



UNIVERSITÀ
DEGLI STUDI
DI UDINE

Università degli studi di Udine

The determinants of individual health care expenditures in the Italian region of Friuli Venezia Giulia: evidence from a hierarchical spatial model estimation

Original

Availability:

This version is available <http://hdl.handle.net/11390/1123046> since 2021-03-22T18:29:28Z

Publisher:

Published

DOI:10.1007/s00181-017-1372-9

Terms of use:

The institutional repository of the University of Udine (<http://air.uniud.it>) is provided by ARIC services. The aim is to enable open access to all the world.

Publisher copyright

(Article begins on next page)

The determinants of individual health care expenditures in the Italian region of Friuli Venezia Giulia: evidence from a hierarchical spatial model estimation

Luca Grassetti and Laura Rizzi

Received: date / Accepted: date

Abstract This work investigates the determinants of health care expenditures, such as drug prescriptions, inpatient care, and outpatient care, of the resident population of the Region of Friuli Venezia Giulia (Italy). The phenomenon of interest is examined here by considering a cross-sectional register-based dataset on individual expenditures exhibiting a cross-classified hierarchical structure. In fact, patients (about 1,000,000) are grouped by general practitioners and municipalities. Does the evidence in individual data analyses support the results of the micro and macroeconomic literature? The adoption of disaggregated data allows us to disentangle the role of the micro and macroeconomic determinants of the expenditures. Moreover, the degree of interdependence between neighbouring municipalities is measured by accounting for the spatial correlation in the error convolution. A feasible two-stage Heckit method has to be adapted to encompass the zero inflation issue, to consider the hierarchical structure of data and to study the spatial diffusion process of the expenditures in the sample selection model framework. The main results on the determinants of health care expenditures at the macro-level are confirmed in our analysis on disaggregated data. On the contrary, however, the substitution effect, which is typically observed in aggregated data, has not been confirmed by the present research. Moreover, the selection process appears to be relevant in drug prescriptions and outpatient care expenditures

Luca Grassetti (orcid.org/0000-0003-1997-8001) and Laura Rizzi (orcid.org/0000-0002-2470-3953)
Dept. of Economics and Statistics - University of Udine, Italy
Via Tomadini, 30/A 33100 Udine (Italy) Tel: +39 0432 249592
Fax: +39 0432 249229
E-mail: luca.grassetti@uniud.it

and a significant spatial correlation in both the selection and the outcome equations emerges from the structure of error components.

Keywords Health Care Expenditures · Health Econometrics · INLA · Multilevel Models · Sample Selection · Spatial Econometrics

JEL CODES: C01, C11, C21, H75, I11, I18

Acknowledgement

The authors would like to thank Prof. Badi H. Baltagi, prof. F. Moscone and all the participants to the “IV Health Econometrics Workshop” for the precious and helpful suggestions and comments. We also would like to thank all the blind reviewers for their fundamental suggestions.

1 Introduction

During the last decades, the relevance of health care expenditures (HCE) has substantially increased, leading to a growing pressure on public budgets and to long-run sustainability issues in countries with public health care services. For this reason, the analysis of HCE growth and of health care drivers has become an object of active research (*Horizon 2020*). Since the seminal paper by Newhouse (1992), the assessment of the determinants of HCEs has been a matter of extensive policy debates. The related literature on the topic has taken a variety of directions, either by studying the demand (use) of health care services or by carrying out the analysis of HCE itself, both at the micro and macro-level. In the former case, studies have focused on the factors affecting the probability of use and/or the amount of expenditures at the individual level (see, for instance, Angulo et al., 2011; Albouy et al., 2010). In the latter case, the determinants of health expenditures have been explored at the aggregate level (as in Lopez-Casasnovas et al., 2005) where the relevance of income, age, education, and general health status has been assessed.

The main objective of this research is to show how the complexity of the available information can be used to develop a detailed policy-making process. In particular, the empirical analysis aims at investigating the extent to which the socio-demographic characteristics, the heterogeneity of health care delivery inputs, and the proxies of individual health status affect individual heterogeneity of HCEs. The joint analysis of the micro and the macroeconomic drivers of HCE in a spatial economics framework represents the main contribution of the present article. In fact, the results of the present analysis can be used to develop an evidence-based policy-making process focusing on the behaviour of general practitioners (GPs hereafter), on the estimated effect of the local administrations, on the results about the spatial diffusion process of the HCE phenomenon, and on the role of the observed expenditure drivers. The database used in the empirical analyses has been derived from the fully developed administrative regional health informative system (SISSR). This ‘regionally-limited’ dataset is considered in order to reduce the unobserved heterogeneity characterising the HCE phenomenon. Moreover, in Italy each region has its own health service management and regional differences reduce the opportunity to study disaggregated HCEs at the national level.

Individual HCE includes expenditures for different health care services, which are strictly connected to each other and provide the grounds for a multivariate characterisation of the expenditure phenomenon. In the present paper, the economic evaluation of individual HCE focuses on drug prescriptions, inpatient care and outpatient care services only. This limitation is due to the availability and reliability of the administrative data collected in the regional health databases. However, these three sources of expenditures cover the largest part of National and Regional Health System budgets. In OECD countries 79.2% of HCEs are determined by these three components (see OECD Health Data 2013 for further details). Several studies have been developed to investigate substitution effects among different types of health care. In some cases, investigations on substitution effects have represented the main purpose of the studies in question (as, for instance, in Helms et al., 1978; Lichtenberg, 2007). However, in the majority of cases substitution has been studied as a side-effect of the changes characterising the health care economic framework (as in Goldman et al., 2006). For instance, these studies show that negative changes in drug consumption are coupled to an increase in emergency room and inpatient hospital care demand. With this research, we want to deepen the analysis of the presence of trade-offs between different kinds of health care services at the patient level. One of the main results of the present analysis is that the typical behaviour observed at the aggregated level is invalidated at the individual level.

The need for health care and the amounts of expenditures are studied here by focusing also on spatial interdependencies and on the concentration of unobservable risk factors. Spatial mechanisms may be caused by different sources and learning effects among municipalities, leading to a clustering of demand of health care. By using individual data, neighbourhood heteroskedasticity can be taken into consideration in order to disentangle the effect of endogenous and exogenous factors (observed at different levels of aggregation). The correlation between spatial units is assessed by considering a clusterisation at the municipality level. Given the hierarchical structure of the data, the usual spatial models such as the spatial lag model and the spatial cross-regressive specification (as given in Anselin, 2002) or the spatially autoregressive model, cannot be easily applied in the present framework. Moreover, the dataset presents an even more articulated hierarchical structure, given that individuals are grouped both in GPs and municipalities. For these reasons, spatial correlation has been introduced at the municipality level in a hierarchical

Bayesian model specification (as in Eibich and Ziebarth, 2014). The most generic model specification considers the convolution of GPs and municipality (unstructured and spatial) effects. The estimation results show that the spatial diffusion process is heterogeneous for different kinds of expenditures. Furthermore, the analysis of the trade-off between these components allows us to identify the relevance of the behaviour of different macro units in relation to the HCE phenomenon. We expect a prevalent GP component in the models for outpatient care and for drug prescriptions. We also expect municipality effects and/or spatial dependencies to arise in the model for inpatient care, as hospitalisation may be affected by economic, demographic, social, labour market, and urbanisation features of the neighbouring units.

The paper is organised as follows. Section 2 summarises some recent empirical evidence on the determinants of health expenditures and on related public policies. Section 3 is devoted to the description of the model specification. Section 4 presents the data structure, while Section 5 provides the estimation results. Comments, concluding remarks and future developments of this research are reported in Sections 6 and 7.

2 The analysis of HCE

According to the OECD Health Data 2014, public HCEs accounted for 7.4% of the Italian Gross Domestic Product (GDP hereafter) in the year 2010 (while the average value for OECD countries reached 6.8%). The economic relevance of these expenditures and the availability of micro and macro datasets on health care has encouraged the development of studies on the drivers of the HCEs.

Several studies focus on macro data analysis and cross-country frameworks in order to identify the determinants of HCE (see, for instance, Gerdtham et al., 1992; Gerdtham and Jönsson, 2000). By exploring the phenomenon at the macro-level, Wang (2009) and Baltagi and Moscone (2010) state that the GDP, the proportion of population over 65 years of age, the level of urbanisation and deprivation along with the number of hospital beds, all have significant effects on expenditures. Aggregate data on HCE have also been considered to study the income effect and the income elasticity in the long-run. In a panel data framework, OECD countries have been studied by Baltagi and Moscone (2010), while the case of the U.S. has been examined by

Freeman (2003). Other macro data studies have been developed within single countries, comparing jurisdictions or regions in order to disentangle local heterogeneity (see Giannoni and Hitiris, 2002; Di Matteo and Di Matteo, 1998).

Thanks to the growing availability of administrative databases and the presence of increasingly efficient information systems, it has been possible to develop micro data studies on the topic (e.g. Deb et al., 2006; Werblow et al., 2007). Individual data are often collected to study the effect of patients' socio-economic characteristics, chronic diseases prevalence, illness severity and general health status on HCE and on health care demand (as in Wong et al., 2011). Typically, these studies concern specific pathologies and are developed in controlled experimental frameworks (e.g. Madden, 2008). Though fewer studies use individual data to evaluate the spatial heterogeneity of the need and the demand of health care (as, for instance, in Wang, 2009). At the micro-level, the analyses in Getzen (2000), Madden (2008) and Deb et al. (2006) identify age, gender, health status, income, and educational level as the main determinants of individual health care demand.

In our study, several variables have been collected at the individual level: age class, sex, health-related fee-exemptions, and the demand for other health services (considered as proxies of co-morbidity and general health status). All these variables are designed in such a way to allow for individual heterogeneity; they are expected to explain the majority of the amount of HCE. The model specification is then completed by including some characteristics of the grouping factors: GPs and municipalities. Foster et al. (1980) and Wildasin (2013) suggest that differences across spatial units can be explained by the characteristics of the municipalities. The context which the analysis is based on is substantially homogeneous in terms of policies, taxes, and general socio-economic conditions. For this reason, only demographic and geographical factors are included in the final model.

The characteristics of our dataset imply that attention needs to be paid to two common issues in microeconomic data analyses: zero-inflation and group specific heterogeneity (due to the hierarchical data structure). Some alternative approaches have been developed in the econometric literature to deal with the former problem. In particular, the Two-Part, sample selection and latent class models (see, for instance, Buntin and Zaslavsky, 2004; Madden, 2008; Deb and Trivedi, 2002, respectively, for the three approaches) can be adopted to this end. All these

alternatives define individual expenditures as the combination of two stochastic processes (one for the presence of health care expenditures and the other for their amount). Mixture models, where data are considered as a combination of a zero mass distribution and of the common cost function, also play a relevant role in this type of analysis (as explained in Deb and Trivedi, 1997). Finally, copula bivariate probit models (Winkelmann, 2012) represent an alternative strategy for the estimation of sample selection models. Given the complex hierarchical structure of the dataset and the computational issues related to the large dataset size, the sample selection model paradigm is adopted here. In particular, the Heckit correction is developed in the two-step estimation procedure (Heckman, 1979). A probit mixed effect model is adopted for the selection process (to overcome the zero-inflation issue), while a linear mixed effects model is used to define the outcome equations. In fact, the peculiar hierarchical structure of the data allows us to consider two distinct non-nested levels of data aggregation: GPs and municipalities. The final model specification accounts for these in a cross-classified random effects framework. The decision to adopt the random effect paradigm instead of the fixed effect one is due to the large number of GPs observed in the sample (the model matrix is considerably larger when considering the fixed effect specification). In particular, the correlation related with the influence of the general practitioners (informal ‘gatekeepers’ for a part of the health services) is accounted for by considering a specific Gaussian term. For the municipality specific error term, the choice of the random effect paradigm is connected with the specific convolution approach selected to model the spatial correlation structure. Moreover, these choices are supported by Bayesian common practice, also, as in this approach, the difference between the fixed and the random effects specification is not as relevant as in the classical inferential framework.

Decentralisation designs require the analysis of spatial patterns of health. For this reason, spatial econometric models have been introduced in health care studies (see Moscone and Knapp, 2005; Rosenberger et al., 2005; Costa-Font and Pons-Novell, 2007). As a matter of fact, statistical tools for spatial analyses have been thoroughly applied to health care data. For instance, Moscone (2011) and Moscone et al. (2007) analyse the prevalence and incidence of diseases and the related costs. Costa-Font and Pons-Novell (2007) examine the potential spillovers of expenditures across the geographical areas of a country. A similar analysis is developed in Baltagi and Moscone (2010) by considering the spillover phenomenon in a cross-country international HCE

comparison. As commonly stated in the cited literature, one of the major assumptions of spatial models is that ‘units sharing an observable characteristic, such as industry or location, also share unobservable characteristics that would lead the regression disturbances to be correlated’ (Moulton, 1990). This conforms well with the existence of some contextual factors of patients (social and economic status, air pollution, exposure to acoustic pollution, distance from a hospital, etc.) which are likely to explain the geographical variability in the use of services. As a matter of fact, neighbouring authorities share common general population characteristics which affect expenditures in health care, and interactions among individuals may lead to a collective behaviour, which empirically represents a spatial correlation pattern. In particular, interactions among individuals may lead to a collective behaviour, which empirically corresponds to a spatial correlation pattern. The spatial correlation can be the outcome of common exogenous environmental conditions of HCE or of the unobserved determinants of interdependence. In the present work, the introduction of the spatial correlation structure in the municipality random component allows us to specify the models by means of a feasible neighbouring system. The observed individuals are, in fact, nested in a relatively small number of municipalities. The presence of latent characteristics which are common to the observational units within the municipalities (such as care habits, the availability of specific community services, etc.) and sometimes to the spatial units themselves justifies the choice of the specific convolution adopted for the municipality level error component. The correlation between the municipalities must then be considered in order to obtain efficient estimates and it often represents a serious issue from the computational point of view. The spatial diffusion process can be characterised by the municipality specific effect (called unstructured effect) and the neighbouring effect associated with the unknown characteristics shared between the municipalities (called spatial effect). The adopted solution considers a Gaussian Random Markov Field (GRMF) specification. The idea of the GRMF is to limit the non-zero elements of the precision matrix (the inverse of the variance covariance matrix) only to the diagonal of the matrix and to the units belonging to neighbourhood. Following the Conditional AutoRegressive (CAR) approach, introduced in Besag (1974), the spatial structure is modelled by considering the random effects in a hierarchical Bayesian model (as, for instance, in Eibich and Ziebarth, 2014). Several different models have been tested for the spatial component, in particular, by focusing on the intrinsic Conditional Autoregressive (CAR) specifications (see

Besag, 1974; Besag et al., 1991, for further details). The possibility of considering a very general error structure is not only important for the interpretation of the model. More specifically, adopting a very general convolution error term for the model specification we can account for different sources of heteroskedasticity and correlations between observed units.

3 The spatial HCE function specification

To our best knowledge, studies on individual health data which also include spatial correlation structures are rare in the health economics literature. Adapting the sample selection model specification (as defined in Leung and Yu, 1996; Madden, 2008) to the cross-sectional hierarchical framework, the model can be finally defined as

$$y_{ij,m} = \begin{cases} \beta' \mathbf{X}_{ij,m} + u_j + v_m + \epsilon_{ij,m} & \text{for } y_{ij,m}^d = 1 \\ 0 & \text{for } y_{ij,m}^d = 0, \end{cases} \quad (1)$$

where $y_{ij,m}$ is the logarithm of HCEs, i , j and m identify individuals, GPs, and municipalities, respectively. The linear predictor $\beta' \mathbf{X}_{ij,m}$ captures the effects of covariates, which are observed variables measured at the different levels of the hierarchical structure and includes the model intercept. The model error is the convolution of the idiosyncratic error term $\epsilon_{ij,m}$, the random effect connected with heterogeneity between GPs (u_j) and the municipality specific random effect (v_m). The process $y_{ij,m}^d$ can be based on the probit generalised linear mixed model. The Heckit method is adopted to check for selection bias. As stated in the introduction, the multivariate HCE phenomenon is studied here by estimating specific equations for each type of expenditure. Correlations between the three HCE sources are partially introduced in the models by means of dummy variables. The adopted estimation algorithm is then characterised by the following selection and outcome equations.

Step 1 The probit model for the presence of a positive expenditure is based on the following linear predictor

$$\eta_{ij,m} = \beta_0 + \beta_1 \mathbf{P}_{ij,m} + \beta_2 \mathbf{GP}_j + \beta_3 \mathbf{MUN}_m + \beta_4 \mathbf{HS}_{ij,m} + u_j + v_m \quad (2)$$

Step 2 Considering Heckman's correction, the specification of a linear model for the subset of positive expenditures only is as follows

$$\begin{aligned} E[\log(\mathbf{HCE}_{ij,m}) | \mathbf{HCE}_{ij,m} > 0] = & \beta_{0j}^* + \beta_1^* \mathbf{P}_{ij,m} + \beta_2^* \mathbf{GP}_j + \beta_3^* \mathbf{MUN}_m + \beta_4^* \mathbf{HS}_{ij,m} \\ & + u_j^* + v_m^* + \lambda \frac{\phi(\hat{\eta}_{ij,m})}{1 - \Phi(\hat{\eta}_{ij,m})} \end{aligned} \quad (3)$$

Equations (2) and (3) consider a set of explicative variables observed at different levels of aggregation. At the patient level, \mathbf{P} and \mathbf{HS} denote socio-economic and health status variables, respectively. The characteristics of general practitioners and the explicative variables at the municipality level are identified by \mathbf{GP} and \mathbf{MUN} , respectively. The exclusion restriction is fulfilled by considering a different set of variables for the two estimation steps.

Both the idiosyncratic error term and the random effect of the GPs are supposed to be independently and normally distributed with zero means and variances σ_ϵ^2 and σ_u^2 , respectively. The municipality level error component (v_m) is supposed to account for the spatial distribution of the phenomenon of the health expenditures and for the correlation between units in a municipality. v_m is finally defined as the convolution of an unstructured Gaussian term (d) and of the spatial component (s) following Besag et al. (1991) approach (obtaining the so-called BYM or convolution model)

$$v_m = d_m + s_m \quad (4)$$

where $d_m \sim N(0, \sigma_d^2)$, $s_m | s_l \sim N\left(\frac{1}{n_m} \sum_{l \in \mathfrak{N}(m)} s_l, \frac{\sigma_s^2}{n_m}\right)$, with $m \neq l$, $\mathfrak{N}(m)$ represents the neighbourhood of m^{th} unit and n_m is the number of neighbours of the observed unit. The spatial units in the considered dataset are located on an irregular grid. For this reason, we finally decided to adopt the iCAR specification, where the correlation between units is intrinsically introduced by defining the mean of the Gaussian distribution of the component. The variance-covariance matrix is given by $\Sigma = \sigma_s Q^{-1}$ where $\text{diag}(Q) = \{n_m\}_{m=1}^M$ and the generic element $Q_{m,l}$ is -1 if $m \neq l$ and m^{th} and l^{th} units are neighbours and 0 otherwise. As introduced previously, the two components of the municipality random effect are included in order to capture the unobserved common characteristics of local administrations and of the neighbourhood effect, respectively. The relative magnitudes of the two components and the trade-off between them can be studied

to characterise the spatial process at play. The obtained information can be useful to describe the kind of spatial dependence observed in the spending process. For each source of expenditure, the error structure has been tested by comparing the DIC (Spiegelhalter et al., 2002) and the WAIC (Watanabe, 2010) measures. In some cases, the information criteria suggest to consider a simplified error structure. Notwithstanding, the estimation of the linear predictor and the relevance of the other error components are not affected by the changes in the error structure. For this reason, the following results are obtained by considering the most general model specification in all the empirical analyses.

In the model specification, the strength of the interactions between spatial units can be affected by the choice of the spatial weighting system. For this reason, several alternative metrics have been tested. In particular, the first and second order contiguity matrices and the neighbourhood based on truncated Euclidean distances are compared with one another. The latter solution results to be the best alternative on the basis of the information criteria. In particular, as suggested by the cited authors, differences larger than or equal to 10 should be considered as significant. The truncation distance is finally calibrated by means of empirical trials. In order to identify the optimal value, the model has been estimated by considering different alternative cut-offs: 10, 15, 20, and 30km ranges. Different distances have been identified as the best solution for the various models involved in the estimation process. In the final model specification, a unique solution is adopted for all the models. In particular, the neighbourhoods are built by considering municipalities within a 15km range. In fact, this solution is the best one in two of the tested cases and the information criteria present acceptable values also for the other models. Moreover, this solution is acceptable also from an empirical point of view as it identifies areas which are sufficiently large but not too large (a larger range would produce neighbourhoods covering most of the region). The average number of neighbouring units is 10.62 with a minimum and a maximum of 0 (only one municipality has a null-neighbourhood) and 28, respectively.

All the analyses have been developed in R (R Core Team, 2017). The Integrated Nested Laplace Approximation approach (as described in Rue et al., 2009) is finally adopted in the model estimation by using R-INLA library (see Martins et al., 2013). When considering this approach, a large number of Bayesian hierarchical models can be estimated by means of direct numerical calculation of posterior densities, thus avoiding the time-consuming simulations.

4 The data

The dataset was obtained by collecting the population regional health databases on all the individuals resident in the Italian region of Friuli Venezia Giulia (FVG). Information on about one million patients was collected for the year 2010. The population data are grouped into 217 regional municipalities and 1,078 GPs. Record linkages and post-editing processes led to a total of 965,579 observations included in this study. The archives taken into consideration collect data on: general practice care, hospital care, drug prescription, and outpatient services. Such archives include vast quantities of data both on pathologies and on health care profiles. Furthermore, individual health care burdens can be derived from information on other kinds of health care (such as home care and emergency services) also. Finally, the population regional registry provides the socio-demographic characteristics of all the resident patients. Some municipality characteristics result from the national statistical institute database (iStat). For each spatial unit the per-capita income, the proportion of citizens presenting the tax returns, the population density and some geographical features are observed (such as altitude classification). The characteristics of GPs finally introduced in the model relate to the number of patients, the percentage of elder patients and a dummy identifying paediatric GPs. Given the complexity of the model and the corresponding computational burden, the estimation is based on a subset of the observed dataset. In particular, the selected sub-sample size is 482,785 and the sampled units are selected by means of a stratified sampling technique.

The following descriptive statistics are based on the sub-sample, while the behaviour of the outcome variables (Section 4.4) is studied by considering the full population.

4.1 The individual data

By introducing individual factors in the model all the differences due to gender, age, and health status can be taken into account. The estimation sub-sample contains 52.00% (251,035) females. The mean and median age is 46.07 and 47, respectively, with standard deviation of 22.92. Information on the age of the patients is included by means of the following classes: class ≤ 24 (102,596 individuals - 21.25%), class [25; 44] (115,182 individuals - 23.86%), class [45; 64] (153,958 individuals - 31.89%) and class ≥ 65 (111,049 individuals - 23.00%). Health status is proxied by

the following set of variables. First, three kinds of exceptions are identified by specific dummies. 0.15, 10.19 and 7.94% of the patients present maternity, disease-related, and generic exceptions, respectively. Patients with particular kinds of diseases and/or those who are physically invalidated are also cared for at home. The home assistance factor identifies three increasing levels of need. No need at all, low, and high level of dependence are observed in the 89.61, 6.90 and 3.49% of the sample, respectively.

4.2 Data on the general practitioners

The patients included in the study are under the care of 1078 general practitioners, who generally represent the first filter of health care demand in Italy. In particular, their behaviour may affect the phenomenon of interest at least with respect to drugs prescriptions and outpatient care services. GPs are not gatekeepers in the Italian health system but their judgements guide individuals' access to health services. GPs who are in charge of young patients (≤ 14 years old) amount to a total of 113 (10.48% of the sample). The average age of patients which GPs are in charge of is used as a proxy of the workload of GPs. This measure takes values in the interval from 1.98 to 81.04 but the average and median are equal to 49.03 and 49.84 (standard deviation is 6.31), excluding the paediatricians. However, we finally decided to consider a dummy identifying the GPs with an average age of patients to be higher than 50 (458 GPs - 42.49% of the sample). Another measure of the GPs' workload is given by the number of patients under their care, which has a mean and median equal to 895.7 and 939, respectively. 50% of the sample lies in the interval [659.5; 1190.0]. The effect of other non-observed characteristics of the macro units is introduced in the model by means of a specific random effects term.

4.3 The municipality data

Individuals are, finally, grouped in 217 municipalities, which represent the lower level of the Italian territorial administrative system. The variables collected at the municipality level are related to socio-economic and geographical features. In particular, population density (expressed in thousands of people spread across the surface of the municipality surface - km^2 - mean 0.187,

median 0.110 and standard deviation 0.274) is included in the study to take into account of the possible heterogeneity between rural and metropolitan areas. Moreover, geographical features are used to distinguish between alpine (58 - 26.73%), hill (44 - 20.28%), coastal hill (6 - 2.76%) and plain (109 - 50.23%) landscapes. The coastal municipalities are 8 (3.69% of the sample). A dummy identifying the province seat (4 overall) is also considered in the analysis. The health system administrative macro areas are also included in the descriptive analysis. The regional territory is divided in six health administrations: area 1 (6 municipalities - 2.765%), area 2 (25 municipalities - 11.521%), area 3 (43 municipalities - 19.856%), area 4 (62 municipalities - 28.571%), area 5 (30 municipalities - 13.825%) and area 6 (51 municipalities - 23.502%). The considered socio-economic characteristics are given by the per-capita income in Euro (mean 12,740, median 12,460 and standard deviation 1,549.52) and the percentage of citizens presenting tax returns (mean 60.53, median 60.38 and standard deviation 3.05). These last two measures can be considered as proxies of social welfare. As a matter of fact, a higher income level increases the likelihood of the use of private health care services (which are excluded from the analysis).

As mentioned previously, the spatial structure of the dataset can be treated only at an aggregated level. As clearly represented in Figure 1, the administrative districts (municipalities) are organised in irregular lattices. Pictures in Figure 1 show the spatial distribution of the multivariate phenomenon observed at the municipality level. This descriptive analysis does not exhibit clear spatial patterns in the HCEs, however: the spatial correlation needs to be considered in the model specification whenever the effect of control variables is taken into consideration. A specific analysis on aggregated data is performed to examine the spatial autocorrelation. The relative results (omitted here for reasons of space) show that the phenomenon presents relevant spatial correlation structures suggesting to include a spatial component in the disaggregated models too.

[Fig. 1 about here.]

4.4 The dependent variable behaviour in the population

The dataset is characterised by a large proportion of null HCEs. Summary statistics in Table 1 show the relationship between this proportion and a set of possible explicative variables. This

evidence supports the choice of the prediction model for the excess of zeros. The analyses developed to investigate this issue at the level of municipalities and GPs reveal relatively homogeneous behaviours. In line with the existing literature (as stated, for instance, in Werblow et al., 2007) a lower percentage of positive expenditures is observed in males, while as expected, a larger percentage of positive expenditures is observed in higher age classes. In particular, an increasing proportion is observed for each source of expenditure (as in Albouy et al., 2010). As in Deb et al. (2006) a positive association is observed between proxies of comorbidities and the presence of health care expenditures. Patient health status is proxied here with the presence of different kinds of fee exemptions and the need for home assistance (commonly associated with the presence of severe chronic diseases). As expected all HCEs are positively associated with bad health status proxies. The comparison between patients assigned to paediatric and non-paediatric GPs states that the former present less non-null expenditures for drug prescriptions and outpatient care. On the contrary, the proportion of positive expenditures for inpatient care is lower for the non-paediatric sub-population. Some small differences can be observed between different administrative areas. The effects of geographical features are deepened considering the classification of municipalities as alpine, hill, coastal hill and plain landscapes.

The descriptive analysis of positive expenditures is reported in Table 2. These summary statistics highlight the effects of individual, GP, and municipality characteristics on HCE heterogeneity. The mean HCEs for drug prescriptions and inpatient care are higher for males, while gender differences in outpatient care expenditure are negligible. With respect to age classes, the observed averages show growing patterns for each kind of HCE. All sources of HCEs are positively affected by the presence of specific fee exemptions and the need for home assistance (as in Buntin and Zaslavsky, 2004). As expected, being under the care of a paediatrician has a negative effect on the average expenditures. The differences between administrative areas are small and, finally, a limited heterogeneity can be observed between different kinds of municipalities.

[Table 1 about here.]

[Table 2 about here.]

As regards the relationship between different sources of HCE, the main results in the literature show a substitutive pattern in health expenditures due to the effect of changes in the

economic framework (as, for instance, in Goldman et al., 2006). Moreover, Lichtenberg (2007) shows that patients who use more and newer drugs present a smaller demand in inpatient care. The association between the presence of the three types of health care expenditures and the correlation between their levels are reported in Table 3. The results of this analysis show that the substitution effect cannot be observed at the disaggregated level. In fact, the positive associations identified by Kendall's tau-b indexes suggest that, at the individual level, the three health care treatments are complements. The same conclusion can be achieved by considering the analysis of the correlation indexes. A more detailed discussion of these effects will be given in the following sections.

[Table 3 about here.]

5 The results of the model estimation

As stated in Section 3, the estimation procedure is carried out separately for each source of HCE. The estimation results of the probit selection models (as defined in equation 2) are given in Table 4 and the results of selection bias-corrected linear models (as defined in equation 3) are summarised in Table 5. The model estimation for the outcome equation (considering only the positive expenditures) is also performed without considering the correction for selection bias. The comparison between the obtained estimates confirms the need for Heckman's correction, except for the analysis of inpatient care expenditures. As stated in the previous section, given the model complexity, the following results are based on a sub-sample of the available dataset. Moreover, in order to be able to compare and interpret the estimation results, all the models consider the general error correlation structure as defined in Section 3.

As a matter of fact, the mixed probit models (see Table 4) reveal a negative effect connected with the gender dummy (the effect is insignificant in the model for inpatient care) and increasing and positive effects for age class. In particular, patients older than 65 years exhibit a significantly greater probability of demand for health care services, especially for drug prescriptions. Moreover, patients cared for at home are more likely to require all types of health services and the probability of positive HCEs increases accordingly to the level of home assistance. Being under the care of a paediatrician leads to a greater probability of drug prescriptions and hospitalisations and a

lower probability of positive outpatient expenditures. The proxy of the workload of GPs (the logarithm of the number of patients) presents a positive significant effect in the selection model for outpatient care, while its effect is negative in the model for hospitalisations. The results regarding the geographical features show that the probability of drug prescription in the sub-population living in the alpine municipalities is significantly higher than the other citizens (alpine is the benchmark level). For all the other sources of expenditures, the coefficients of this factor are insignificant. The effect of the logarithm of per capita income is significantly negative in the model for drug prescriptions, while its effect on outpatient care is not significant. The log-income can be considered as a proxy of well-being as it reduces the probability of drug prescriptions. The presence of disease-related exemptions generally corresponds to a strong positive effect on all the considered sources of health expenditures. On the contrary, patients with a generic fee exemption show a lower probability of demand for inpatient and outpatient cares. The coefficients for the last three dummy variables confirm the positive association between the demands for different health care services. This result reveals that different health care services cannot be considered as direct substitutes. In other words, patients facing greater needs for health care are generally more likely to require all types of care services. As the values of σ_u suggest, the GP random effects are non-negligible in the probit models for all the considered kinds of expenditures. The models excluding the GP components present DIC and WAIC values higher than the ones reported in Table 4. In particular, the WAIC values are 506228.68, 290310.24, and 528850.67, for drug prescriptions, inpatient, and outpatient care, respectively (DIC values present coherent results). Looking at the municipality error convolution term, the spatial components are relevant for all types of HCE, while the unstructured municipality components seem to be negligible. Once again the information criteria can be used to evaluate their significance. For the models excluding the unstructured municipality components, the values of the Watanabe criteria are 501102.43, 284553.50, and 525571.89 for the three kinds of expenditures, respectively. Only the model for inpatient care presents a counter-intuitive result. The information criterion for the restricted model is close to the one of the full model, leading to inconclusive results. In all the estimated models the spatial component can be considered as significant.

[Table 4 about here.]

The estimation results of the linear mixed models (see Table 5) for drugs and outpatient care expenditures assess the selection bias given the significant negative effect of the inverse Mills ratios. This means that patients with positive HCEs are also more likely to present greater HCEs (the classical models would produce downward-biased estimates). The selection process is irrelevant only in the inpatient care analysis. Being a male has a positive effect on the mean expenditure in drugs and inpatient cares, while a negative significant effect is observed in the outpatient care model. The effect of age (the reference class is 'less than 24 years of age') is confirmed as positive and increasing in all the HCEs (with a particularly high effect on drug consumption). Generally speaking, patients under the care of paediatricians have a lower average level of expenditures. The effect of disease-related exemption is significant and positive for drugs and outpatient care expenditures, while the presence of a generic fee-exemption has a negative effect on them. The higher the level of home assistance, the higher is the average level of generic health care expenditures. A particular behaviour can be observed in the outpatient care model. In this case patients with a higher need for home assistance present an average expenditure lower than patients with a low level of dependence. This uneven result can be related to the fact that people with higher dependence are often unable to access the outpatient care service. At the municipality level, the logarithm of per-capita income shows a positive effect on the spending capability of patients in the outcome model for outpatient care. Finally, all the sources of health expenditures are positively affected by the presence of the other kinds of HCEs. This result means that, at the individual level, the different kinds of HCEs are complementary. In fact, the considered services are commonly contemporarily activated when the patients need for health care, and consequently, the different kinds of expenditures are jointly observed. With regard to the random components of the model, relevant heterogeneity at the GP level is generally observed in the expenditures. Considering the comparison of information criteria, only the model for inpatient cares shows a small difference between the full model specification and the one excluding the GPs effect. The WAIC measures are, in fact, 1102491.52, 123987.92, and 949930.95 for the drug prescriptions, the inpatient and the outpatient cares, respectively. The standard deviations for the unstructured effect of municipalities are always very small. By adopting the WAIC measure, the exclusion of these effects is suggested for the inpatient care (WAIC=123973.84) and for the outpatient care (WAIC=948182.57) models (the information cri-

teria are, in fact, close to the full model ones). On the contrary, the component is relevant in the drug prescription (WAIC=1097420.39). The WAIC measures for the models excluding the spatial error component are 1097435.62, 123989.52, and 948170.95, respectively. The spatial diffusion processes are relevant for drug prescriptions and for inpatient care while the best model specification for outpatient care is the one which considers only the unstructured municipality effect.

[Table 5 about here.]

The estimated standard deviations for the convolution effect at the municipality level ($d + s$) can be studied to highlight the presence of different kinds of spatial behaviours for the components of the HCE. In particular, drug prescriptions are characterised by a low unstructured municipality effect and a relatively larger spatial component, meaning that the observed municipalities present similar behaviours but the neighbouring effect is relevant for both the selection and the outcome equations. For the inpatient care, the variance of the spatial component is again larger than the unstructured one in the selection equation, while the two effects present closer magnitudes in the outcome equation. Finally, the outpatient care analysis shows that the variance of the spatial component overcomes the unstructured one in both the equations. The significance of the spatial components in drug prescriptions and in outpatient care can be associated with the presence of health units coordinating the supply of medical care in the territory. This kind of information can be used to build an evidence-based policy making process and to evaluate its feasibility and effectiveness.

6 Discussion

This paper focuses on the assessment of HCE determinants within a micro data framework. The use of disaggregated data allows for the combination of the advantages of both micro and macro-level studies. The outcome of sample selection model estimation confirms some well-known results on health expenditure drivers, such as gender, age, and health care needs. The relevant role of individual health status is supported by the positive correlation between different sources of expenditures (e.g. Werblow et al., 2007). This result supports also the idea that different types of health care services are not substitutes when individual demand is taken into consideration and

increasing levels of expenditures for all types of services reveal higher levels of need. Generally speaking, the individual health status presents significant effects in both the selection and level models. Variables included as proxies of co-morbidity (such as fee exemptions, home assistance level and dummies identifying the presence of other kinds of expenditures) show positive effects in all models, except for hospitalisation services. We can assess a partial trade-off between HCE types even if the demand for care changes consistently within the considered expenditure classes. As the results show positive relationships between different HCEs, we can conclude that the substitution effect observed in many empirical studies is not confirmed at the individual level.

The hierarchical structure of our database allows for the assessment of positive spatial patterns in individual expenditures as highlighted in the prior literature (e.g. Moscone et al., 2007, for the macroeconomic data analysis). Selection models on the different sources of HCE present relevant spatial component effects, due to the relevant role of the proximity to the health care providers. Models on the amount of expenditures, however, point out relevant spatial effects when drug prescriptions and inpatient care expenditures are taken into consideration. The Gaussian unstructured municipality specific effect is irrelevant in most of the considered models. In particular, the neighbourhood spatial effect is stronger than the municipality specific one, revealing the presence of territorial common behaviour in most of the considered models. The estimation results show that the GPs affect the selection process for all of the expenditures. As regards the level models, the error components of GPs are significant only for drugs and outpatient expenditures. The role of GPs on drug prescription and outpatient care expenditures has been assessed in the models for both selection and level equations.

The probability of access to the health care system and the level of expenditures strongly depend on the characteristics of the population, both at individual and area levels. As a matter of fact, the estimates of the models exhibit a significant relationship between HCEs and age-class, gender, as stated also by Costa-Font and Pons-Novell (2007), and per-capita GDP at the municipality level. At the GP level, a higher probability of drug prescriptions and hospitalisations and a lower probability of outpatient visits are observed for paediatric patients. However, the corresponding effect on the amount of expenditures is negative, as expected, given the characteristics of the population under their care. Finally, the introduction of the inverse Mills ratio reveals the relevance of the selection process in the models for drugs and outpatient care.

An even more important feature of the estimated models concerns the analysis of the spatial distribution of the municipality random effects. In Figure 2 the posterior means of random effects for the selection (left panels) and outcome (right panels) equations show that the behaviour in the three kinds of expenditures is heterogeneous. Relevant spatial effects are observed for drug prescriptions and outpatient care in the selection process, while the effects in the outcome equations are all negligible.

[Fig. 2 about here.]

Generally speaking, the obtained results support the main indications in the literature but they are obtained at the disaggregated level. This specific characteristic allows us to study the HCE phenomenon in a limited context, such as that of the population in the FVG Italian region. By adopting the introduced approach, researchers and, more interestingly, policy makers can focus on the regional level when investigating the behaviour of the budget components. Moreover, the effect of the considered HCE drivers on the three main sources of health care expenditures can be evaluated separately. The richness of the available information allows us to disentangle spatial effects by controlling the phenomenon for socio-economic and health-related sources of heteroskedasticity.

7 Conclusions

To account for all the substantial aspects of our dataset, a cross-classified hierarchical spatial sample selection model is proposed and studied in-depth. The model is applied to the HCE data observed on the population residing in the Italian region of FVG. The application of hierarchical spatial models to the individual-level analysis of HCE allows us to examine this phenomenon by using the original disaggregated data.

From the decision makers' point of view, the assessment of HCE determinants can be relevant at both the micro and macro-level. More specifically, this kind of evidence leads to a better understanding of the health care economic phenomenon, while also showing that most of the significant effects in the disaggregated estimated models are connected with individual specific variables. However, the contextual variables (at both general practitioner and municipality levels) are influential. As a matter of fact, a relevant heterogeneity between GPs has been identified.

Moreover, a significant spatial effect can be observed throughout the regional territory. After accounting for all these effects, the estimation results can be used to define an individual risk profile and to evaluate the spending behaviour of patients and of local units at the aggregate level.

The estimated models point out some interesting and well-established results, but also other drivers of HCE need to be included in the analysis (i.e. more accurate measures of individual health status, social deprivation at the municipality level and GP performance measures). Therefore, this research can be considered as a first attempt towards the analysis of individual expenditures at the population level.

As discussed in Section 1, the specification and the estimation procedures of the models are developed through a feasible and naïve approach. The model could be developed in a more general sample selection framework (at least with a fixed-effects specification) and a proper multivariate specification could be taken into consideration. Preliminary analyses show that correlations between different expenditures have negligible effects on model estimation, due to the large amount of available observations.

References

- Albouy, V., L. Davezies, and T. Debrand (2010). Health expenditure models: A comparison using panel data. *Econ Model* 27(4), 791–803.
- Angulo, A. M., R. Barberán, P. Egea, and J. Mur (2011). An analysis of health expenditure on a microdata population basis. *Econ Model* 28(1), 169–180.
- Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agr Econ* 27(3), 247–267.
- Baltagi, B. H. and F. Moscone (2010). Health care expenditure and income in the OECD reconsidered: Evidence from panel data. *Econ Model* 27(4), 804 – 811.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J R Stat Soc B Met* 36(2), 192–236.
- Besag, J., J. York, and A. Mollié (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann I Stat Math* 43(1), 1–20.

- Buntin, M. B. and A. M. Zaslavsky (2004). Too much ado about two-part models and transformation?: Comparing methods of modeling medicare expenditures. *J Health Econ* 23(3), 525–542.
- Costa-Font, J. and J. Pons-Novell (2007). Public health expenditure and spatial interactions in a decentralized national health system. *Health Econ* 16(3), 291–306.
- Deb, P., M. K. Munkin, and P. K. Trivedi (2006). Bayesian analysis of the two-part model with endogeneity: application to health care expenditure. *J Appl Econ* 21(7), 1081–1099.
- Deb, P. and P. K. Trivedi (1997). Demand for medical care by the elderly: a finite mixture approach. *J Appl Econom* 12(3), 313–336.
- Deb, P. and P. K. Trivedi (2002). The structure of demand for health care: latent class versus two-part models. *J Health Econ* 21(4), 601–625.
- Di Matteo, L. and R. Di Matteo (1998). Evidence on the determinants of canadian provincial government health expenditures: 1965–1991. *J Health Econ* 17(2), 211–228.
- Eibich, P. and N. R. Ziebarth (2014). Examining the structure of spatial health effects in germany using hierarchical bayes models. *Reg Sci Urban Econ* 49, 305–320.
- Foster, C. D., R. Jackman, and M. Perlman (1980). *Local government finance in a unitary state*. London: George Allen & Unwin.
- Freeman, D. G. (2003). Is health care a necessity or a luxury? Pooled estimates of income elasticity from us state-level data. *Appl Econ* 35(5), 495–502.
- Gerdtham, U.-G. and B. Jönsson (2000). International comparisons of health expenditure: theory, data and econometric analysis. *Handb Health Econ* 1, 11–53.
- Gerdtham, U.-G., J. Sogaard, F. Andersson, and B. Jönsson (1992). An econometric analysis of health care expenditure: a cross-section study of the OECD countries. *J Health Econ* 11(1), 63–84.
- Getzen, T. E. (2000). Health care is an individual necessity and a national luxury: applying multilevel decision models to the analysis of health care expenditures. *J Health Econ* 19(2), 259 – 270.
- Giannoni, M. and T. Hitiris (2002). The regional impact of health care expenditure: the case of italy. *Appl Econ* 34(14), 1829–1836.

- Goldman, D. P., G. F. Joyce, and P. Karaca-Mandic (2006). Varying pharmacy benefits with clinical status: the case of cholesterol-lowering therapy. *Am J Manag Care* 12(1), 21–28.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* 47(1), 153–161.
- Helms, J., J. P. Newhouse, and C. E. Phelps (1978). Copayments and demand for medical care: the california medicaid experience. *Bell J Econ* 9(1), 192–208.
- Leung, S. F. and S. Yu (1996). On the choice between sample selection and two-part models. *J Econ* 72(1), 197–229.
- Lichtenberg, F. R. (2007). The impact of new drugs on us longevity and medical expenditure, 1990–2003: Evidence from longitudinal, disease-level data. *Am Econ Rev* 97(2), 438–443.
- Lopez-Casasnovas, G., J. Costa-Font, and I. Planas (2005). Diversity and regional inequalities in the Spanish “system of health care services”. *Health Econ* 14(S1), S221–S235.
- Madden, D. (2008). Sample selection versus two-part models revisited: The case of female smoking and drinking. *J Health Econ* 27(2), 300–307.
- Martins, T. G., D. Simpson, F. Lindgren, and H. Rue (2013). Bayesian computing with inla: new features. *Comput Stat Data An* 67, 68–83.
- Moscone, F. (2011). Geographical variations in expenditure of learning disability services in england. *Appl Econ* 43(23), 2997–3005.
- Moscone, F. and M. Knapp (2005). Exploring the spatial pattern of mental health expenditure. *J Ment Health Policy Econ* 8(4), 205.
- Moscone, F., M. Knapp, and E. Tosetti (2007). Mental health expenditure in england: A spatial panel approach. *J Health Econ* 26(4), 842 – 864.
- Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *Rev Econ Stat* 72(2), 334–338.
- Newhouse, J. P. (1992). Medical care costs: how much welfare loss? *J Econ Perspect* 6(3), 3–21.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rosenberger, R. S., Y. Sneh, T. T. Phipps, R. Gurvitch, et al. (2005). A spatial analysis of linkages between health care expenditures, physical inactivity, obesity and recreation supply. *J Leisure Res* 37(2), 216–235.

- Rue, H., S. Martino, and N. Chopin (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *J R Stat Soc B* 71(2), 319–392.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002). Bayesian measures of model complexity and fit. *J R Stat Soc B* 64(4), 583–639.
- Wang, Z. (2009). The determinants of health expenditures: evidence from us state-level data. *Appl Econ* 41(4), 429–435.
- Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *J Mach Learn Res* 11(Dec), 3571–3594.
- Werblow, A., S. Felder, and P. Zweifel (2007). Population ageing and health care expenditure: a school of ‘red herrings’? *Health Econ* 16(10), 1109–1126.
- Wildasin, D. (2013). *Urban public finance*. London: Routledge.
- Winkelmann, R. (2012). Copula bivariate probit models: with an application to medical expenditures. *Health Econ* 21(12), 1444–1455.
- Wong, A., P. H. van Baal, H. C. Boshuizen, and J. J. Polder (2011). Exploring the influence of proximity to death on disease-specific hospital expenditures: a carpaccio of red herrings. *Health Econ* 20(4), 379–400.

List of Figures

- 1 Spatial distribution of the health care expenditure components. The expenditures are represented in terms of logarithms along with the map of Friuli Venezia Giulia. The obtained values are represented on a graded scale. Lighter shades stand for higher values and darker shades for lower values. 27
- 2 Estimated spatial diffusion processes for the selection and outcome equations for the three kinds of HCEs. The municipalities' specific effects are represented on a graded scale. Lighter shades stand for higher values and darker shades for lower values. 28

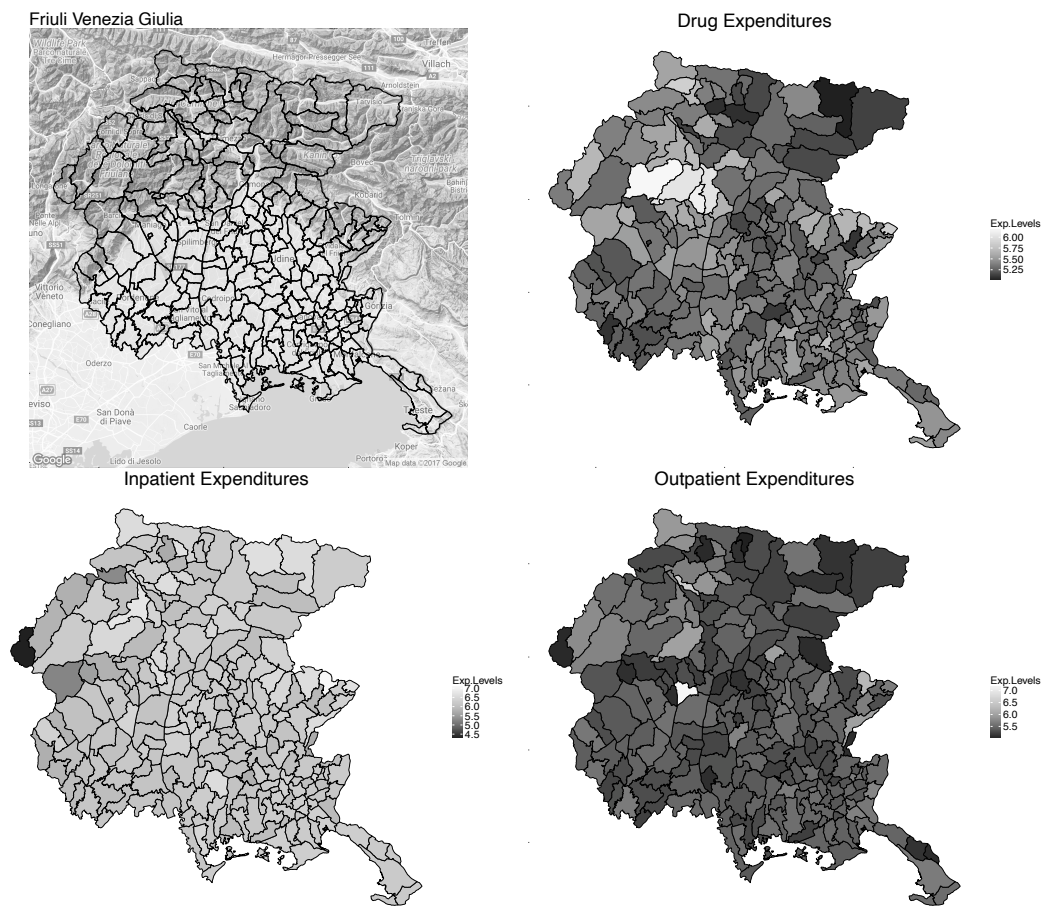


Fig. 1 Spatial distribution of the health care expenditure components. The expenditures are represented in terms of logarithms along with the map of Friuli Venezia Giulia. The obtained values are represented on a graded scale. Lighter shades stand for higher values and darker shades for lower values.

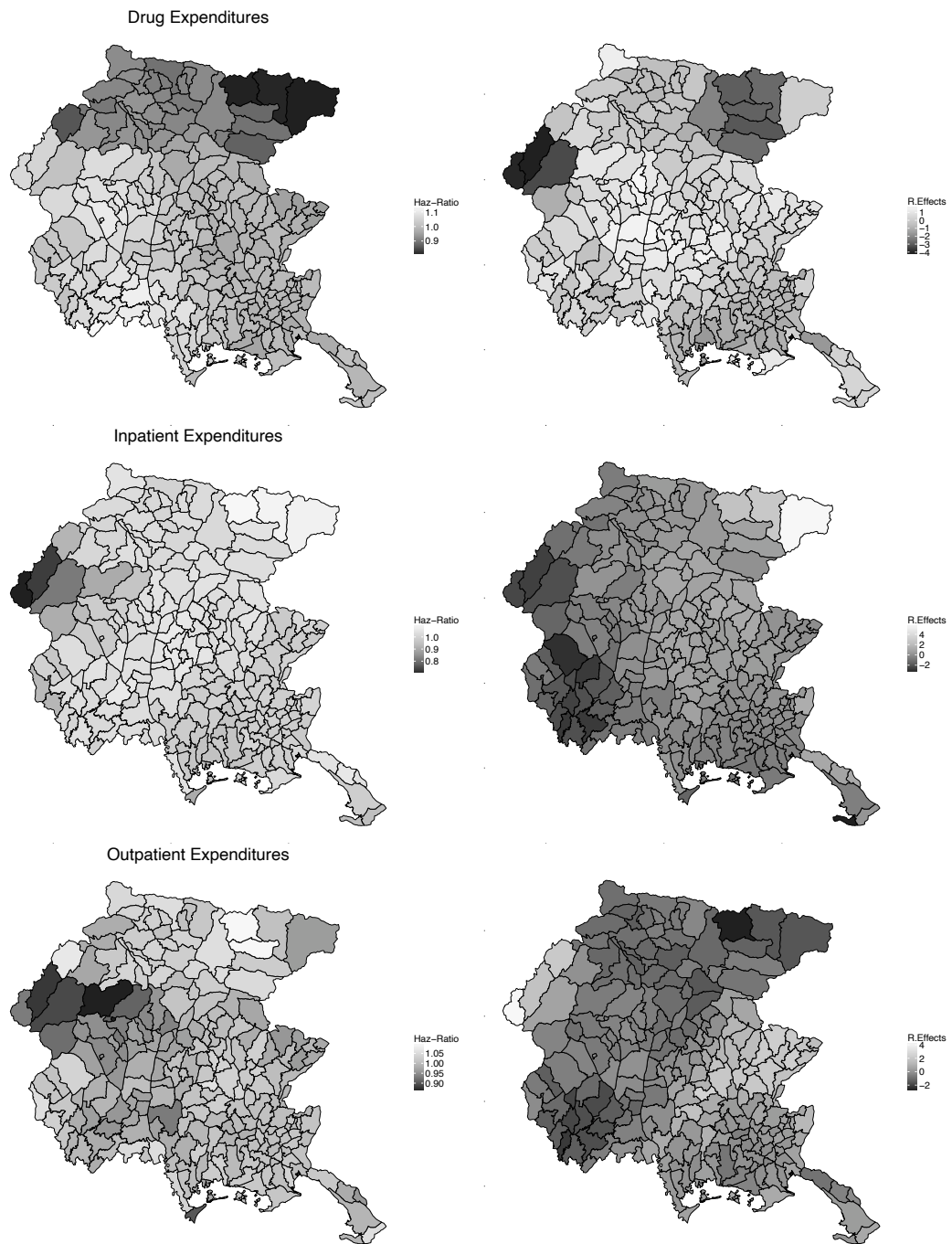


Fig. 2 Estimated spatial diffusion processes for the selection and outcome equations for the three kinds of HCEs. The municipalities' specific effects are represented on a graded scale. Lighter shades stand for higher values and darker shades for lower values.

List of Tables

1	Conditional percentages of positive health care expenditures in year 2010. Data refers to 965,579 individuals clustered in 217 municipalities and in 1,078 GPs.	30
2	Conditional means of positive health care expenditures in year 2010. Data refers to 965,579 individuals clustered in 217 municipalities and in 1,078 GPs.	31
3	The relationship between different kinds of health care expenditures. Kendall's tau-b and Spearman correlation indexes are reported above and below the matrix diagonal, respectively. The p-values of correlation tests and asymptotic standard error of association measures are reported in brackets.	32
4	First step probit mixed models Bayesian estimation results (standard deviations in brackets).	33
5	Second step linear mixed models estimation results (standard errors in brackets).	34

Table 1 Conditional percentages of positive health care expenditures in year 2010. Data refers to 965,579 individuals clustered in 217 municipalities and in 1,078 GPs.

Variables	Percentage of individual expenditures			Number of observations
	Drug prescriptions	Inpatient care	Outpatient care	
<i>General</i>	66.029	9.884	64.939	482,785
<i>Gender</i>				
Female	70.979	10.707	69.973	251,035
Male	60.667	8.993	59.486	231,750
<i>Age Class</i>				
Under 25	46.521	7.741	47.581	102,596
25-44	53.484	8.543	56.715	115,182
45-64	70.498	8.376	68.051	153,958
Over 65	90.867	15.346	85.190	111,049
<i>Exemptions</i>				
Maternity	77.479	18.414	79.887	706
Generic Exemption	91.497	16.922	86.634	49,208
Chronic Disease Exemption	92.818	18.504	89.389	38,348
<i>Home assistance level</i>				
0	63.175	8.583	62.326	432,609
1	88.336	18.998	85.531	33,341
2	95.183	25.287	91.298	16,835
<i>GP type</i>				
General Practitioner	67.081	9.661	67.001	446,010
Paediatrician	53.275	12.596	39.932	36,775
<i>GPs' patient average age</i>				
Average age > 50	68.435	9.887	68.152	211,434
<i>Administrative area</i>				
1	66.992	9.972	65.808	93,393
2	66.986	10.097	66.454	54,305
3	64.744	10.252	65.213	24,952
4	64.907	9.962	64.684	134,196
5	66.752	9.588	66.037	40,447
6	66.727	9.552	63.962	116,795
<i>Geographical Features</i>				
Alpine	64.501	10.097	64.782	24,561
Hill	65.956	9.765	64.342	79,317
Coastal Hill	66.677	10.010	65.479	96,281
Plain	65.962	9.856	64.936	282,626
Non Coastal	65.760	9.827	64.700	370,360
Coastal	66.915	10.072	65.726	112,425

Table 2 Conditional means of positive health care expenditures in year 2010. Data refers to 965,579 individuals clustered in 217 municipalities and in 1,078 GPs.

Variables	Mean values of individual expenditures		
	Drug prescriptions	Inpatient care	Outpatient care
<i>Number of observations</i>	318,778	47,720	313,515
<i>General</i>	334.158	4864.568	391.071
<i>Gender</i>			
Female	321.661	4508.773	402.275
Male	349.995	5323.443	376.795
<i>Age Class</i>			
Under 25	69.556	2120.352	199.545
25-44	141.206	3392.810	311.992
45-64	313.711	5253.770	375.476
Over 65	599.105	6698.714	561.777
<i>Exemptions</i>			
Maternity	461.625	3196.354	426.026
Generic Exemption	745.983	6790.499	645.691
Chronic Disease Exemption	842.692	6945.422	705.916
<i>Home assistance level</i>			
0	275.197	4045.862	341.583
1	614.508	7387.859	670.555
2	824.499	8250.801	740.657
<i>GP type</i>			
General Practitioner	352.895	5202.907	399.221
Paediatrician	48.022	1717.251	225.225
<i>GPs' patients average age</i>			
Average age > 50	379.680	5411.597	411.071
<i>Administrative area</i>			
1	364.372	5049.899	436.063
2	355.578	5239.557	400.979
3	313.611	4937.244	346.338
4	329.872	4824.647	368.934
5	329.928	5077.274	360.811
6	314.927	4560.255	391.440
<i>Geographical Features</i>			
Alpine	332.657	5014.569	355.043
Hill	324.350	5020.706	397.255
Coastal Hill	362.009	4991.153	448.015
Plain	327.446	4764.007	372.914
Non Coastal	325.069	4819.741	377.709
Coastal	363.581	5008.662	434.401

Table 3 The relationship between different kinds of health care expenditures. Kendall's tau-b and Spearman correlation indexes are reported above and below the matrix diagonal, respectively. The p-values of correlation tests and asymptotic standard error of association measures are reported in brackets.

	Drugs Exp.	Inpat. Exp.	Outpat. Exp.
<i>Drugs Exp.</i>	–	0.227(< 0.001)	0.198(< 0.001)
<i>Inpat. Exp.</i>	0.334(< 0.001)	–	0.158(< 0.001)
<i>Outpat. Exp.</i>	0.294(< 0.001)	0.233(< 0.001)	–

Table 4 First step probit mixed models Bayesian estimation results (standard deviations in brackets).

	Presence of positive Health Care Expenditures					
	Drug Prescriptions		Inpatient Care		Outpatient Care	
<i>Intercept</i>	-0.137	(0.009)	-1.965	(0.071)	-1.019	(0.459)
<i>Female</i>	–		–		–	
<i>Male</i>	-0.161	(0.004)	–		-0.175	(0.004)
<i>Age ≤ 24</i>	–		–		–	
<i>Age 25 – 44</i>	0.127	(0.006)	0.226	(0.011)	0.017	(0.007)
<i>Age 45 – 64</i>	0.497	(0.006)	0.080	(0.010)	0.192	(0.006)
<i>Age ≥ 65</i>	0.979	(0.007)	0.199	(0.011)	0.493	(0.008)
<i>Gen. Exemption</i>	–		-0.123	(0.017)	-0.242	(0.015)
<i>Dis. Exemption</i>	0.392	(0.010)	0.234	(0.018)	0.551	(0.017)
<i>Home Assist. Lev.1</i>	–		–		–	
<i>Home Assist. Lev.2</i>	0.301	(0.010)	0.333	(0.009)	0.275	(0.010)
<i>Home Assist. Lev.3</i>	0.427	(0.016)	0.492	(0.012)	0.336	(0.015)
<i>Paediatrician</i>	0.204	(0.014)	0.628	(0.016)	-0.436	(0.014)
<i>N. Patients</i>	–		-0.050	(0.010)	0.068	(0.009)
<i>Alpine</i>	–		–		–	
<i>Hill</i>	-0.100	(0.015)	–		–	
<i>Coastal Hill</i>	-0.097	(0.017)	–		–	
<i>Plain</i>	-0.096	(0.011)	–		–	
<i>Income</i>	-0.014	(0.001)	–		0.034	(0.048)
<i>Dummy Drugs Exp.</i>	–		0.322	(0.007)	0.774	(0.004)
<i>Dummy Inpat. Exp.</i>	0.339	(0.008)	–		0.807	(0.009)
<i>Dummy Outp. Exp.</i>	0.679	(0.004)	0.690	(0.007)	–	
Variiances of error components						
(with their 95% confidence intervals)						
σ_u	0.128	[0.128, 0.128]	0.098	[0.108, 0.091]	0.109	[0.110, 0.107]
σ_d	0.047	[0.047, 0.047]	0.017	[0.032, 0.010]	0.029	[0.030, 0.029]
σ_s	0.107	[0.107, 0.107]	0.071	[0.114, 0.047]	0.103	[0.105, 0.101]
DIC	502967.37		284551.75		525584.99	
WAIC	502931.86		284553.78		525587.92	

Table 5 Second step linear mixed models estimation results (standard errors in brackets).

	Health Care Expenditures					
	Drug Prescriptions		Inpatient Care		Outpatient Care	
(Intercept)	3.295	(0.048)	7.177	(0.150)	3.123	(0.363)
<i>Female</i>	–		–		–	
<i>Male</i>	0.189	(0.006)	0.062	(0.008)	–0.068	(0.005)
<i>Age ≤ 24</i>	–		–		–	
<i>Age 25 – 44</i>	0.358	(0.011)	0.345	(0.020)	0.279	(0.008)
<i>Age 45 – 64</i>	1.130	(0.016)	0.551	(0.018)	0.397	(0.008)
<i>Age ≥ 65</i>	1.818	(0.023)	0.738	(0.020)	0.592	(0.011)
<i>Gen. Exemption</i>	–0.117	(0.015)	–		–0.224	(0.013)
<i>Dis. Exemption</i>	0.764	(0.016)	–		0.425	(0.015)
<i>Home Assist. Lev.1</i>	–		–		–	
<i>Home Assist. Lev.2</i>	0.325	(0.010)	0.266	(0.020)	0.210	(0.008)
<i>Home Assist. Lev.3</i>	0.483	(0.013)	0.390	(0.028)	0.111	(0.010)
<i>Paediatrician</i>	–0.351	(0.020)	–0.371	(0.036)	–0.407	(0.018)
<i>Income</i>	–		–		0.144	(0.038)
<i>Dummy Drugs Exp.</i>	–		0.191	(0.02)	0.313	(0.017)
<i>Dummy Inpat. Exp.</i>	0.253	(0.009)	–		0.605	(0.012)
<i>Dummy Outpat. Exp.</i>	0.356	(0.018)	0.188	(0.036)	–	
<i>Inverse Mills Ratio</i>	–0.498	(0.046)	–0.028	(0.058)	–0.272	(0.036)
Variations of error components (with their 95% confidence intervals)						
σ_u	0.141	[0.152, 0.133]	0.037	[0.061, 0.023]	0.105	[0.112, 0.100]
σ_d	0.020	[0.039, 0.012]	0.033	[0.057, 0.022]	0.017	[0.036, 0.011]
σ_s	0.092	[0.146, 0.063]	0.049	[0.098, 0.029]	0.049	[0.096, 0.029]
σ_ϵ	1.351	[1.355, 1.348]	0.886	[0.891, 0.879]	1.100	[1.102, 1.098]
DIC	1097416.10		123973.62		948192.17	
WAIC	1097406.89		123972.77		948182.40	