it-1}$, $\widehat{n}_{it-1} = 0$, $\widehat{n}_{it} = n^*_{it} = n^*_{it+1}$, $\widehat{n}_{it+1} = 0$, $\widehat{n}_{it+2} = n^*_{it+2} = n^*_{it+3}$, etc. The extreme age-structure, i.e. there are either only young or only old physisicans in practice also implies that turn-over of phyisicians is not continuous but occurs fully at every second period. For this type of equilibrium, the total number of physicians within a period t, in which entry takes place, then depends on the expectations for the subsequent period $t+1$. We therefore have

$$\frac{dn^*_{it}}{dx_{it}} = \frac{n^*_{it}}{\widehat{W}_i}\left(\frac{\partial Y_{it}}{\partial x_{it}} + \delta \frac{dY_{it+1}}{dx_{it}}\right) = \frac{n^*_{it}}{\widehat{W}_i}\frac{dW_{it}}{dx_{it}},$$

with $\frac{dW_{it}}{dx_{it}}$ as described by (13)-(15). Note that in this case, $\frac{dn^*_{it}}{dm_{it}} > 0$ and $\frac{dn^*_{it}}{d\mu_{it}} > 0$ suggesting that the current supply of physicians increases in the rates of net migration and of reproduction as these correspond to higher levels of future population.

Similar arguments apply for the cross-effects $\frac{d^2 n^*_{it}}{dkdx_{it}}$, $x_{it} \in \left\{\ell\overline{\lambda}_{it}, \underline{\lambda}_{it}\right\}$. For an equilibrium with full generational overlap we obtain $\frac{dn^*_{it}}{dx_{it}} = \frac{(1+\delta)n^*_{it}}{\widehat{W}_i}\frac{d^2 Y_{it}}{dkdx_{it}}$, with the last derivative as given in (6) and (7), and for an equilibrium with no overlap we obtain $\frac{dn^*_{it}}{dx_{it}} = \frac{n^*_{it}}{\widehat{W}_i}\frac{d^2 W_{it}}{dkdx_{it}} = \frac{n^*_{it}}{\widehat{W}_i}\left(\frac{d^2 Y_{it}}{dkdx_{it}} + \delta\frac{d^2 Y_{it+1}}{dkdx_{it}}\right)$. We can summarise as follows.

Corollary 2 *(i) Expectations about future population determine the current supply of physicians if and only if the entry pattern involves no (or incomplete) generational overlap. (ii) Under no generational overlap the current number of physicians is determined the current and future demographic structure as determinants of life time income. (iii) Under full generational overlap, the current number of physicians is determined solely by the current demographic structure as determinant of current income.*

In particular, it is ture that for an equilibrium with full generational overlap, the current rate of net migration and the current rate of reproduction should have no influence on the current number of physicians. A significant positive impact would therefore lends support for an equilibrium with no (or limited) generational overlap rather than with full (or seizable) generational overlap. This is, indeed, what we find from our estimates.

4 Empirical Analysis

As mentioned in the introduction, since the second half of the nineties it is increasingly reported that the number of physicians decreases at the regional level, because less young physicians enter these local markets compared to those physicians who retire. According to the National Association of Statutory Health Insurance (SHI) the age structure of physicians is driven by large baby-boomer cohorts. As shown in figure 1 the age structure in 1993 differs significantly from that in 2004. The 1993er "iceberg" is skewed right and the number of physicians who retire is small. In 2004 the "iceberg" has reached the age of retirement. Between 1993 and 2004 the number of young physicians who enter the market has decreased. Related to section 3.3 it seems that the equilibrium with no (or limited) generational overlap rather than with full (or sizeable) generational overlap is empirically, on average, of major importance.

figure 1 about here

Within the period we considered in the estimates, 1995 - 2004, the number of physicians has increased at the macro level and, additionally, changed markedly at the regional level. Figure 2 provides information on the percentage change of the number of physicians between 1995 and 2004 in Germany disaggregated into 439 regions. It should be mentioned, however, that the figure depict the total number of physicians registered by the SHI located in a region. Particular disciplines might differ from the average development. The dark red areas suffer from a reduction of physician supply of up to 26% between 1995 and 2004. This applies in particular to many East German regions, but also to some West German rural areas. The orange colored regions experience stagnation or a slight increase in the provision of physician services.

figure 2 about here

On the other hand, most of the West German regions experience an increase in the supply of physicians, as the light and dark green colored areas depict. As mentioned in the introduction, the question emerged, of what are the causes for these increasingly unequal distribution of physicians at the regional level. An empirical fact is that in the same period the average age of the population in rural areas and in particular in East Germany has increased rapidly and, additionally, the population in these regions declined faster than in other regions.

Using various econometric estimates we therefore analyze in this section the effects of regional demographic change on the regional supply of physicians. In particular, we examine whether the age structure of the population has an impact on the supply of physicians. As argued in the theoretical section, the assumption that aging affects the local decision of physicians has three components: the age structure itself, net regional migration, and the difference between births and deaths (subsequent labelled as natural balance). Due to data availability we cannot separate fertility and mortality effects.

Regional aging depends not only on the current age structure but in particular on regional migration, which can intensify or mitigate aging at the regional level. Likewise, a change in fertility affects the regional age structure, but in contrast to migration only via the share of children and a delayed cohort effect. The latter, in turn, can be intensified or mitigated by migration. This very simple characterisation of regional population developments point out that all three effects should be considered in the estimates. In addition, it seems to be useful to analyse first the statistical relationship between migration flows, natural balance, and the age structure of the population at the regional level.

In figure 3 the relationship between overall migration and migration by families (30 to 50 years old and up to 17 years old), young (18 to 29 years old), and retired (65 years and older) people is displayed. For regional overall net migration we have observations over the period 1995 to 2004, while for the subgroups we have data only for 2003 and 2004. Therefore a comparison of the subgroups with overall net migration will help interpreting the estimates, in which we cannot consider the migration subgroups.

figure 3 about here

Two aspects are visible. First, all groups are positively related to overall migration.[14] Second, the young have a higher regional mobility, and the lowest mobility can be observed for retired people. While both findings are not surprising, they are important for the analysis in the econometric section. This is because we will argue that net migration is related to physicians expectation about future income. This means that positive net migration is mainly driven by migration of the young which are, in turn, potential future patients. A second point is that migration then slowed the aging process.

Figure 4 shows the statistical relation between the migration flow groups and regional natural balance (births minus deaths). With respect to correlation it is again the group of the young that is more important ($r = 0.25$). While the family age group is is also positively correlated with natural balance ($r = 0.18$), migration of retired person is uncorrelated ($r = -0.05$) with the difference births and deaths. In addition, the figure suggests that regions with a negative natural balance also often suffer from negative net migration of the young, and vice versa. Similar to migration flows we argue

[14] The correlation coefficients are 0.64 for the young, 0.54 for families, and 0.42 for retired persons.

that a positive natural balance is related to physicians expectation about potential future patients.

figure 4 about here

We now turn to the population shares of those who are less than 20 years (20minus) and those who are 60 years and older (60plus). In contrast to the data used in the figures 3 and 4 the underlying data cover the period 1995 to 2004. The population share 60plus is negatively correlated with net migration ($r = -0.32$), while the share 20minus is positively related to net migration flows ($r = 0.33$). Young people exhibit a higher regional mobility and, hence, will be more gravitated to economically more attractive regions than the elderly. This explains why net migration is positively related to the share of the young and negatively to the share of the 60plus. Put differently, the correlation pattern is consistent with the finding that regional mobility decreases with age. For natural balance we find a similar pattern. The difference between births and deaths is positively correlated with the population share 20minus ($r = 0.50$) and negatively with the share 60plus ($r = -0.62$). The difference between net migration and natural balance on the one hand and the age related population shares on the other hand is that the latter reflects the current patient structure. This means that in the econometric analysis we should find significant effects of all variables in the long run. However, we expect that in the short run only the age structure matters.

4.1 Data

Ideal for our analysis would be the use of micro data that cover physicians socio-demographic characteristics and information on his income and patient structure. However, such a data set is currently not available, and we build our analysis on regional information. More precisely, we use data for 439 districts in Germany. The data cover the years 1995 to 2004 with the exception of 1999. In this year the SHI has reorganised the registry of physicians at the regional level and do not provide information at the regional level. Hence, we have 3951 observation available for our estimates.

The data used are taken from the INKAR data set provided by the Federal Office for Civil Engineering and Regional Development (BBR). The dependent variable is the number of physicians that include general practitioners as well as specialists like, for example, ophthalmologists, otorhinolaryngologist, paediatricians, and orthopaedist. We aggregate the different specialisations to have more variation in the data. This is because the SHI has imposed in some regions a freeze of settlement for certain branches. However, in exceptional cases they slackening the imposed freeze. Since the SHI neither provide information on the degree of stoppage of settlements nor on the exceptional cases at the regional level, we decided to overcome this specialists related potential shortage in variation using the number of physicians within a region. This ensures that the data vary from year to year in each region. According to the SHI the freeze of settlement for certain branches remains unchanged in the considered period in most cases. In addition, more attractive regions have a higher probability of being afflicted with a physician sattlemtent-stop for certain specialisations. With respect to our estimates this has two consequences: First, the causes for the regional unequal distribution of physicians are potentially underestimated. Second, spatial effects on the distribution of physicians are potentially overestimated. We get back to both points below. In section 4.3 we also use the number of physicians per 100,000 population, which has frequently been used in analyses similar to ours. However, as we will ague below this variable exhibits to some extent a measurement bias, and for this reason we prefer to use the number of physicians as dependent variable.

As we have outlined in the previous section, demographic change bears on the regional physician supply by way of two channels: Changes in the age-structure and change in the population size. Changes in the size of the population are captured by the flow variables net migration and the difference between births and deaths (natural balance) in a region. The size of a population increases if both variables are positive. Our main interest, however, is in the effect of a changing age structure. We approximate regional population ageing by the population share of people 60 years and older (share 60plus) and the population share of people less than 20 years (share 20minus). As pointed out in the theoretical section, it is possible that the effects of ageing on physician supply vary according to the character of the region (urban vs. rural). We control for these effects by interacting the age groups (60plus and 20minus) with an index of rurality, as provided by the BBR. This index is subaggregated into agglomeration areas, municipalized areas, and rural areas. Within these area groups there are up to four different classes differing in population and population density. Altogether there are nine levels of the index, with higher levels corresponding to a greater degree of 'rurality'. See the Appendix for further details. The advantage of using this index is that it captures information on population size, population density, and a sphere of influence of (supraregional) cities.

The remaining set of regional control variables comprise: population size, population density, GDP per capita, unemployment rate, employment rate, share of foreigners, share of welfare recipients, share of school leaver with higher education entrance qualification, share of school leaver without school leaving certification for a secondary education, share of new houses for up to two families and new flats in the total stock of flats, tourist accommodation per 100,000 population, and cars per 1,000 population. We will discuss the expected effects in the next section when we present the results.

4.2 Estimation of the Basic Model

At the macro level physician supply is usually measured in per capita terms, i.e. by physician density. However, if ageing at the regional level is intensified by interregional migration, this measure could comprise a measurement error. This is possible if, for example, younger and older people differ in their geographical mobility. It is observed empirically that younger people leave rural areas more often than older people. Outmigration on the part of the young has two distinct implications for a region. On the one hand, the corresponding decline in population implies that the physician density increases for a given number of physicians. At the same time the share of older people increases, generating a positive, yet spurious, correlation with physician density. According to our data physician density is correlated negatively with the population share 20minus (-0.57) and positively with the population share 60plus (0.26). Given that physician location choices are made long term, physician density then comprises a measurement error.

To make this more clear we decompose the physician density (D) into the real term that is based on decisions of physicians (y) and a measurement error due to interregional migration (ϵ). The regression that should approximate the true model $D = y + \epsilon = \alpha + \beta x + u + \epsilon$ (with u as idiosyncratic error term) yields a biased parameter β if X and ϵ are correlated. For example, population density and the population share 60plus are correlated negatively (-0.31), which is in line with the finding that net migration and the population share 60plus are negative correlated as stated above. Since a declining population increases (at least in the short run) the physician density, we can conclude that the parameter that corresponds to the share of 60plus is biased positively. Therefore, in this section we use only the number of physicians as dependent variable. In the robustness section we run additional regressions with the physician density as dependent variable, conditional on the same set of covariates, to account for this bias.

In the theoretical section we have argued that physician supply is, among other things, related to net migration flow (which increases in the net migration rate), natural balance (which increase in the fertility rate), and the age structure of the population in a region. We first seek to establish whether the underlying relationships can be estimated even if we do not control for other observed or unobserved influencing factors. We then extend the specification by different possibilities to account for unobserved heterogeneity. Hence, we will start with the following specification:

$$\begin{aligned} S_{it} &= \beta_0 + \beta_1 60plus_{it} + \beta_2 20\min us_{it} + \beta_3 migration_{it} \\ &\quad + \beta_4 natural_{it} + \alpha_i + \eta_{jt} + \epsilon_{it} \end{aligned} \tag{16}$$

where S is the number of physicians, α_i are individual (regional) fixed effects and η_{jt} are state level time fixed effects, and ϵ_{it} is an error term. The term η_{jt}

mainly captures changes in the remuneration system, which will be arranged at the state level. Note that German states (Bundeslaender) comprise of a whole number of regions (Kreise).

The results are provided in table 1. Regressions (1)-(3) involve increasingly more controls for unobserved heterogeneity, where regression (1) contains no controls at all, regression (2) controls for regional and period fixed effects, and regression (3) controls in addition for fixed effects at state level. We consistently find that physician supply decreases in the share 60plus and increases in net migration and the natural balance. This corresponds well with our model, according to which population growth, fertility or migration driven, will generally increase the physician's future demand and, thus, provide incentives to locate within this region. A higher current share of older people, certainly leads to the expectation of lower future demand, a disincentive for location. Whether or not the treatment of older patients is relatively profitable or not cannot be inferred, only that possible current profits from treating older patients would be overcompensated by the expectation of future losses in demand and income.

table 1 about here

The effect of the age share 20minus is not robust with respect to changes in specification and changes sign from negative to positive once unobserved heterogeneity is controlled for. From the more reliable estimations (2) and (3) it follows that young populations provide a positive stimulus for physician supply. This may be for young patients being relatively profitable, but it also embraces the expectation that currently young populations guarantee a high demand well into the future. From these simple regressions it follows that both channels of demographic change, namely ageing and population decline, affects physician decisions negatively.[15]

In the following we add three elements to our estimation. Firstly, we now include a number of control variables to analyze if the estimated effects of the demographic change remain significant, if other determinants of physician supply are considered. Secondly, in table 1 we provide standard errors that are robust to heteroskedasticity and autocorrelation. We now calculate standard errors that are additionally robust to contemporaneous cross-sectional correlations in the error terms following Driscoll and Kraay (1998). According to Driscoll and Kraay spatial correlations among cross-sections may arise for a number of reasons, ranging from observed common shocks such as terms of trade oil shocks, to unobserved contagion or neigh-

[15] In table 1 as well as in subsequent tables we provide the results of the Maddala-Wu panel data unit root test. Although the test rejects the null hypothesis clearly, we do not that much attach importance to the results because of the relatively short time span and the missing information for 2000.

borhood effects.[16] Thirdly, we examine in greater detail the relationship between the demographic and geographic make-up of a region. We have argued before, that in particular the effect of age structure on physicians' location incentives may importantly be shaped by the degree of rurality. Therefore we additionally include interactions of the age groups with our proxy for rurality discussed in the data section.

One argument is, that it is more difficult for rural areas to attract or retain young people or those in the working age population. Hence, if ageing happens in these regions it seems to be obvious that this process is not easily reversed. Furthermore, we have argued in the theoretical section that the degree of rurality may affect the relative profitability of treating different age-groups. In particular, we have conjectured that the treatment of old - and similarly perhaps for very young - patients within a rural context arguably involves a larger share of provision by way of house visits. In particular in rural areas these are likely to be costly for physician due to long travelling times.

We now estimate

$$\begin{aligned} S_{it} &= \beta_0 + \beta_1 60plus_{it} + \beta 60plus_{it} \times rurality_i \\ &\quad + \beta_3 20\min us_{it} + \beta_4 20\min us_{it} \times rurality_i \\ &\quad + \beta_5 migration_{it} + \beta_6 natural_{it} + \varphi' X_{it} + \alpha_i + \eta_{jt} + \epsilon_{it} \end{aligned} \tag{17}$$

where, in addition to specification (**??**), we have included the interaction of population share 60plus with the rurality variable ($60plus_{it} \times rurality_i$), the interaction of population share 20minus with the rurality variable ($20\min us_{it} \times rurality_i$), and a vector of control variables, X_{it}.

The models presented in table 2 differ with respect to the restriction $\beta_2 = \beta_4 = 0$ in the regression (1). According to regression (1) ageing affects physician supply negatively, which is in accordance with the results presented in table 1. The additional interaction with rurality in regression (2) shows that the negative effect of the population 60plus on physician supply is particularly pronounced for rural areas. In fact, within an urban context the reverse may well be true, as the pure effect of the share 60plus is now positive. As the (negative) effect of ageing on the expected future demand should not vary too much with the regional context, this strongly hints at the fact that rurality reduces the relative profitability of treating old patients. As is well known from the empirical literature, older patients (60plus) exhibit far higher consultation rates and, thus, generate a higher

[16] Driscoll and Kraay (1998) argue that the presence of such spatial correlations in residuals complicates standard inference procedures that combine time-series and cross-sectional data since these techniques typically require the assumption that the cross-sectional units are independent. When this assumption is violated, estimates of standard errors are inconsistent, and hence are not useful for inference.

demand for physicians (see e.g. Pohlmeier and Ulrich 1995, Dusheiko et al. 2002, Dormont et al. 2006, Juerges 2007). Whereas meeting this demand appears to be relatively profitable within an urban context, it is unprofitable within rural settings. According to our results the relative profitability of treating old patients is given only in regions with a very high population density and cities with more than 100,000 inhabitants.

table 2 about here

When taken across all regions alike, the share of the young population 20minus does not have a significant effect on physician supply (see regression (1)). However, if we consider in addition the interaction with rurality, we find that a high share of the young is most attractive in urban regions: whereas the direct effect of the share of 20minus is significantly positive, the interaction with rurality is significantly negative. The positive direct effect is consistent with a young population being a good indicator for a high future demand. The fact that the attractiveness of young patients weakens within rural areas either hints at the fact that treating the higher demand from young ages within a rural context exposes the physician again to additional costs.[17] Alternatively, it may hint at the expectation that young populations may not stay within rural areas but rather migrate elsewhere. Based on our estimates the relative profitability of treating young patients is true for the subaggregate agglomeration regions.

Net migration and natural balance are significant positive in both specifications. These two variables being proxies for the expected future demand, confirms once more that physician supply is driven not only by the size and age-structure of the current population but also by the expectations about future population structure.

The main reason for why we can interpret the two population flow effects in terms of expectations is that population size has a significant positive effect in of itself. As such, this is not surprising since the number of physicians is related to the number of inhabitants in the region. But to the extent that changes in population over time are measured directly, the population balance would suggest that at least one of the three determinates - population size, net migration or natural balance - should turn out to be insignificant. The fact that all three variables are significant then hints at the additional role of expectation over and above the mere population accounting. In section 4.3.3 we apply a dynamic model that allows us to conclude whether this effect is important even in the short run. In this case, we argue, that expectations play only a minor role.

[17] The studies on the intensity of use of physician services by Pohlmeier and Ulrich (1995), Dusheiko et al. (2002), Dormont et al. (2006) and Juerges (2007) consistently reveal a U shaped age-pattern, thus suggesting a relatively high demand for physician services both for the old and youngest ages.

According to the literature the effect of population density is positive, because higher densities indicate on average, a lower need for travelling either by patients or by physicians and thus a higher demand and/or greater profitability of provision. In our case the corresponding parameter is not significant different from zero. One reason for this finding might be that we control for the age distribution of the population, which is, in turn, correlated with population density. Another explanation is that population density in East Germany decreases in nearly all regions in the considered period. Physicians, however, do not abandon their practice promptly, since the location choice is a long-run decision and in most cases they persevere until they retire.

For the GDP per capita and the employment rate we argue that they are a proxy for the standard of living and the sound condition of the labor market. In addition, higher per capita income might also be related to a larger share of private patients, which are more attractive to physicians. At first glance, the estimated effect for the unemployment rate might be surprising, because, for example, from a labour economists perspective a negative effect would be due. However, in the health economics literature the unemployment rate is a proxy for morbidity in the working age population and hence, the expected effect is positive. The same applies to the share of welfare recipients. Hence, the effects are as expected and in accordance with the literature.[18].

For the share of foreigners we find no significant effect, nor for the share of upper secondary school leavers. As expected, the share of school leavers without a formal education affect the number of physicians in a region negatively. In as far as this hints at a relatively poor educational environment, it would reduce the residential utility of physicians who care for the development of their own children. The two variables that measure the development of new buildings for living should capture the attractiveness of regions. The share of new houses is related only to one and two family houses and is therefore directly related to an increase in the number of families. The share of new flats in the stock of flats measures the general activities of housebuildung. In as far as such construction activities hint at the simultaneous development of a regional infrastructure (schools, shopping, etc) they provide an additional measure of a region's attractiveness. In a far as development is in particular directed at families, this may provide an additional hint that such a region may be particularly attractive for young physicians (upon the point of their location choice).

The capacity of tourist accommodation hints at the attractiveness of such a region both in general and for holiday makers. The effect on physician supply can be expected to be positive, both because the region is attractive

[18] See Stewart (2001) for a relatively recent survey on the positive correlation between unemployment and morbidity.

and because of the additional demand generated by holiday makers. Finally, cars per 1000 population acts as a measure of geographical mobility. On the one hand, mobility should have a positive effect on physician supply: potential patients find it easier to visit the physician and, thus, exhibit a greater (expected) contact frequency. On the other hand, given that we control for the population size and density as well as for the age structure this variable measures the possibility to consult physician outside the region, e.g. in order to visit a specialist within a city rather than the local general practitioner. In this case the expected effect is negative, and it seems that this effect is stronger in our estimates.

Since the main focus in our analysis is on the age structure, we now take a closer look at the overall effect of the population shares including the rurality interaction effect according to our estimates presented in table 2. Including this interaction effect the population share of those 20 years and younger affects physician decisions positively in agglomeration regions, but negatively in urbanised and rural areas. The overall effect (including the interaction) of the population share 60plus has no significant effect in independent cities with more than 100,000 inhabitants. In any other region of the applied classification the population share of those who are 60 years and older is negatively related to the local decision of physicians. Figure 5 shows the results graphically. The light yellow areas are the independent cities with more than 100,000 inhabitants, the region where the share 60plus has no effects on the supply of physicians. The light and dark yellow areas are the agglomeration regions, in which the share of 20minus is positively related to physician supply. The orange (urbanised regions) and red coloured areas (rural areas) are those who suffer from an increase of both shares, related to the local decision of physicians.

figure 5 about here

In 2004 about 54% of German citizen live in the orange and red coloured regions. Hence, half of the German population is danger of being potentially under-doctored, because of the unfavourable age structure effect on physicians local decision. It is important to not, however, that it is not the goal of the paper to identify under-doctored regions. Rather our objective is to analyse the causes for regional inequality in physician supply with a special emphysis in demographic variables.

4.3 Further Estimates

4.3.1 Physician Density

As mentioned in section 4.2, we will provide two further models that should help to assess the importance of our results. In the first case we estimate the specifications of equation (17), but with physician density as dependent variable. As mentioned in the data section, it is rather common in the literature to use this variable as a dependent instead of the number of physicians. The subsequent analysis thus makes our results more comparable with other studies and allows to analyse if our general conclusions will change.

table 3 about here

Using the physician density as dependent variable in equation (17) yields the results displayed in table 3. With respect to the effects of the population age structure we get different results. The population share 60plus is not significant in both specifications, and the share of the young population is no longer significant in regression (2), compared to regression (2) in table 2. The population flow variables, net migration and natural balance, are still significant, although the population size is not considered on the right hand side of the equation. We can conclude that the measurement bias in the dependent variable is apparently large enough to affect our results (e.g., share 60plus, share 20minus, and population density). Hence, the general conclusions based on the results provided in table 5 differ significantly from that based on the preferred model presented in section 4.2.

4.3.2 Spatial Effects

Related to section 4.2 it is possible that physicians consider in their regional choice decision the provision of physician services in the neighbor regions. Put differently, market tightness in a preferred region can direct the decision on the final location choice. In this case one can expect a positive relation between the number of physicians (or physician density) in neighbor regions and the local region.

Although we are principally interested in long run effects, we consider this alternative approach, because it additionally allows us to differentiate between expectations concerning the future and short-term utility/profit maximisation. Put differently, we argue that only the latter effect should remain significant in this specification. In addition, it is important to analyze the consequences of existing spatial effects for the measured effects of the population age structure, identified in section 4.2.

To analyze spatial effects econometrically it is necessary to consider also the time dimension. Otherwise the estimated effects would be biased because of the omitted correlation between spatial and time lagged effects. Hence, we have to choose a spatial and time dynamic model. This allows to analyse both short-term effects of the demographic determinants in the local area and long-term effects of the market for physicians in the local and surrounding regions. Put differently, the dependent variable will be explained by a local (time) and spatial lagged effect, and deviations from these effects due to changes in the exogenous variables.

In order to generate spatially lagged counterparts of the dependent variable, we construct a spatial weight matrix indicating the contiguity of regions. We define contiguity between two regions as regions sharing a common border. The corresponding spatial weight matrix W is therefore a symmetric 439×439 matrix.

Usual fixed effects estimators that include spatial and time dynamic effects of the dependent variable are biased.[19] On this account, we use a spatial and time dynamic data approach with both regional and time fixed effects as suggested by Lee and Yu (2007) and Yu et al. (2008). In this case, the parameters for the time lagged and spatial lagged values of the dependent variables will be estimated using a quasi-maximum likelihood estimator that is extended by a bias correction. To avoid biased estimates for the lagged effects of the dependent variables, Lee and Yu (2007) developed a data transformation approach that has the same asymptotic efficiency as the quasi-maximum likelihood estimator when n is not relatively smaller than T.

The model has the following general specification:

$$S_{nt} = \gamma_0 S_{n,t-1} + \lambda_0 W_n S_{nt} + X_{nt}\beta_0 + \alpha_{n0} + \eta_{t0} l_n + E_{nt} \qquad (18)$$

where $S_{nt} = (s_{1t}, s_{2t}, \ldots, s_{nt})'$ and $E_{nt} = (\epsilon_{1t}, \epsilon_{2t}, \ldots, \epsilon_{nt})'$ are $n \times 1$ column vectors and the residual ϵ_{it} is i.i.d. across i and t with zero mean and variance σ_0^2. W_n is an $n \times n$ spatial weights matrix which is nonstochastic and generates the spatial dependence between cross sectional units s_{it}, X_{nt} is an $n \times k_x$ matrix of nonstochastic regressors, α_{n0} is $n \times 1$ column vector of individual fixed effects, η_{t0} is a scalar of time effect and l_n is $n \times 1$ column vector of ones. W_n is row normalized from a symmetric matrix, which ensures that all weights are between 0 and 1, and weighting operations can be interpreted as an average of the neighboring values.

We will use this model for both dependent variables the number of physicians and the physician density. Hence, the left hand side variable Y in equation (18) equals these variables. $S_{n,t-1}$ is the time lagged dependent variable and $W_n S_{nt}$ is the spatial lagged dependent variable. The matrix X_{nt} includes all right hand side explanatory variables and hence, also those of main interest: $60plus_{it}$, $60plus_{it} \times rurality_i$, $20\min us_{it}$, $20\min us_{it} \times rurality_i$, $migration_{it}$, and $natural_{it}$. In addition the matrix contain the average spatial lagged population size, when the number of physicians is the dependent variables.

Table 4 provides the results using this estimation technique with the number of physicians as well as the physician density as dependent variables. We only show the results for the important variables.[20]. While regressions (1) and (3) are the specifications without the interaction of age groups and

[19] See, for example, Nickell (1981) with respect to the asymptotic bias of OLS estimation using the time lagged effect and, for example, Kelejian and Prucha (1998) for biased OLS estimates when spatial lagged effects are considered.

[20] Complete results including the remaining control variables are available upon request.

rurality, regressions (2) and (4) contain these interaction for the shares of the young and old in the regional population.

The time lagged effects has to be interpreted carefully, because of two reasons. First, the considered period is short and second, and probably more problematically, we have no information on the number of physicians and physician density for 1999, as mentioned in the data section. However, our interest is directed towards the spatial effect, and a missing year is not problematically here.

The spatial lagged effect of the dependent variable is significant positive in the estimates for physician density, but it is not when the number of physicians is considered as dependent variable. This means that the physician density in a particular region is positively related to the physician density in neighboring districts. One could argue that high densities of physicians in neighboring regions are likely to be associated with strong competition there. However, as argued in section 4.2 physician density comprises a measurement error if physician location choices are made long term. With respect to the spatial lagged effect the expected bias is positive. Hence, we must conclude that the insignificant effect in regressions (1) and (2) provide unbiased evidence that physician decisions are not related to physicians in neighboring districts.

As mentioned in the data section, more attractive regions have a higher probability of being afflicted with a physician sattlemtent-stop for certain specialisations. Hence, the spatial effects on the distribution of physicians are potentially overestimated. Since the effects are significant only for physician density and not for the number of physicians, we can carefully conclude that the bias is small. Given this is true, we therefore can also conclude that the estimated effects for the causes of regional unequal distribution of physicians (table 2) suffer from a negligible bias, too. However, in case of doubt these parameters are underestimated which means that unbiased estimates would yield even stronger effects.

Related to the number of physicians the effects of changes in population shares are qualitatively similar to that in table 2, with the exception of the significance level of the pure share effects in regression (2). In comparison with table 3 we see that the share 20minus is significant at the 1% level, when the interactions are excluded. However, the are no qualitative differences when the interactions with rurality are considered.

Hence, we can conclude that the young as well as the old are negatively associated with rural areas. That is, the higher the rurality level is the relative more profitable is the working age population. This is consistent with our previous findings. The difference to the results in table 2 allows to conclude that the share of young affects expectations about future income. In contrast to this, the share of the old seems to be primary correlated with physicians expectations concerning short-term utility/profit maximisation.

table 4 about here

For net migration and natural balance we find that the latter is significant only in regressions (3) and (4) and the former only in regression (2). In addition, all parameters have declined and have a lower significance level, compared to table 2 and table 5. A careful conclusion is that the population flows thus measure in fact mainly expectations. For regression (2) we can conclude further that net migration has a somewhat shorter horizon than natural balance. In (3) and (4) we cannot argue in a similar way due to biased estimates: negative net migration leads to population decline and this, in turn, increases the physician density in the short run. In fact, net migration and physician density are negatively correlated ($r = -0.1$). For natural balance we could argue in the same. However, in contrast to net migration the correlation between natural balance and physician density is to low for an apparently statistical bias in the estimates. The effects of natural balance is significant in (3) and (4) because the parameters of the covariates in the physician density regression exhibit lower standard errors, compared to the regressions with the number of physicians as dependent variable.

5 Conclusions

Population ageing is widely expected to come with an increased per capita demand for (ambulatory) physician services. Everything else equal regions with high population shares of old persons should then be particularly attractive on economic grounds for the location of physician practices and should therefore exhibit high physician to population ratios.

At the regional level in Germany, ageing is not only shaped by low or declining fertility rates but also by outmigration. This is because younger people exhibit a higher regional mobility than older people. Hence, ageing may be accelerated by outmigration in particular as it becomes progressively more difficult to attract young people into regions in which the average age is increasing towards (or at) high levels. According to our data the largest population share of the age cohort 60plus in 2004 lives in the region Hoyerswerda (33.4%) followed by the region Görlitz (32.9%); two regions in East Germany. Between 1995 and 2004 the share of 60plus has increased by 89% in Hoyerswerda and 41% in Görlitz. However, there are other regions, in particular in the Eastern part of Germany, that experience ageing in a similar manner. According to the results, physician supply is negatively related to the population share of 60+ within rural areas, while it is not significantly related to the share within urban regions. Hence, many rural regions with a high share of the elderly population are in particular danger of being under-doctored in the future, if the identified effects in our regressions are in fact a causal.

Additionally, we adopt the hypothesis that demographic and geographic characteristics do not determine in isolation the attractiveness of a region from a physician's perspective but interact in a particular way. More specifically, we posit that the high demand for services by elderly patients may be less attractive for a physician if it has to be served within a rural context. Long travel times and poor availability of public transport deter frail elderly patients from attending the physician's practice, implying a high demand for home visits as compared to an urban context. Furthermore, due to the long travel times the provision of home visits is more costly for the physician in rural areas. Thus, a high demand from old populations may be served at a relatively low cost within an urban region but only at a high cost within a rural context. Given that the interaction of age-related population shares and the rurality variable (in nine levels) is adequate to measure this effect, we find that the share of the elderly has an increasing negative effect on physician supply, the higher the level of rurality.

At regional level, demographic developments and geographical location imply a framework with both an intertemporal and a spatial dimension. Given that physicians' location choices are mostly long-term decisions, the intertemporal aspect implies that not only the current population structure will matter but also the expected population. In fact, we find that net population growth (births – deaths) and net migration to constitute significant determinants for physician supply besides the population and its age structure in of themselves. Furthermore, spatial interactions are not important. More specifically, the econometric results do not confirm our hypothesis whereby the propensity to locate in a particular region is positively related to the number of physicians in neighbouring districts.

The age structure of physicians is characterised by large baby-boomer cohorts. In 2007 the share of baby-boomers (between 42 and 65 years old) is 85.7 %. Hence, ageing happens to physicians, too. From this it follows that a large number of physicians will retire in the next 15 years and this will touch the medical provision in rural areas and in particularly in the eastern part of Germany. According to the German Medical Association about 13% of the East German regions are medically "underprovided" in 2007, and more than half of the hospitals in East Germany have problems to fill vacancies in medical employment. According to our results demographic change will amplify these problems, as it is known from statistical projections that population decline and population aging in these regions will go on for the next 20 years.

Hence the important question is, how to attract new and young physicians to rural areas? Since the speed of population ageing is much faster in rural regions, policy should provide incentives to move into these regions not only on equity grounds but also because a worsening provision of basic goods and services, such as health care, may reduce even further the development prospects (e.g. by way of job creation) for such regions. In addition, many rural regions in East Germany have a rural contiguity neighbor, which means that the positive spatial effect disappears here.

6 References

Dormont, B., Grignon, M., Huber, H., 2006, Health expenditure growth: reasssessing the threat of ageing, *Health Economics* 15, 947-963.

Driscoll, J.C., Kraay, A.C., 1998, Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data, *Review of Economics and Statistics* 80, 549-560.

Dusheiko, M, Gravelle, H., Campbell, S., 2002, Inequality in Consultations with General Practitioners, CHE Technical Paper 24.

Gravelle, H., 1999, Capitation contracts: access and quality, *Journal of Health Economics* 18, 315-340.

Hingstman, L., Boon, H., 1989, Regional Dispersion of Independent Professionals in Primary Health Care in the Netherlands, *Social Science & Medicine* 28, 121-129.

Juerges, H., 2007, Health Insurance Status and Physician-Induced Demand for Medical Services in Germany: New Evidence from Combined District and Individual Level Data, DIW Discussion Paper 689.

Kelejian, H.H., Prucha, I.R., 1998, Estimation of Spatial Regression Models with Autoregressive Errors by Two-Stage Least Squares Procedures: A Serious Problem, *International Regional Science Review* 20, 103-111.

Kopetsch, T., 2007, Arztdichte und Inanspruchnahme aerztlicher Leistungen in Deutschland. Eine empirischeUntersuchung der These von der angebotsinduzierten Nachfrage nach ambulanten Arztleistungen [Physician density and use of physician services in Germany. An empirical investigation into the thesisof supplier-induced demand], *Schmollers Jahrbuch.*

Kopetsch, T., Munz, H., 2007, Analyse des Niederlassungs- und Wanderungsverhaltens von Vertragsaerzten und -psychotherapeuten in Deutschland [Analysis of location and migration choices by registered physicians in Germany], Berlin: Kassenaerztliche Bundesvereinigung.

Kraft, K., von der Schulenburg, J.-M., 1986, Co-Insurance and Supplier-Induced Demand in Medical Care: What Do We Have to Expect as the Physician's Response to Increased Out-of-Pocket Payments, *Journal of Institutional and Theoretical Economics* 142, 360-379.

Lee, L.-F., Yu, J., 2007, A Spatial Dynamic Panel Data Model with Both Time and Individual Fixed Effects, Ohio State University, Department of Economics, mimeo.

Nickell, S.J., 1981, Biases in Dynamic Models with Fixed Effects, *Econometrica* 59, 1417-1426.

Newhouse, J.P., Williams, A.P., Bennett, B.W., Schwartz, W.B., 1982, Does the geographical distribution of physicians reflect market failure?, *Bell Journal of Economics* 13, 493-505.

Nuscheler, R., 2003, Phyisian Reimbursement, Time Consitency, and the Quality of Care, *Journal of Institutional and Theoretical Economics* 159, 302-322.

Pohlmeier, W., Ulrich, V., 1995, An Econometric Model of the Two-Part Decisionmaking Process in the Demand for Health Care, *Journal of Human Resources* 30, 339-361.

Stewart, J., 2001, The impact of health status on the duration of unemployment spells and the implications for studies of the impact of unemployment on health status, *Journal of Health Economics* 20, 781-796.

Yu, J.; de Jong, R.; Lee, L.-F., 2008, Quasi-Maximum Likelihood Estimators for Spatial Dynamic Panel Data with Fixed effects when Both n and T are Large, *Journal of Econometrics* 146, 118-134.

7 Appendix

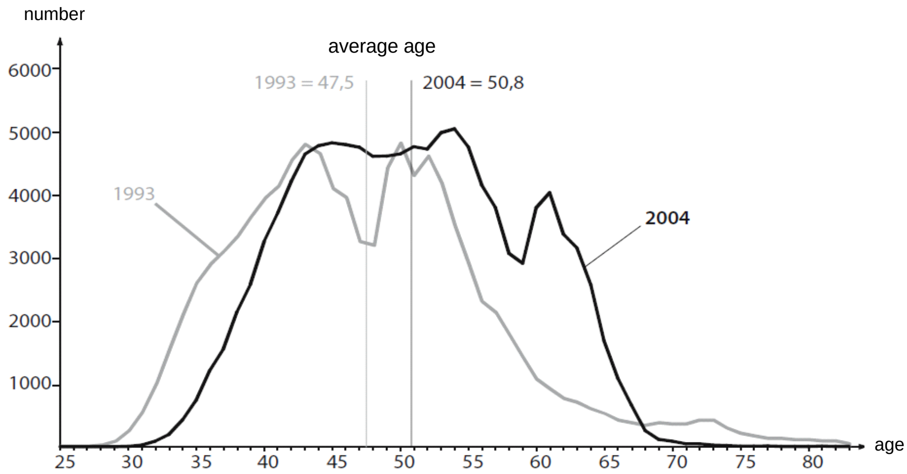

Figure 1: Age Structure of SHI-Physicians in 1993 and 2004 (source: SHI)

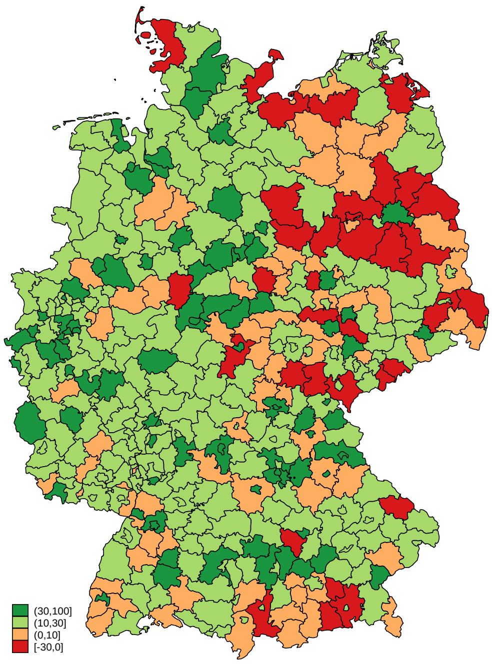

Figure 2: Change in Regional Physician Supply in Germany

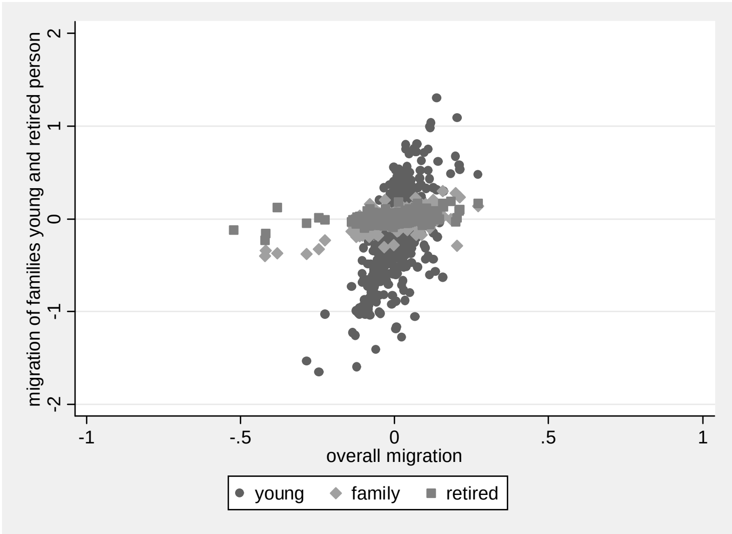

Figure 3: Net Migration Flows by Age Groups

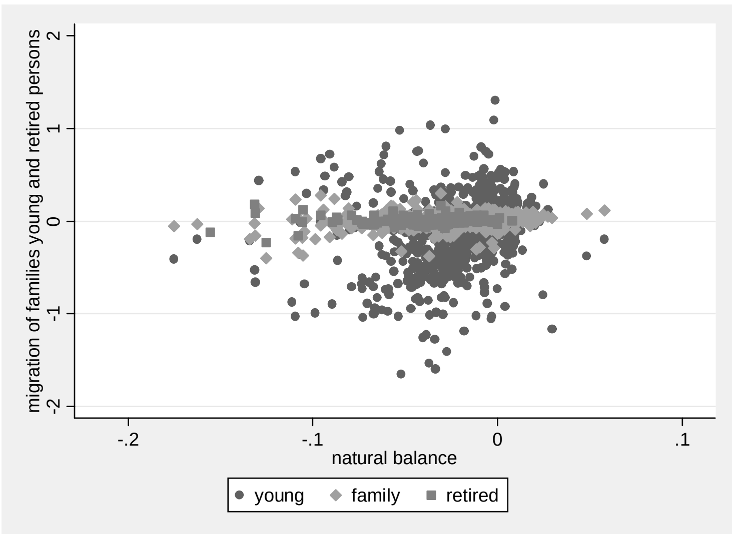

Figure 4: Natural Balance and Migration Flows

Table 1: Demographic Change and the number of Physicians - Basis Specification

	(1)	(2)	(3)
share 60plus	-32.730‡	-6.582‡	-5.570‡
	(5.787)	(1.611)	(1.470)
share 20minus	-79.004‡	12.106‡	6.649†
	(7.813)	(1.681)	(2.805)
net migration	1.148	2.185‡	1.755‡
	(0.769)	(0.240)	(0.225)
natural balance	18.082‡	16.634‡	12.509‡
	(2.714)	(2.014)	(1.517)
regional FE		✓	✓
time FE		✓	
state level time FE			✓
R^2	0.101	0.233	0.699
Maddala-Wu test	1304.9‡	1211.0‡	1257.3‡

Notes: Dependent variable: number of physicians; number of observations 3951; robust standard errors in parenthesis; Maddala-Wu test: unit root test with the null hypothesis of non-stationarity; ‡ 1% significance level; † 5% significance level.

Table 2: Demographic Change and the Number of Physicians

	(1)		(2)	
share 60plus	-5.320‡	(2.083)	4.131†	(1.796)
share 60plus×rurality			-2.216‡	(0.243)
share 20minus	1.489	(2.708)	15.17‡	(4.344)
share 20minus×rurality			-3.264‡	(0.304)
net migration	1.297‡	(0.099)	1.389‡	(0.114)
natural balance	8.930‡	(2.594)	7.892‡	(2.346)
population size	0.002‡	(0.0002)	0.002‡	(0.0002)
population density	0.053	(0.135)	0.009	(0.132)
GDP per capita	0.999‡	(0.308)	0.748‡	(0.288)
employment rate	2.452‡	(0.525)	2.112‡	(0.487)
unemployment rate	1.268‡	(0.353)	1.204‡	(0.276)
share of welfare recipients	0.743‡	(0.140)	0.680‡	(0.132)
share of foreigners	-0.390	(0.578)	0.320	(0.525)
up. second. school leavers	-0.268	(0.244)	-0.399	(0.278)
no formal education	-1.009‡	(0.238)	-0.998‡	(0.259)
share of new houses	1.287‡	(0.175)	1.191‡	(0.171)
new flats	0.564‡	(0.180)	0.414‡	(0.155)
tourist accommodation	0.073‡	(0.022)	0.047	(0.027)
cars per habitants	-0.073‡	(0.021)	-0.067‡	(0.017)
individual FE	✓		✓	
state level time FE	✓		✓	
R^2	0.732		0.738	
Maddala-Wu test	1820.8‡		1785.9‡	

Notes: Dependent variable: number of physicians; number of observations 3951; Driscoll & Kraay robust standard errors in parenthesis; Maddala-Wu test: unit root test with the null hypothesis of non-stationarity; ‡ 1% significance level; † 5% significance level.

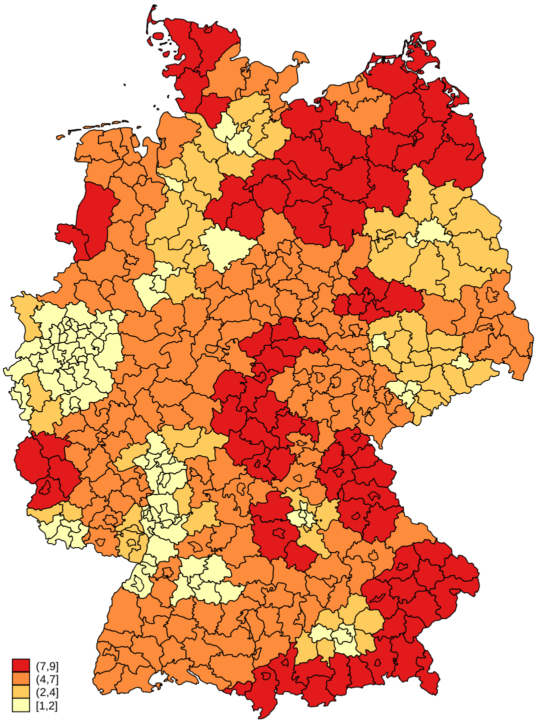

Figure 5: Age Structure of the Population and Physician Supply

Table 3: Demographic Change and Physician Density

	(1)		(2)	
share 60plus	-0.415	(0.522)	0.582	(0.323)
share 60plus×rurality			-0.238‡	(0.069)
share 20minus	-1.172	(1.680)	1.610	(1.700)
share 20minus×rurality			-0.542‡	(0.087)
net migration	0.457‡	(0.031)	0.475‡	(0.032)
natural balance	2.849‡	(0.775)	2.804‡	(0.768)
population density	-0.059†	(0.024)	-0.066‡	(0.025)
GDP per capita	0.603‡	(0.151)	0.587‡	(0.158)
employment rate	0.291	(0.168)	0.242	(0.166)
unemployment rate	0.451‡	(0.116)	0.408‡	(0.107)
share of welfare recipients	0.069	(0.039)	0.063	(0.039)
share of foreigners	1.397‡	(0.210)	1.491‡	(0.224)
up. second. school leavers	0.096	(0.107)	0.090	(0.113)
no formal education	-0.027	(0.077)	-0.025	(0.079)
share of new houses	0.435‡	(0.037)	0.432‡	(0.038)
new flats	0.205‡	(0.065)	0.178‡	(0.062)
tourist accommodation	0.058‡	(0.006)	0.054‡	(0.005)
cars per habitants	-0.042‡	(0.009)	-0.040‡	(0.009)
regional FE	✓		✓	
state level time FE	✓		✓	
R^2	0.696		0.698	
Maddala-Wu test	1541.2‡		1588.4‡	

Notes: Dependent variable: physician density; number of observations 3951; Driscoll & Kraay robust standard errors in parenthesis; Maddala-Wu test: unit root test with the null hypothesis of non-stationarity; ‡ 1% significance level; † 5% significance level.

Table 4: Spatial and Time Dynamic Model

dependent variable	physicians		physician density	
	(1)	(2)	(3)	(4)
time lagged dep. var.	0.972‡	0.956‡	0.872‡	0.869‡
	(0.012)	(0.012)	(0.013)	(0.013)
spatial lagged dep. var.	0.023	0.014	0.076‡	0.050‡
	(0.013)	(0.013)	(0.019)	(0.019)
share 60plus	-3.027‡	0.480	-0.377	0.495
	(0.985)	(1.247)	(0.285)	(0.359)
share 60plus×rurality		-0.813‡		-0.200‡
		(0.174)		(0.050)
share 20minus	-1.408	4.354	-0.945†	0.639
	(1.706)	(2.398)	(0.491)	(0.702)
share 20minus×rurality		-1.271‡		-0.340‡
		(0.305)		(0.089)
net migration	0.144	0.199†	-0.020	-0.010
	(0.096)	(0.096)	(0.029)	(0.029)
natural balance	0.416	0.2277	0.495†	0.425†
	(0.746)	(0.748)	(0.217)	(0.220)
population size	0.001	0.001		
	(0.001)	(0.001)		
spatial lagged population size	0.002	0.002		
	(0.001)	(0.001)		
additional controls	✓	✓	✓	✓
regional FE	✓	✓	✓	✓
state level time FE	✓	✓	✓	✓

Notes: Number of observations 3951; standard errors in parenthesis; ‡ 1% significance level; † 5% significance level.

Table 5: Summary Statistics

	Mean	Std. Dev.	Min	Max
physicians	271.53	433.10	51	7894
physician density	141.96	47.54	68	391
share 60plus	23.44	2.72	13.8	33.4
share 60plus×rurality	126.24	61.68	18.5	271.8
share 20minus	21.38	2.32	15	29
share 20minus×rurality	116.89	57.84	15	239.4
net migration	2.16	7.66	-43.1	57.2
natural balance	-1.63	2.66	-10.2	7.1
population size	186160.3	214296.7	35499	3471418
population density	509.49	656.79	40	4024
GDP per capita	22.77	9.32	10.1	85.4
employment rate	48.11	15.29	20.9	139.1
unemployment rate	11.66	5.34	3.0	31.4
share of welfare recipients	28.45	16.41	3.4	138
share of foreigners	6.95	4.85	0.1	28.9
up. second. school leavers	22.22	7.94	0	52.2
no formal education	9.28	2.70	1.4	26
share of new houses	88.80	9.81	29.2	100
new flats	12.31	7.42	0	69.3
tourist accommodation	36.38	49.71	0.6	581.1
cars per habitants	528.09	51.65	350	959
rurality	5.39	2.52	1	9

Notes: Number of observations 3951.

Table 6: Rurality

Type I: Agglomeration Regions	
1	Independent Cities with more than 100,000 inhabitants
2	Districts with at least 300 inhabitants per square kilometer
3	Districts with at least 150 inhabitants per square kilometer
4	Districts with less than 150 inhabitants per square kilometer
Type II: Urbanised Regions	
5	Independent Cities with more than 100,000 inhabitants
6	Districts with at least 150 inhabitants per square kilometer
7	Districts with less than 150 inhabitants per square kilometer
Type III: Rural Areas	
8	Districts with at least 100 inhabitants per square kilometer
9	Districts with less than 100 inhabitants per square kilometer

Notes: The criteria for Type I regions is that they have a concentrated hinterland. Type III regions are defined by a low number of inhabitants per square kilometer. The remaining regions are merged to Type II areas. In contrast to the Type III regions they have a higher urbanisation degree, a rudimental metropolitan centre, and a higher density.