Thünen-Series of Applied Economic Theory

Thünen-Reihe Angewandter Volkswirtschaftstheorie

Working Paper No. 105

Demographic and Geographic Determinants of Regional Physician Supply

by

Michael Kuhn and Carsten Ochsen

Universität Rostock

Wirtschafts- und Sozialwissenschaftliche Fakultät
Institut für Volkswirtschaftslehre
2009

Demographic and Geographic Determinants of Regional Physician Supply

Michael Kuhn[*]   Carsten Ochsen[†]

July 17, 2009

Abstract

Against the backdrop of an ongoing debate in most countries about the geographic (mal-)distribution of physician practices, we develop a theoretical and empirical framework to analyze how physician supply at regional level depends on demographic (population size, age structure, fertility and migration) and geographic determinants. Particular attention is given (i) to local population change as a predictor of future demand for physician services, (ii) to the way in which the age-structure of the (potential patient) population and regional structure interact in shaping the profitability of treating the local population, and (iii) to cross-regional correlations in physician supply. Using regional data for Germany, we examine econometrically the determinants of regional physician supply. We find it to be negatively related to both the population share 60+ and the population share 20- in rural areas. While both population shares tend to have a less negative impact in urban areas, a pronounced positive effect arises only for the share 20- in regions with agglomeration character. Net migration and natural balance turn out to be significant positive as long-run determinants only, indicating thus their role as predictors of future demand. On average, cross-regional spillovers in demand do not seem to be important determinants of regional supply.

Keywords: age structure, physician supply, regional population ageing, regional migration, overlapping generations, 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

The (right) geographical distribution of physicians keeps on being an issue for debate in health care systems as varied as the managed care system in the US, the UK's National Health Service and the German health care system with its corporatist structure.[1] In Germany, this debate has seen a swing of tide during the late 1990s and early 2000s. Set against the general view of ubiquitous excess supply - at least in Western Germany, there is now a growing flow of reports about a declining number of office-based physicians at regional level. This comes with increasing concerns about physician shortages in rural regions, mostly but not exclusively in Eastern Germany. Initially, the decline in physician numbers was limited to general practitioners, but this trend now extends to other specializations. According to some projections there may be substantial under-supply in rural areas if this development continues.[2] Over the time span 1995-2004, however, the number of office-based physicians has continued to increase by about 30 % at the aggregate level, suggesting rather large shifts in the regional distribution of physicians. It is the aim of this study to identify some of the causes underlying this development and, in particular, to understand the role of regional population change. At a more general level, this study seeks to shed some light on the way in which the demographic and geographic make-up of a region (and its neighbors) shapes the local supply of physicians over time.

Population ageing is widely expected to come with an increased per capita demand for (ambulatory) physician services. According to a naive argument regions with high population shares of elderly persons should then be particularly attractive on economic grounds for physician practice and should therefore exhibit high physician densities (i.e. high numbers of physicians per capita). However, cursory evidence suggests that this may not be the case. In Germany it appears that rural regions with a high share of the elderly population are in particular danger of being under-doctored. In these regions ageing may therefore lead to a widening gap between the demand and supply of health care, a situation potentially warranting policy intervention. Similar concerns are raised in other countries, e.g. in the US (Cooper et al. 2002).

We construct an overlapping generations model in which physicians commit to a practice location when young and then provide services to the local population over their working lives. Within this model we identify three particular reasons for why the view that regions with large elderly populations

---

[1] Some recent contribution are Kopetsch (2004) and Klose and Uhlemann (2006) for Germany; Hann and Gravelle (2004) and Elliott et al. (2006) for England (and Wales); Iversen and Kopperud (2005) for Norway; and Cooper et al. (2002) and Rosenthal et al. (2005) for the US.

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

are attractive for physicians may be too simplistic. First, the profitability of offering services to different age-groups does not only vary with the demand from these groups but also with the profit margin earned on each patient. The latter depends on the reimbursement rate and the marginal treatment cost, which likely varies with age. If reimbursement rates are imperfectly adjusted to differences in treatment costs and if the treatment of older patients is more costly, then it is no longer clear that a larger share of elderly patients raises profits. Furthermore, it is likely that treatment costs and demand are not only sensitive to age but that this sensitivity varies systematically with the regional circumstances. For instance, long travelling distances and poor availability of public transport within rural (as opposed to urban) settings may render the treatment of elderly (and frail) patients more costly, both as more home visits are required and as the rendering of a home visit is more costly for the physician. Thus, we expect the relative profitability of different age-groups to vary with the position of a region on the urban-rural spectrum.

Second, as physicians typically set up their location for a long period - in many instances for the remainder of their working lives - their location choice should not only reflect current but also future profit opportunities. Physicians' expectations about the future demand for their services are strongly linked to population change. Hence, regions are attractive if they experience population growth through births or migration. In contrast, regions with a high share of the elderly may turn out to be less attractive prospects. Although they promise high current levels of demand, all other demographic rates equal, they suggest a decline in demand.

Third, the supply of physician practices is prone to vary with the entry pattern of physicians over time and space. We examine how the pattern of entry shapes the impact of expectations on the supply of practices and show that the expectations about future population developments matter only when the pattern of physician entry is cyclical.

Finally, we would expect that spatial interactions play a role in shaping regional physician supply. For instance, one could argue hat the propensity to locate in a particular region is positively related to the supply of physicians in neighboring districts. Tough competition within a particular region should increase the relative attractiveness of neighboring regions. However, countervailing effects arise in the presence of patient mobility, where a large number of physicians in a region may attract patients from neighboring regions and, thus, reduce the local demand there. We should stress that a detailed analysis of the spatial interactions stretches well beyond the scope of this paper, which is more focused on the role of population structure and population flow. Thus, we omit cross-regional spillovers from our theoretical analysis. However, by employing spatial estimation techniques we seek to control for these effects in our empirical analysis.

We use panel data at regional level to examine empirically the relation-

ship between the local physician supply and its demographic and geographic determinants. Using annual data for 439 regions at the district level in Germany for the period 1995 to 2004 we examine how the population share of individuals aged 60+ and the share of individuals aged 20- affect the local supply of physicians. 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 data estimator to analyze the effects of the age structure of the population, population flows, and population shrinking. In addition, we apply a spatial and time dynamic panel data estimator to find out if spatial interactions in physician supply are of importance.

Our results show that in rural areas physician supply is negatively related to both the population share 60+ and the population share 20-. Whereas both population shares tend to have a less negative impact in urban areas, a pronounced positive effect arises only for the share 20- in regions with agglomeration character. Our results also indicate that expectations about future population, operationalized by population flows, drive the current supply of physicians. Both net migration and the natural balance (i.e. the difference between births and deaths) are positively related to the current number of physicians resident in a region even when controlling for the current size and density of the regional population. For the dynamic specification with time and spatial lagged variables, we find that the effects of the population shares remain more or less in place but that the population flows now turn out insignificant. This is consistent with the population flows representing expectations about future demand. Furthermore, this result is consistent with the highly cyclical pattern of physician entry in Germany, where very little entry is taking place (and will take place) up to the point at which the baby-boom cohorts of physicians are going to retirement. Finally, it turns out that physician supply is not significantly correlated across regions. With respect to the regional distribution of physicians this is an important result, as it indicates that the impact of regional specific entry regulations on the supply of physicians is limited.

The remainder of the paper is organized as follows. The next section provides a brief survey of the literature on regional physician supply. In section 3, we develop a theoretical model of regional physician supply in order to clarify the role of demographic and geographic variables. Section 4 contains the empirical analysis and section 5 concludes.

2 Literature

The literature on the geographic distribution of physicians is predominantly empirical and somewhat eclectic in terms of the modelling approach. Newhouse et al. (1982) examine the location patterns of US physicians by esti-

mating the probability of finding at least one practice (in a certain specialty) established within a town depending on the size of the town.[3] As expected larger towns are more likely to attract a practitioner within any given specialty. As more specialized physicians require larger catchment areas they are more prone to settle in larger cities. In turn, this implies that general practitioners are over-represented in urban areas, where the substitute services of (the more specialized) internists are less available. These location patterns have been confirmed by Dionne et al. (1987) for Canadian data, and by Rosenthal et al. (2005), who re-estimate the original Newhouse et al. (1982) study with more recent data.[4] Their results show that population growth has triggered a diffusion, albeit incomplete, of specialists into more rural areas. While the analytic approach and the results of these studies are well in line with location theory, they are essentially (comparative) static in nature and do not examine the role of population structure or population flows.

A number of studies regress physician density within a certain region on a set of demographic, geographic and economic covariates. The findings with regard to age-structure are somewhat mixed. Using cross-sectional data from Switzerland, Kraft and v.d. Schulenburg (1986) find a positive yet insignificant impact of the population share 55+. Using cross-sectional data from Germany, Kopetsch (2007) and Jürges (2007) find a significant and positive impact of the population share 50+ and 65+, respectively.[5] Hingstman and Boon (1989), however, provide contrary results.

Analyzing the regional dispersion of primary health care practitioners (GPs, dentists, physiotherapists, midwives and pharmacists) in the Netherlands, Hingstman and Boon (1989) identify a significant negative effect of the share of elderly on the densities of all practitioners other than midwives. Interestingly, 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 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 old who require intensive treatment.

All of the above studies rely on cross-sectional analysis. On the one hand, this exposes them to the potential problem of unobserved regional

---

[3]In a separate estimation they also establish a one to one relationship between the change in the number of physicians and the change in population size.

[4]See also Newhouse (1990) for an 'intermediate' follow-up with US data.

[5]These studies estimate physician densities as integral part of an analysis of supplier induced demand. Kraft and v.d. Schulenburg (1986) estimate physician density jointly with various measures of treatment intensity. Kopetsch (2007) and Jürges (2007) estimate physician density as an instrument to be used in the estimation of treatment intensity.

heterogeneity; on the other hand, it poorly reflects the intertemporal nature of the market for physician services, in particular with respect to the evolution of demand. Foster and Gorr (1992) and Nocera and Wanzenried (2008) provide panel data estimations of physician density for the US and Switzerland, respectively. While these studies can trace out the evolution of the densities in response to population growth, they lack yet again data on population structure or on population flows.[6]

We conclude with the observation that none of the studies surveyed examines in a satisfactory way the spatial interaction in physician location choice. Rosenthal et al. (2005) find for the US that physician densities exhibit little variation with respect to regional types for general (and family) practitioners. However, there is a distinct regional pattern of densities for specialists, suggesting the presence of patient flows from rural into urban areas if the latter are sufficiently close. Surprisingly, spatial correlations appear to have attracted relatively little attention in the economic literature on the geographic distribution of physicians.[7] Although they do not lie at the core of our work, we employ spatial econometric techniques to control for cross-regional correlations and the scope for patient flows.

3 Theoretical Framework

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

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 parametrized by τ_i, they earn an income $Y\left(n_i, \ell_i, \overline{\lambda}_i, \underline{\lambda}_i, k_i, \tau_i\right)$. Finally, $U(k_i)$ denotes the physician's preferences for the amenities of a region described by k_i.[8] Assuming additive separability of income, we can

---

[6] Hurley (1991) and Bolduc et al. (1996) provide microeconometric analyses of physicians' choices of their practice location. Their analyses, too, do not control for the regional population structure.

[7] The presence of inter-regional patient flows is a similar issue when it comes to the definition of hospital markets. See e.g. Dranove et al. (1992) and Gaynor et al. (2005) for modelling options.

[8] 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 may have a preference for intermediate population densities - and thus against living in over-crowded or under-

then write

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

for a physician's overall utility from practicing in region i.[9]

3.1 Determinants of current physician income

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

The cost per unit of service $c^a = \gamma\left(a, k\right) x^a$ increases in the quality/intensity of the service, x^a, and varies with the patient's age a and the character of the region, k, according to the function $\gamma\left(a, k\right)$. Generally it may be assumed that the provision of services to the elderly and to the young is relatively resource intensive, as their treatment requires more time and as higher levels of morbidity for these groups render it necessary more frequently to provide a service by way of a home visit.[10] Furthermore, the cost-weights $\gamma\left(a, k\right)$ are likely to vary with the regional structure k. While we do not wish to place an assumption on the sign $\gamma_k\left(a, k\right)$ for any of the age-groups, we assume that the additional cost for treating the youngest and oldest relative to the middle-aged increases with the degree of rurality. This is plausible, as for rural regions with their longer travel times (i) home visits have to be carried out more frequently (as the share of the sick patients who are willing and able to travel decreases with travel time/distance) and (ii) the cost for the physician of carrying out a home visit is higher. But then as the incidence of home visits is higher amongst the elderly (i.e. age-group $a = 3$) and the young (age group $a = 1$) these groups are relatively more costly to treat in

---

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.

[9] The important role of income in driving a physician's location choice has been established e.g. by Hurley (1991) and by Bolduc et al. (1996) who find (average) income elasticities of 1.05 and 1.11, respectively.

[10] While we could not identify empirical evidence as to the age-profiles of treatment cost, the age-sex workload curve underlying the primary care funding allocation in the English NHS represents a U-shaped relationship (Department of Health 2004: fig. 1).

a rural context.[11] Thus, we assume

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

Demand $q^a = \phi(a,k)\,x^a\widetilde{\ell}^a$ from patient group a is given as the product of demand per patient $\phi(a,k)\,x^a$ and number of patients $\widetilde{\ell}^a$ within age group a. The demand per patient increases in the quality/intensity of medical services, x^a, offered to this group. Furthermore, it is assumed to vary with age and with the character of the region according to the function $\phi(a,k)$. We assume a U-shaped age-profile for the demand for physician services, implying that $\phi(2,k) < \min\{\phi(1,k),\phi(3,k)\}$.[12] 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.[13] It does not seem unreasonable to assume that due to their lack of mobility this effect is more pronounced for the youngest and oldest patients. Thus, we assume

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

The physician chooses the vector of quality levels $\{x^1,x^2,x^3\}$ so as to maximize income Y. From the first-order conditions $\frac{\partial Y}{\partial x^a} = (\tau^a - c^a)\,\phi(a,k)\,\widetilde{\ell}^a - \gamma(a,k)\,q^a = 0$ for $a = 1,2,3$ we obtain the set of optimal quality levels $\{x^{1*},x^{2*},x^{3*}\}$, where $x^{a*} = \frac{\tau^a}{2\gamma(a,k)}$.[14] We can thus determine the (optimized) value of practice income as $Y = \sum_a (\tau^a)^2 \frac{\phi(a,k)}{4\gamma(a,k)}\widetilde{\ell}^a$ from which we

_____

[11] Using discrete choice analysis, Scott (2001) identifies a strong willingness to pay on the part of English general practitioners (GPs) to avoid out of hours care. Giuffrida and Gravelle (2001) analyze the supply and demand for GP night visits, when to some extent GPs can manage demand. Similar to previous findings (referenced there) they find from an analysis of English panel data that the demand for night visits increases significantly and strongly with the share of children and with the population share 75+. The population density and the number of GPs per capita have a positive, albeit insignificant, effect on the demand (and supply) of night visits. This suggests (weakly) that the longer travelling times implied by lower population densities and the higher costs of making night visits in practices with large list sizes may imply a higher cost of undertaking night visits.

[12] We refer here to the unconditional pattern of use including age-related changes in morbidity. Such unconditional profiles are plotted e.g. in Department of Health (2004: fig. 1), Morris et al. (2005: fig. 1) or Dormont et al. (2006: fig. 2a). Note, however, that for the English data used by Morris et al. (2005) the U-shape in the age profile of general practitioner consultations only applies to males, whereas for females unconditional use is monotonically increasing with age. Conditional age-profiles are typically found to be U-shaped; see e.g. Ulrich and Pohlmeier (1995), Dusheiko et al. (2002) and Jürges (2007).

[13] The positive relationship between accessibility and usage of physician services is well documented (e.g. Dusheiko et al. 2002, Arcury et al. 2005, Iversen and Kopperud 2005, Thode et al. 2005).

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

readily obtain

$$\frac{\partial Y}{\partial \widetilde{\ell}^a} = (\tau^a)^2 \, \frac{\phi(a,k)}{4\gamma(a,k)} > 0, \tag{1}$$

As physicians choose service provision in a way that guarantees a positive mark-up per patient this implies that a greater number of patients, $\widetilde{\ell}^a$, within any age-group 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 $\widetilde{\ell}^a = \lambda^a \widetilde{\ell}$, 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. 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

$$\frac{\partial Y}{\partial \widetilde{\ell}} = \sum_a \frac{\partial Y}{\partial \widetilde{\ell}^a} \lambda^a = \sum_a (\tau^a)^2 \, \frac{\phi(a,k)}{4\gamma(a,k)} \lambda^a > 0, \tag{2}$$

$$\frac{\partial Y}{\partial \ell} = \frac{1}{n} \frac{\partial Y}{\partial \widetilde{\ell}} > 0, \qquad \frac{\partial Y}{\partial n} = -\frac{\widetilde{\ell}}{n} \frac{\partial Y}{\partial \widetilde{\ell}} < 0, \tag{3}$$

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.[15] Furthermore, we obtain

$$\frac{\partial Y}{\partial \overline{\lambda}} = \left(\frac{\partial Y}{\partial \widetilde{\ell}^3} - \frac{\partial Y}{\partial \widetilde{\ell}^2}\right) \widetilde{\ell} = \left(\frac{(\tau^3)^2 \phi(3,k)}{4\gamma(3,k)} - \frac{(\tau^2)^2 \phi(2,k)}{4\gamma(2,k)}\right) \widetilde{\ell} \tag{4}$$

$$\frac{\partial Y}{\partial \underline{\lambda}} = \left(\frac{\partial Y}{\partial \widetilde{\ell}^1} - \frac{\partial Y}{\partial \widetilde{\ell}^2}\right) \widetilde{\ell} = \left(\frac{(\tau^1)^2 \phi(1,k)}{4\gamma(1,k)} - \frac{(\tau^2)^2 \phi(2,k)}{4\gamma(2,k)}\right) \widetilde{\ell}. \tag{5}$$

Whether or not a larger share of older patients, $\overline{\lambda}$, contributes to a higher practice income then depends on the income per patient, $\frac{(\tau^a)^2 \phi(a,k)}{4\gamma(a,k)}$, gained for group 3 relative to the group of middle-aged patients (group 2). 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

---

[15] In our model physicians do not compete for patients, implying that a growing number of rival physicians only leads to a reduction in the number of patients per practice. In a model of oligopolistic competition (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 the number of practices.

In one of the few studies testing for the determinants of physician income Dormont and Samson (2009) identify a significant negative relationship between the density of general practitioners and their income.

the reimbursement system and on the regional structure. According to our assumptions both unit demand and the costs per unit of service are higher for older patients, i.e. $\phi(3,k) > \phi(2,k)$.and $\gamma(3,k) > \gamma(2,k)$. It follows that old patients are more profitable in a system for which the fees are age-adjusted in such a way that $\tau^a = \pi\gamma(a,k)$ induces the same level of service quality for all age-groups: $\frac{\partial Y}{\partial \overline{\lambda}} = \pi^2 \left(\frac{\gamma(3,k)\phi(3,k)}{4} - \frac{\gamma(2,k)\phi(2,k)}{4} \right) \widetilde{\ell} > 0$. In contrast, a uniform fee $\tau^a = \tau$ would generate the same mark-up per patient, $\tau^a - \gamma(a,k)x^{a*} = \frac{\tau^a}{2} = \frac{\tau}{2}$, so that $\frac{\partial Y}{\partial \overline{\lambda}} = (\tau)^2 \left(\frac{\phi(3,k)}{4\gamma(3,k)} - \frac{\phi(2,k)}{4\gamma(2,k)} \right) \widetilde{\ell} > 0 \Leftrightarrow \frac{\phi(3,k)}{\phi(2,k)} > \frac{\gamma(3,k)}{\gamma(2,k)}$. Here, older patients are more profitable if and only if the higher unit demand more than compensates the higher 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 impact of age-structure on the supply of physicians (and implicitly on practice income) depends on the degree of rurality. Assuming that reimbursement rates do not vary across regions, the cross effect between age-structure and rurality is described by the cross derivatives

$$\frac{\partial^2 Y(\cdot)}{\partial k \partial \overline{\lambda}} = \left(\frac{\partial Y}{\partial k \partial \widetilde{\ell}3} - \frac{\partial Y}{\partial k \partial \widetilde{\ell}2} \right) \widetilde{\ell} \tag{6}$$

$$= \left[\frac{\partial Y}{\partial \widetilde{\ell}3} (\varepsilon_\phi(3,k) - \varepsilon_\gamma(3,k)) - \frac{\partial Y}{\partial \widetilde{\ell}2} (\varepsilon_\phi(2,k) - \varepsilon_\gamma(2,k)) \right] \frac{\widetilde{\ell}}{k}$$

$$\frac{\partial^2 Y(\cdot)}{\partial k \partial \underline{\lambda}} = \left(\frac{\partial Y}{\partial k \partial \widetilde{\ell}1} - \frac{\partial Y}{\partial k \partial \widetilde{\ell}2} \right) \widetilde{\ell} \tag{7}$$

$$= \left[\frac{\partial Y}{\partial \widetilde{\ell}1} (\varepsilon_\phi(1,k) - \varepsilon_\gamma(1,k)) - \frac{\partial Y}{\partial \widetilde{\ell}2} (\varepsilon_\phi(2,k) - \varepsilon_\gamma(2,k)) \right] \frac{\widetilde{\ell}}{k}.$$

where $\varepsilon_\phi(a,k) := \frac{\phi_k(a,k)k}{\phi(a,k)} < 0$ and $\varepsilon_\gamma(a,k) := \frac{\gamma_k(a,k)k}{\gamma(a,k)}$ 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 *Suppose there exists a degree of rurality $k = \widehat{k}^{a'} \in [k_{\min}, k_{\max}]$, $a' = 1, 3$ for which $\frac{\partial Y}{\partial \widetilde{\ell}a'} = \frac{\partial Y}{\partial \widetilde{\ell}2}$. If $\varepsilon_\phi(a',k) - \varepsilon_\phi(2,k) < \varepsilon_\gamma(a',k) - \varepsilon_\gamma(2,k)$ then $\frac{\partial^2 Y(\cdot)}{\partial k \partial \overline{\lambda}} \big|_{k=\widehat{k}^3} < 0$ and $\frac{\partial^2 Y(\cdot)}{\partial k \partial \underline{\lambda}} \big|_{k=\widehat{k}^1} < 0$, implying that $\frac{\partial Y(\cdot)}{\partial \overline{\lambda}} < 0 \Leftrightarrow k > \widehat{k}^3$ and $\frac{\partial Y(\cdot)}{\partial \underline{\lambda}} < 0 \Leftrightarrow k > \widehat{k}^1$.*

Starting from a situation where old and young patients generate the same income as patients from the intermediate age group, then rurality tends to decrease the relative profitability of old and young patients if for the intermediate age-group ($a = 2$) demand is less responsive with respect to rurality (as measured by the difference of elasticities) and/or if the cost per unit of service is less (more) responsive when rurality raises (lowers) cost. In this case old and young patients would turn out to be profitable relative to

the intermediate age-group only within the least rural regions. The impact of the share of oldest and (youngest) patients on physician income would then change with the degree of rurality from positive within urban areas to negative within rural areas. Although we cannot measure the elasticities, Proposition 1 corresponds well with our empirical results in section 4.4.2. In an estimation where long-run effects are picked out by a lagged dependent variable, we are able to identify a short term cross effects between the age-shares $\overline{\lambda}$ and $\underline{\lambda}$, respectively, and the index of rurality k, which is negative in its impact on the number of physicians. As the latter is positively correlated with physician income, this implies a negative cross-effect 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 (Taylor and Leese 1998, Elliott et al. 2006, Kopetsch and Munz 2007). It is then 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 therefore not only seek to assess the current demand (depending on current population size and age-structure) but also make predictions about future demand. Consider thus a set-up, where physicians practice for two-periods and face a distinct population structure in each period. Thus, for any two subsequent periods t and $t+1$ we can write a physician's expected utility as

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

where the second period utility is discounted by a factor δ. In the following we examine the development of the population from which the physician recruits patients.

Assume that individuals (including physicians) 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 can thus be described by the following OLG model comprising the three age groups $a = 1, 2, 3$, as defined in the previous section. 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 $\ell_t = \sum \ell_t^a$.

11

Furthermore, define

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

as the shares of the population 60+ and 20-, 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.[16] Substituting $\ell_t^2 = (1 - \overline{\lambda}_t - \underline{\lambda}_t) \ell_t$, $\ell_t^1 = \underline{\lambda}_t \ell_t$ and $M_t = m_t \ell_t$ we can express the population in period $t+1$ in terms of the demographic make-up $\{\ell_t, \overline{\lambda}_t, \underline{\lambda}_t, \mu_t, m_t\}$ in period t

$$
\begin{aligned}
\ell_{t+1}^3 &= \left[\rho_t \left(1 - \overline{\lambda}_t - \underline{\lambda}_t \right) + \omega_t m_t \right] \ell_t & (9) \\
\ell_{t+1}^2 &= \left[(1 - \rho_t) \left(1 - \overline{\lambda}_t - \underline{\lambda}_t \right) + \underline{\lambda}_t + (1 - \omega_t) m_t \right] \ell_t & (10) \\
\ell_{t+1}^1 &= \mu_t \left(1 - \overline{\lambda}_t - \underline{\lambda}_t \right) \ell_t. & (11)
\end{aligned}
$$

Population size and age-structure in period $t+1$ can thus be expressed as functions of the demographic make-up in period t

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

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,

$$
\begin{aligned}
& Y \left(n_{t+1}, \ell_{t+1}(\cdot), \overline{\lambda}_{t+1}(\cdot), \underline{\lambda}_{t+1}(\cdot), k_{t+1}, \tau_{t+1} \right) \\
=\; & Y' \left(n_{t+1}, \ell_t, \overline{\lambda}_t, \underline{\lambda}_t, \mu_t, m_t, k_{t+1}, \tau_{t+1} \right) =: Y_{t+1}
\end{aligned}
$$

---

[16] 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 characterized 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.

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

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

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

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

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

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

as obtained from (9)-(12).∎

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_t^2, 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_{t+1}^3 and ℓ_{t+1}^1, respectively, but it increases the size of the intermediate age-group, ℓ_{t+1}^2. The impact on practice income thus depends on the relative profitability of these age-groups, on the birth rate, μ_t, and on the the share of individuals 40-59 within age-group 2, ρ_t.

The impact of the rate of net migration, m_t, on the size of the future age-groups ℓ_{t+1}^3 and ℓ_{t+1}^2 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 aggregate rate of net-migration. 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 on average migration streams are of the same direction and that migration (in or out) takes place predominantly at the beginning of working life, implying that ω_t is positive and small.[17] For this case, we obtain $\frac{dY_{t+1}}{dm_t} > 0$. The effect of the fertility rate on future income is unambiguously positive, as the young age-group expands for a constant size of the other age-groups.

The cross effects between rurality and age-structure

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

are given by weighted sums of the impact of rurality on the marginal profitability of the various patient groups, $\frac{\partial Y}{\partial k \partial \widetilde{\ell}^a}$. If rurality lowers the marginal profitability of all patient groups, i.e. if $\frac{\partial Y}{\partial k \partial \widetilde{\ell}^a} < 0$ for all a, then $\frac{d^2 Y_{t+1}}{dk\,d\overline{\lambda}_t} > 0$ is always true. Under the same conditions $\frac{d^2 Y_{t+1}}{dk\,d\underline{\lambda}_t} > 0$ is true if $\frac{\partial Y}{\partial k \partial \widetilde{\ell}^3} - \frac{\partial Y}{\partial k \partial \widetilde{\ell}^2} = \frac{1}{\widetilde{\ell}} \frac{\partial^2 Y}{\partial k \partial \widetilde{\lambda}} < 0$. However, this is not guaranteed and ultimately these effects should be subject to empirical investigation.

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

---

[17] See figure 2 in section 4.1 and the surrounding discussion.

through changes in present and future income

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

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

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

where for $z \in \{\ell, \overline{\lambda}, \underline{\lambda}, \mu, m\}$ we have $\frac{dY_t}{dz_t} = \frac{\partial Y}{\partial z}$ for $z \in \{\ell, \overline{\lambda}, \underline{\lambda}\}$ as defined in (3)-(5) and $\frac{dY_{t+1}}{dz_t}$ as defined in Lemma 1. The current size and age-structure of the population $\{\ell_t, \overline{\lambda}_t, \underline{\lambda}_t\}$ affects a physician's discounted income both in the current and future period. Future income depends on the current demographic make-up through the implied changes in the size and structure of the patient population. This latter argument also applies 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 lifetime income and utility. A similar clear-cut argument applies to the rates of fertility and net migration, provided in the latter case that $|\omega_t|$ small. As they imply a larger 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 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 of $\overline{\lambda}_t$ on a physician' life time income within rural regions. However, even where the treatment of elderly patients is relatively profitable for low levels of rurality, i.e. even where $\frac{dY_t}{d\overline{\lambda}_t} > 0$, this may be 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. Similarly, the impact on life-time income of the youth share, $\underline{\lambda}_t$, is ambiguous for all types of regions. Thus, while theory helps us to identify the sources of the various trade-offs regarding the demographic determinants of life-time income, it yields the plausible yet unspectacular insight that the effect of population ageing/shrinking $\{d\ell_t < 0, d\overline{\lambda}_t > 0, d\underline{\lambda}_t < 0, d\mu_t < 0\}$ on a representative physician's lifetime income is usually ambiguous. It will be part of our empirical exercise to shed some light on the actual relationships.

15

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 $\overline{W}_t \geq 0$.[18] Furthermore, we assume that patient mobility across regions is too low as to generate significant cross-regional market overlap from the perspective of the representative physician. In this case, we can write $W_{it}(n_{it}, En_{it+1})$, as defined in equation (8), as a physician's expected lifetime utility when settling in region i at time t. Here, En_{it+1} denotes the number of rival physicians expected during the second period of practice. Assuming that physicians are free in their location choice, entry into region i takes place as long as $W_{it}(n_{it}, En_{it+1}) \geq \max\left\{W_{i't}(n_{i't}, En_{i't+1}), \overline{W}_t\right\}$, for any other region $i' \neq i$. As is readily checked, we have $\frac{\partial W_{it}(n_{it}, En_{it+1})}{\partial n_{it}} < 0$, implying stability of an entry equilibrium, and $\frac{\partial W_{it}(n_{it}, En_{it+1})}{\partial En_{it+1}} < 0$. An entry equilibrium is then defined by $W_{it}(n_{it}^*, En_{it+1}) = \overline{W}_t \quad \forall i$.[19]

Defining $\widehat{Y}_{it} := \sum_a (\tau_{it}^a - \gamma(a,k) x_{it}^{a*}) \phi(a,k) x_{it}^{a*} \lambda_{it}^a \ell_{it}$ as the total income from physician services within region i in period t and $\widehat{W}_{it} := \overline{W}_t - [U(k_{it}) + \delta U(k_{it+1})]$ we can rewrite the equilibrium condition as $\frac{\widehat{Y}_{it}}{n_{it}^*} + \delta \frac{\widehat{Y}_{it+1}}{En_{it+1}} = \widehat{W}_{it}$. Noting that $\widehat{W}_{it}$ can be interpreted as the compensating income required by a physician to take up practice in region i at time t, it follows that in equilibrium physicians enter up to the point at which their (discounted) life-time income equals their compensating income.[20] We then obtain

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

as the equilibrium number of physicians in region i and period t. As is readily checked n_{it}^* increases in present 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 (necessary) compensating income. In order to simplify the following analysis, we assume $k_{it} = k_{i+t} = k_i$ and $\overline{W}_t = \overline{W}$ for all t and therefore $\widehat{W}_{it} = \widehat{W}_i$. 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 polar configurations of entry equilibrium as follows.[21]

---

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

[19] For simplicity, we ignore here the integer issue.

[20] Likewise, $\widehat{W}_{it} - \widehat{W}_{i't}$ is the compensating income required by a physician to practice in region i rather than in region i'.

[21] There exist other equilibrium patterns, depending on the period $t + \overline{t}$ at which zero entry takes place. As we show in the appendix, for zero entry at period

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 $t \leq \infty$. 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*

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

Proof: We drop the regional index i for convenience. To prove (i) consider a sequence of periods $\widehat{t} = t, t+1, ..., t+\bar{t}-1$, with $\bar{t} \in [1, \infty]$. The assumption $\frac{\widehat{Y}_t}{\widehat{n}_{\widehat{t}-1}} + \delta \frac{\widehat{Y}_{\widehat{t}+1}}{E\widehat{n}_{\widehat{t}+1}} > \widehat{W}$ implies that there is some entry taking place in all periods, i.e. $\widehat{n}_{\widehat{t}} > 0$ for all $\widehat{t}$. We can thus write $n_{\widehat{t}}^* = \frac{\widehat{Y}_{\widehat{t}} n_{\widehat{t}+1}^*}{\widehat{W} n_{\widehat{t}+1}^* - \delta \widehat{Y}_{\widehat{t}+1}}$. Assuming that $t+\bar{t}$ is the first period without entry we have $n_{t+\bar{t}-1}^* = \frac{\widehat{Y}_{t+\bar{t}-1} E n_{t+\bar{t}}}{\widehat{W} E n_{t+\bar{t}} - \delta \widehat{Y}_{t+\bar{t}}}$. Substituting recursively, we obtain

$$
\begin{aligned}
n_t^* &= \prod_{\widehat{t}=t}^{t+\bar{t}-1} \widehat{Y}_{\widehat{t}} E n_{t+\bar{t}} \left[\frac{(-\delta)^{\bar{t}} - 1}{(-\delta)-1} \widehat{W} \prod_{\widehat{t}=t+1}^{t+\bar{t}-1} \widehat{Y}_{\widehat{t}} E n_{t+\bar{t}} + (-\delta)^{\bar{t}} \prod_{\widehat{t}=t+1}^{t+\bar{t}} \widehat{Y}_{\widehat{t}} \right]^{-1} \\
&= \widehat{Y}_t E n_{t+\bar{t}} \left[\frac{1-(-\delta)^{\bar{t}}}{1+\delta} \widehat{W} E n_{t+\bar{t}} + (-\delta)^{\bar{t}} \widehat{Y}_{t+\bar{t}} \right]^{-1}
\end{aligned}
$$

. Assuming $\bar{t} = \infty$, we then obtain $(-\delta)^{\bar{t}} = 0$ and, therefore, $n_t^* = \frac{(1+\delta)\widehat{Y}_t}{\widehat{W}}$.

To prove (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, so that $\widehat{n}_{t-1} = \widehat{n}_{t+1} = 0$. In this case, we obtain $E n_{t+1} = n_t^*$ 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 within any given period, implying that there are always two generations of physicians in practice . The total number of

---

$t + \bar{t} \geq t + 1$, the equilibrium number of physicians in period t is given by, $n_{it}^* = \widehat{Y}_{it} n_{it+\bar{t}} \left[\frac{1-(-\delta)^{\bar{t}}}{1+\delta} \widehat{W}_i n_{it+\bar{t}} + (-\delta)^{\bar{t}} \widehat{Y}_{it+\bar{t}} \right]^{-1}$, where $n_{it+\bar{t}}$ is the number of physicians in period $t+\bar{t}$. As by assumption no entry takes place in period $t+\bar{t}$, we have $n_{it+\bar{t}} = \widehat{n}_{it+\bar{t}-1}$.

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). For a variation in variable $z_{it} \in \left\{ \ell_{it}, \overline{\lambda}_{it}, \underline{\lambda}_{it}, \mu_{it}, m_{it} \right\}$ we then obtain

$$\frac{dn_{it}^*}{dz_{it}} = \frac{1+\delta}{\widehat{W}_i} \frac{d\widehat{Y}_{it}}{dz_{it}} = \frac{(1+\delta)\, n_{it}^*}{\widehat{W}_i} \frac{dY_{it}}{dz_{it}},$$

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 $\frac{dY_{it}}{d\mu_{it}} = \frac{dY_{it}}{dm_{it}} = 0$. Thus, for an equilibrium with full generational overlap expectations over the future population have no bearing on the current number of physicians.[22]

However, it is by no means guaranteed that physicians of every generation enter at all times. For instance, entry may take place only every second period. Such a situation arises 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 $t+1$, depressing the level of income Y_{it+1} in the first period of practice, 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. 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 unprofitable for members of generation $t+1$, implying that $\widehat{n}_{it+1} = 0$. If the same applies to generation $t-1$, then a cyclical pattern of entry may arise 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 lack of generational overlap implies an extreme age-structure, where there are either only young or only old physicians in practice. Hence, the turn-over of physicians is not continuous but occurs fully at every second period. Within a period t, in which entry takes place, the total number of physicians then depends on the expectations for the subsequent period $t+1$. Specifically,

$$\frac{dn_{it}^*}{dz_{it}} = \frac{n_{it}^*}{\widehat{W}_i} \left(\frac{dY_{it}}{dz_{it}} + \delta \frac{dY_{it+1}}{dz_{it}} \right) = \frac{n_{it}^*}{\widehat{W}_i} \frac{dW_{it}}{dz_{it}},$$

with $\frac{dW_{it}}{dz_{it}}$ 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 reproduction, where both imply a larger future population.

Similar arguments apply for the cross-effects $\frac{d^2 n_{it}^*}{dk dz_{it}}$, $z_{it} \in \left\{ \overline{\lambda}_{it}, \underline{\lambda}_{it} \right\}$. For an equilibrium with full generational overlap we obtain $\frac{dn_{it}^*}{dz_{it}} = \frac{(1+\delta) n_{it}^*}{\widehat{W}_i} \frac{d^2 Y_{it}}{dk dz_{it}}$,

---

[22] 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 $\widehat{Y}_{it-1}$.

with the last derivative as given in (6) and (7), and for an equilibrium with no overlap we obtain $\frac{dn^*_{it}}{dz_{it}} = \frac{n^*_{it}}{\widehat{W}_i}\frac{d^2W_{it}}{dkdz_{it}} = \frac{n^*_{it}}{\widehat{W}_i}\left(\frac{d^2Y_{it}}{dkdz_{it}} + \delta\frac{d^2Y_{it+1}}{dkdz_{it}}\right)$. We can summarize 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 affected by the current and future demographic structure as determinants of life time income. (iii) Under full generational overlap, the current number of physicians is affected solely by the current demographic structure as determinant of current income.*

In particular, it is true 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. Finding empirically a significant positive impact would therefore lend support for an equilibrium with no (or limited) generational overlap rather than full (or sizeable) generational overlap. This is, indeed, what we find from our estimates.

4 Empirical Analysis

4.1 Regional Population Change and the Supply of Office-Based Physicians in Germany

In the time span covered by our data, 1995-2004, Germany has witnessed an increase by 30% in the total number of office-based physicians. However, the change in physician numbers during these years has exhibited a large variation. Figure 1 provides information on the percentage change of the number of ambulatory physicians between 1995 and 2004 in Germany disaggregated into 439 regions. The dark red areas suffer from a reduction of physician supply of up to 26%. This applies in particular to many Eastern German regions, but also to certain rural areas in Western Germany. The orange colored regions experience stagnation or a slight increase in the number of physicians.

figure 1 about here

On the other hand, large parts of Western Germany have experienced an unabated increase in the supply of physicians, as depicted by the light and dark green colored areas. This begs the question as to what are the causes for these divergent trends in the supply of physicians at the regional level. One candidate explanation relates, of course, to population change, where over the period under examination rural areas, particularly those in Eastern Germany, have experienced a rapidly shrinking population and, at

19

the same time, a strong increase in the mean age. Especially in Eastern Germany, the decline in population size over the 1990s has been driven by strong outmigration of young people into Western German regions that offer better employment prospects.

Regional Population Change: As we theoretical analysis has shown, population change could potentially impact on physician supply through three channels: changes in the age structure, in net migration, and in the natural balance. A preliminary analysis of our population data shows how the population flows are related to the age-structure of the population. As expected, we find that the natural balance is positively correlated with the population share 20- ($r = 0.50$) and negatively with the share 60+ ($r = -0.62$). Similarly, the population share 60+ is negatively correlated with net migration ($r = -0.32$), while the opposite is true for the share 20- ($r = 0.33$). This is consistent with an economy in which migration occurs predominantly amongst younger individuals. Immigration (emigration) then tends to boost the share of the young (old) in the overall population.[23] While our data for 1995-2004 does not allow us to distinguish migration streams by age, we can nevertheless confirm the above conjecture by considering the more detailed data on migration that is available for the years 2003 and 2004. For these years the data contains information on migration by families (individuals aged 30 to 50 years and up to 17 years), by the young (18 to 29 years old), and by the retired (65 years and older). In figure 2, we plot for each region the levels of net migration found for these sub-groups against the overall level of net migration (per thousand in the corresponding population).

figure 2 about here

Two aspects emerge. First, the level of net migration within each age-group is positively correlated with overall net migration.[24] Hence, on average the migration flows of all age-groups take on the same direction. In terms of our theoretical model (see section 3.2), this implies that the parameter $\omega_t \in [0, 1]$ can, in fact, be interpreted as the share of the old in total (net) migration. Second, the young exhibit the highest regional mobility, whereas the lowest mobility can be observed for retired people. As it turns out the retired make up for a share of about 8% of all migrants. Thus, we can

---

[23]One striking such example is the state of Mecklenburg-Vorpommern in the North-Eastern part of Germany. According to population forecasts (Statistische Ämter 2007) for the period 2005-2020 the population will shrink from 1.7 million to 1.5 million due to emigration at an annual rate of 3% and due to a birth deficit of roughly 0.5% per annum. At the same time the dependency ratio (population 65+ per 100 individuals aged 20-64) is forecast to increase from 31 to 46. For the whole of Germany the dependency ratio is expected to increase from 32 to only 39, which illustrates the strong incidence of migration on population ageing in Mecklenburg-Vorpommern.

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

infer that ω_t takes on a very low level, as we have assumed throughout our analysis in section 3.2. The observation that migration is predominantly driven by young(er) individuals also implies that net migration biases the process of population ageing. Specifically, immigration mitigates population ageing, whereas emigration intensifies it.

From the theoretical analysis we would expect that age structure and population flows should have independent effects on the supply of physicians in spite of their strong correlation. This is because the age-shares reflect current and future patient populations and as such have a short-term and long-term effect. The flow variables, in contrast, reflect future demand for physician services and should therefore only have a long-term effect, but one that is independent from the effect related to the age-shares. In the econometric analysis we should therefore find significant effects of all variables in the long run, whereas only age structure should matter in the short run.

Office-Based Physicians: In Germany, the group of office-based primary care physicians comprises both general practitioners and specialists (e.g. internists, gynecologist, ophthalmologists, pediatricians, orthopedists, etc.). The large majority of physicians working in ambulatory care hold an affiliation with the statutory health insurance (SHI). In 2003, SHI-affiliated physicians made up for 117,600 out of 132,400 office based-physicians. SHI covers about 88% of the population (in 2003).[25] The payment for services delivered to SHI patients is determined in two steps. First, sickness funds allocate a negotiated budget to the physician association(s) at state (Länder) level. The physician association(s) then remunerate physicians on a fee-for service basis where the reimbursement rates are adjusted by a point value system.[26] Leaving further details of the payment process aside, two things are worthy of note: (i) the fees for service are typically neither adjusted for age nor for other patient characteristics; and (ii) while SHI fees may vary across states (Länder) they apply uniformly across all regions (Kreise) within each state. We do not have data about the reimbursement system. However, in our estimation we control for state-level time effects, which allows us to pick up spatial and temporal variation in the fees. Ambulatory physicians also provide services to the 9% of the population with private health insurance. While the remuneration for private patients is typically more generous, the fee structure reflects the one of SHI. Lacking data on the distribution of insurance status amongst the regional populations, we would expect this to be picked up by the regional fixed effects and possibly by the variable on regional income per capita, which we would expect to be

---

[25]For further information on the physician workforce and some of the institutional detail see Busse and Riesberg (2004: pp. 96-99).

[26]For a more detailed description of the reimbursement system see Busse and Riesberg (2004: pp. 177-183).

positively related to the share of the privately insured.[27]

In principle, SHI affiliated physicians are free to chose their location of practice.[28] However, since 1993 entry regulations apply at the level of physician specialty (including general practice). Specifically, regions for which an excess supply at specialty level has been established are closed to new entries within this specialty but not to replacements. Furthermore, the opening of a new practice may be granted if a special need is constated, these exceptions occurring at increasing rates.[29] Since regional-level data is available neither on the presence or absence of closures (by specialty) nor on exceptional permissions we have opted to overcome the problem related to upper bounds of physician supply within certain specialities by using the aggregate number of physicians within a region. This guarantees a sufficient year-to-year variation of the data for each region. According to the National Association of SHI Physicians (Kassenärztliche Bundesvereinigung) the closure to entry at specialty level has been fairly stable, i.e. where implemented for a specialty it has remained in place for much of our observation period 1995-2004. Furthermore, there is evidence that the more attractive regions have a higher probability of being closed to new entry for certain specialties. With respect to our estimates this has three consequences: First, the causes for an unequal geographic distribution of physicians are potentially underestimated. Second, positive spatial correlations in the distribution of physicians are potentially overestimated. Third, given that entry closures have remained in place within a region over the whole period of observation, to some extent this is captured by regional fixed effects. We get back to these points in section 4.4.2 below.

Finally, it is instructive to consider some preliminary information on the age-distribution of the physician population. According to the National Association of SHI Physicians the age structure of office-based physicians is driven by large baby-boomer cohorts. As shown in figure 3, the age distribution in 1993 differs significantly from that in 2004.[30] In 1993 the distribution is skewed right and the number of retiring physicians is small, whereas in 2004 the "iceberg" has shifted to the right, implying that the large cohorts are reaching the age of retirement. Over the same time span, the number of young physicians entering the market has decreased markedly.

---

[27]In Germany, private health insurance is only accessible to civil servants, the self-employed or to individuals who earn a salary in excess of a threshold. SHI insurance is compulsory for all other salaried workers and employees.

[28]About 75% of office-based physicians work single-handed.

[29]An excess supply is established if the current physician density (by specialty) exceeds 110% of the average (across regions of the same classification) density (by specialty) in the year 1990. This definition is obviously unrelated to patient need, however defined. See Busse and Riesberg (2004; pp. 124-125), Kopetsch (2004) and Klose and Uhlemann (2006) for a more detailed discussion of the German system of entry regulation.

[30]No data on the age distribution of physicians is available covering the full period 1995-2004. Neither are there data available at regional level.

Altogether the skewed age-distribution and the associated uneven patterns of entry and retirement suggest an equilibrium with limited generational overlap (at regional level) rather than one with full generational overlap. According to Corollary 2 we would therefore conjecture that expectations about future demand, as measured by age structure and population flows, should have an influence on the geographic distribution of physicians.

figure 3 about here

4.2 Data

We build our empirical analysis on regional panel data provided by the Federal Office for Civil Engineering and Regional Development (BBR) as part of the INKAR data set. This data covers 439 districts (Kreise) in Germany (corresponding to NUTS III level) over the time span 1995 to 2004. We have to exclude the year 1999, for which no information on the number of office-based physicians is available at regional level due to a reorganization of the registry. Hence, we have 3951 observation available for our estimates. The dependent variable is the aggregate number of SHI-affiliated physicians. Despite the loss of information regarding the distribution of physicians by specialty, using the aggregate number of physicians improves our estimation, as we benefit from greater variation in the data (see above). In section 4.4.1 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.

Demographic change bears on the supply of physicians by way of both changes in the age-structure and changes in the population size. We directly control for population size but capture additional effects, related to expectations, by including the flow variables net migration and natural balance at regional level. Our main interest, however, lies with the effect of a changing age structure. We approximate the age structure by including both the population share of people 60 years and older (share 60+) and the population share of people less than 20 years (share 20-). From our theoretical analysis we know that it is possible that the effects of population ageing on physician supply vary with the character of the region (urban vs. rural). We control for these effects by interacting the age groups (60+ and 20-) with an index of rurality, as provided by the BBR. This index is subaggregated into agglomeration areas, municipalized areas, and rural areas. Within each of these area groups there are up to four sub-groups, differing in population and population density. Altogether there are nine levels of the index, with higher levels corresponding to an increasing degree of rurality. See the Appendix for further details. The advantage of this index is that it aggregates information on population size, population density, and possible spillover

effects from supraregional cities.

The remaining set of regional control variables comprises: population density, GDP per capita, unemployment rate, employment rate, share of foreigners, share of welfare recipients, share of school leavers qualified for higher education, share of school leavers without degree, 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 of these variables in the next section when we present the results.

4.3 Estimation of the Basic Model

The supply of physicians at macro-level is typically measured in density terms (i.e. physicians per capita). In the presence of interregional migration, however, the use of physician density as a dependent variable may lead to measurement error. According to the empirical evidence discussed in section 4.1 younger people have a greater propensity to migrate. Consider thus a region subject to intense emigration on the part of young individuals. On the one hand, the corresponding decline in population size implies that the physician density increases for a given number of physicians. On the other hand, the share of older people increases, generating a positive yet spurious correlation with physician density. According to our data physician density is, indeed, correlated negatively with the population share 20- (-0.57) and positively with the population share 60+ (0.26). Given that the supply of physicians responds only with a lag to changes in population size, physician density then comprises a measurement error.

To illustrate this point more formally, we decompose physician density (D) into a measure of real supply (S) and a measurement error due to interregional migration (ϵ). The regression $D = S + \epsilon = \alpha + \beta X + u + \epsilon$ (with u as idiosyncratic error term) that should approximate the true model yields a biased parameter β if the explanatory variable X is correlated with ϵ. Since a declining population boosts the physician density (at least in the short run), we can conclude that the parameter corresponding to the share 60+ is biased upwards. For this reason we prefer to use the number of physicians as dependent variable (while directly controlling for population size). In order to gain some insights into the effects of the bias, we rerun in section 4.4.1 the regressions with physician density as the dependent variable.

We have shown in sections 3.2 and 3.3 how the supply of physicians is related to the net migration flow (which increases in the net migration rate), the natural balance (which increases in the fertility rate), and the size and age structure of the population. 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 in different ways, accounting for unobserved heterogeneity. Hence, we will

24

start with the following specification:

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

with $i = 1, 2, ..., I$ regions, $j = 1, 2, ..., J$ states and $t = 1, 2, ..., T$ periods. Here, S denotes the number of physicians, α_i are regional fixed effects, η_{jt} are state level time effects, and ϵ_{it} is an error term. The term η_{jt} mainly captures variation in reimbursement, which is determined at state level. Recall that German states (Bundesländer) comprise 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 for regional fixed effects and state level time effects. We consistently find that physician supply decreases in the share 60+ 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 future demand for physician services and, thus, provide incentives to locate within a region. A higher current share of older people 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 income gains from treating a larger share of 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 20- is not robust with respect to the 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 imply that young patients are relatively profitable, but it may also speak to the expectation that young populations guarantee a high demand well into the future. From these simple regressions it follows that population ageing, implying a shift in the age-distribution towards the higher age-groups, and population decline both bear negatively on the supply of physicians.[31]

In the following, we add three elements to our estimation. First, we examine in greater detail the relationship between the demographic and geographic make-up of a region. We have argued before that the effect of

[31] 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 clearly rejects the null hypothesis, we should not overrate this because of the relatively short time span and the missing information for 1999.

age structure on physician supply may be shaped by the degree of rurality. Therefore we include interactions of the age-shares with the index of rurality that was presented in section 4.2. Second, we now include a number of control variables to examine if the estimated effects of population change remain significant once other determinants of physician supply are considered. Third, while 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 these authors spatial correlations among cross-sections may arise for a number of reasons, ranging from observable common economic shocks to unobserved contagion or neighborhood effects.[32] We now estimate

$$
\begin{aligned}
S_{it} \;=\; & \beta_0 + \beta_1 60plus_{it} + \beta_2 60plus_{it} \times rurality_i \quad\quad (17)\\
& + \beta_3 20\min us_{it} + \beta_4 20\min us_{it} \times rurality_i \\
& + \beta_5 migration_{it} + \beta_6 natural_{it} + \varphi' X_{it} + \alpha_i + \eta_{jt} + \epsilon_{it},
\end{aligned}
$$

where, in addition to the specification (16), we have included the interaction of the population share 60+ with the rurality variable ($60plus_{it} \times rurality_i$), the interaction of the population share 20- with the rurality variable ($20\min us_{it} \times rurality_i$), and a vector of control variables, X_{it}. The regression results are summarized in table 2.

Age structure: According to regression (1) (in table 2) a higher share of the elderly 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 60+ 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 60+ is now positive. The finding that an elderly population become progressively less attractive for more rural regions is well in line with our conjecture that rurality reduces the relative profitability of treating old patients. However, we cannot say at this point whether (or how) the degree of rurality also modifies the (negative) expectations about future demand that are formed in the presence of older populations.

> *table 2 about here*

When considering all types of regions alike, the share of the young population 20- does not have a significant effect on physician supply (see regression (1)). However, when adding the interaction with rurality, we find

---

[32]Driscoll and Kraay (1998) argue that the presence of such spatial correlations in residuals complicates standard inference procedures for combined 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, the estimates of standard errors are inconsistent and hence not useful for inference.

that a high share of the young is most attractive in urban regions: whereas the direct effect of the share 20- is significant positive, the interaction with rurality is significant negative. The positive direct effect is consistent with a young population being a good indicator for a high future demand. The lower attractiveness of young patients within rural areas hints either at the treatment of young patients being more costly within a rural context or at the expectation that young populations may not stay within rural areas but rather migrate elsewhere.

Comparing, within regression (2), the effects on physician supply of the two age-shares, we find that while the interaction terms are not significantly different from each other, the base effect of the age-share 20- amounts to the threefold of the base effect found for the age-share 60+. This observation is consistent with the role of expectations. While a large age-share 60+ leads physicians to expect a reduction in demand, the opposite is expected for a large age-share 20-. Thus, while greater rurality bears negatively on the treatment of both youngest and oldest patients much in the same way and while the current treatment of both youngest and oldest patients is profitable in agglomerations (only), the negative (positive) expectations about future demand bias downwards (upwards) the incentive to locate in regions with a high share of oldest (youngest) population. This explains why a large share 60+ is never attractive (apart perhaps from the most metropolitan region with $k = 1$),[33] whereas a large share 20- is attractive for a range of agglomerated regions ($k = 1, 2, 3, 4$).

Population flows: Net migration and natural balance are significant positive in both specifications. These two variables being proxies for the expected future demand, the result 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 and of itself. As such, this is not surprising as 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 determinants - population size, net migration or natural balance - should turn out to be insignificant. The statistical significance of all three variables then hints at the additional role of expectations stretching beyond mere population accounting. Recall from Corollary 2 that the expectations about future demand, which are implied by net migration flows and by the natural balance, should only affect the current number of physicians within an entry equilibrium with limited generational overlap as opposed to an equilibrium with sizeable generational overlap.

---

[33]The overall effect including the interaction term with $k = 1$ is positive yet insignificant.

Our estimates are thus consistent with an unbalanced entry process, where large entry cohorts at some point block the market until they retire. Indeed, we know from figure 3 that this is a likely case due to the large baby-boom cohorts being still active. We will return to the role of expectations in section 4.4.2, where we apply a dynamic model that allows us to identify short-run effects. As expected, the flow variables then lose their significance.

Control variables: According to the literature the effect of population density on physician supply should be positive because higher densities indicate on average, a lower need for travelling either by patients or by physicians and thus a higher demand for and/or greater profitability of treatments. In our case the corresponding parameter is not significantly different from zero. One reason for this finding might be that we control for the age distribution of the population, which, in turn, is correlated with population density. The GDP per capita and the employment rate reflect the standard of living and the condition of the labor market. In addition, higher per capita income is also related to a larger share of private patients, which are more attractive prospects for physicians. More surprisingly perhaps, we also find a significant positive coefficient for the unemployment rate. One explanation is that the unemployment rate acts as proxy for morbidity in the working age population and hence indicates a greater demand for physician services.[34] A similar argument applies to the share of welfare recipients. The share of foreigners and the share of upper secondary school leavers turn out to be insignificant. As expected, the share of school leavers without formal degree exerts a negative influence on the number of physicians. In as far as the variable hints at a relatively poor educational environment, it implies a lower residential utility for physicians who care for the development of their own children. The two variables measuring the development of new residential buildings should also reflect the amenity value of a region.[35] In addition, however, these variables signal an increasing population and thus an increasing demand for physician services. The stock of tourist accommodation hints at the attractiveness of a region for holiday makers. The effect on physician supply should thus be positive both because the region is attractive and because of the demand for physician services generated by holiday makers. Finally, the number of cars per 1,000 residents acts as a measure of geographical mobility. On the one hand, mobility should have a positive effect on physician supply as potential patients find it easier to visit the physician and, thus, exhibit a greater demand. On the other

---

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

[35]The variable 'share of new houses' relates to single and double family homes. The variable 'share of new flats in the stock of flats' measures general property development. In as far as this comes with the simultaneous development of a regional infrastructure (schools, shopping, etc.) it provides an additional measure of a region's attractiveness for young physicians (upon the point of their location choice).

hand, given that we control for population size and density as well as for the age structure, the mobility variable also measures the possibility to consult physicians outside the region. In this case the expected effect is negative, and this effect dominates in our estimates.

The main focus of our analysis being on the age structure, we now take a closer look at the overall effect of the population shares on the supply of physicians according to the estimates presented in table 2. Taking into account the interaction with rurality, the population share 20- increases the supply of physicians in agglomeration regions (type I regions) but lowers it in urbanized and rural areas (type II and III regions). Overall, the population share 60+ has no significant effect in independent cities with more than 100,000 inhabitants, i.e. the most urbanized regions. For all other types of region the share 60+ is negatively related to the number of physicians. Figure 4 depicts the results graphically. The light yellow areas are the independent cities with more than 100,000 inhabitants, where the share 60+ has no effect on the supply of physicians. The light and dark yellow areas are the agglomeration regions, in which the share 20- is positively related to physician supply. The orange (urbanized regions) and red areas (rural areas) are those where both the share of the young and the share of the old are negatively related to the number of physicians.

figure 4 about here

In 2004 about 54% of the population in Germany lived in the orange and red colored regions. Hence, about half of the population lives in regions in which an unfavorable age structure (biased towards the elderly or the very young) dampens the supply of physicians. Rural regions (i.e. the red colored regions) with their relatively high shares of the elderly are thus particularly exposed to a risk of becoming under-doctored. A judgement as to whether or not this risk is materializing, and to what extent, lies beyond the scope of this paper.

4.4 Further Estimations

4.4.1 Physician Density

In this subsection we (re-)estimate the different specifications of equation (17), using physician density as dependent variable. This makes our results more comparable to the findings of other studies, which typically employ physician density. Moreover, we can assess to what extent the presumed measurement error in relation to physician density has a bearing on our results.

table 3 about here

From the results displayed in table 3 we can glance that the measurement bias in the dependent variable is apparently large enough to affect at least some of the estimated coefficients, in particular those related to age structure.[36] The population share 60+ is no longer significant in either of the regressions (1) or (2), and the share of the young population is no longer significant in regression (2), as compared to regression (2) in table 2. The population flow variables, net migration and natural balance, retain their positive significance. This speaks for a rather robust impact of expectations on the supply of physicians.

4.4.2 Spatial Effects

There are good reasons to believe that the regional supply of physicians is not only correlated over time but also across space. In this section we provide a set of estimations which take account of spatial correlations between neighboring regions. Using this more complex approach allows us to address a number of aspects related to the spatial distribution of physician supply.

First, during the entry process physicians are likely to compare regions regarding their attractiveness as a location for setting up practice. By compressing income, competition between physicians is reducing the attractiveness of any particular region. But then a tight market within an otherwise preferred region is likely to direct a physician's location choice to a neighboring region. In this case one should expect a positive correlation in the supply of physicians across neighboring regions. Second, and along a similar line of argument, the closure of a region to entry into a particular specialty should induce the specialists concerned to locate their practice in a neighboring region. If (and only if) the effect of such entry restrictions is substantial we should then find a significant positive spatial correlation in the number of physicians. Third, patient mobility may affect the distribution of physicians if it induces significant patient flows out of or into certain regions. To the extent that physicians within a given region are able to attract patients from neighboring regions we would expect that the population in neighboring regions is positively correlated with the number of physicians in the local region. However, at the same time this should imply a negative correlation in the numbers of physicians across neighboring regions.

Fourth, as we will explain in greater detail below, the spatial estimation approach requires the inclusion of a time lagged dependent variable. While this is done primarily for technical reasons, this feature of the model conveniently allows us to identify the short run effects on the supply of physicians of the demographic (and other) variables. This is because, when included as a regressor, the time lagged dependent variable captures the time trend

---

[36]Also note that population density now takes on significance but with the 'wrong' sign.

in the supply of physicians. Any deviations from this trend must necessarily be of a short-term nature. Thus, if we find significant effects for some of the demographic variables these would relate to their current impact on physician supply only. We would therefore expect that the population flow variables lose their significance as they relate to expectations only. Any significant effects related to the age-shares (and/or their interactions with rurality) would then reflect the (short-term) profitability of the respective patient groups.

Finally, we should stress that an inclusion of the spatial correlations is important in its own right to arrive at unbiased estimations of the demographic make-up. In that regard the spatial estimations also amount to a robustness check of the results found in section 4.3.

When analyzing spatial effects econometrically it is necessary to consider the time dimension as well. Otherwise the estimated effects are biased because of the omitted correlation between spatial and time lagged effects. Hence, we have to employ a spatial and time dynamic model. This allows us to examine both short-term effects of the demographic determinants in the local area and long-term linkages in supply between neighboring 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 any two regions as being contiguous if they share a common border. The corresponding spatial weight matrix W is therefore a symmetric 439×439 matrix.

Common fixed effects estimators that include spatial and time dynamic effects of the dependent variable are biased.[37] On this account, we use a spatial and time dynamic data approach with both regional and time fixed effects as suggested by Yu et al. (2008) and Lee and Yu (2009) .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 (2009) developed a data transformation approach that has the same asymptotic efficiency as the quasi-maximum likelihood estimator when the regional dimension I is not relatively smaller than the time dimension T. The model has the following general specification:

$$S_{It} = \gamma_0 S_{I,t-1} + \lambda_0 W_I S_{It} + X_{It}\beta_0 + \alpha_{I0} + \eta_{t0} l_I + E_{It} \qquad (18)$$

where $S_{It} = (s_{1t}, s_{2t}, \ldots, s_{It})'$ and $E_{It} = (\epsilon_{1t}, \epsilon_{2t}, \ldots, \epsilon_{It})'$ are $I \times 1$ column vectors and the residual ϵ_{it} is i.i.d. across i and t with zero mean

---

[37] 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.

and variance σ_0^2. W_I is an $I \times I$ spatial weights matrix which is nonstochastic and generates the spatial dependence between cross sectional units s_{it}. X_{It} is an $I \times k_x$ matrix of nonstochastic regressors, α_{I0} is an $I \times 1$ column vector of individual fixed effects, η_{t0} is a scalar of time effects, and l_I is an $I \times 1$ column vector of ones. W_I 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 averaging of 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 S in equation (18) equals these variables. $S_{I,t-1}$ is the time lagged dependent variable and $W_I S_{It}$ is the spatial lagged dependent variable. The matrix X_{It} 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 $natural_{it}$. When the number of physicians is the dependent variable, the matrix also contains the average spatial lagged population size.

Table 4 provides the results with the number of physicians and physician density, respectively, as the dependent variables. We only show the results for the demographic variables.[38]. While regressions (1) and (3) are the specifications without the interaction of age groups and rurality, regressions (2) and (4) contain these interactions. For two reasons, the time lagged effects has to be interpreted with caution. First, the period of observation is short and second, and probably more problematically, we have no information on the number of physicians (or physician density) for the year 1999. For the spatial effect a missing year is not problematic.

We first discuss the results for the number of physicians as dependent variable.[39] The spatial lagged effect of the number of physicians is insignificant in this case, which implies that on average there are no systematic spatial interactions driving the location of physicians. Whether this implies the absence of significant spatial spillovers in the supply of physicians or whether positive and negative spillovers just balance out on average cannot be inferred without further analysis. One important implication of this finding is that entry controls at regional (and specialty) level do not systematically affect our estimates. If physicians are exposed to entry restrictions within a particular region one would expect them to locate in neighboring regions. An insignificant spatial lagged effect then means that the entry restrictions in place do not systematically affect our estimates for the aggregate number of physicians at regional level. We can therefore conclude that the estimates provided in table 2 suffer from a negligible omitted variable

---

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

[39]Due to the high values for the time lagged effect in regressions (1) and (2) we additionally apply the Maddala-Wu unit root test on the number of physicians. We test different specifications with and without drift or trend and include one lagged difference term. We can reject the null hypothesis of a unit root at the 1% level in all specifications.

bias.[40]

The effects of the age-shares on the supply of physicians are qualitatively similar to those reported in table 2 with the one exception that the pure age-share effects in regression (2) lose significance within the time-space dynamic framework. In particular, our finding remains valid that both the population 60+ and the population 20- tend to be less attractive in rural regions. The difference to the results reported in table 2 is that now we estimate only the short-term effects related to the profitability of the current population. This allows us to conclude that the treatment of older patients is less profitable on average than the treatment of middle-aged patients, with relative profitability declining in the degree of rurality. The latter result also applies to the treatment of young patients; however, as the treatment of these patients is relatively profitable in urban settings, the average effect is not significantly different from zero.

The variables related to the population flows, net migration and natural balance, lose their significance, with the one exception that net migration retains significance at the 5% level in estimation (2). This notwithstanding, the values of all coefficients have declined and have a lower significance as compared to table 2. The lack of significance with regard to their short-term effects is consistent with our conjecture that the population flows bear on the supply of physicians via the expectations about future demand and via the expectations only.[41] Finally, the effect of population size across regions is insignificant in both specifications (1) and (2). This means that on average the effects of patient flows on the supply of physicians is balanced out. Therefore, while we deem it important to have controlled for possible systematic effects of demand spillovers between regions, it is difficult to infer much more from this result without a more detailed analysis of patient flows.

table 4 about here

Regressions (3) and (4) in table 3 have been run with physician density as the dependent variable. The spatial lagged effect of the dependent variable is significant positive, implying that physician densities are positively correlated across regional boundaries. One could argue that a high physician density in a particular region is associated with strong competition, which induces entrants to locate in neighboring regions. However, we have argued in section 4.3 that physician density comprises a measurement error if location choices are made for the long term. With respect to the spatial

---

[40] Furthermore, remember that if such bias were present its likely direction would imply an underestimation of the demographic effects.

[41] Given its significance at 5% level we may also conclude from regression (2) that the impact of net migration on the demand for physician services materializes earlier than the impact of the natural balance.

lagged effect the expected bias is positive.[42] Hence, we must conclude that the insignificant effect in regressions (1) and (2) provides unbiased evidence that physician supply is not significantly correlated across regions. Regarding the age-shares, we obtain results qualitatively similar to those reported in table 3.[43] As expected, the flow variables mostly lose their significance.[44]

5 Conclusions

Population ageing is widely expected to come with an increased per capita demand for (ambulatory) physician services. For a policy-maker who is interested in ensuring an adequate provision of health services, it is then important to know whether the supply of physicians is sufficient for this demand to be met. Typically, population change within a country is not uniform but varies across different regions. Here, the geographic distribution of the population - both in terms of size and in terms of the age-structure - is shaped, in particular, by inter-regional migration. As it is predominantly the young who migrate, the process of population ageing may be accelerated or slowed down depending on whether there is emigration or immigration at regional level. This begs the second policy question as to how the distribution of physicians across regions responds to geographically differentiated population change and whether the outcome reflects the spatial distribution of needs.

At regional level, demographic developments and geographical location thus imply a framework with both an intertemporal and a spatial dimension. We have examined both theoretically and empirically how the supply of physicians across a set of different regions responds to (differential) population change. In so doing we have identified a number of aspects that should be borne in mind when thinking about the provision of physicians services over space and time. First, in as far as physicians make long-term decisions about their location and base these on the (expected) profitability from offering treatments to the resident population, the demographic make-up of a region becomes an important determinant not only of the current income but also of the expected income from the provision of services to the future population. Indeed, we find strong empirical support for the important role of expectations in that both net migration and the natural balance are significantly and positively related to the current number of physicians resident in a region even when controlling for the current size and density of the regional population. In addition, these variables turn out to be insignificant (or less significant) as short-term predictors of supply. We also

---

[42] The underlying assumption is that the population change within neighboring regions is more likely to be positively correlated than not.

[43] Note that the age share 20- is now significant at 5% level.

[44] Only net migration remains significant at 5% level in regression (3), but in comparison to the estimation without spatial and time lags (see table 3) the coefficient is much lower.

find effects of the age-shares of the population on physician supply which are broadly consistent with the fact that they reflect both a long-run and short-run dimension. For instance, a larger share of the elderly amongst the current population may reflect a high demand for services today but leads to the expectation of a smaller population in the future.

Second, the relative profitability of treating different patient groups varies with the character of a region. In particular, we find evidence that treating both the young (20-) and old (60+) patient populations becomes progressively less profitable relative to treating the middle-aged when the regional character changes from urban to rural. While our data do not allow us to identify the causal mechanism, this evidence is consistent with a health care system in which the higher treatment costs of the oldest and youngest patients (e.g. due to the higher incidence of costly home visits) within a rural context are not fully compensated by the payment system. As it turns out the loss of relative profitability with higher degrees of rurality is similar for the young and old populations. Interestingly, however, the (long-run) marginal effect on physician supply of a shift of population from the youngest to the oldest age group is strongly negative for any kind of region. This again bears indirect evidence on the importance of expectations, where a large share of the young (old) population indicates the scope for population growth (decline).

Third, the way in which physician supply responds to changes in the local population depends crucially on the entry pattern of physicians. In particular, the expectations about future demand (and thus the future population) become the more important as a predictor of the current number of physicians, the more uneven is the process of physician entry over time. Thus, our finding that expectations regarding future population matter is consistent with the highly cyclical pattern of physician entry in Germany, where very little entry is taking place (and will take place) up to the point at which the baby-boomer physicians are going to retirement.

Fourth, we find that the physician supply and the patient population in neighboring regions do not have a significant impact on the local supply of physicians. Thus, we find no support for the conjecture that the propensity to locate in a particular region is somehow systematically related to the number of physicians in neighboring districts (neither positively where physicians seek to evade market tightness in certain preferred regions, nor negatively if there is strong competition for patients across regions). Neither do we find a systematic impact of patient flows across regions. These findings also indicate hat our results are not systematically driven by the potential closure to entry into certain specialties within certain regions. This notwithstanding, a more detailed analysis of the role of spatial interactions clearly provides scope for future research.

Over the past decade or so, in Germany the extent of population ageing at regional level has not so much been driven by low or declining fertility

rates but by inter-regional migration flows. Over our period of observation 1995-2004 sustained migration has occurred predominantly from rural regions in Eastern Germany to agglomerations in the West. The outmigration of predominantly young people has, thus, contributed to a strong degree of population ageing in the rural (and less agglomerated) regions in the East. According to our data the largest population share of the age cohort 60+ in 2004 lived in the Eastern German regions Hoyerswerda (33.4%) and Görlitz (32.9%). Between 1995 and 2004 the share of 60+ has increased by 89% in Hoyerswerda and by 41% in Görlitz. These developments are reflective of many other regions, in particular in the Eastern part of Germany, that experience ageing in a similar manner. According to our estimates, physician supply is strongly and negatively related to the population share 60+ within rural or less agglomerated areas, while this is not the case within metropolitan areas. Likewise, a high share of the population 20- is significantly positively related with physician supply in metropolitan areas. Provided our estimations have identified a causal effect, then they suggest the scope for an increasingly divergent development in the supply of physicians. While there is a real danger of many rural regions with a high share of the elderly population turning out to be under-doctored in the future, at the same time agglomerations with relatively young populations may turn out to be over-doctored.

In the mean time, the medical profession itself has been undergoing a considerable process of ageing. In 2007 the share of baby-boomers (between 42 and 65 years old) amongst the population of physicians was 85.7 %. From this it follows that a large number of physicians will retire over the next 20 years or so, and this is likely to affect significantly the medical provision in rural areas and in particularly in the Eastern part of Germany. According to the German Medical Association (Bundesärztekammer) about 13% of the Eastern German regions have already been found to be 'under-doctored' in 2007, and more than half of the hospitals in Eastern Germany have problems to fill vacancies in medical employment. In addition, the numbers of medical students and graduates have been declining over the past years, while increasingly attractive job offers from abroad have been sharpening the shortage of young physicians. A pessimistic view is therefore that real problems related to the adequacy and distribution of physician services may, indeed, only be about to surface. Whether and how health policy-making may ameliorate these problems is an issue which has been discussed elsewhere (e.g. Newhouse 1990, Bolduc et al. 1996, Hann and Gravelle 2004, Iversen and Kopperud 2005, Elliot et al. 2006, Klose and Uhlemann 2006, Kopetsch 2006). We may conclude with the summarical observation that short-term incentives for opening practice in under-doctored regions may well prove to be inadequate unless they are tied in with (i) a reimbursement system that is properly adjusted for demographic and geographic heterogeneity and (ii) long-term regional policies that address or otherwise account for regional

heterogeneity.

6   References

Abraham, J.M., Gaynor, M.S., Vogt, W.B., 2005, Entry and Competition in Local Hospital Markets, NBER Working Paper 11649.

Arcury, T.A., et al., 2005, The Effects of Geography and Spatial Behavior on Health Care Utilization among the Residents of a Rural Region, *Health Services Research* 40, 135-155.

Bolduc, D., Fortin, B., Fournier, M.-A., 1996, The Effect of Incentive Policies on the Practice Location of Doctors: A Multinomial Probit Analysis, *Journal of Labor Economics* 14, 703-723.

Cooper, R.A. et al., 2002, Economic and Demographic Trends Signal an Impending Physician Shortage, *Health Affairs* 21, 140-154.

Department of Health, 2004, Global Sum Allocation Formula, London.

Dionne, G., Langlois, A., Lemire, N., 1987, More on the geographical distribution of physicians, *Journal of Health Economics* 6, 365-374.

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

Dormont, B., Samson, A.-L., 2008, Medical Demography and Intergenerational Inequalities in General Practitioner's Earnings, *Health Economics* 17, 1037-1055.

Dranove, D., Shanley, M., Simon, C., 1992, Is Hospital Competition Wasteful?, *RAND Journal of Economics* 23, 247-262.

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.

Elliott, B., et al., 2006, Adjusting the General Medical Services Allocation Formula to Reflect Recruitment and Retention Difficulties, Final Report, University of Aberdeen: Health Economics Research Unit.

Foster, S.A., Gorr, W.L., 1992, Federal Health Care Policy and the Geographic Diffusion of Physicians: A Macro-Scale Analysis, *Policy Sciences* 25, 117-134.

Giuffrida, A., Gravelle, H., 2001, Inducing or Restraining Demand: The Market for Night Visits in Primary Care, *Journal of Health Economics* 20, 755-779.

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

Hann, M., Gravelle, H., 2004, The Maldistribution of General Practitioners in England and Wales: 1974-2003, *British Journal of General Practice* 54, 894-898.

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

Hurley, J.E., 1991, Physicians' Choices of Specialty, Location and Mode: A Reexamination within an Interdependent Decision Framework, *Journal of Human Resources* 26, 47-71.

Iversen, T., Kopperud, G.S., 2005, Regulation versus Practice - the Impact of Accessibility on the Use of Specialist Health Care in Norway, *Health Economics* 14, 1231-1238.

Jürges, 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.

Klose, J., Uhlemann, T., 2006, Fehlallokation in der vertagsärztlichen Versorgung: Abbau und Vermeidung von Über- und Unterversorgung [Misallocation in the Provision of SHI Contract Physicians: Reducing and Avoiding Excess and Shortfalls in Supply], *GGW* 3, 7-17.

Kopetsch, T., 2004, Die ärztliche Profession in Deutschland: Eine Darstellung der zahlenmäßigen Entwicklung in der Vergangenheit, eine Bestandaufnahme des Status quo sowie ein Ausblick in die Zukunft [The Medical Profession in Germany: Sketching Past Developments, Taking Stock of the Status-Quo, and Providing an Outlook], *Journal of Public Health* 12, 304-320.

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 thesis of supplier-induced demand], *Schmollers Jahrbuch*.

Kopetsch, T., Munz, H., 2007, Analyse des Niederlassungs- und Wanderungsverhaltens von Vertragsaerzten und -psychotherapeuten in Deutschland [Analysis of the Location and Migration Choices by SHI Contract 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., 2009, A Spatial Dynamic Panel Data Model with Both Time and Individual Fixed Effects, *Econometric Theory*, forthcoming.

Morris, S., Sutton, M., Gravelle, H., 2005, Inequity and Inequality in the Use of Health Care in England: an Empirical Investigation, *Social Science & Medicine* 60, 1251-1266.

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

Newhouse, J.P., 1990, Geographic Access to Physician Services, *Annual Review of Public Health* 11, 207-230.

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.

Nocera, S., Wanzenried, G., 2008, On the Dynamics of Physician Density: Theory and Empirical Evidence from Switzerland, *Swiss Journal of Economics and Statistics* 104, 247-282.

Nuscheler, R., 2003, Physician Reimbursement, Time Consistency, 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.

Rosenthal, M.B., Zaslavsky, A., J.P. Newhouse, 2005, The Geographic Distribution of Physicians Revisited, *Health Services Research* 40, 1931-1952.

Scott, A., 2001, Eliciting Preferences for GPs' Pecuniary and Non-Pecuniary Job Characteristics, *Journal of Health Economics* 20, 329-347.

Statistische Ämter, 2007, Bevölkerungs- und Haushaltsentwicklung im Bund und in den Ländern: Demographischer Wandel in Deutschland [Statistical Offices: Development of Population and Households at Federal and State Level: Demographic Change in Germany], vol. 1.

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.

Taylor, D.H., Leese, B., 1998, General practitioner turnover and migration in England 1990-94, *British Journal of General Practice* 48, 1070-1072.

Thode, N. et al., 2005, Einflussfaktoren auf die ambulante Inanspruchnahme in Deutschland [Determinants of the usage of ambulatory care in Germany], *Bundesgesundheitsblatt* 48, 296-306.

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.
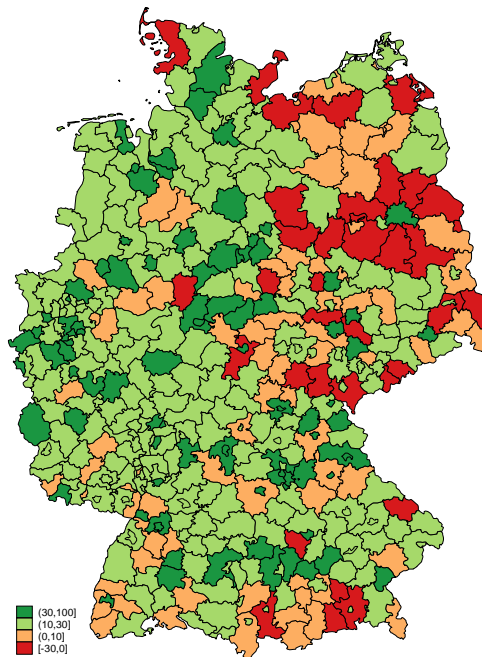
7 Appendix

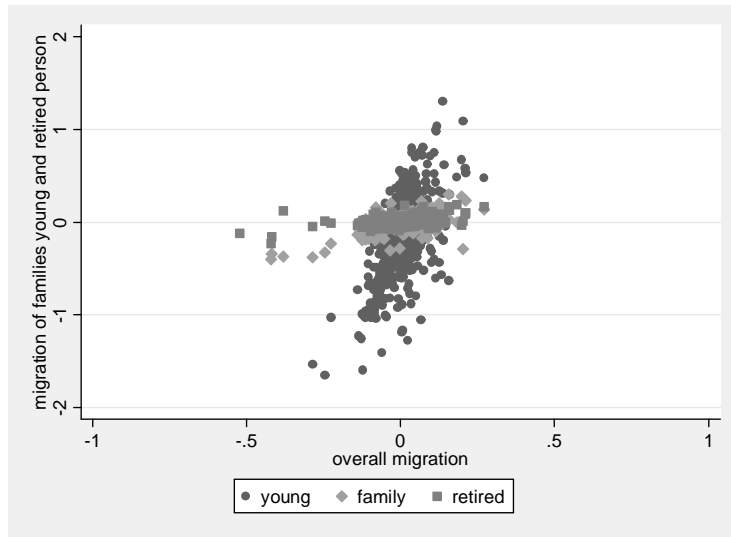Figure 1: Change in Regional Physician Supply in Germany
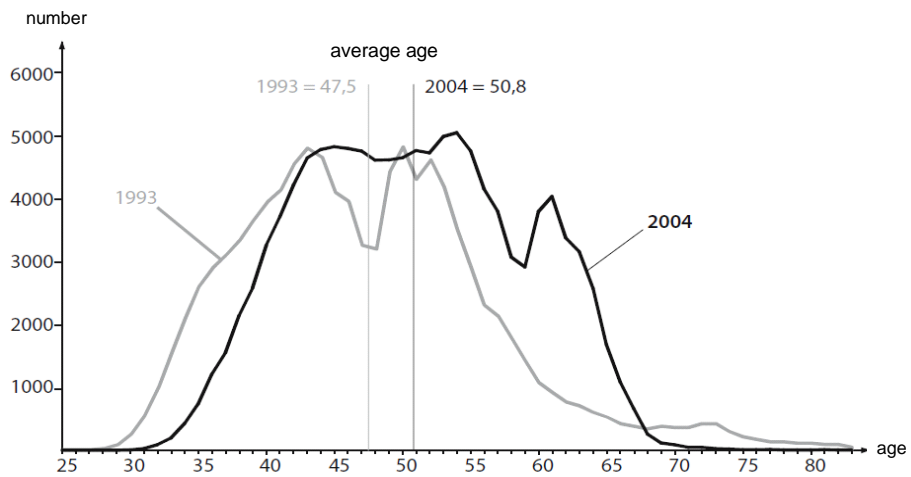
Figure 2: Net Migration Flows by Age Groups

Figure 3: Age Structure of SHI-Physicians in 1993 and 2004 (source: SHI)

41

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

	(1)	(2)	(3)
share 60plus	-5.012‡	-6.653‡	-3.554‡
	(0.940)	(1.766)	(1.482)
share 20minus	-30.592‡	12.116‡	4.923†
	(1.544)	(1.710)	(2.511)
net migration	0.754‡	2.181‡	1.866‡
	(0.281)	(0.241)	(0.220)
natural balance	9.655‡	16.667‡	11.611‡
	(0.924)	(2.236)	(1.464)
population size	0.002‡	0.002‡	0.002‡
	(0.0001)	(0.0002)	(0.0002)
regional FE		✓	✓
time FE		✓	
state level time FE			✓
R^2	0.929	0.233	0.716
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 degree	-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 residents	-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.

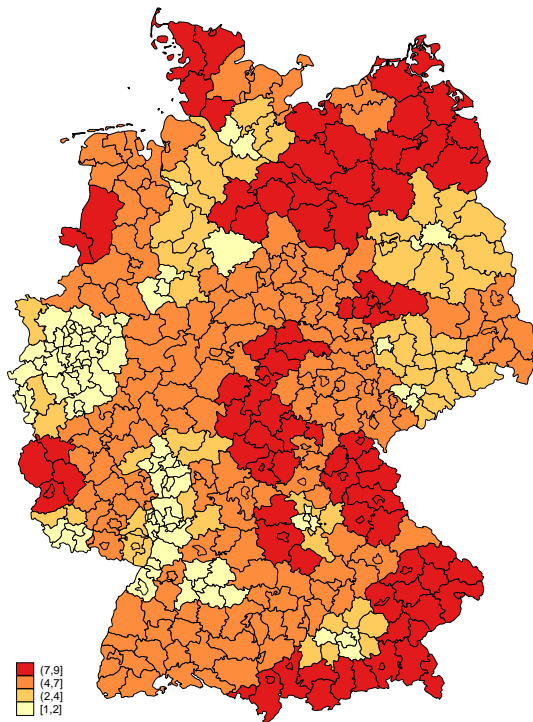|	
■	(7,9]
■	(4,7]
■	(2,4]
■	[1,2]

Figure 4: Estimated Effects of the Population Age Structure on 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 degree	-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 residents	-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 degree	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 residents	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 residents

2 Districts with at least 300 residents per square kilometer

3 Districts with at least 150 residents per square kilometer

4 Districts with less than 150 residents per square kilometer

Type II: Urbanized Regions

5 Independent Cities with more than 100,000 residents

6 Districts with at least 150 residents per square kilometer

7 Districts with less than 150 residents per square kilometer

Type III: Rural Areas

8 Districts with at least 100 residents per square kilometer

9 Districts with less than 100 residents per square kilometer

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