TESIS DOCTORAL

AÑO 2022

Essays on Spatial Econometrics and Economics

Maryna Makeienko

PROGRAMA DE DOCTORADO EN ECONOMÍA INTERUNIVERSITARIO (DEcIDE)

Director: **Mariano Matilla García**

Co-Directora: **Paloma Úbeda Mollá**

Acknowledgements

First, I am very grateful to my doctoral thesis supervisor Mariano Matilla-García for the scientific and pedagogical talent which he dedicated to me over the last years. I want to thank him for the passion of the researcher that he could raise in me and for showing me how to notice the key points in everything I was doing without diminishing the importance of routine and details in scientific work. Finally, I appreciate the fact that in the course of joint work, he formed the basis of a scientist in me and taught me how to look for evidence with interest and find exciting new approaches. I am indebted to him for the moments when he guided and supported me and did not let me give up.

I also want to thank my co-supervisor Paloma Úbeda Molla for her patience, her ability to support me in the course of our work, for instilling a systematic approach and attention to details in me and for teaching me how to see interrelations and formulate thoughts in a proper manner.

This thesis is a result of the Doctorado en Economia, Interuniversity doctoral Programme (DECIDE) and UNED, these two institutions, which demonstrate the highest scientific and professional standards, have made me feel like part of a scientific community and a great team. The Departamento de Economía Aplicada y Estadística of the Facultad de Ciencias Económicas y Empresariales de la UNED made this work possible by providing us with financial help needed to get the Database upon which my thesis is based.

I am grateful to Victoria for her friendship, support and faith in me and for helping me with my first steps in the scholastic environment.

At a personal level, my deepest gratitude is to my family and my parents for instilling in me a love of books and a passion for learning, teaching me how to set and achieve goals, and their love and support in all my aspirations. It would have been impossible to get to this point without their unwavering support.

I also express my profound thanks to all my teachers and friends who have shaped me as a person and have always supported me throughout my life.

Contents

1. Introduction

Prices have always been (and continue to be) central in empirical and theoretical economics. Understanding price dynamics and price formation have been challenging in the economic literature since its origin. In this thesis, in general terms, we focus on the formation of two prices: house prices and salaries.

Needless to say that many papers, books and theses have been written on these central prices. Therefore the reader has the right to ask what the (novel) point of view is to justify reading this manuscript, which has entertained the authors for more than three years; and accordingly we have the debt of answering them.

We have been interested in the “permanent components or variables” that might explain house prices and unemployment. “Permanent component” is understood to be a characteristic variable that is almost impossible to modify. In the case of the house market, the permanent variable that we are especially concerned with is the (geographical) location of the house. This is a permanent and important variable that can clearly explain (or alter) the price under ceteris paribus conditions. Notice that one can only change the location if the house is deconstructed and built somewhere else.

In the case of the labor market, we are concerned with personality traits. Indeed, personal traits for someone at the “entrance” of the labor market are really permanent characteristics of the labor supply and might be critical to form their entry (or exit) into (or from) the labor market and also for salary formation. The personality of an individual, captured via the Big Five, and their labor market characteristics are the subject of study in this thesis.

In order to study the role of these so-called permanent variables in both markets, adapted methodological tools and proper data sets are required. From the methodological point of view, we elaborate on the paper by Garcia-Cordoba et al. (2019), which develops a spatial econometric test for linear and nonlinear spatial structures. In this seminal paper, the authors delineate a statistical procedure based on the entropy concept to determine whether a cross-sectional data set contains a leading deterministic component in the form of either a trend or a chaotic non-linear process, building on the previous studies of these authors, but then within a time-series context. It is easy to imagine what is meant by a trend in the time domain of a spatial panel, but difficult to imagine what is meant by its counterpart in the cross-sectional domain of such a panel. The authors apply their test to the well-known cigarette demand data set of Baltagi (2005). Among other variables, this data set is formed by real per capita sales of cigarettes in 46 US states over the period 1963–92. When applied to the 46 time-averages observed for each state, the authors found no evidence of a spatial deterministic component. When applied to each cross-section of observations in each year, they again found no evidence of a spatial deterministic component for 17 of the 30 years. Apparently, no deterministic spatial trend is needed in addition to the commonly imposed spatial lags of the dependent variable, explanatory variables or the error term (Elhorst (2013)). This evidence is reassuring since previous literature using this data set has not uncovered any evidence of strong spatial dependence, which is conceptually akin to stochastic non-stationarity in the spatial dimension. See Ciccarelli and Elhorst (2018) for contrasting evidence in another data set. Tests for weak spatial dependence are available (Bhattacharjee and Holly (2013),Pesaran (2015)), and see Ditzen (2018) for a practical application and Stata programs. This paper by Garcia-Cordoba et al. (2019) could generate research interest in testing for weak spatial dependence in the presence of a leading deterministic component, similar to time-series tests for unit roots in the presence of drift and/or trend.

Chapter 2 and Chapter 3 build on this seminal paper. Chapter 2 includes the introduction to the theoretical part of the spatial models, incorporating both traditional spatial models and our newly-developed delta-models that include the geographical position of the unit in them. We also present the scheme that introduces the delta-test (which we will see later) as an important step in analyzing spatial data. Later on we introduce three different housing price datasets. These datasets become the base of the complex spatial analysis, which includes both spatial model analysis and the analysis of the possibly-present deterministic part. All the steps taken in Chapter 2 are performed to answer the important question of whether there might be a simple model that, taking into consideration only the geographical position of the unit, might help us control the spatial dependence better than currently existing procedures and models. Moreover, Chapter 2 adds on the existing problem of introducing and specifying weight matrices.

Chapter 3 becomes the sequel of the previous Chapter, where we extend the theoretical part of the spatial models, introducing other approaches to the trend analysis. However, the important part of this Chapter becomes the practical analysis of different types of datasets with different characteristics and economic backgrounds. The analysis itself follows the same steps as in Chapter 2, however we introduce one more delta model and apply the procedure to a higher number of datasets. This type of analysis allows us to capture the main characteristics of the datasets, where we can control both deterministic structure and spatial structure of the data. One would like to find the deterministic part of different datasets, that might be useful to develop a generalized method of using each model specifications.

Chapter 4 deals with another permanent characteristic, namely, personality traits and their effects on the labor market. Given the nature of the problem, no elements of spatial analysis are required, however what is certainly critical is the quality of the data and its econometric treatment. In this regard, we would like to emphasize the role of the personal characteristics in the labor situation of each individual. Our main goal in this analysis is to examine the role of personality traits in determining the success of unemployed workers seeking a job and their success in the labor market, measured by a number of unemployment spells. We would like to analyze whether the personality traits are major determinants of job transitions and unemployment status. Based on the previous findings, we might find significant relations between the labor market outcomes and psychological personal characteristics of the person.

Finally, we conclude both research lines and present the possible future paths to follow with different datasets and further development of the approaches offered.

The Appendix contains the tables, schemes and full tables for all the chapters.

2. Spatial Trends and Spatial Econometric Structures: The case of housing prices

Spatial, organizational or social interactions among economic agents are common in economics. Anselin, 2002 lists the following terms used to name these interactions: social norms, neighborhood effects, peer group effects, social capital, strategic interaction, copy-catting, yardstick competition and race to the bottom, etc. In particular, he highlights two situations of competition between companies justifying the use of a spatial or interaction model. In the first case, the decision of an economic agent (e.g. a company) depends on the decision of the other agents (its competitors). One example is provided by companies competing with each other by quantity (Cournot competition). Firm i wishes to maximize its profit function $\pi(q_i, q_{-i}, x_i)$ by taking into account its competitors' production levels q_{-i} and its own characteristics x_i which determine its costs. The solution to this maximization problem is a reaction function such as $q_i = R(q_{-i}, x_i)$. In the second case, the decision of an economic agent depends on a scarce resource.

Using the same example of an industrial firm, the profit function is written $\pi(q_i, s_i, x_i)$ with s_i being a scarce resource (which can be natural, for example uranium, or otherwise, for instance, an electronic component manufactured by a single firm). Quantity s_i , which will then be consumed by the company, depends on the quantities consumed by the other companies and therefore on their production q_{-i} . This brings us back to the previous reaction function. This example shows that the use of an interaction model is micro-founded and that the concept of neighborhood is not necessarily spatial. Depending on the industrial sector, a company's competitors will be those that show proximity in terms of distance (services to individuals, supermarkets) or products sold (Coca-Cola and Pepsi). It is emphasized that these two situations lead to the implementation of the same spatial or interaction model. They are equivalent from an observational point of view. The Data Generating Processes (DGP) are different but provide the same observations. Simple cross-section data are not enough to identify the source of the interaction (strategic quantity competition or resource competition in our example) but they can only confirm its presence and assess its strength. As with conventional econometrics, the effects identified by the model and the data still need to be considered.

In addition, externalities or neighborhood effects are commonly taken into account (or controlled) using spatial variables such as distance (e.g. to the nearest competitor) or indicators aggregated by geographical zone (e.g. number of competitors). This type of variable can be interpreted as having spatial lag (i.e. function of observations in neighboring zones), with an *a priori* definition of neighborhood relations. Spatial econometrics therefore justify and foster the widespread use of these empirical choices.

2.1. Trends in time series

Spatial trends have not played a prominent role in explaining and understanding how the outcomes in one geographical location are related to the outcomes in nearby locations (regions, countries or points in space). This is especially evident in the spatial econometric literature. This absence contrasts with the role that time trends have played in time series econometrics to explain economic outcomes that are close in temporal terms. The trend shows the general tendency of the data to increase or decrease during a long period of time.

Mathematically, a time series is given as

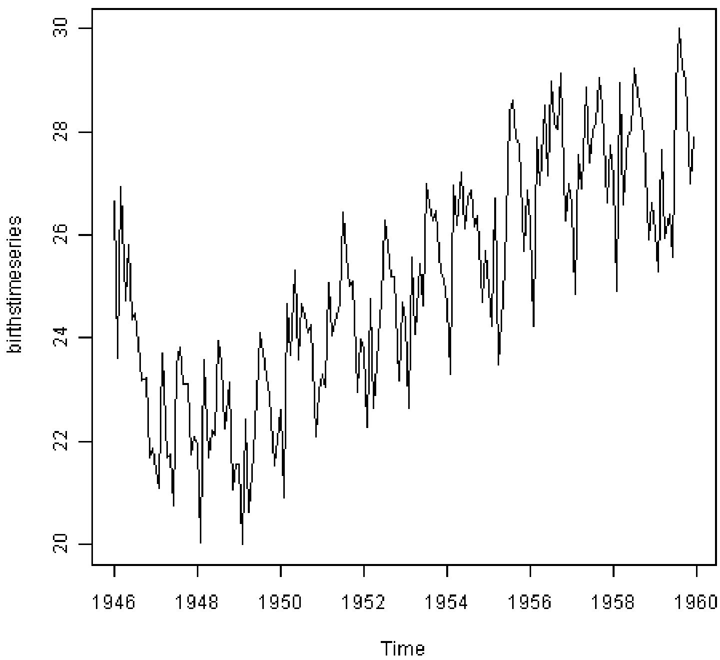
$$y_t = f(t)$$

Here, y_t is the value of the variable under study at time t. If the population is the variable under study at the various time periods $t_1, t_2, t_3, \ldots, t_n$, then the time series is:

$$t : t_1, t_2, t_3, \ldots, t_n$$
$$Y_t : Y_{t1}, Y_{t2}, Y_{t3}, ..., Y_{tn}$$

Thus, when analyzing time series, the economists supposed that the dependence of the outcomes in different points of time might exist. The resulting model is: $Y_t = \beta_0 + \beta_1 Y_{t-1} + ... + \beta_n Y_{t-n} + \gamma X + \varepsilon$. This type of models makes it possible to include the outcomes of the variable in the nearest points, in terms of time. Determining the nearest point in time is easy, as one can define any necessary lag and include it in the analysis. The graphic example of time series trend is presented in 2.1.1

Figure 2.1.1.: Time series trend



However, it becomes a challenge when one tries to include the dependence lag in the spatial data. Difficulties arise from the way the data generating process is formed. The data generating process for a conventional cross-sectional non-spatial sample of n independent observations y_i is introduced as

$$y_i = X_i\beta + \varepsilon_i$$
$$\varepsilon_i \sim N(a, b)$$

where X_i is a $1 \times k$ vector of explanatory variables and β is a $k \times 1$ vector of parameters. It suits for linear regression models with mean $X_i\beta$ and a random component ε_i. $(E(x_i x_j) = E(x_i)E(x_j) = 0)$

Spatial dependence has a dependence model similar to that of time series, as values observed in one location depend on the values of the neighboring observations in the nearby locations.

Suppose $i = 1$ and $j = 2$ are two neighborhoods, then a DGP is:

$$\begin{aligned} y_i &= \alpha_i y_j + X_i\beta + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2), \quad i = 1 \\ y_j &= \alpha_j y_i + X_j\beta + \varepsilon_j \quad \varepsilon_j \sim N(0, \sigma^2), \quad j = 2 \end{aligned}$$

These are simultaneous DGPs where y_i depends on y_j and vice versa. Under standard econometric modeling, it is impossible to model spatial dependency.

2.2. Spatial data solution

2.2.1. Spatial Weight matrix

However, including the nearest spatial point in the analysis is not that trivial. It is not evident how traditional econometrics can deal with spillover effect, while spatial econometrics have proposed (Anselin (1988); LeSage (2005)) a well-known method to capture outcomes that might depend on outcomes in “nearby” locations but not those further away.

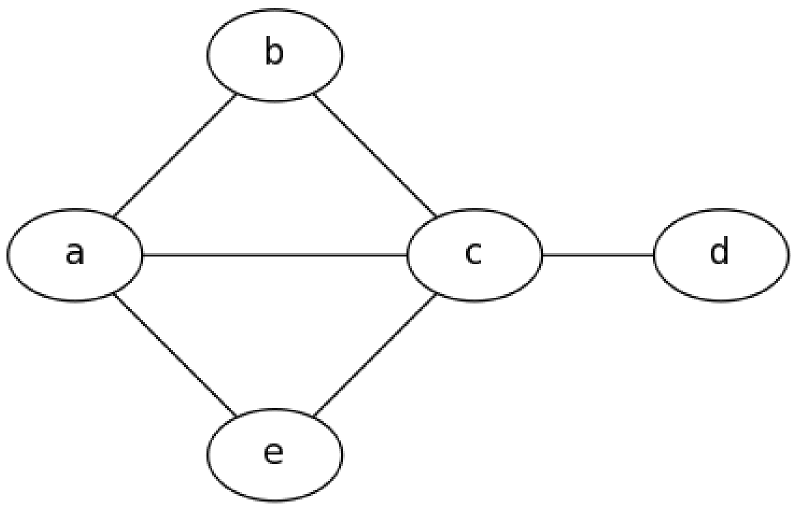
To do so, a simple way to capture these restrictions is to define a W matrix to reflect spatial connectivity among neighbors, a so-called spatial weight matrix, which has served as a basis for different econometric model specifications that explicitly incorporate spatial lags. It imposes a structure in terms of what the neighbors are for each location and assigns weights that measure the intensity of the relationship among pairs of spatial units. What is more, the spatial weight matrix must not necessarily be symmetric.

Let n be the number of spatial units. The spatial weight matrix, W, a $n \times n$ (with n being a number of spatial units) positive symmetric and non-stochastic matrix with element w_{ij} at location i, j. The values of w_{ij} or the weights for each pair of locations are assigned by some preset rules which define the spatial relations between locations. By convention, $w_{ij} = 0$ for the diagonal elements. The values in the cells of W comprise an explicit hypothesis about the strength of interlocation connection (typically towns, regions, or countries). There are two main approaches to construct a weight matrix:

1. Contiguity.
2. Distance-based.

The example of the contiguity weight matrix is presented in 2.2.1, where we take into account the direct connection of node i with a node j.

Figure 2.2.1.: Spatial weight matrix



We observe that a is connected to b, c and e, while c is connected to every node in the network. Thus, following this rule that

$$w_{ij} = \begin{cases} 1 & \text{if } j \text{ has a direct connection with } i \\ 0 & \text{otherwise} \end{cases}$$

W will take the form of

$$W = \begin{array}{cc} & \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \\ e \end{array} & \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

Going back to Tobler's First Law of Geography, we know that everything depends on everything else, but closer things more so.

Thus, in our analysis, we use the second approach of constructing the weight matrix, namely the method of k-nearest neighbors (i.e. we explicitly limit the number of neighbors).

$$w_{ij} = \begin{cases} 1 & \text{if centroid of } j \text{ is one of the } k \text{ nearest centroids to that of } i \\ 0 & \text{otherwise} \end{cases}$$

A typical example, although there are many others that we will define later, is the SAR model $Y = \beta X + \rho WY + \varepsilon$ where Y is the dependent variable, X contains explanatory variables of Y, ε stands for the error term and W is the weight matrix or matrix of connections. Another classic model is SEM model $Y = \beta X + u$, where $u = \rho W\varepsilon + \nu$, i.e, there is spatial autocorrelation in the errors. Other spatial econometric models share a common denominator, namely, space enters into the equation through W.

The use of the weight matrix W has been a controversial issue over the past few years. The two main and most severe critiques are McMillen (2012) and Gibbons and Overman (2012).

McMillen's critique is based on the fact that introducing space via the error or the dependent variable is to control for unknown sources of spatial dependence, while most of the literature takes a leap of faith by considering or viewing the model as a correct parametric form. The problem emphasized by McMillen is that W is subject to a severe identification problem that is only overcome by imposing a simple structure on W, despite the fact that this structure is never known *a priori*. He states that the possible explanation for it is that the location that is taken into account in the weight matrix is not the best predictor, which contradicts all previous assumptions that economists proposed.

Furthermore, as argued by Gibbons and Overman (2012) , the use of W in the dependent variable and/or in the error structure might be pointless for identifying causal links and it could also be easily biased because W might be endogenous. Another related source of problems is that the specification of W is often arbitrary, as this selection is made *a priori*. It usually depends on the user's judgment regarding every spatial lag of X, which are just neighbour averages that are almost always very highly mutually correlated. Even this weak identification depends on the strong assumption that W is correctly specified, so that higher order spatial lags of dependent variables provide additional information (e.g., they satisfy the exclusion restrictions required to make them valid instruments).

The problem of selecting a weight matrix among the different possibilities is a problem of model selection. In fact, different weight matrices result in different spatial lags of the endogenous or the exogenous variables included in the model. Different equations with different regressors amount to model selection problems, even when the weighting matrix appears in the equation of the errors. Moreover, these different specifications are generally impossible to distinguish without assuming prior knowledge about the true data generating process that we often do not possess in practice. This decision is extremely important because if matrix W is misspecified in some way, parameter estimates are likely to be biased and they will be inconsistent in models that contain some spatial lag, as stated in Mur et al. (2011). Furthermore, the consequences for evaluating effects of policy decisions can be serious if model specification is not conducted properly.

Now all causal parameters are identified (only three parameters but an infinitely large number of spatial lags). The fact that spatial econometricians argue that parameters are identified is because they assume that W is known and represents real-world linkages. While neighborhood effect researchers state that the true W is almost never known.

If the exact structure is not known, than identification breaks, as $w_i'W$ can better capture the connections between the neighbors, than $w_i^{'}$ (if x has effect up to 5 km, but $w_i^{'}$ incorrectly restricts effects to within 2 km.). So if the exclusion restrictions on $w_i^{'}W$ are invalid, then these spatial lags are not suitable as instruments, nor as sources of identification.

If W is assumed as known (not idempotent), then there is an estimation problem as the interactions are likely to be highly correlated. (a “weak instruments/identification” problem, because there is little independent variation (and hence little additional information) in the higher order spatial lags). In theory, the degree of collinearity between spatial lags depends on sample size, sampling frame and how W changes as observations are added. In practice, in large samples (and using standard $w_i^{'}$), $w_i^{'}X$, $w_i^{'}WX$ etc. are likely to be highly correlated for the simple reason that they are a weighted average (and consistent estimates of the mean) of x_i in some neighborhood of i - the parameters are likely to be imprecisely measures or severely biased.

Serious problems also arise if there is spatial correlation in the unobserved components u_i. This may happen because of sorting (unobservably similar agents tend to be co-located), common unobserved shocks or causal linkages between neighbours´ unobserved characteristics. For simplicity, assuming that neighborhood exogenous characteristics ($Xw_i^{'}$) do not directly affect outcomes: $y_i = \rho w_i^{'}y + x_i^{'}\beta + w_i^{'}X\gamma + u_i$. Estimation provides two coefficients which identify β but do not allow separate identification of ρ and γ. How can you distinguish between something unobserved and spatially correlated driving spatial correlation in y from the situation where y is spatially correlated because of direct interaction between outcomes?

In traditional spatial econometric models , it is the assumption that most standard W matrices are not idempotent, which allows identification (Gibbons and Overman (2012)).

2.2.2. Spatial models, parametric approach

Ideally, spatial economic theories should provide the researcher with sufficient *a priori* information to enable the construction of fully specified spatial econometric models. In such a situation, the researcher can make an unambiguous choice from a wide range of possible model specifications and appropriate econometric/statistical methods in accordance with various criteria such as unbiasedness, consistency, efficiency, etc. Unfortunately, this is not the common situation in (spatial) economics.

As a consequence, researchers from the social sciences are confronted with substantial specification uncertainty.

Starting with the OLS model, spatial econometrics literature has developed models that treat three different types of interaction effects among units:

(i) endogenous interaction effects among the dependent variables,

(ii) exogenous interaction effects among the explanatory variables, and

(iii) interaction effects among the error terms.

$$Y = \alpha \iota_N + X\beta + \varepsilon, \tag{2.2.1}$$

where Y represents an $N \times 1$ vector consisting of one observation on the dependent variable for every unit in the sample $(i = 1, ..., N)$, N is an $N \times 1$ vector of ones associated with the constant term parameter, X denotes an $N \times K$ matrix of explanatory variables associated with the $K \times 1$ parameter vector, and $\varepsilon = (\varepsilon_1, ..., \varepsilon_N)^T$ is a vector of independently and identically distributed disturbance terms with zero mean and variance σ^2. Since 2.2.1 is commonly estimated by ordinary least squares (OLS), it is often referred to as the OLS model. As we mentioned before, a number of models exist, where space enters into the equation through W(2.1). Taking into account the number of existing models, economists propose different approaches when it comes to choosing the best-fitting spatial model.

2.2.2.1. Top-down approach

The so-called top-down approach 2.2.2.1 consists of starting from the GNS spatial model (2.2.2), we refer to it as the General Nesting Spatial (GNS) model since it includes all types of interaction effects.

The model in 2.2.2.1 that includes all possible interaction effects takes the form:

$$Y = \rho WY + \alpha \iota_N + X\beta + WX\theta + u, \; u = \lambda Wu + \varepsilon \tag{2.2.2}$$

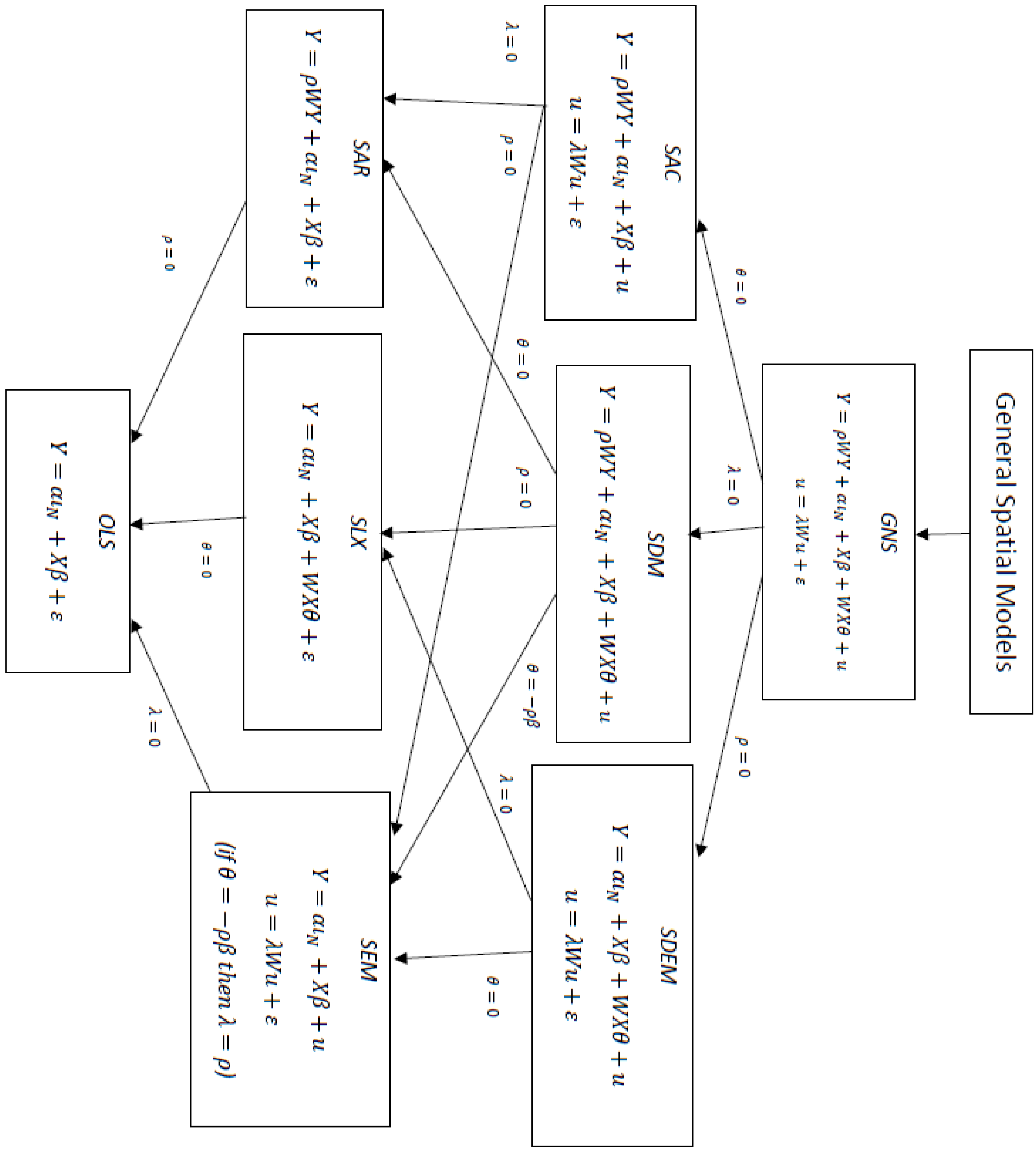


Figure 2.2.2.: Top-down approach in spatial models

Table 2.1.: Spatial model specification

GNS	$Y = \rho WY + \alpha\iota_n + X\beta + WX\theta + u$ $u = \lambda Wu + \varepsilon$
SAC	$Y = \rho WY + \alpha\iota_n + X\beta + u$ $u = \lambda Wu + \varepsilon$
SDM	$Y = \rho WY + \alpha\iota_n + X\beta + WX\theta + \varepsilon$
SDEM	$Y = \alpha\iota_n + X\beta + WX\theta + u$ $u = \lambda Wu + \varepsilon$
SAR	$Y = \rho WY + \alpha\iota_n + X\beta + \varepsilon$
SLX	$Y = \alpha\iota_n + X\beta + WX\theta + \varepsilon$
SEM	$Y = \alpha\iota_n + X\beta + u$ $u = \lambda Wu + \varepsilon$ if $\theta = -\rho\beta$, then $\lambda = \rho$
OLS	$Y = \alpha\iota_n + X\beta + \varepsilon$

As we already know, W, the spatial weights matrix, is a positive $N \times N$ matrix that describes the structure of dependence between units in the sample. The variable WY denotes the endogenous interaction effects among the dependent variables, WX the exogenous interaction effects among the explanatory variables, and Wu the interaction effects among the disturbance terms of the different observations. The scalar parameters ρ and λ measure the strength of dependence between units, while θ, like β, is a $K \times 1$ vector of response parameters. The other variables and parameters are defined as in model2.2.1. Since the GNS model incorporates all interaction effects, models that contain less interaction effects can be obtained by imposing restrictions on one or more of the parameters (shown next to the arrows in 2.2.2.1). Both frequently used, but also largely neglected models are included. In particular, the SLX model is generally overlooked in spatial econometrics literature. Various methods can be applied to estimate spatial econometric models such as Maximum Likelihood (ML), Instrumental Variables or Generalized Method of Moments (IV/GMM), and Bayesian methods.

A mechanism to choose the best-fit spatial model (spatial mechanism). After running the spatial models' regressions, one of the criteria used to choose is the likelihood ratio (LR) test based on the log-likelihood function values of the different models. The LR test is based on minus two times the difference between the value of the log-likelihood function in the restricted model and the value of the log-likelihood function of the unrestricted model: $-2 \times (logL_{restricted} - logL_{unrestricted})$. This test statistic has a Chi squared distribution χ^2_n with n degrees of freedom equal to the number of restrictions imposed. The election rule states that if $LR_{test} > \chi^2_n$, then the unrestricted model performs better than the restricted one. Using this criterion we can make a comparison of the models, as detailed below.

- OLS vs SLX
- OLS vs SAR
- OLS vs SEM
- SAR vs SAC
- SEM vs SAC
- SLX vs SDM
- SEM vs SDM
- SAR vs SDM
- SLX vs SDEM
- SAR vs SDEM
- SEM vs SDEM

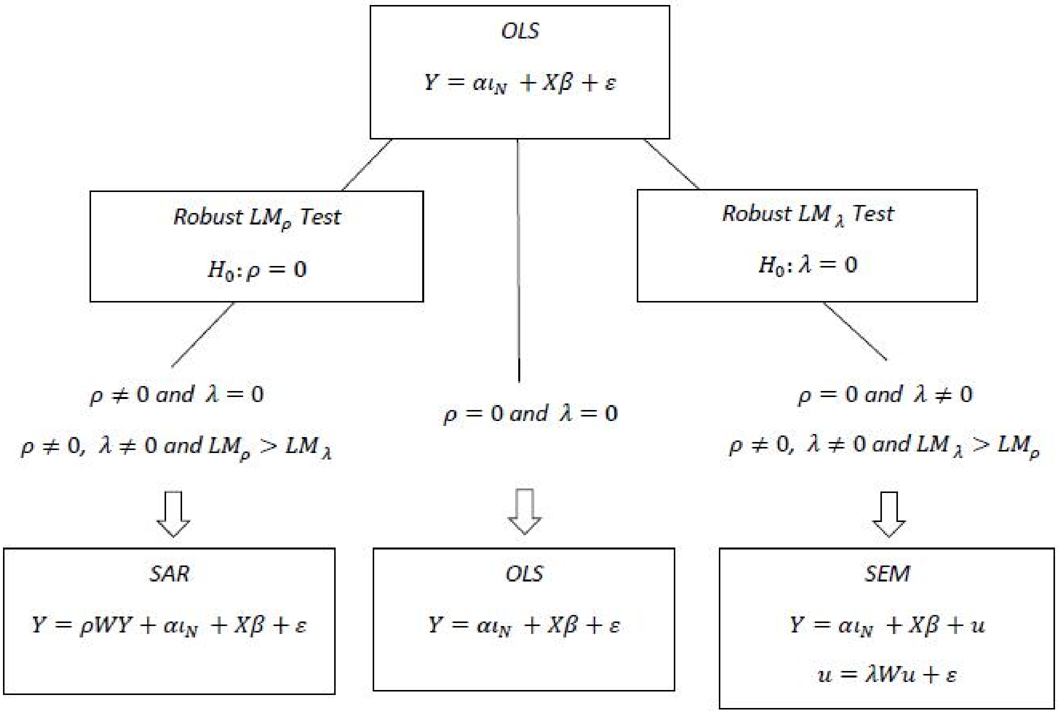
Other models can not be compared among themselves with LR test, as they are not nested. The only two models that are not nested but can be compared are SAR and SEM models. The criteria used to make a comparation are the Lagrange multiplier tests that make it possible to choose between spatial or non-spatial model according to the mechanism we explain below. The rest of the models can only be subjectively compared, checking the significance of the estimated coefficients and the structure of the model.

2.2.2.2. Bottom-up approach

The first so-called bottom-up approach (2.2.3) consists of starting with the non-spatial model (see Le Gallo (2002) for a summary). The Lagrange multiplier tests (

Anselin (1988)for the SAR and SEM model specification tests, robust to the presence of other types of spatial interactions), then make it possible to choose between SAR, SEM or non-spatial models. This approach was widely-favored until the 2000s because the tests developed by Anselin et al. (1996) are based on the residuals of the non-spatial model. They are therefore inexpensive from a computational point of view. Florax et al. (2003) have also shown, using simulations, that this procedure was the most effective when the real model is a SAR or SEM model.

Figure 2.2.3.: Bottom-up approach



There is extensive literature on how the coefficients of each of the interaction effects can be estimated. Considerably less attention has been paid to the interpretation of these coefficients. Many empirical studies use the point estimates of the interaction effects to test the hypothesis as to whether or not spillovers exist. Only recently, thanks to the work of LeSage and Pace (2009), researchers started to realize that this may lead to erroneous conclusions, and that a partial derivative interpretation of the impact from changes to the variables of different model specifications represents a more valid basis for testing this hypothesis.

Parametric methods are helpful in a lot of cases, however, they become unfeasible in the simultaneous presence of different sources of model misspecification, such as substantial spatial dependence, nonlinear relationship of spatially correlated independent variables, unobserved spatial heterogeneity, spatially varying relationships, and common factors. (Basile and Minguez (2018)) That leads to the impossibility of obtaining consistent and efficient estimates. Thus, a number of non-parametric and semiparametric frameworks, that are more flexible to be able to deal with the problem of spatial dependence, have been developed.

2.2.3. Semiparametric approach and Splines

Spatial econometric frameworks that include parametric methods appear to be unfeasible when another source of model misspecification appears. The latter can include substantial spatial dependence, nonlinear relationship of spatially correlated independent variables, unobserved spatial heterogeneity, spatially varying relationships, and common factors. Though non-parametric methods have already gained a great popularity in time series analysis, their usage in spatial econometrics is still scarce. Some contributions of Basile and Minguez (2018), Montero et al. (2012) attempt to promote a more flexible estimation framework to address this problem.

Nonparametric and semiparametric models are attractive alternatives to parametric variations because they admit at the start that the structure of a true model is unknown. This type of models can be used to carry out hypothesis testing and they can be easily implemented.

Recently, Geniaux and Martinetti (2018) have introduced a new class of models, called MGWR-SAR (Mixed Geographically Weighted Regression Simultaneous Auto Regressive models), where the regression parameters and the spatial dependence coefficient can vary over space. In its most general form, the MGWR-SAR is specified as:

$$y = \rho(x_{s_1}, x_{s_2}; h)Wy + X^*\beta^* + \beta(x_{s_1}, x_{s_2}; h)X + \varepsilon$$

where y is the $N-$vector of the continuous dependent variable, X^* is a matrix of k_1 exogenous explanatory variables entering the model linearly (i.e. with spatially stationary coefficients β^*), while X is a matrix of k_2 exogenous explanatory variables with non-stationary coefficients $\beta(x_{s_1}, x_{s_2}; h)), x_{s_1}, x_{s_2}$ are spatial coordinates, W is the spatial weights matrix, ρ the spatial spillover parameter and ε is an i.i.d. error vector.

In this way, they relax the hypothesis that the spatial parameter ρ and the regression parameter β are constant over the coordinate space. The value of these parameters, in fact, depends on the coordinates. The parameters $\rho(x_{s_1}, x_{s_2})$ and $\beta(x_{s_1}, x_{s_2})$ are only required to be spatially smoothed. The use of the Spatial Two-Stage Least Squares (S2SLS) technique is proposed for the estimation of these types of models. In particular, a 5-step approach is used, a local linear estimator (a variant of the GWR) and Cross Validation for the selection of the bandwidth parameter.

A characteristic of this approach is that it only considers spatial parameter heterogeneity (i.e. parameter heterogeneity over the space of coordinates), while neglecting the possibility of pure nonlinearities (i.e. parameter heterogeneity over the domain of the explanatory variable). However, it remains very important to assess the existence of pure nonlinearities in the relationship between the response variable and the covariates. Moreover, keeping the spatial autocorrelation parameter (ρ) constant over space is a valid option: in that case, the feedback effects of spatial autocorrelation have a clearer definition and the interpretation of direct and indirect effects is easier.

Another branch of the spatial econometric literature has proposed Spatial Autoregressive Semiparametric Geoadditive Models as a means of simultaneously dealing with different critical issues typically encountered when using spatial economic data; namely, spatial dependence, spatial heterogeneity and unknown functional form (Basile et al. (2014)). This approach combines penalized regression spline (PS) methods Eilers et al. (2015) with standard spatial Autoregressive models (such as SAR, SEM, SDM and SLX) which we mentioned above. An important feature of these models is that they make it possible to include within the same specification: (i) spatial autoregressive terms to capture spatial interaction or network effects; (ii) parametric and nonparametric (smooth) terms to identify nonlinear relationships between the response variable and the covariates; and (iii) a geoadditive term, i.e. a smooth function of the spatial coordinates, to capture a spatial trend effect, that is, to capture spatially autocorrelated unobserved heterogeneity.(Basile and Minguez (2018)) The structural form of the Penalized-Spline Spatial Lag model (PS-SAR) is:

$$
\begin{aligned}
y &= \rho W y + X^* \beta^* + f_1(x_1) + f_2(x_2) + f_3(x_3, x_4) \\
&\quad + f_4(x_1) z + ... + h(x_{s_1}, x_{s_2}) + \varepsilon
\end{aligned}
$$

where y is a continuous univariate output variable, Wy its spatial lag, $X^* \beta^*$ is the linear predictor for any strictly parametric component (including the intercept, all categorical covariates and eventually a set of continuous covariates) and $f_k(.)$ are unknown smooth functions of univariate continuous covariates or bivariate interaction surfaces of continuous covariates, capturing nonlinear effects of exogenous variables. Which of the explanatory variables enter the model parametrically or non-parametrically may depend on theoretical priors or can be suggested by the results of model specification tests (Kneib et al. 2009). $f_4(x_1)z$ is a varying coefficient term, where z is either a continuous or a binary covariate. The term $h(x_{s_1}, x_{s_2})$ is a smooth spatial trend surface, i.e. a smooth interaction between latitude and longitude. It allows us to control for unobserved spatial heterogeneity, which is a primary task when dealing with spatial data. When the term $h(x_{s_1}, x_{s_2})$ is interacted with one of the explanatory variables (e.g., $h(x_{s_1}, x_{s_2})x_1$), it allows us to estimate spatially varying coefficients (like in the GWR model). Finally, ε are iid normally distributed random shocks.

This model reflects the notion of spatial dependence that consists of two parts: (i) a spatial trend due to unobserved regional characteristics, which is modeled by the smooth function of the coordinates, and (ii) global spatial spillover effects, which are modeled by including the spatial lag of the dependent variable. The introduction of the spatial lags of the exogenous (X) variables results in what can be called the Penalized-Spline Geoadditive Spatial Durbin Model (PS-SDM).

The Spatial Error Geoadditive Model (PS-SEM) proposed by Montero et al. (2012) augments the PS model by including a spatial autoregressive error term, while leaving the systematic part unchanged:

$$
\begin{aligned}
y &= X^* \beta^* + f_1(x_1) + f_2(x_2) + f_3(x_3, x_4) \\
&\quad + f_4(x_1) z + ... + h(x_{s_1}, x_{s_2}) + u \\
u &= \lambda W u + \varepsilon \; \varepsilon \sim \text{iid } N(0, \sigma_\varepsilon^2)
\end{aligned}
$$

where λ is a spatial autoregressive parameter.

Next, we will discuss the types of splines we are going to use in our analysis. We consider the following configurations of the nonparametric part:

$$Spline : f(z) = (a, b) \tag{2.2.3}$$

where $f(z)$ is fully nonparametric and is limited to longitude and latitude variables;

$$C{-}spline : f(z) = \beta_0{+}\beta_1(x{-}x_0){+}\beta_2(x{-}x_0)^2{+}\beta_3(x{-}x_0)^3{+} \sum_{s=1}^{S} \delta_s(x{-}x_s)^3 D_s \tag{2.2.4}$$

where the spline simply adds a set of interaction terms between dummy variables and cubic terms to a standard cubic function, and where S is equal intervals ranging from $x_0 = min(x)$ to $x_S = max(x)$ and there is a dummy variable D_s indicating whether x is greater than x_s. C-spline is used as an analogy approximation to the G-model we introduced before. This allows us to make better comparison of the models that are in the same analysis line. Lastly, a Fourier based spline of the form

$$F - spline : f(z) = \beta_0 + \beta_1 z + \beta_2 z^2 {+} \sum_{j=1}^{J} (\gamma_j sin(jz) + \lambda_j cos(jz)) \tag{2.2.5}$$

where $z = 2\pi(x{-}min(x))/(max(x){-}min(x))$.

It should be recalled that splines and series regression are based on the mathematical theory of the approximation of functions. Particularly, spatial-econometricians that are concerned with approximating the conditional expectation function, find the Weierstrass-Stone theorem which states that any continuous function can be uniformly well approximated by a polynomial of sufficiently high order, under mild regularity conditions, very useful. There are mathematical results that point out that when the true conditional expectation function is smoother, it is possible to approximate it with a fewer number of series terms. This explains why other spline methods like B-splines or P-Splines can be used instead of (or together with) the ones we have selected. The central point is the same one as in the delta-models that are introduced in the following section, which is considering basic coordinates can be a first step to control for spatial relationships. One or more of these simple structures (i.e., a family of models) can approximate a spatial trend even in the case of a nonlinear spatial trend.

2.3. Delta test

The delta-test, that we briefly describe below, tests for the null of a non-stochastic leading term in a spatial dataset $\{X_s\}_{s\in S}$ where S is a set of coordinates. To do so the spatial realization $\{X_s\}_{s\in S}$ is embedded in an m-dimensional space:

$$X_m(s_0) = (X_{s_0}, X_{s_1}, \ldots, X_{s_{m-1}}) \text{ for } s_0 \in S$$

where $N_s = \{s_1, \ldots s_{m-1}\}$ are the $m-1$ nearest neighbors to s_0. A symbolization map is then defined $f : \{X_s\}_{s\in S} \hookrightarrow \mathbb{R}^m \to \Gamma$ as:

$$f(X_s) = (\mathcal{I}_{ss_1}, \mathcal{I}_{ss_2}, \ldots, \mathcal{I}_{ss_{m-1}}) \tag{2.3.1}$$

where $\mathcal{I}_{ss_j}$ is an agreement indicator function of being above or below the median at locations s and s_j, Γ is the set of 2^{m-1} different vectors of dimension $m-1$ with entries in the set $\{0, 1\}$, where we refer to each symbol by σ_i. Obviously, it is required that the spatial process X_s has a finite median, otherwise the test cannot be applied, which is not a very strict limitation. Then the relative frequency, p_σ, of each symbol is computed from the data, and the associated entropy of the dataset is calculated: $h(\Gamma) = -\sum_{\sigma\in\Gamma} p_\sigma \ln(p_\sigma)$. The delta-test consists of estimating the behavior of a function of the difference between entropies $h^{\mathcal{W}_{j+1}}(\Gamma) - h^{\mathcal{W}_j}(\Gamma)$ where $\mathcal{W}_j$ and $\mathcal{W}_{j+1}$ are sets of symbols chosen at random from Γ. Under the null of a non-stochastic spatial structure, that difference does not increase with the number of symbols considered.

Particularly, the delta-test is implemented by testing if $\alpha_1 = 0$ in the following regression

$$dh^{\mathcal{W}_j}(\Gamma) = \alpha_0 + \alpha_1 j + \varepsilon_j, \quad \text{for} \quad j = 1, 2, ...k - 1 \tag{2.3.2}$$

where

$$dh^{\mathcal{W}_j}(\Gamma) = \frac{h^{\mathcal{W}_{j+1}}(\Gamma) - h^{\mathcal{W}_j}(\Gamma)}{\log \frac{j+1}{j}}.$$

As shown in Garcia-Cordoba et al. (2019), the delta statistic is a test well-suited to detecting simple and complex spatial trends. Provided with the delta-test, $(dh{-}test)$ in the following tables, we can supplement the spatial analysis by applying the test to the spatial raw data. In case of a rejection of the null hypothesis of the non-stochastic spatial leading term, the possibility of specification of a scenario with spatial (deterministic) trends opens up for the econometric modeler. A natural way for modeling this situation from an econometric point of view is by using what we call restricted semiparametric regression:

$$Y = \alpha\iota_n + X\beta + f(a, b) + \varepsilon \tag{2.3.3}$$

where each element on vector Y is a continuous output variable in a given location and $X\beta$ contains all explanatory variables (i.e., a set of explanatory variables that can include categorical variables and where vector β collects fixed parameters). The important nonparametric part $f(a, b)$ is restricted to geographic functions of longitude and latitude, a, b, respectively. At this point, according to the rejection of the null hypothesis of the non-stochastic spatial leading term, there is no evidence for introducing a weight matrix (W) into the model, neither in the parametric part $X\beta$ nor in the nonparametric one.

Several comments are important in this respect. The previous family of models aims to ascertain whether a specification of space via latitude and longitude might serve to control for spatial heterogeneity, once the researcher has had statistical evidence of a spatial trend. At this stage, prior to the use of a given W weight matrix, we wonder if considering some form of geographical variables in the model is enough to correctly estimate vector β. This will avoid the severe consequences in estimation and inference (about β) of not considering spatial heterogeneity when it really exists, as occurs in many fields. Notice also that the family of models (2.3.3) will not be the object of the main critiques that spatial econometrics has received by scholars, upon which we have commented in the previous section.

The delta-test can be used as a diagnostic tool helping in the model selection procedure. Consider a model that erroneously omits some form of spatial dependence

$$Y = \alpha\iota_n + X\beta + u,$$

we understand that the omission can be in form of a linear spatial dependence or in the form of a spatial trend. An example of the former is

$$u := WX\theta + \varepsilon,$$

while the latter can be of the form

$$u := \mu f(a, b) + \varepsilon.$$

How to choose between these specifications is far from straightforward. As shown in Garcia-Cordoba et al. (2019), the delta-test can be used to distinguish between them if the test is applied to the residuals of the misspecified model, that is, if it is applied to $\hat{u}$. In the case of a true spatial dependence via W, the delta-test will tend to point out that no spatial trend is found in the residuals, and therefore the researcher will have to deal with a statistically correct specification of the model (this will probably be done through well-known models in the spatial econometrics literature, as we indicate later in this paper). In this regard, we will expect that Moran's I test will correctly indicate spatial autocorrelation in the residuals. On the contrary, the delta-test will highlight that a spatial trend is omitted if the true spatial dependence comes in the form of a non-stochastic geographic spatial structure (spatial trend). Obviously, the researcher should now take a different modeling strategy, as he/she has put forward a statistically compatible spatial trend. In other words, the proposal of some form of $f(a, b)$ should be required.

Our procedure consists of specifying the model using the previous diagnostics' tools. Particularly we firstly run delta-test on the raw data to check for the existence of a deterministic structure and Moran's test to check if there is a spatial autocorrelation in the data we use. If delta-test cannot reject the statistical existence of a spatial trend, we introduce a geographical additive model of the form given in (2.3.3). In particular, we consider and study two forms for the restricted nonparametric part, $f(a, b)$. The first way (that we will refer to as delta-model strategy) is to restrict $f(a, b)$ to be low-degree polynomials of coordinates, which is inspired by the practice of including powers of t-time in time-series modeling:

- $f_A(a, b) = a + b + a^2 + b^2 + ab$
- $f_B(a, b) = a + b + a^2 + b^2$
- $f_C(a, b) = a + b + ab$
- $f_D(a, b) = a^2 + b^2 + ab$
- $f_E(a, b) = a^2 + b^2$
- $f_F(a, b) = a + b$
- $f_G(a, b) = a^3 + b^3$

We will use letters $A, B, ..., G$ to indicate the model specification we refer to. For example, by Model B we will mean $Y = \alpha\iota_n + X\beta + f_B(a, b) + \varepsilon$.

The procedure we follow to make our analysis is presented in Figure 2.3.1

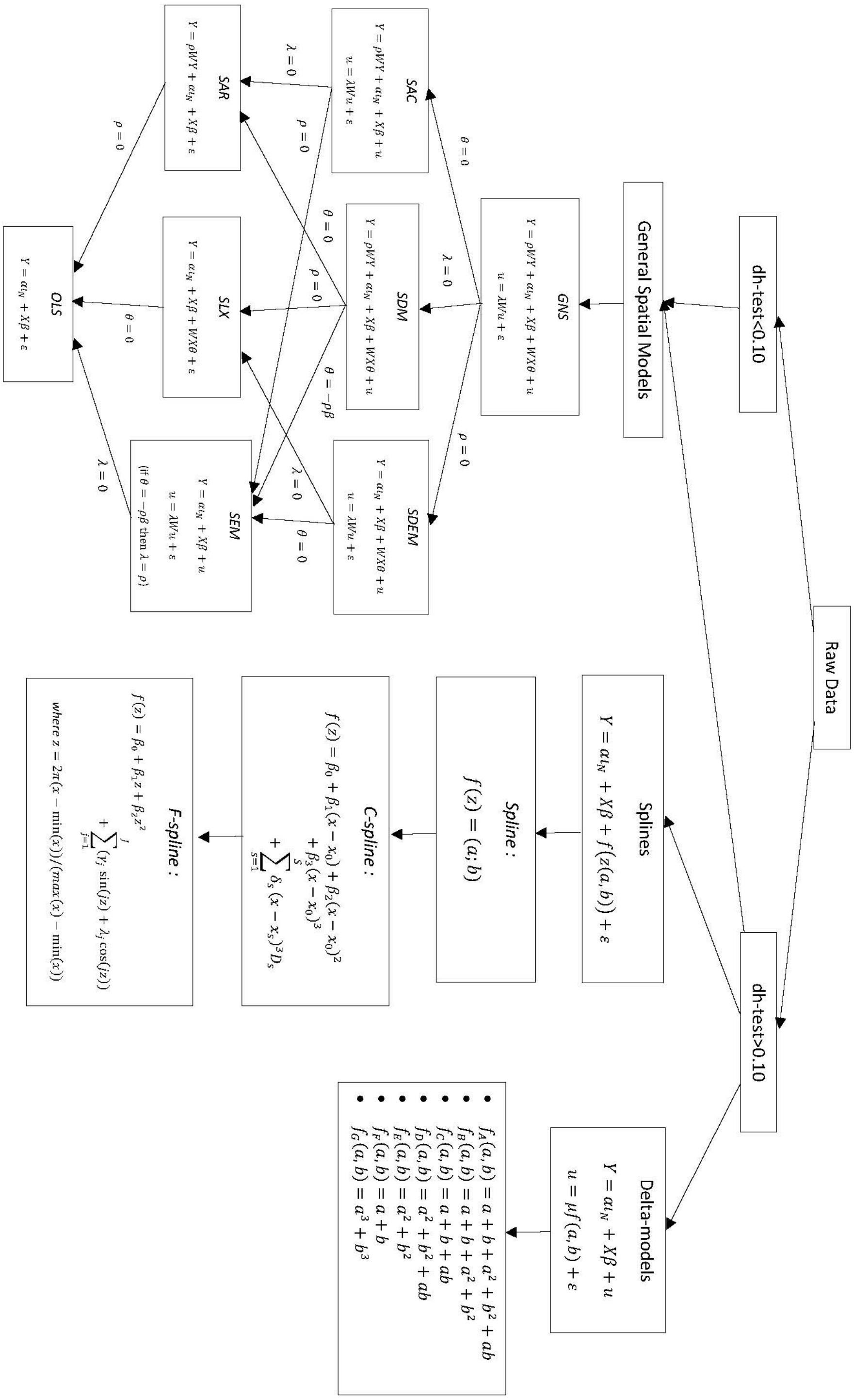


Figure 2.3.1.: Procedure of choosing a model

2.4. Software

The first software programme used in the analysis performed is the toolbox for spatial econometric models written by LeSage and Pace (2009) in MATLAB (MATLAB 2017), some functions, also in MATLAB, are used to estimate static and dynamic spatial panel data models developed by Elhorst (2013).

Another software programme is R:

- it is a well-tested free application with a growing number of packages in all statistical fields (spatial analysis included);
- it has a huge community of users;
- the possibility to combine functional programming with object-oriented programming (Chambers (2014)) allows the developers to build new packages making use of the existing ones;
- it allows the estimation of most of the spatial econometric models detailed in this chapter including both parametric models (for cross-sectional and static panel data) and semiparametric models.

The R packages spdep and sp (Pebesma and Bivand (2005), Bivand et al. (2013)) facilitate the creation, transformation and manipulation of spatial objects, neighborhood matrices and the computation of descriptive measures of spatial autocorrelation. Moreover, the package spdep allows researchers to estimate the whole set of cross-sectional spatial autoregressive models detailed in Sect.2.2 including SAR, SEM, SDM, SLX and SAC models using either ML or GMM estimation in an efficient way. Furthermore, this package also permits us to compute the marginal effects and make inference on their values. Focusing on semiparametric spatial data models,McMillen (2015, 2012) has written the McSpatial package which includes routines to estimate nonparametric and conditionally parametric versions of spatial linear regression and spatial models with a binary dependent variable. It mainly uses kernel techniques to perform the nonparametric estimations.

2.5. Modelling Housing Prices via Hedonic Models

Numerous economists have empirically considered that the price of a house or building or of a plot of land is determined by the characteristics of the property itself (e.g., its size, appearance, features like patios or rooms or fireplaces, and condition of the property), as well as characteristics of its surrounding environment (e.g., if the neighborhood has a high crime rate and/or is accessible to schools and a downtown area, the level of water and air pollution, or the value of other homes close by). To model the relation between price and those characteristics, several hedonic models have been used. Hedonic pricing is a model which identifies price factors according to the premise that price is determined both by internal characteristics of the object of the sale and external factors affecting it. From our point of view, the relevance of estimating a hedonic price model is to estimate the extent to which each factor (characteristic) affects the price of the home. For this reason we focus our study on well-studied data on housing prices in Baltimore, Boston and Toledo, as these are perfect examples of the market where space, along with other characteristics, is directly introduced. A house is a commodity with a fixed location and its value greatly depends on its location and market characteristics. Over the past years researchers have made a great contribution to the housing prices analysis, taking spatial and temporal factors into account. Thus, we complement the existing analysis, including the possibility of spatial trends in the data.

House prices and characteristics from Baltimore were firstly studied in Dubin (1992). The data comes from 1978 multiple listings for Baltimore (Maryland) that contain structural descriptors of the house, the sale price, and the address. Each house is assigned coordinates by locating the address on the Maryland coordinate system. The dependent variable is the selling price of the house and the exogenous variables include the attributes of the structure of the house and its sale: Number of rooms (NROOM), information on whether the house is a detached unit (=1) or not (DWELL), number of bathrooms (NBATH), whether the house has a patio (PATIO), fireplace (FIREPL), air conditioning (AC) and a basement (BMENT), number of stories (NSTOR), number of car spaces in garage (GARAGE), age of dwelling (AGE), a dummy variable that takes a value of 1 if the dwelling is located in Baltimore county (CITCOU), lot size (LOTSZ), interior living area (SQFT), and a month of 1978 in which the house was sold (STIME).

Table 2.2.: Spatial models Baltimore Housing Prices

	OLS	SAR	SEM	SLX	SAC	SDM	SDEM	GNS
Constant	7.16	-0.91	6.95	8.78	-0.05	13.05	3.18	16.66**
Nroom	0.29	0.53	0.75	0.27	0.08	0.29	0.22	0.47
Dwell	6.30**	5.74**	6.96***	6.33**	4.46**	6.35***	7.52***	6.72***
Nbath	6.08***	5.37***	6.57***	5.57***	4.08**	5.59***	6.87***	6.72***
Patio	9.40***	6.92***	7.55***	5.25*	7.39***	5.21***	9.81***	5.63**
Firepl	10.69***	8.21***	8.95***	8.93***	8.46***	8.88***	9.48***	7.59***
AC	8.20***	7.15***	7.34***	7.18***	7.12***	7.17***	7.14***	7.25***
Bment	3.81***	3.78***	3.59***	3.36***	3.77***	3.36***	3.34***	3.02***
Nstor	-4.36	-5.08*	-3.93	-2.41	-5.82**	-2.44	-1.20	-3.13
Gar	5.53***	5.65***	5.23***	5.74***	5.71***	5.74***	4.58***	4.86***
Age	0.00	0.04	-0.01	0.01	0.06	0.00	0.03	0.01
Citcou	12.46***	10.12***	12.43***	10.87***	9.08***	10.86***	13.42***	10.32***
Lotsz	0.04**	0.02	0.03	0.02	0.02	0.02	0.04***	0.01
Sqft	0.37*	0.38*	0.26	0.25	0.47**	0.25	0.08	0.18
WNroom				0.02		0.02	-2.07***	-4.37**
WDwell				-3.29		-3.27	-1.30	-5.39
WNbath				-1.55		-1.70	17.85***	-11.20***
WPatio				-7.21*		-7.28*	14.31***	2.92
WFirepl				10.96*		10.75*	31.64***	9.62*
WAC				17.45***		17.02***	26.05***	-1.10
WBment				5.36		5.19	16.69**	-0.82
WNstor				0.35		0.27	8.19***	4.81
WGar				7.69		7.59	0.71	0.22
WAge				2.24		2.07	9.73**	-0.02
WCitcou				0.04		0.04	0.11	-6.87
WLotsz				-1.71		-1.71	26.68***	0.07*
WSqft				0.10**		0.10**	0.14***	0.48
rho		0.27***			0.36***	0.00		0.45***
lambda			0.31***		-0.34		0.96***	-0.71***
Log-Likelihood	-926.53	-817.7957	-751.92	-895,15	-816.07	-802.20	-778.74	-797.55
Moran(p-value)	0.02	0.09	0.02	0.46	0.39	0.46	0.36	0.92
dh-test(p-value)	0.02	0.05	0.12	0.01	0.12	0.12	0.13	0.12
AIC	1.88	1.67	1.53	1.84	1.66	1.51	1.66	1.65
	OLS	SAR	SEM	SLX	SAC	SDM	SDEM	GNS

Using the Elhorst methodology we can see that the log-likelihood function value of the OLS model increases from -926.53 to -895.15 when this model is extended to include exogenous interaction effects (WX), known as the SLX model. The LR-test of the SLX model versus the OLS model takes the value of 62.76 with 13 degrees of freedom (df), while the 5 % critical value is 22.36. This implies that the OLS model needs to be rejected in favor of the SLX model. However, if the OLS is extended to include endogenous interaction effects (WY) or interaction effects among the error terms (Wu), leading us to SAR and SEM models respectively, the log-likelihood function value increases even more, even though in these two cases only one interaction effect is added to the model. Whether it is this SAR or SEM model that better describes the data is difficult to say, since these two models are not nested. One solution is to test whether the spatial lag model or the spatial error model is more appropriate to describe the data, provided that the OLS model is taken as point of departure. We use the classic and robust LM tests. Both the classic and the robust tests are based on the residuals of the OLS model and follow a Chi squared distribution with 1 degree of freedom. Using the classic tests. both the hypothesis of no spatially lagged dependent variable and the hypothesis of no spatially autocorrelated error term must be rejected at five per cent significance; the LM test for the spatial lag amounts to 21.88 and for the spatial error to 3.71. When using the robust tests, the hypothesis of no spatially lagged dependent variable must still be rejected, whereas the hypothesis of no spatially autocorrelated error term can no longer be rejected; the robust LM test for the spatial lag amounts to 19.01 and for the spatial error to 0.83. This indicates that on the basis of these robust LM tests the spatial lag model is more appropriate (though Moran test shows us that there is spatial correlation in the residuals both in SEM and SAR).

Another solution is to consider the SAC model(Moran test on residuals is 0.39). which considers both endogenous interaction effects and interaction effects among the error terms, and therefore nests both the SAR and SEM models. The SAC model produces coefficient estimates of the WY and the Wu variables. and we can see that the coefficients of WY are significant in both SAC and SAR models. Similarly. the LR-test of the SAC model versus the SAR model takes the value of 31.19 with 1 df, and the LR-test of the SAC model versus the SEM model the value of -100.57 with 1 df, while the 5 % critical value in both cases is 3.84. This implies that SAR model needs to be rejected in favor of SAC model, but we cannot say which of the models SAC or SEM is better.

A different way to look at the SAR, SEM and SLX models, on their turn, is to consider the SDM model(Moran test on resiudals is 0.46). since the SDM model nests these three models. The SDM appears to outperform the SLX model (LR-test 185.89, 1 df. critical value 3.84), the SAR model (LR-test 31.19, 13 df. critical value 22.36). but not the SEM model (LR-test -100.57, 13 df, critical value 22.36). Alternatively, one might consider the SDEM model (Moran test on residuals is 0.36) which also nests the SLX and the SEM models. The SDEM model also appears to outperform the SLX model (LR-test 232.81, 1 df, critical value 3.84), SAR model (LR-test 78.10, 13 df, critical value 22.36) but not the SEM model (LR-test. -53.65, 13 df, critical value 22.36). Whether it is the SDM model or the SDEM model that better describes the data is difficult to say, since these two models are not nested. Unfortunately, estimation of the GNS model(Moran test on residuals is 0.92) which nests these two models does not provide an answer. The increase of the log-likelihood function value when estimating this model is not clear, so on the basis of the results reported in 2.2 it is impossible to draw any conclusion as to whether it is SEM, SDEM, SDM or GNS that best describes the data. However, taking into account that rho is not significant in the SDM model one can say that it would be more rational to use GNS model in that case. For our purposes, we have chosen SEM and SDEM models to compare its results with other models of the analysis, as it is not clear which of them it is better to use, and the results of dh-test differ depending on the model.

Moran's test on the data on Baltimore housing prices gives a clear evidence of the spatial autocorrelation (Table 2.3). The delta-test on the raw data confirms the presence of deterministic structure, which gives evidence in favor of running restricted semiparametric analysis, including a spatial trend. Following the modelling proposal given in the previous section, we firstly model the deterministic part by using the so-called delta-models. Results for models A and B are clearly in favor, as the estimated model controls for the spatial heterogeneity of the data: both delta-test and Moran's test indicate that spatial structure is controlled with models A and B. In other words, our estimation of the extent to which each characteristic of the house affects the house price is safely estimated, which was one of the main concerns as initially claimed. The same conclusion is reached if we opt for some spline based methods.

If instead we model according to models in Table (2.1), we find that the best spatial models for our data are SEM and SDEM models. This conclusion is reached by using the method described in the Appendix, where we also present several outputs in the Table in Appendix. One interesting conclusion of the results given in Table 2.3, is that neither SEM nor the SDEM models are able, according to delta-test results, to remove the previously found spatial trend. In other words, the residuals of these models are compatible with a deterministic structure that has not yet been removed. For this reason, the results seem to indicate that restricted semiparametric models work better in this case, as they let us get rid of the spatial structure of the model and thus get more credible results on the estimates. The practical implications for Baltimore housing prices are mainly relative to the partial effects of several explanatory variables, but not to the list of significant variables, nor to the signs.

Table 2.3.: Model comparison for Baltimore Housing Prices

	OLS	Model A	Model B	SEM	SDEM	C-spline	F-spline
Constant	7.16	2799.08	2864.02*	6.95	3.18	-17690.00	-48.30
Nroom	0.29	0.21	0.21	0.75	0.22	0.33	0.29
Dwell	6.30**	6.09**	6.12**	6.96***	7.52***	6.79**	6.95**
Nbath	6.08***	5.71***	5.75***	6.57***	6.87***	5.82**	5.89**
Patio	9.40***	7.94***	7.98***	7.55***	9.81***	8.16**	7.94**
Firepl	10.69***	8.92***	8.95***	8.95***	9.48***	8.80***	8.66***
AC	8.20***	6.26***	6.25***	7.34***	7.14***	6.91**	6.93**
Bment	3.81***	3.71***	3.69***	3.59***	3.34***	3.36**	3.41***
Nstor	-4.36	-4.63*	-4.60*	-3.93	-1.20	-4.19	-4.08
Gar	5.53***	5.15***	5.14***	5.23***	4.58***	4.93**	4.94**
Age	0.00	0.01	0.01	-0.01	0.03	-0.02	-0.03
Citcou	12.46***	10.72***	10.74***	12.43***	13.42***	16.55***	16.83***
Lotsz	0.04**	0.03*	0.03*	0.27	0.04***	0.04*	0.04*
Sqft	0.37*	0.31	0.31	0.26	0.08	0.29	0.28
Moran(p-value)	0.02	**0.25**	**0.29**	0.02	0.36	**0.32**	**0.34**
dh-test (p-value)	0.02	**0.02**	**0.02**	0.12	0.13	**0.02**	**0.01**
AIC	1.88	1.86	1.85	1.53	1.56	5.1	5.09
	OLS	Model A	Model B	SEM	SDEM	C-spline	F-spline

#of embedding dimensions m=6
dh-test p-value on raw data 0.13
***, **, * =coefficient estimates that are significant at the 0.01, 0.05 and 0.1 level respectively.

The second dataset includes data for census tracts in the Boston Standard Metropolitan Statistical Area (SMSA) in 1970. This data was used by Pace and Gilley (1997). It is interesting mainly because it considers environmental characteristics as hedonic determinants of the price. The Boston sample contains 506 census tracts, (one observation per census tract) on 14 non-constant independent variables (excluding tracts containing no housing units or comprised entirely of institutions). These variables include crime rate (CRIME), proportion of area zoned with large lots (ZONING), proportion of non-retail business area (INDUSTRY), location contiguous to the Charles River (CHARLESR), levels of nitrogen oxide (NOXSQ), average number of rooms (ROOMS), proportion of structures built before 1940 (HOUSEAGE), weighted distances to the employment centers (DISTANCE), an index of accessibility (ACCESS), property tax rate (TAXRATE), pupil-teacher ratio (PUPIL/TEACHER), black population proportion (BLACKPOP) and lower status population proportion (LOWCLASS).

The results of Moran's test and delta-test on the data on Boston housing prices, give a clear evidence of the spatial structure. After estimating spatial trends, the results (Table 2.4) show that the deterministic part has been removed, but there still is a spatial structure according to Moran's I.

Therefore, modeling according to models given in Table 2.1 is justified. All the results are reported in the Table in Appendix.

Running the dh-test for the data, the presence of a spatial trend in the data is demonstrated (dh-test (p-value) =0.13).

In this case, the best choice would be GNS and SAC models which correct for spatial structure in the sense that both Moran and delta-tests statistically indicate that spatial structure has been controlled, see Table 2.4. As happened with the previous data set, the list of relevant explanatory variables is common to all the models. Variations are again on the partial effects.

Table 2.4.: Model comparison for Boston Housing prices

	OLS	Model A	Model B	SAC	GNS	C-spline	F-spline
Constant	36.46***	46.69***	46.78***	28.09***	25.74***	704400.00**	-274.30
Crime	-0.11***	-0.13***	-0.13***	-0.13***	-0.15***	-0.12***	-0.12***
Zoning	0.05***	0.03**	0.03**	0.04***	0.03**	0.04**	0.04**
Industry	0.02	0.02	0.02	-0.01	-0.02	0.03	0.03
Charlesr	0.02***	2.68***	2.69***	-0.27	-0.1	2.42**	2.15*
Noxsq	-17.77***	-21.80***	-21.83***	-17.80***	-19.23***	-20.18***	-19.33***
Rooms	3.81***	3.74***	3.74***	4.24***	4.66***	3.62***	3.58***
Houseage	0.01***	-0.02	-0.02	-0.02	-0.03**	-0.02	-0.02
Distance	-1.48***	-3.22***	-3.24***	-1.62***	-1.61***	-2.37***	-2.45***
Access	0.31***	0.36***	0.36***	0.32***	0.31***	0.36***	0.35***
Taxrate	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***
Pupil/Teacher	-0.95***	-1.02***	-1.02***	-0.64***	-0.57***	-0.99***	-0.95***
Blackpop	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***
Lowclass	-0.52***	-0.53***	-0.54***	-0.43***	-0.39***	-0.53***	-0.53***
Moran(p-value)	0	0	0	**0.46**	**0.54**	0	0
dh-test (p-value)	0.04	0.03	0.03	**0.01**	**0.01**	0.03	0.03
AIC	3.52	3.49	3.49	2.89	2.88	3.1	3.09
	OLS	Model A	Model B	SAC	GNS	C-spline	F-spline

#of embedding dimensions m=7
dh-test p-value on raw data 0.20
***, **, * =coefficient estimates that are significant at the 0.01, 0.05 and 0.1 level respectively

The third dataset is the dataset on housing prices in Toledo, Ohio. It includes data on average house values for 98 census tracts along with 10 explanatory variables and latitude-longitude coordinates. Explanatory variables include the neighborhood quality (Neighborhood), net lot square feet (Lot sqft), total square feet living area (Total sqft), family room conditions (Family room), recreation room conditions (Rec room), air conditioning (Air cond), number of bathrooms (Baths), condition of the house (Condition), garage condition (Garage condition) and the age of the house (Age).

As happened with the other two datasets, Moran's test and delta-test reported spatial structure. As regards restricted semiparametric models, it is worth mentioning that both delta models and spline models are able to control for both types of spatial dependences, as can be observed from Table 2.5, which can be completed with information found in the Appendix. This is similar to what happened with the Baltimore housing dataset and model. The difference with respect to the Toledo model is that there is a spatial model (from the set of models given in Table 2.1) that also controls for spatial heterogeneity, and in this sense the SDEM model is preferred to the SAC model. If the modeller is uncomfortable with or suspicious of using a given W matrix because of some of the reasons explained in the introduction, we would recommend using a restricted semiparametric model approach. From the practical point of view, it is remarkable that for this dataset the variable “condition” has no statistical relevance if semiparametric models are adopted, while it has impact if the SDEM model is selected. Something similar happens with the fact that the house has or does not have a recreation room.

Table 2.5.: Model comparison for Toledo Housing Prices

	Model A	Model G	SAC	SDEM	C-spline	F-spline
Constant	1252522880	1224987.63**	8564.30	-7219.56	3.299e+08	3.445e+06*
Neighborhood	-2165.26	-2377.64	-2867.48	-708.78	-1348.00	-1162.00
Lot sqft	0.56	0.38	0.30	0.19	0.55	0.54
Total sqft	2.67	4.78	6.91	4.29	3.95	3.86
Family room	25736.93***	30321.64***	30718.27***	20990.69**	29010.00**	28620.00**
Rec room	-5686.75**	-2600.6*	-1944.34	-4953.70	-4645.00	-5318.00
Air cond	-22861.58	-34788.76**	-37519.17***	-19710.70	-27840.00	-25680.00
Baths	19030.72***	17723.07***	16616.57***	18988.56***	18410.00***	18650.00***
Condition	5442.27	7472.95	8674.14**	8044.16**	5434.00	4823.00
Garage condition	-9686.32***	-10585.34***	-10936.93***	-11136.00***	-9691.00***	-9479.00***
Age	-261.17**	-188.05*	-150.07	-235.34***	-257.00*	-265.50*
Moran(p-value)	**0.89**	**0.88**	0.63	**0.37**	**0.51**	**0.53**
dh-test (p-value)	**0.01**	**0.02**	0.10	**0.03**	**0.01**	**0.01**
AIC	2.08	2.08	2.00	2.00	17.6	17.6
	Model A	Model G	SAC	SDEM	C-spline	F-spline

#of embedding dimensions m=5
dh-test p-value on raw data 0.11
***, **, * =coefficient estimates that are significant at the 0.01, 0.05 and 0.1 level respectively

2.6. Conclusions

The focus in this chapter has been on cross-sectional spatial data modelization. The chapter contributes, on one hand, to the use of specification tests in order to assess the robustness of the results, once it is recognized that our ability to accurately model spatial data is very limited. On the other hand, it also contributes to the debate around the W weight matrix in formulating spatial econometrics models. There are several challenges when econometrically modeling economic relations for which space can have a relevant role. The combination of missing correlated over space variables along with many sources of potential nonlinearities has led scholars to consider that conventional parametric spatial models are not necessarily the best modelling strategy. We firstly argue that it is critical to use specifications tests in order to validate results. This explains why we propose (and develop on) (i) the use of a nonparametric test for spatial structure of unknown (either linear or nonlinear) form along with other available tests, and (ii) the use of geographically restricted semiparametric models that assume that the true model is unknown. To put it differently, we argue that, supported by powerful statistical tools, semiparametric models and estimators could be used before going through conventional spatial lag models. This chapter has studied hedonic price functions and estimates because housing price formation has been a central object for the implementation of policy decisions and therefore model specification may be particularly important. The three studied cities are independent from each other. Regarding specifications issues, the main conclusion is the following: spatial dependence structure (potentially of unknown form) can be controlled by restricted semiparametric models that do not use W matrix specification, however there might be spatial data relationships for which spatial structure can only be controlled at the cost of assuming a W matrix of connections. Given the potential consequences of a wrong specification, those models that avoid a W matrix are preferred. However, if the modeller requires the use of a weight matrix (as might happen), according to the results, robust specification tests are recommendable to be used to choose the best parametric model. Several lines of further research can be proposed. From a broarder point of view, this paper has contributed to the spatial econometric literature around open concerns for the scholars, in particular to the literature on spatial dependence bias, on functional form bias and on spatial heterogeneity bias. However, other concerns remain open for further research, namely, the potential interaction between geoadditive terms and covariates of particular interest for researchers. From the applied side, this study has centered on housing prices, however there are other well studied cross-sectional datasets (along with their economic relationships) for which space is a natural source of variation and explanation that has kept the attention of geographical modellers, for instance, industry location and knowledge spillovers, among others. In this vein, some models could be revised to double-check and understand the existence of remaining traces of spatial dependence.

3. Comparative Applied Research on Spatial Deterministic trends and Spatial Econometric Structures

Spatial trend concept was proved to be useful in order to depict the systematic variations of the phenomenon concerned over a region based on geographical locations. We use six different geographical datasets to check if there exist potential leading deterministic spatial components and whether we can econometrically model spatial economic relations that might contain unobserved spatial structure of unknown form. Hypothesis testing is conducted with a symbolic-entropy based non-parametric statistical procedure, proposed in Garcia-Cordoba et al. (2019), which does not rely on prior weight matrices assumptions. Geographically restricted semiparametric spatial models are taken from the previous chapter to perform a modeling strategy for cross-sectional data sets. The main question to be responded is whether the models that merely incorporate space coordinates might be sufficient to capture space dependence when applied to different types of data. Moreover, it is important to study what intrinsic characteristics of the economic problem or the dependent variable itself make feasible (and optimal) to use the methodological approach developed in the previous chapter.

The structure of the chapter goes as follows. Part 1 supplement more theoretical points on Spatial Trends and Spatial Econometric structures that we did not mention in the previous chapter. It is important to comment these structures, even though we are not using this methods later on with our data. Part 2 describes in detail the datasets, specific characteristics of each dataset and general conclusions on the analysis of each of it. Finally, Part 3 concludes and opens the path of the possible further research.

3.1. Spatial Trends and Spatial Econometric Structures

The previous chapter was based on explaining the general concept of trends and the main solution that is proposed to deal with spatial data. However, there are more ways to perform a spatial trend analysis, and more spatial econometric structures that should be mentioned. This part presents more theoretic concepts(based onZ (2017)) in the analysis of spatial data.

3.1.1. Spatial Trend Analysis

Any natural phenomenon or its similitude occurs extensively over a region, and therefore, its recordings or observations at different locations pose some questions as, for instance, are there relationships in the form of trends between phenomena in various locations? In such a question, the time is as if it is frozen and the phenomenon concerned is investigated over the area and its behavioral occurrence between the locations. It can be described theoretically, introducing a trend term, but its quantification necessitates objective methodologies.

Three-dimensional statistical techniques help to obtain maps of variable concerned provided that the two geographic coordinates are given at the measurement points. This procedure is referred to as the trend surface analysis in the statistics literature. It is also referred to as the multivariate statistical analysis. Its basis is to match a surface similar to ordinary regression analysis but in three-dimensional space. The same restrictive assumptions as in the ordinary regression analysis are also valid in the trend surface fitting. There are further difficulties in the spatial trend surface search such as the paucity of spatial data and extensive computation requirements. For the success of trend surface fitting uniform data distribution is necessary. The significance of trend surface analysis is to separate the spatial behavior of the phenomenon into two components as the deterministic component in terms of trend surface and the residuals, which are deviations of the measurements from the fitted trend surface and they are the uncertain (random, stochastic) component. The uncertain components are representatives of local sites, whereas the trend surface is the regional behavior of the phenomenon concerned. Trend analysis separates the ReV into two complementary components, namely regional nature of deterministic variations and local fluctuations around the regional component. The regional and local components are dependent on the scale of the ReV. In any trend analysis there are three variables. In any spatial analysis there are three general components for the application of a convenient methodology.

(1) The basis of any spatial analysis is two basic deterministic variables such as easting and northing or longitude and latitude variables that provide locations of measurement variable at a set of locations,

(2) Decomposition of the spatial variable first to a general regional deterministic part, which can be expressed by any mathematical function,

(3) The stochastic (uncertain) part, which constitutes a set of deviations of the measurements from the corresponding trend surface value.

In general spatially variable event may include gradual monotonic trends or even abrupt changes (jumps) due to externally effective phenomenon. It is by now well understood that the global warming leading to climate change imprints an increasing trend into global temperature data. Abrupt changes may also take place as a result of sudden or short duration exogenous impacts, such as volcanic eruptions, earthquakes, sudden changes in monitory rates (devaluation) and alike.

The main purpose of spatial trend analysis is to decompose regional variable into subcomponents such as trends, abrupt changes, stochastic, and entirely independent error terms so that the construction of a suitable model by the synthesis of these components provides opportunity to predict the variable concerned at non-measurement sites.

It is possible to obtain regular grid points from irregular measurement sites by fitting a surface to available data, which can be achieved either globally or locally over the study area. In the former case there is a single functional form of the trend surface in addition to the stochastic nature of the residuals and the latter case is just the repetition of global procedure on pieces of subareas within the study area.

Another version of the local surface fitting is to consider the neighbor points to reach to locally representative trends. However, the most widely used procedure is the global trend surface search, for this purpose the spatial variable is approached by a polynomial expansion of the geographic coordinates, and the coefficients of the polynomial function are estimated from available measurements by means of the least squares method, which relies on the sum of the squared deviations minimization from the trend surface. After the identification of the trend surface the sum of the trend surface value at a site and the residual is equal to the measurement value.

Fixation of a trend surface to given set of spatial variable measurements at irregular sites separates the whole measurements into two components, namely trend and residual values. Trend values are collection of deterministic quantities, but the residuals are uncertainty parts.

So far explained numerical solution of spatial variable through the differential expression rule for evolution of the spatial variable, it is noted that regular and uniform distributions of the coordinate variables are necessary.

Apart from the temporal tendencies there are also spatial trends over a region on the basis of easting (longitude) and northing (latitude), which provides information about regional variability of the phenomenon concerned. For this purpose, it is necessary to have measurements at a set of different locations. Even a single record at each measurement station is enough for spatial trend and variation appreciation. Again visual inspection and assessment of spatial data is recommended prior to the application of any detailed scientific procedure. The initial visual helps to set the foundations of a convenient methodology for the spatial evaluation of data. Prior to any quantitative evaluation of the spatial data at the hand the following points provide assistance. (1) The sampling locations are characterized by coordinates, X and Y, preferably on a scaled map. The spatial variable can be shown by Z. In general, the data locations are irregularly distributed, but in any new study, if possible, data positions are selected better at the nodes of regular nets. The measurement locations may already been such as the existing well locations (water or oil), meteorology stations, urban areas, etc.

The statistical spatial analysis is entirely different from the numerical solution coordinate system. The following points are very specific for statistical spatial analysis and distinct from the mathematical numerical solutions. (1) The measurement locations are irregularly (unevenly) scattered over the study area. For instance if the city centers are thought of a country, their coordinates are not regularly located, and therefore, for mapping purposes it is necessary to reduce the irregularity to a regular mesh, which is the first step in any mapping procedure in software, (2) There is not mathematically known spatial regularity in the spatial records on any natural, environmental, economic and social studies, (3) There may be statistically identifiable spatial regularity within the spatial event, and it is the main purpose to identify linear or nonlinear surface trends, (4) The subtraction of each spatial data value from the corresponding spatial trend value provides a set of shifted values with zero arithmetic average. These are referred to as the residuals or stochastic part.

In the spatial assessment of available data by scientific methodologies it is necessary to make simplifying assumptions and idealizations so as to be able to suggest a valid model for the spatial variation representation. The basic assumptions are homogeneity and isotropy, which can be decided on by comparison of the numerical quantities in a set of spatial measurement sites.

The main purpose is to model the spatial behavior of natural, environmental, and economic phenomena. The trend surface passes through rather uncertain and complex spatial data scatter over a region. Its application can be achieved as geographic information for continuous events in space and the measurements must be at cardinal levels. The basic principle in trend surface analyses is matching a continuous surface to the available spatial data through a regression function. In order to facilitate the spatial trend concept the best example is a topographic map where the independent variables are longitudes and latitudes with spatial variable as altitude (elevation from the mean sea level). For this purpose, the spatial topographic variability must be sampled at n n sites that are irregularly distributed in the study area. To reach to the final trend surface there are following three steps that should be completed in sequence.

(1) Model selection and parameter estimation: If possible, with an expert view, one can guess the most convenient linear or polynomial mathematical form for the spatial trend component of the spatial variable measurements. In doing so one should keep in mind that the trend surface should explain as much as the regional variability in terms of spatial variance.

(2) Model validation: After the model parameter estimations the model (regression) function should be then applied to an independent set of sample points for validation purpose by taking into consideration cross-validation,

(3) Model estimations of spatial variable: The developed model after the execution of the two previous steps, the model now can be used for spatial variable estimations at any desired point within the study area. One should be cautious at this stage to extend (extrapolate) model estimations outside the study area.

The spatial trend surface fit to a set of spatial variable measurements can be achieved by the least squares technique. The surface must be such that it minimizes the variance of the surface with respect to the input values. The fitted surface rarely coincides with some of the measurement points, but it is susceptible to outliers in the data. Trend surface analysis is used to find general tendencies of the sample data, rather than to model a surface precisely and completely.

One can obtain regular grid points from irregular measurement sites by fitting a surface to available data. It can be achieved either globally or locally over the study area. In the former case there is a single functional form of the trend surface in addition to the stochastic nature of the residuals and the latter case is just the repetition of global procedure on pieces of subareas within the study area. Another version of the local surface fitting is to consider the neighbor points to reach to locally representative trends. However, the most widely used procedure is the global trend surface search for this purpose the spatial variable is approached by a polynomial expansion of the geographic coordinates, and the coefficients of the polynomial function are estimated from available measurements by means of the least squares method, which relies on the sum of the squared deviations minimization from the trend surface. After the identification of the trend surface the sum of the trend surface value at a site and the residual is equal to the measurement value.

3.1.2. Spatial Structures

3.1.2.1. Horizontal plane

This corresponds to the case when all the trend surface points have the same (constant) spatial value as in Fig. 6.8. Such a surface in the form of a plane provides homogeneity and isotropy of the spatial variable as for its spatial trend component is concerned.

$$z(x, y) = c$$

Horizontal planes

Within the spatial variable there may not be any trend component but a sudden (abrupt) jump as in Fig. 6.9. In such a case, there are two no spatial trend regions with sudden change (upwards or downwards) in between. The horizontal continuity in Fig. 6.9 is disrupted by a discontinuous feature (cliff, fault, facies change, boundary, etc.)

$$z_u(x,y) = c_{LU} + z_L(x,y)$$

Inclined Trend Plane

This is the most commonly thought and in practical applications frequently employed spatial trend form, which is in the form of an inclined plate(include only linear contributions from the coordinate variables). $z(x,y) = a_0 + a_1 x + a_2 y$

Inclined Trend Planes abrupt change with inclined spatial trend surface together in a spatial data structure(lower and upper trend surfaces)

$$z_L(x,y) = a_0 + a_{1L}x + a_{2L}y y z_U(x,y) = a_0 + a_{1U}x + a_{2U}y$$

Curved Trend Surface

- $z(x,y) = a_0 + a_1 x + a_2 y + a_3 xy$
- $z(x,y) = a_0 + a_1 x + a_2 y + a_3 xy + a_4 x^2 + a_5 y^2$
- $z(x,y) = a_0 + a_1 x + a_2 y + a_3 xy + a_4 x^2 + a_5 y^2 + a_6 y^3 + a_8 xy^2 + a_9 x^2 y$

All the trend surfaces are smooth surfaces, which are generated artificially according to aforementioned mathematical expressions. However, natural surfaces are not in this form, but perhaps it is a mixture of such smooth surfaces piece by piece.

Random Surface

In some cases of the natural, environmental, economic or social spatial data, there may not be any spatial dependence among the measurement values and in this case the surface is in the form or random variations. For instance, rough sea surface is a valid example for such a spatial surface.

Each individual measurement site represents a very considerable area around it. Logically, measurement at any individual site will have an area of influence or in isotropy case a radius of influence around it, but there is no physically or data based objective criterion for the definition of such an area, but one can find the quantitatively its magnitude from SDF, which is an indicator of spatial variable uncertainty (probabilistic, statistical, stochastic) dependence that provides visual and quantitative information about the dependence between any two locations. The dependence can be measured by covariance provided that the uncertainties are distributed according to the Gaussian (normal) PDF, otherwise semivariogram, cumulative semivariogram or point cumulative semivariogram functions should be used. Having a 3D map, one can suppose an existing trend and thus, starting with a 3rd order polynomial and continue to increase the degree of the polynomial up to 7 and select among them the one with the least sum of square residuals squares. we should keep in mind that the higher the degree of the polynomial the rougher gets the surface.

Spatial Correlation Parameter Calculation

1. Find the set of actual distances among each pair of sites, hence, if there are n sites there will be $n(n-1)/2$ different distances,
2. Find the squared differences among each pair of the spatial variable. The same number of squared differences will be obtained,
3. Plot distances versus squared differences of the spatial variable, and hence an irregular graph of the squared differences variation will be reached,
4. Plot the successive cumulative sums of the square distances along the distance sequence. The result will be another graph that shows the change of squared difference accumulation by distance,
5. The last value in this graph is the maximum squared distance summation, and it is very significant for RDF calculation,
6. Divide each cumulative squared difference values in the last graph by this maximum value. The result will be the change of scaled cumulative squared difference values by distance, where the values on the vertical axis changes between zero and one,
7. Finally, subtract scaled cumulative squared differences from one and the resulting graph will appear in the form of decreasing trace (Basile et al. (2014)).

Planer Trend Regression Analysis

In majority of cases a spatial trend is a planer surface and rarely is it in the form of curvature surface of the second order degree usually over geographical (longitude and latitude) directions. X and Y are most of the time longitude (easting) and latitude (northing) directions, whereas the Z direction is for the spatial date values. Linear trend surface analysis is also called as “spatial interpolation” method. The trend surface provides a means of spatial interpolation possibility. The classical trend surface methodology is a way of fitting the entire surface with a linear or polynomial equation with parameters, which are estimable from the given data set by means of the least squares technique. For this purpose, in many works trend surfaces are the only basic tool as maps for communication in any scientific domain spatial variable concerned. After the analysis the trend model statistical model coefficients are estimated and the final product is presented as a contour map, which is the same as the preparation of topographic maps. In the spatial trend analysis methods as presented in this section RDF is not taken into consideration explicitly.

The mathematical form of planar trend surface has the linear contributions of the geographical coordinates, i.e., data measurement point longitude and latitude or easting and northing direction values as x and y. If the spatial variable values are shown by z, then the planar model can be expressed as follows. $z = a_0 + a_1x + a_2y$ This expression represents deterministic trend surface without any measurement error. However, addition of the error (uncertainty) component, e, to this equation gives the representative positions of the data values with uncertainty. The uncertain expression can be written as, $z = a_0 + a_1x + a_2y + \epsilon$. We use OLS, as we want to achieve the minimum of the sum of error terms.

Polynomial Trend Regression Analysis

The polynomial trend surface mathematical formulation relates the geographic variables x and y to the spatial variable value, z as,

$$z(x, y) = a_0 + a_1x + a_2y + a_3xy + a_4x^2 + a_5y^2$$

After the determination of model parameters the positions of local or global maxima and minima points can be obtained after simple algebraic calculations. The locations of local maximum and minimum can be obtained by taking partial derivative with respect to x and y, as, $\frac{\partial z}{\partial x} = a_1 + 2a_3x + a_4y = 0$ and $\frac{\partial z}{\partial y} = a_2 + a_4y + 2a_5y = 0$. This point may show trend surface highest, lowest or inflection point depending on the data features. The point to be cared for in any trend surface analysis is that there must not be model parameter numbers more than the data number. A good practical rule is that there must not be more than one third of data number model parameter.

Polynomial trend analysis is the basis of the methodology we use to apply the delta-test to different types of the datasets. Delta-models specified in the previous chapter are used as the alternative to the classic approach using the spatial matrix. We restrict the function of the geographical position $f(a, b)$ to be a low-degree polynomials of coordinates, which is inspired in the practice in time-series modeling.

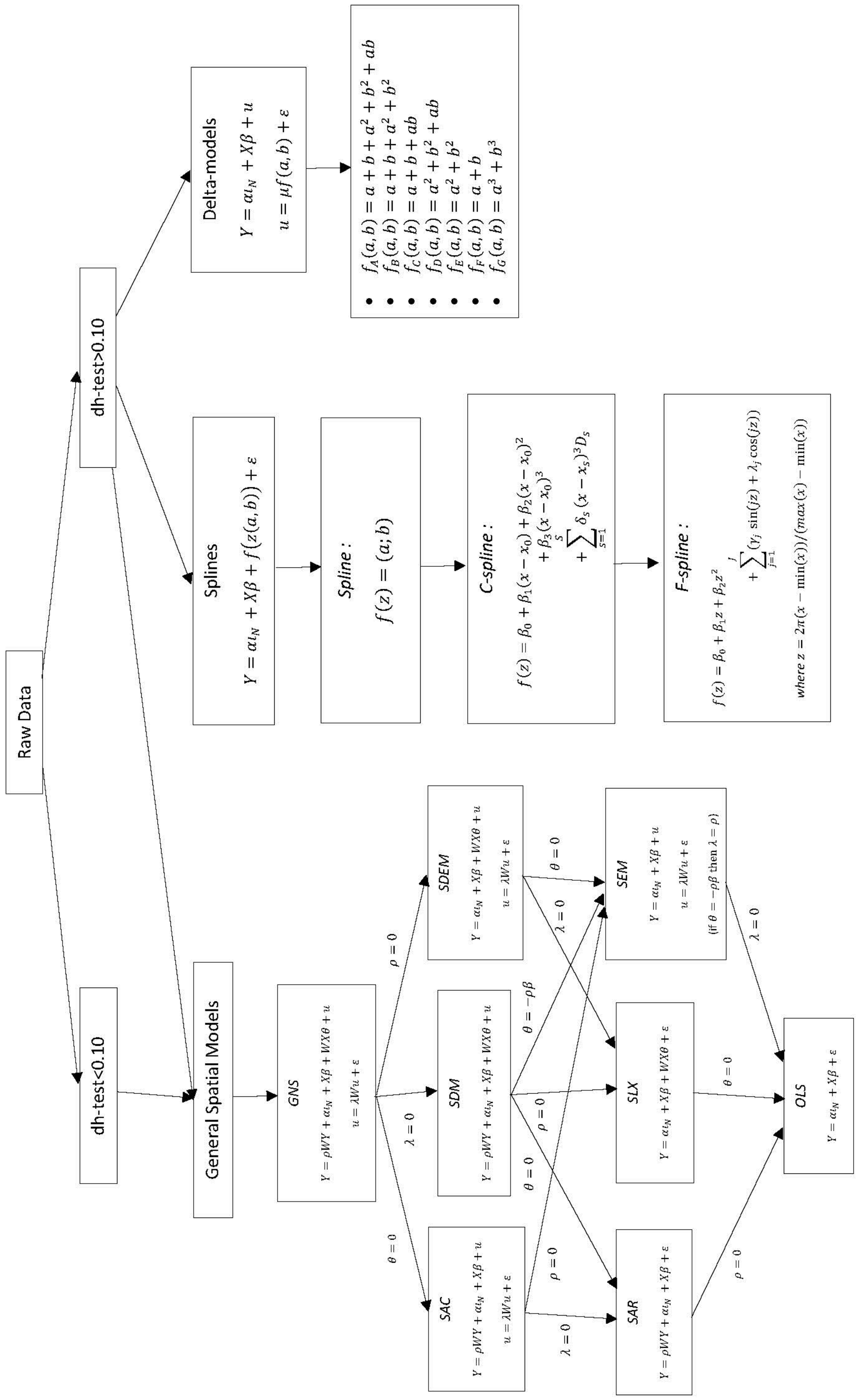
In this way, we can compare different model specifications, both in classic and delta approaches.

3.2. Databases and General results

This section of analysis is based on six different datasets. Each of the dataset includes the full information on the object of the analysis, where a certain relation can be found. Apart from the special characteristics of the units, every dataset includes the information of the geographical position(longitude and latitude) of the units described. Thus, we have a possibility to compare the general characteristics of data analyzed to produce a better methodology of specifying a trend methodology (including a delta test usage). The process of choosing the best model and main steps of the analysis are based on the scheme presented in the Figure3.2.1 [1] As mentioned before, we present the results of only 7 datasets in total, however, more than 15 different datasets with similar characteristics were previously analyzed. Taking into account, that the first step of our analysis reveals the existence(or absence) of the spatial dependence in the raw dataset, using Moran's I test, we have found that only 7 out of 15 datasets presented the existence of spatial autocorrelation in it. The second step is to check the existence of deterministic component in the data, using the dh-test. It might be the case, that the data presents the existence of spatial dependence, but not the deterministic part.[2] Nevertheless, we present all the datasets where the existence of spatial dependence was confirmed. Two of them resulted having no deterministic component, still we use these datasets to additionally analyze probable common characteristics to be taken into account for our further research.

[1] The explicit explanation of the steps and models can be seen in detail in the previous chapter.
[2] Case of Chicago Airbnb and Earthquake datasets.

Figure 3.2.1.: Procedure of choosing a model



3.2.1. Chicago Airbnb prices

This first dataset includes data on the Airbnb prices in Chicago.The data were collected on October 3rd, 2015 and includes 77 observations from 2008 to 2015.It includes response rate, acceptance rate, review rating, price per included guest, room type (1 is entire home/apartment, 2 is private room, and 3 shared room), number of Airbnb spots. The socioeconomic indicators are percentages by community area: households below poverty, housing crowed, under 18 or over 64 years old (dependency), aged 25+ without high school diploma, and unemployed above 16 years old. Also per capita income and hardship index are included . These indicators were built for the period 2008 – 2012. The crime data include the number of crimes (battery, burglary, gambling, homicide, kidnapping, robbery, stalking, homicide, and theft, among others; murders with data for each victim are not included) and thefts from October 2014 to September 2015 (one year before the Airbnb data). Population by community area based on Census 2010 data.

The first step taken was to check if there exists any deterministic component in the data. In this case, the result is negative, there is no deterministic part that could be controlled[3]. Thus, we do not proceed with the whole analysis and pass to the next dataset.

3.2.2. Earthquake

This dataset contains data about the earthquakes that hit the center of Italy between August and November 2016. The data was taken from the National Earthquake Information Center (NEIC), that determines the location and size of all significant earthquakes that occur worldwide and disseminates this information immediately to national and international agencies, scientists, critical facilities, and the general public.

The NEIC compiles and provides to scientists and to the public an extensive seismic database that serves as a foundation for scientific research through the operation of modern digital national and global seismograph networks and cooperative international agreements. The NEIC is the national data center and archive for earthquake information. This dataset includes a record of the date, time, location, depth, mag-

[3] dh-test p-value on raw data 0.02

nitude, and source of every earthquake with a reported magnitude 5.5 or higher since 1965 with 8087 observations.

Same as in the case of the previous dataset, the results of the delta-test shows no evidence of deterministic component in the data structure[4]. Taking this into account, we proceed with the next datasets, where we will be able to take all the necessary steps of the analysis.

3.2.3. NUTS2

Next dataset is analyzed by estimating a number of growth regression models on a sample of 249 NUTS 2 regions belonging to the enlarged Europe (EU 27). We start from the linear specification of the neoclassical growth model proposed byMankiw et al. (1992):The dependent variable is the per-worker income growth rate, $gy = T - 1(lny_T - lny_0)$, computed for the 1990–2004 period. The model predicts that gy is higher in the economies with higher rates of investment in physical and human capital (s_K and s_H, respectively), lower initial conditions, lny_0, and lower effective depreciation rates $(n+g+d)$, with n the working-age population growth rate, g the common exogenous technology growth rate and d the rate of depreciation of physical capital assumed identical in all economies. β_k (with $k = 1, .., 4$) are unknown parameters to be estimated and ε is an error term assumed to be identically and independently distributed (iid). Basic data to measure these variables come from the EUROSTAT Regio and Cambridge Econometrics databases, which include information on real gross value added, employment, investment and tertiary education. We measure per worker income levels, y, as the ratio between total real value added and total employment; the physical capital accumulation rate, s_K, as the average share of gross investments on real gross value added; the human capital accumulation rate, s_H, as the percentage of a region's working population that is in the tertiary level of the education process; n is the average growth of total employment. Finally, we assume, as it is usual, that $(n+g+d)$ is equal to 0.05 (see Mankiw et al. (1992)). Furthermore, as mentioned before, we include the information on longitude and latitude as well to control for the geographical position of the unit analyzed.

[4] dh-test p-value on raw data 0.001

Table 3.1.: Results for NUTS2 Dataset

	SAC	GNS	C-spline	F-spline	Model E	Model G
Constant	0,096***	0,080***	0.055	-0.018	0,098***	0,088**
Human capital	-0,003	-0,003	-0.001	-0.001	0,001	0.001
GDP	0,002**	0,002*	0.001	0.001	-0,001	-0,001
Population Growth	0,019**	0,015*	-0.004	-0.004	0,001	-0,002
Physical Capital	0,022***	0,021***	0.033***	0.033***	0,036	0,036***
Moran(p-value)	0.68	0.72	0.01	0.01	0	0
dh-test(p-value)	0.01	0.02	0.13	0.12	0.11	0.09
	SAC	GNS	C-spline	F-spline	Model E	Model G

#of embedding dimensions m=6
dh-test p-value on raw data 0.15
***, **, * =coefficient estimates that are significant at the 0.01, 0.05 and 0.1 level respectively.

Moran's test on the data on the NUTS2 dataset, gives a clear evidence of the spatial autocorrelation (Table 3.1). The delta-test on the raw data confirms the presence of deterministic structure, that gives an evidence in favor of running restricted semiparametric analysis, including spatial trend. Following the modeling proposal given in the previous Chapter, we firstly model the deterministic part by using the so called delta-models. Results for model G are clearly in favor, as controlling for spatial trend is concerned(p-value =0.09). However, based on the Moran test results we cannot be sure that the estimated model controls for the spatial heterogeneity of the data. The same conclusion is reached if we opt by some spline based methods.

If instead we model according to the classic spatial models, we find that the best spatial models for our data are SAC and GNS models. This conclusion is reached by using the method described in the the previous chapter(an additionally added in the Appendix) and we also present several outputs in Tables in Appendix. One interesting conclusion of the results given in Table 3.1, is that neither SEM nor the SDEM models are able, according to delta-test results, to remove the previously found spatial trend. In other words, the residuals of these models are compatible with a deterministic structure that have not been yet removed. For this reason, results seem to point that restricted semiparametric models work better in this case, as they let us get rid of the spatial structure of the model and thus get more credible results on the estimates. The practical implications for NUTS2 dataset are mainly relative to the partial effects of several explanatory variables, but not to the list of significant variables, nor to the signs. in general.

3.2.4. GECON

Another dataset is based on the G-Econ data. The G-Econ research project is devoted to developing a geophysically based data set on economic activity for the world. Current dataset used in the performed analysis (GEcon 4.0) is now publicly available and covers "gross cell product" for all regions for 1990, 1995, 2000, and 2005 and includes 27,500 terrestrial observations. The basic metric is the regional equivalent of gross domestic product. Gross cell product (GCP) is measured at a 1-degree longitude by 1-degree latitude resolution at a global scale. This dataset includes such characteristics as:

- Gross cell product, 2005 US $ at market exchange rates, 2000
- Distance to coast (km)
- Elevation (km)
- Distance to major navigable lake (km)
- Distance to major navigable river (km)
- Distance to ice-free ocean (km)
- Distance to navigable river (km)
- Vegetation category
- Grid cell population, 2000
- Average precipitation, prior data
- Soil category
- Average temperature, prior data
- Geographical position(Longitude and Latitude)

This dataset is interesting mainly because of the complete information on the geographical characteristics, that might be important when analyzing data with spatial components.

Table 3.2.: Results for G-Econ Dataset

	SAC	GNS	C-spline	F-spline	Model A	Model G
Constant	4,198***	3,252***	2,597***	10,73**	2,664	3,494***
Distance to coast (km)	107,520***	139,548***	-567,8	-549,8	-516,009	-471,904
Distance to coast (km)	-0,004**	-0,004**	0,001	0,001	-0,001	-0,001
Elevation (km)	0,001	0,001	-0,002***	-0,002***	-0,002***	-0,001*
Dist. to mn lake (km)	0,001	0,01***	-0,001***	-0,001***	-0,001	-0,001**
Dist. to mn river (km)	-0,001	-0,01*	-0,001	0,001	0,001	0,001
Dist. to ice-free ocean (km)	-0,108***	-0,139***	0,568	0,055	0,516	0,472
Dist. to navigable river (km)	-0,001***	-0,001***	0,001	0,001	-0,001**	-0,001***
Veg. category	0,018	0,019	0,047*	0,047*	0,063***	0,056***
Grid cell population, 2000	0,001***	0,001***	0,001***	0,001***	0,001***	0,001***
Avg precipitation, prior data	-0,001**	-0,001**	0,001	0,001	0,001	0,001**
Soil category	0,003	0,003	0,003	0,004*	0,002	0,003
Avg temperature, prior data	-0,204***	-0,199***	-0,097***	-0,089***	-0,168***	-0,191***
Moran(p-value)	0.82	0.87	0	0	0	0
dh-test(p-value)	0.36	0.35	0.42	0.43	0.29	0.42
	SAC	GNS	C-spline	F-spline	Model A	Model G

#of embedding dimensions m=6
dh-test p-value on raw data 0.42
***, **, * =coefficient estimates that are significant at the 0.01, 0.05 and 0.1 level respectively.

As in the previous dataset, Moran's test on the data on the GEcon dataset, gives a clear evidence of the spatial autocorrelation (Table 3.2.4). Moreover, the presence of deterministic structure is confirmed by running a delta-test(p-value=0.42). We apply restricted semiparametric analysis, including spatial trend. In this case, the best choice would be GNS and SAC models that correct for spatial structure in the sense that Moran's test statistically indicates that spatial structure has been controlled, see Table 3.2.4. However, none of the delta-models is able to control the deterministic part of the data. As happened with the previous data set, the list of relevant explanatory variables is common to all the models. Variations are again on the partial effects.

3.2.5. California housing prices

Next dataset is the most common dataset on housing prices. Though we have already used the data on housing prices in the previous chapter, it might be useful to compare the characteristics between different type of datasets. This is the dataset used in Geron (2017), that contains information from the 1990 California census and pertains to the houses found in a given California district and some summary statistics about them based on the 1990 census data. The variables we use are as follows:

- Housing median age
- Total room number
- Total bedrooms number
- Population
- Households
- Median income
- Median house value
- Proximity to the ocean(km)
- Geographical position(Longitude and Latitude)

We got the information on the variables in using all the block groups in California from the 1990 Census. In this sample a block group on average includes 1425.5 individuals living in a geographically compact area. Naturally, the geographical area included varies inversely with the population density. We computed distances among the centroids of each block group as measured in latitude and longitude. We excluded all the block groups reporting zero entries for the independent and dependent variables. The final data contained 20640 observations on 9 characteristics.

Table 3.2.5 presents the results of the analysis performed.

Table 3.3.: Results for California housing prices Dataset

	SLX	SAC	C-spline	F-spline	Model D	Model H
Constant	32398,866***	30653,921***	-517600000*	1704000***	-1975883,03***	-660179826,1***
Housing median age	1149,584***	1212,721***	1145,00***	1127,00***	1172,236***	1141,883***
Total room number	-9,949***	-7,288***	-7,97***	-7,85***	-8,067***	-7,071***
Bedroom number	75,520***	53,326***	115,80***	116,1***	117,242***	89,731***
Population	-30,679***	-30,274***	-38,76***	-38,79***	-37,416***	-38,025***
Households	76,439***	84,744***	45,19***	44,04***	40,836***	67,839***
Median income	36562,699***	36141,689***	4025,00***	4016,00***	40327,951***	39785,866***
Moran(p-value)	0	0	0	0	0	0
dh-test(p-value)	0.09	0.05	0.09	0.11	0.09	0.09
	SLX	SAC	C-spline	F-spline	Model D	Model H

#of embedding dimensions m=13
dh-test p-value on raw data 0.11
***, **, * =coefficient estimates that are significant at the 0.01, 0.05 and 0.1 level respectively.

Running the dh-test for the data, we get that there is a spatial trend in the data(dh-test (p-value) =0.11). Following the same steps as before, we can conclude, that neither delta models, nor classic spatial models can control both spatial heterogeneity and deterministic part. However, even we cannot control for spatial heterogeneity, there are several models that do good controlling the deterministic part of the data. Thus, SLX, SAC and GNS are classic models and Model D and Model H are delta models that can control the spatial trend of the data. However, the results of this analysis do not permit us make any conclusions, as none of the models return the clean result, controlling the spatial part of the data.

3.2.6. Texas Airbnb prices

The following dataset has a lot in common with the dataset on Chicago Airbnb prices. However, this database is about Airbnb spots, socioeconomic indicators, and crime by community area in Texas. Sharing economy and vacation rentals are among the hottest topics that has touched millions of lives across the globe. Airbnb has been instrumental in this space and currently operating in more than 191 countries. Hence, we decided to use the available data on Airbnb housing. The dataset we use[5] contains more than 18,000 property listings from Texas, United States. Given below are the data fields:

[5]The Airbnb data was extracted by PromptCloud's Data-as-a-Service solution

- Rate per night
- Number of bedrooms
- City
- Joining month and year
- Geographical position(Longitude and Latitude)
- Property description
- Property title
- Property URL

All these characteristics give some details about the kind of property included in the Airbnb dataset, however, we include only the number of bedrooms and geographical position in our analysis.

Following the same steps as before, we obtain the results below in the Table 3.2.6

Table 3.4.: Results for Texas Airbnb Dataset

	SDEM	GNS	C-spline	F-spline	Model E	Model H
Constant	-123,379***	-116,113***	46224.69**	2315.13***	1212,858***	-1071891,504***
Bedroom number	172,849***	168,827***	172.88***	173.03***	173,408**	172,535***
Moran(p-value)	0	0,30	0	0	0	0
dh-test(p-value)	0.31	0.25	0.39	0.39	0.39	0.39
	SDEM	GNS	C-spline	F-spline	Model E	Model H

#of embedding dimensions m=6
dh-test p-value on raw data 0.41
***, **, * =coefficient estimates that are significant at the 0.01, 0.05 and 0.1 level respectively.

Moran's test on the data on the Airbnb dataset, gives a clear evidence of the spatial autocorrelation (Table 3.2.6). The delta-test on the raw data confirms the presence of deterministic structure(p-value =0.41), that permits us to run a restricted semiparametric analysis, including spatial trend. Following the analysis performed before, we find the following results. None of the delta models can deal with spatial heterogeneity and deterministic part of the data. On the contrary, classic GNS model can control spatial heterogeneity of the data, but not the spatial trend of the data. All other classic and spline models cannot control the spatial part of the data.

3.3. Conclusions

The general contribution of this work is the exploitation of the analysis techniques and how they perform in different environments. Once we recognized the limited ability to accurately model spatial data, it is important to explore how different analysis techniques perform once applied to different types of data. This allows us to make the process of analysis more efficient and precise, when trying to overcome the problems we might face when processing spatial data. We add to the importance of specifications tests usage in order to validate general results. This chapter has studied different types of spatial data to be able to highlight some common characteristics both for datasets where the spatial part is controlled by delta models, and other datasets where classic models perform better when addressing the spatial part.

The general results of the analysis performed allows us to draw some conclusions. First, neither delta models, nor classic spatial models can control the spatial component of the data in all the types of data we have chosen for our analysis. Second, delta models do better with the data that have some specific theoretical model behind, as in the case of NUTS2 data, while classic spatial models perform better with the data that have some detailed geographic information, as in the case of GEcon dataset. Taking into account that the Texas dataset has a lot in common with the datasets we used in the previous chapter, one would expect delta models to do a good job in controlling for a spatial trend, however, the GNS model resulted more efficient. This could be due to the lack of data in this dataset, as adding some more characteristics might help delta models in controlling for the trend. Other datasets have not presented any clear evidence in favour of classic or delta models.

The next steps of our research might include the application of the methodology developed to datasets with more detailed characteristics. Moreover, a step-by-step analysis might be considered, repeating the same analysis when adding characteristics one by one. This is one of the ways to detect crucial characteristics of the observations, that can help us to control both the spatial heterogeneity and spatial deterministic part.

4. Unemployment and personality traits: Is there any connection?

Researchers have become increasingly interested in understanding the relation between the personality of an individual and her labor market characteristics. In this paper, we examine how Big Five personality traits are related to employment status and number of unemployment spells of each individual. Using data from the National Longitudinal Study of Adolescent to Adult Health in the United States, we find that Conscientiousness has a significant negative and Neuroticism has a significant positive effect on the number of the unemployment spells of each person. The employment status of the person has a significant negative dependence on Neuroticism. These dependence are robust to controlling for early life background and other sociodemographic and economic factors.

4.1. Unemployment: Major Issues

Unemployment is a widely studied part of the labor market. In the study of the causes of the transitions from employment to unemployment and vice versa, economists have usually used institutional, educational and socioeconomic variables. The differences between the individuals in these characteristics are referred to as individual heterogeneity. It is regarded as being important in the analysis. However, the major part of the individual heterogeneity, from an econometric point of view, is usually considered as unobservable (Borghans et al. (2008)). However, some unobserved individual characteristics may play an important role in explaining the labor success of the individual. One factor not widely considered in labor economics is the individual personality, which is stated to be *a set of characteristics of behaviors, cognitions, and emotional patterns that evolve from biological and environmental factors* (Corr and Matthews (2009)). We refer to this characteristics as to “non-cognitive skills”, that are any skills that are not cognitive, such as memory, attention, planning, language and intellectual skills. Non-cognitive skills include emotional maturity, empathy, interpersonal skills and verbal and non-verbal communication. While psychologists and sociologists have focused on the analysis of non-cognitive skills, economists have not explored much the importance of personality on the economic outcomes. Education level and socio-demographic background were considered more relevant for labor market success or financial outcomes of the person, than personality traits. Moreover, previously it was difficult to analyze the issue of personality due to the lack of methodology and suitable data.

However, over the last decades, the number of economists interested in non-cognitive skills has increased significantly. They have investigated the way personality may describe financial outcomes, job performance and even unemployment status of the individual. This type of analysis became possible due to the development of Five-Factor model of personality (McCrae and Costa (1985)) which describes personality in terms of five broad factors. During this period of time the views of many personality psychologists have converged regarding the structure and concepts of personality. Thus, the taxonomy, most researchers agreed on is the Big Five personality taxonomy that comprises five personality traits: Extroversion, Openness to new experiences, Conscientiousness, Agreeableness and Neuroticism. Psychologist Lewis Goldberg referred to these as the ‘Big Five’ factors of personality, and developed the International Personality Item Pool (IPIP) - an inventory of descriptive statements relating to each trait. Within each factor, a set of individual traits relate to more specific aspects of personality. This is the methodology we will keep up with while analyzing the individual traits of the subjects of our database.

Previous studies suggest that Big Five traits are linked to certain unemployment (or employment) outcome, such as probability of finding a job, employment status, number of unemployment spells and the cumulative unemployment. The result referred to Extroversion and Conscientiousness appear to be associated with positive employment outcomes, such as higher probability of finding a job, a lower number of unemployment spells, etc. Contrarily, Neuroticism and Openness had negative patterns, increasing the probability of being unemployed and leading to a greater duration of an unemployment spell.

This paper adds to an existing literature a more complete analysis of the American labor market. Our analysis is based on an Add Health panel data that provides information on the individual traits, their occupational status and the number of times of being unemployed. Using the Big Five personality taxonomy mentioned before, our main idea is to evaluate how personality traits are linked to the one employment status, the transition from one status to another and to the number of unemployment spells.

4.2. Previous Research

4.2.1. Labor market

Unemployment occurs when a person who is a participant of the labor force and is actively searching for employment is unable to find a job. High levels of unemployment have been a worldwide problem during the last decades. According to the International Labor Organization[1], decent work deficits remain widespread: the global economy is still not creating enough jobs.

Until recently there has been a tendency to regard unemployment in less developed countries as a symptom of underdevelopment which would disappear as development proceeds. But experience shows that this is not so. On the contrary, countries that are undergoing rapid economic growth are still facing the increasing unemployment. Certain economists state that unemployment is largely a feature of advanced economies, such as USA. Though the U.S. economy, for example, is currently experiencing the longest period of sustained growth, according to Bureau of Labor Statistics in July 2020, the real unemployment rate was 10.2%.(while natural rate of unemployment in U.S. is about 4.1%).

The consequences of unemployment for the individual can be devastating. It is well documented the negative effects on the subjective well-being of the individual and their family and also on mental health. It has been found, that the loss of work generally represents a failure in life and can be extremely harmful to well-being (e.g. Frey and Stutzer (2002)). In addition to the loss of earnings, unemployment represents a loss of purpose and can erode an individual's identity and sense of self-worth (J.B (1995)). Additionally, it can be difficult to recover psychologically from unemployment (Lucas (2004)). Paul, 2009 in his meta-analysis found that

[1] https://www.ilo.org/, General Reports

unemployed persons showed significantly more symptoms of distress and impaired well-being than employed persons did.

The economists tried to understand the differences in the labor market, analyzing whether the background characteristics, personal characteristics and soft skills can influence the employment status of each individual. The analyses performed came to the conclusion that the level of unemployment depends on the sex of the person, its race and on its education and labor experience. One of the studies by Ismail (2011) used the data of graduates of both public and private institutions in Malaysia from 2001 to 2004. Malaysian data shows that in order to get a job, the graduates must have a good command of English and other soft skills such as analytical thinking, intelligence, independence, leadership, communication and computer skills and work experience. The results obtained show that male graduates have more chance of employment than females. Nevertheless, females with a lower education level are more likely to be employed. However, they are rather be employed in the low qualification jobs, such as clerical staff, laborers and operators that do not require any leadership qualities.

Analyzing German labor market, Heineck (2011) found, using data from the German Socioeconomic Panel(SOEP)[2] that unemployment is associated with elder people or not having German nationality (for males), with not having a vocational degree (for females) and living in East Germany. Furthermore, the strongest predictor of the unemployment appears to be the cumulated prior unemployment experience.

Analyzing US job market, Mayer (2010) used the data of CPS for US department of Labor for the 12-month period from July 2010 to June 2011 to analyze a long-term unemployment (more than 99 weeks). The analysis showed that men, older people, married and minorities were more likely than women to be long term unemployed. Almost 40% of the unemployed workers had been employed in construction, leisure and hospitality, or manufacturing. Unemployed workers in construction and manufacturing were more likely than other unemployed workers to experience the long-term unemployment.

As we specified before, racial difference matters as well, when analyzing unemployment. Ritter and Taylor (2011) using the data from the 1979 National Longitudinal Survey of Youth, show that black men have far higher numbers of weeks unem-

[2]The Socioeconomic Panel (SOEP) study is a wide-ranging, nationally representative longitudinal study of private households across Germany that was launched in 1984.

ployed. They also found that there exist a smaller but substantial difference for Hispanic men, and the racial differences in unemployment are somewhat lower for women than for men.

Nunez and Livanos (2010), used micro-data from the 2005 European Union's Labor Force Survey (including 15 member states) and centered mostly on the analysis of educational level and employment. It was found that higher education increases the chances of employment. Similarly, higher education was also found to have a (more moderate) impact on avoiding long-term unemployment. These findings provide a positive view about graduates' employability at a time that many country specific studies suggest the opposite.

Going deeper in the country-specific analysis, Brauns and Steinmann (1999) analyzing the data for 3 countries, Germany, Great Britain and France, found that the results clearly show an inverse relationship between the level of general education achieved and the relative risks of unemployment in all three societies. They show for all three countries that young people's risk of unemployment is strongly related to their educational (non-)achievement. Unemployment rates are typically highest among school-leavers with compulsory education, and lowest among graduates from higher education. Despite substantial cross-country differences in national unemployment rates, the absolute rates faced by the lowest and the highest qualified school-leavers, are fairly similar. This implies that in all three countries, tertiary education provides significant advantages, and compulsory education only major disadvantages with respect to labor market integration.

According to Kettunen (1997) and his analysis of Finnish micro-economic data, education has a positive effect on re-employment probability up to about 13-14 years of education. However, the possibility of getting an acceptable offer decreases toward the highest levels of education. Individuals with a master's, licentiate or doctor's degree have problems in finding acceptable[3] offers.

Theodossiou and Zangelidis (2009) used the data from the European Community Household Panel on six European countries (UK, Finland, Germany, France, Spain and Greece). The results obtained show that, although men and women exhibit overall similar job separation patterns, when the turnover destination is examined men appear to be more mobile across jobs whereas women are more likely to exit to

[3]acceptable offer is a subjective term, that means that the person accepts the job offer even if it does not correspond to the education level or the field.

unemployment. In addition, education is estimated to have a significant impact on turnover decisions, primarily for women. Low educated women have lower job-to-job (JJ) transition probabilities but are more likely to exit to unemployment compared to the other groups, highly educated women and men of both educational categories. All groups have similar mobility behavior, although highly educated men in some cases display higher JJ mobility and lower job-to-un-employment (JUE) turnover probability. Furthermore, unemployment reduces the JJ transition probability of both male and female workers of all educational levels, while the evidence suggests a pro-cyclical response in the JUE transitions of the less-educated males and a counter-cyclical response in the JUE transitions of the less-educated females. Finally, the country-specific analysis of the turnover behavior of men and women suggests that, regardless of institutional and other labor market differences across the six European countries of interest, overall there are remarkable similarities in individuals' labor market mobility.

As far, we have only concentrated on socio-demographic and cognitive characteristics in relation with labor outcomes of the individual. Below, we are going to present some non-cognitive characteristics used in the literature to explain labor outcomes.

In this work we will focus on the Big Five traits to explain the number of unemployment spells the individual has during her labor activity, transitions from employment to unemployment status and vice versa and the probability to be employed in general. In Section 2.2 we present the Big Five traits, the way they are measured and used in the analysis, Section 3 introduces the econometric strategy. Finally, Section 4 presents general results of our analysis and Section 5 concludes and gives some ideas for further research.

4.2.2. Personal characteristics in an unemployment context

Several decades of the research in the field of personality traits made it possible to develop a widely shared taxonomy of traits, known as the Big Five, that is based on a factor analysis of observer and self-reports of behaviors. Big Five traits summarize a large number of distinct, more specific, personality facets(Almlund et al. (2011)). Table 4.2.2 presents the descriptions of each trait.

Table 4.1.: The Big Five Traits

Trait	Definition
Openness to Experience (Intellect)	The tendency to be open to new aesthetic, cultural, or intellectual experiences.
Conscientiousness	The tendency to be organized, responsible, and hardworking.
Extroversion	An orientation of one's interests and energies toward the outer world of people and things rather than the inner world of subjective experience; characterized by positive affect and sociability.
Agreeableness	The tendency to act in a cooperative unselfish manner
Neuroticism (Emotional Stability)	A chronic level of emotional instability and proneness to psychological distress. Emotional stability is predictability and consistency in emotional reaction, with absence of rapid mood changes.

Big Five framework is among the most important psychological characteristics given their predictive power for many consequential economic outcomes(Barrick et al. (2001)). Among the Big Five taxonomy, Neuroticism, as well as Agreeableness, are related to weaker extrinsic career success and job performance (Heineck (2011); Nandi and Nicoletti (2014)). The Big Five personality characteristics that have been related to favorable labor market outcomes include Openness to new experiences (Nandi and Nicoletti (2014)) Conscientiousness and Extroversion (Barrick et al. (2001); Judge et al. (1999);Prevoo and ter Weel (2015)). On the other hand, Extroversion and Neuroticism were found to be negatively related to the wage of the person.(Heineck (2011))

Thanks to a wide range of psychological literature, it became possible to include the individual characteristics into economic analysis. Thus, the effect of personality on earnings, labor status and other economic outcomes, such as financial distress, labor market participation, occupational choices and job seeking has been analyzed.

Barrick and Mount (1991), presented a meta analysis on personality traits and job performance. It stated the consistent result for Conscientiousness, and found that conscious people, independently of the occupation, have higher probability of being employed. However, Extroversion was a valid predictor only for managers and sellers, as this kind of job requires high empathy and interaction with other individuals.

De Fruyt and Mervielde (1999) using the data on Belgium students from different faculties of the University of Ghent and from all sections of the Industrial Engineering School in Ghent, obtained a strong evidence on the fact that Extroversion and Conscientiousness were valid predictors of the employability of the individual, while Neuroticism and Openness predicted the unemployment. Same results on Conscientiousness were found by Egan et al. (2017) in the analysis of adolescent personality and unemployment using the British cohort study[4].

According to the study of Jyväskylä Longitudinal Study of Personality and Social Development(JYLS) in Finland by Viinikainen and Kokko (2012), Openness is related to an increased number of unemployment spells, and an increased cumulative unemployment at the prime working age. But they stated that the reason of this might be because individuals with higher Openness enter into unemployment spells more frequently – not because their unemployment spells would be particularly long. Extroversion and Agreeableness were associated with reduced cumulative unemployment and Extroversion with a reduced number of unemployment spells. What is more, Neuroticism was associated with a decreased probability of unemployment exit, meaning longer duration of single unemployment spells.

Uysal and Pohlmeier (2011) using the German SOEP database, found that Conscientiousness and Neuroticism have a strong impact on the instantaneous probability of finding a job, where the former has a positive effect and the latter has a negative effect. The direction of the effect on the subsequent employment duration is the opposite. Meanwhile, Openness eases finding a job only for female unemployed workers or those with migration background. Later on, Cuesta and Budria-Rodriguez (2012) analyses this dataset with different techniques and find that individuals with high Extroversion and Agreeableness are more likely to be unemployed. Engelhardt (2017) using the same database, focus their analysis only on Conscientiousness and Agreeableness and their relation to employability. The results obtained showed that individuals with low scores in the dimensions of Conscientiousness and Agreeableness have a higher probability of being unemployed, longer unemployment duration, and experience more status changes between employment and unemployment. Results suggests that personality is an important determinant of women's risk of unemployment, but for men personality is more a matter of job keeping.

[4]BCS is a nationally-representative study of 17,000 children born in Britain in a single week in 1970, contains self-reported personality measures at age 16-17 and month-by-month employment data spanning January 1986 to April 2009.

Identifying the psychological characteristics that help people find and retain employment could help us get a better understanding of the differences in the labor market on an individual level. Our next step will be analyzing which traits influence the employment outcomes in the American labor market.

4.3. Data and descriptive statistics

4.3.1. Data

We use the data from the restricted-use version of the National Longitudinal Study of Adolescent to Adult Health (Add Health) database that studies social, economic, psychological and physical situation of a nationally representative sample of adolescents in the United States during 5 different waves of the Study (1994-2018). Wave I questionnaire was administered in 1994-1995 to a representative group of 20745 students in grades 7 to 12. It was followed by a wide range of in-home interviews in 1996(Wave II), 2001-2002(Wave III), 2008(Wave IV) and 2016-2018(Wave V). Our major interest lies mostly in the Wave IV data that includes the information on the individual's labor situation that we will use as the dependent variables in our analysis. In our study we will use the employment status of the person, the number of times of being unemployed, and the individual transition in the labor market. Besides the individual information on the person (i.e. gender, age, race), its family characteristics and its position in the labor market, it includes a wide range of questions on the psychological characteristics of each person. The answers on the questionnaire on personal characteristics presented in Wave IV allowed us to construct Big Five traits of each person to be used in the analysis. The information on psychological characteristics is only present in Wave IV database, thus we assume that Big Five characteristics do not change over time, as has been documented in different Cobb-Clark and Schurer (2012). Wave V is used to compare the labor situation of the person with the one in Wave IV and its transition in the labor market. However, we use the information from Wave I to control for basic characteristics of the person, such as race, gender, health issues(ADHD)[5], etc. to include them in the analysis together with the variables of the socio-demographic and economic background. Combining the datasets for each of the three waves, we

[5]General health situation is analyzed for Wave IV and Wave V separately and is taken from a subjective questionnaire: In general, how is your health?

get a full database of about 10914 individuals with the wide range of labor, personal and economic information.

As we already mentioned, Wave IV data is used to get the information to create the variables on Big Five personality traits, that we will use in our analysis. We use a 20-item short-form version of the International Personality Item Pool-Five-Factor Model (i.e., the Mini-IPIP, Donnellan et al. (2006)). The detailed procedure of creating Big Five variables and the corresponding questionnaire can be found in the Appendix.

Data from Wave V is used to complete the analysis of the labor situation of each individual. We create 4 different transition dummy variables(dependent variables) to check if the individual:

- was employed in Wave IV and is employed in Wave V
- was employed in Wave IV and is unemployed in Wave V
- was unemployed in Wave IV and is employed in Wave V
- was unemployed in Wave IV and is unemployed in Wave V.

Apart from that, we take into account the occupation of the person, that capture 24 categories from the Standard Occupational Classification (SOC)[6] codes of employment. The general description of every occupation can be found in the Appendix.

4.3.2. Descriptive Statistics

Table 4.2 reports the summary statistics on the dependent variables we are going to use about the demographic, socioeconomic status and early life background factors of the young adults in the Add Health sample in Wave I, Wave IV and Wave V . Dependent variables that are going to be used in our analysis are: number of unemployment spells, employment status(employed-unemployed) in Wave IV and Wave V, and the transition variables we mentioned before. We see that above 94%

[6]https://www.bls.gov/soc/
The 2018 Standard Occupational Classification (SOC) system is a federal statistical standard used by federal agencies to classify workers into occupational categories for the purpose of collecting, calculating, or disseminating data. All workers are classified into one of 867 detailed occupations according to their occupational definition. To facilitate classification, detailed occupations are combined to form 459 broad occupations, 98 minor groups, and 23 major groups. Detailed occupations in the SOC with similar job duties, and in some cases skills, education, and/or training, are grouped together.

of the population were employed in Wave IV and the same percentage in Wave V. Additionally, above 72% were employed in both Wave IV and Wave V, that means, that their employment status during these two years(from 2008 to 2016) did not change. It might be a case, that the person was unemployed somewhere during this period, though we are not able to control this.

Table 4.2.: Descriptive Statistics (Dependent variables)

	Wave IV	**Wave V**
Dependent variables		
Number of unemployment spells	0.49 (1.52)	–
Employed	0.94 (0.24)	0.94 (0.24)
Transition variables		
Employment-Employment	–	0, 72 (0.45)
Unemployment-Unemployment	–	0.01 (0.09)
Unemployment-Employment	–	0.04 (0.20)
Employment-Unemployment	–	0.04 (0.20)
Observations	10,914	10,914

Std.Errors given in parenthesis.

Table 4.3 presents the descriptive statistics on personality traits and general demographic controls we are going to use in our analysis. We observe, that the individuals on average score high in Conscientiousness and Agreeableness. The average age of the sample of the Wave V was 37.94 years old. Our data has less men than women (43%), above half of the respondents had ever been married by the time of Wave IV survey(51.29%) and more than 60% by the time of Wave V. Broken down by race and ethnic origin, 66% were white, 21% Black, 14% Hispanic, 6% Asian, and 8% of other race. However, in our analysis we divide the population into two subgroups, white and non-white population. By the time of Wave V, most of the individuals(68%) were college graduates and about 28% graduated from high school.

Table 4.3.: Descriptive Statistics (Explanatory variables)

	Wave I	**Wave IV**	**Wave V**
Personality traits			
Conscientiousness	–	3.68 (0.68)	–
Neuroticism	–	2.59 (0.69)	–
Extroversion	–	3.30 (0.77)	–
Agreeableness	–	3.84 (0.59)	–
Openness	–	3.63 (0.61)	–
Demographic controls			
Female(%)	57.56	–	–
White(%)	66.18	–	–
Black(%)	20.90	–	–
Hispanic(%)	14.53	–	–
Asian(%)	6.57	–	–
Other race(%)	8.28	–	–
Age	16.04 (1.78)	29.05 (1.75)	37.94 (1.89)
Ever married by Wave(%)	0.55	51.29	69.82
Urban(%)	33.37	–	–
Siblings	–	2.91 (2.35)	2.79 (2.09)
Education			
Less than high school(%)	–	6.26	4.25
High school graduate(%)	–	57.90	27.89
College graduate(%)	–	35.84	67.86
Observations	10,914	10,914	10,914

Std.Errors given in parenthesis.

Regarding personal income, the individual average in Wave IV was in the salary range from \$25.000 to \$30.000, while in Wave V went up to the section from \$30.000 to \$40.000 thousand dollars. The detailed statistics on more descriptive variables can be found in the Appendix.

As we mentioned in the previous part, Barrick and Mount (1991) found that the validity of a personal characteristic can vary depending on the occupation of the person. First, we would like to see if there exist any differences in personal characteristics depending on the occupation of the person. Figure 4.3.1 presents the differences in Big 5 characteristics for different occupations. Thus, we can see that people that work in management, business and social services score higher in Agreeableness, but get rather low score in Neuroticism.[7]This result goes in line with the ones presented in Mount (1998). It stated that the situation in which Agreeableness appears to have high predictive validity is in jobs that involve considerable interpersonal interaction, particularly when the interaction involves helping, cooperating and nurturing others. In fact, in those settings, Agreeableness may be the single best personality predictor. At the same time, individuals who work in management and business operations score highly in Conscientiousness and Extroversion, however score rather low on Neuroticism. The results we got coincides with the results by (Barrick and Mount, 1991; Mount (1998)) where, Extroversion has been found to be related to job performance in occupations where social interactions is an important function of the job. In such jobs, such as sales and management, being sociable, gregarious, assertive, energetic and ambitious is likely to contribute to success on the job. Moreover, according to Coenen et al. (2021), “higher Openness to Experience, lower Extroversion, lower Neuroticism, and lower Agreeableness are related to stronger preferences and specialization towards STEM[8]”. These findings go in line with the results we obtain from our data, as can be seen from the tables below.

[7]Our data include occupational groups that have same working characteristics(i.e. managers and business operations, personal care and health support) that perform no statistical differences in the results obtained. The differences emphasized in text are statistically significant.

[8]Science, Technology, Engineering and Mathematics

Figure 4.3.1.: Big 5 characteristics by occupation

General groups taken from the SOC classification system.

Descriptive analysis in Table 4.4 shows that males have higher scores on Openness, while females get higher scores on Conscientiousness, Agreeableness, Neuroticism and Extroversion.

Table 4.4.: Big Five Comparative Statistics (gender)

	Male	**Female**	**t-test**	**Wilcoxon rank sum test**
Openness	3.73 (0.62)	3.56 (0.60)	0.00[9],[10]	0.00
Conscientiousness	3.61 (0.66)	3.73 (0.69)	0.00	0.00
Extroversion	3.27 (0.77)	3.33 (0.77)	0.01	0.00
Agreeableness	3.66 (0.61)	3.97 (0.55)	0.00	0.00
Neuroticism	2.43 (0.66)	2.71 (0.69)	0.00	0.00
Number of individuals	4,629	6,272		

Std.Errors given in parenthesis.

The comparison of unemployed and employed individual in both Waves (IV and V) shows that unemployed individuals score statistically less on each of the Big 5 traits, except on Neuroticism, (see Table 4.5 for detailed comparative summary statistics on the personality traits).

Table 4.5.: Big Five Comparative Statistics (Employment status)

	Unemployed		**Employed**		**t-test**		**Wilcoxon rank sum test**	
	IV	**V**	**IV**	**V**	**IV**	**V**	**IV**	**V**
Openness	3.60 (0.62)	3.56 (0.58)	3.64 (0.61)	3.65 (0.61)	0.11	0.01	0.05	0.01
Conscientiousness	3.59 (0.69)	3.62 (0.66)	3.68 (0.67)	3.69 (0.67)	0.01	0.01	0.01	0.01
Extroversion	3.25 (0.77)	3.20 (0.79)	3.32 (0.77)	3.32 (0.77)	0.04	0.01	0.03	0.00
Agreeableness	3.75 (0.60)	3.71 (0.63)	3.84 (0.59)	3.84 (0.59)	0.01	0.00	0.01	0.00
Neuroticism	2.79 (0.71)	2.79 (0.71)	2.56 (0.68)	2.56 (0.68)	0.00	0.00	0.00	0.00
Number of individuals	585	586	8,895	9,128				

Std.Errors given in parenthesis.

4.4. Empirical methods

In this section we study the relationship between personality traits and labor variables. Taking into account the results of the previous studies, we assume that the unemployment outcomes on the American market and unemployment status are highly dependent not only on the background of the person, but also on its personal

[9]The results for t-test and Wilcoxon test represents the p-value that was obtained for this tests.
[10]We use two-tailed tests in our analysis.

characteristics. In this case we analyze how the differences in Big Five characteristics can influence the number of the unemployment spells that a person has during her labor life and its employment status in different periods of time. Moreover, we analyze how personal characteristics can affect the probability of staying (un)employed, as well as changing it from one period of time to another(e.g. being employed before and being unemployed now and vice versa).

In our analysis the dependent variable y_i is a binary outcome variable(transition variable from employment to unemployment or vice versa, having an unemployment spell, etc.). To deal with a binary dependent variable, we will use the Probit model. The model

$$y_i = \beta_0 + \beta_n(\text{personality}_{ni}) + \gamma_k X_{ki} + \varepsilon_i$$

, where

- y_i is the dependent variable, that differs depending on the stage of analysis performed. We will describe the details of each analysis in the next section before presenting the results.
- **personality**$_{ni}$ includes Big Five characteristics, for $n = 1, ..., 5$ for each of the personal characteristic measured by this test
- X_{ki} is the set of the background characteristics of the person, that we mentioned before, age, gender, race, etc. The full list of background characteristics can be found in the Appendix.

To analyze the probability of having a particular number of unemployment spells for each individual, we will use a generalization of Probit model, so-called Ordered Probit model. It is applied to the cases, when an ordinal dependent variable has more than two outcomes. In our case, the individual can have from 0 to 50 of unemployment spells. An index model for a single latent variable y^* is

$$y_i^* = \beta X_i + \epsilon_i$$
$$y_i = j \text{ if } \alpha_{j-1} < y_i^* < \alpha_j$$

We make things simple here, as X_i include personality characteristics and other explanatory variables to make a notation more simple.

Thus, the probability that the individual i will have a number of unemployment spells j is

$$P_{ij} = P(y_i = j) = P(\alpha_{j-1} < y_i^* < \alpha_j) = \Phi(\alpha_j - \beta X_i) - \Phi(\alpha_{j-1} - \beta X_i)$$

In the next section we will present the results obtained using these approaches.

4.5. General Results

In this section, we will present the effect of personal traits on the labor market outcomes using three different approaches. First, we analyze how personal characteristics can influence the possibility of being employed(unemployed) in each of the Waves (Wave IV and Wave V). The second step is to analyze the probability to stay employed (unemployed) in both Wave IV and Wave V and the probability of changing the status, depending on the personal characteristics of each individual. Next, we will analyze the probability of having at least one unemployment spell[11].And finally, we will analyze whether personal characteristics influence on the number of unemployment spells of the person. We have 4 different sets of background characteristics that we include as the control variables in the analysis on different levels:

- Basic: Gender, age, race, education, living in urban area, civil status of the person
- Math note+Health: Math note score, if ever had been diagnosed with ADHD[12] and depression.
- Income+Number of siblings: number of siblings and personal income.

Unemployment status

Our first step is to check whether Big Five traits influence the unemployment status of the person. The dependent variable is the dummy that is equal to 1 when the person is employed and 0 otherwise (keeps for both Wave IV and Wave V). Table 4.6 shows the results on Probit regression. From the first column of the table we can

[11] Dummy variable, that is equal to 0 if the person has never been fired or left the job, and equals to 1 otherwise.
[12] ADHD-Attention deficit hyperactivity disorder

see that Conscientiousness and Agreeableness have a significant positive effect on the probability of being employed in Wave IV, while Neuroticism has a significant negative effect. However, only Neuroticism is robust to the inclusion of control variables. Thus, we can conclude that Neuroticism has a negative and significant effect on the probability of being employed in Wave IV. Every additional point that a person score in Neuroticism lowers the probability of being employed by 11% (all controls included).

Table 4.6.: Employed in Wave IV

	(1)	**(2)**	**(3)**	**(4)**
Conscientiousness	0.06^{*}	0.04	0.01	-0.02
	(0.03)	(0.03)	(0.04)	(0.04)
Extroversion	0.01	0.03	0.03	-0.02
	(0.03)	(0.03)	(0.03)	(0.04)
Agreeableness	0.10^{***}	-0.01	-0.01	$-0,06$
	(0.04)	(0.04)	(0.04)	(0.05)
Neuroticism	-0.21^{***}	-0.17^{***}	-0.14^{***}	-0.11^{***}
	(0.03)	(0.03)	(0.04)	(0.04)
Openness	-0.01	-0.05	-0.05	-0.04
	(0.04)	(0.04)	(0.04)	(0.05)
Controls: Basic		X	X	X
Controls: Basic+Math Note+Health			X	X
Controls: All Controls				X
Observations	9,442	8,704	8,539	8,202

Std.errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The same results are obtained for Wave V. When we only include Big Five traits as explanatory variables, Agreeableness has a significant positive and Neuroticism has a significant negative effect on the probability of being employed in Wave V. Nonetheless, we see that only Neuroticism keeps its significance when we add more control variables in the regression, but it looses its significance when we add the data on personal income. As we can see from the detailed table in the Appendix, personal income has a positive and significant effect on the probability of being employed in Wave V. The probability of being employed when having an additional point in Neuroticism is lower than in Wave IV.

Table 4.7.: Employed in Wave V

	(1)	**(2)**	**(3)**	**(4)**
Conscientiousness	0.03 (0.03)	0.02 (0.04)	-0.04 (0.04)	-0.04 (0.04)
Extroversion	0.03 (0.03)	0.03 (0.03)	0.02 (0.03)	0.01 (0.04)
Agreeableness	0.13^{***} (0.04)	0.02 (0.05)	0.01 (0.05)	-0.05 (0.05)
Neuroticism	-0.20^{***} (0.03)	-0.13^{***} (0.04)	-0.08^{**} (0.04)	-0.01 (0.04)
Openness	0.04 (0.04)	-0.01 (0.05)	0.01 (0.05)	0.02 (0.05)
Controls: Basic		X	X	X
Controls: Basic+Math Note+Health			X	X
Controls: All Controls				X
Observations	9,679	7,291	7,143	7,080

Std.errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Moreover, from the detailed table in the Appendix, we can highlight the same patterns for both Wave IV and Wave V in terms of the background characteristics. For example, education, race and the fact of being married have a positive significant effect on the probability of being employed. While the subjective estimation of its own health has a significant negative effect.

Transitions in labor market

The next analysis we made is based on the transition variables. When we analyze the probability of staying unemployed in both waves (Table 4.8), we see that Agreeableness has a significant positive effect, while Neuroticism has a negative effect.

Table 4.8.: Staying Unemployed

	(1)	**(2)**	**(4)**	**(6)**
Conscientiousness	-0.09	-0.07	-0.05	-0.01
	(0.06)	(0.07)	(0.07)	(0.09)
Extroversion	0.01	0.03	0.05	0.01
	(0.05)	(0.06)	(0.06)	(0.08)
Agreeableness	-0.16^{**}	0.05	0.07	0.06
	(0.06)	(0.08)	(0.09)	(0.10)
Neuroticism	0.24^{***}	0.22^{***}	0.19^{***}	0.09
	(0.06)	(0.07)	(0.07)	(0.08)
Openness	-0.05	-0.04	-0.05	0.01
	(0.07)	(0.08)	(0.08)	(0.10)
Controls: Basic		X	X	X
Controls: Basic+Math Note+Health			X	X
Controls: All Controls				X
Observations	10,853	8,177	8,004	7,363

Std.errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Focusing now in the analysis of the transition from employment in Wave IV to unemployment in Wave V we observe from Table 4.9, that Agreeableness appears to have a significant negative and Neuroticism a significant positive effect, but both loose their significance when control variables are added in the analysis. Thus, this results are not robust.

Table 4.9.: Transition from Employment to Unemployment

	(1)	**(2)**	**(3)**	**(4)**
Conscientiousness	-0.01	-0.01	0.05	0.07
	(0.04)	(0.04)	(0.05)	(0.05)
Extroversion	-0.03	-0.03	-0.02	-0.03
	(0.03)	(0.04)	(0.04)	(0.05)
Agreeableness	-0.13^{***}	-0.06	-0.05	-0.02
	(0.04)	(0.05)	(0.06)	(0.07)
Neuroticism	0.11^{***}	0.04	0.01	-0.07
	(0.04)	(0.05)	(0.05)	(0.05)
Openness	-0.03	0.01	-0.02	-0.02
	(0.04)	(0.05)	(0.06)	(0.06)
Controls: Basic		X	X	X
Controls: Basic+Math Note+Health			X	X
Controls: All Controls				X
Observations	8,718	6,642	6,520	6,040

Std.errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4.10 presents the results of the analysis on the transition variable for staying employed(employed to employed). When the regression only includes Big Five traits as control variables, we find that Conscientiousness, Extroversion and Openness have a significant positive effect on the probability of being employed in both waves, while Neuroticism affects it negatively. Yet, when we include controls on the background of the person, Conscientiousness and openness loose their significance. Neuroticism and extroversion keep their opposite signs and significance. It looses its significance once the information on the personal income is included as the explanatory variable.

Table 4.10.: Staying Employed

	(1)	**(2)**	**(3)**	**(4)**
Conscientiousness	0.03^{*}	0.04	-0.01	-0.03
	(0.02)	(0.02)	(0.02)	(0.03)
Extroversion	0.05^{***}	0.05^{**}	0.05^{**}	0.03
	(0.02)	(0.02)	(0.02)	(0.02)
Agreeableness	-0.03	-0.02	-0.03	-0.02
	(0.02)	(0.03)	(0.03)	(0.03)
Neuroticism	-0.18^{***}	-0.10^{***}	-0.06^{**}	-0.02
	(0.02)	(0.02)	(0.02)	(0.03)
Openness	0.06^{***}	-0.02	-0.02	-0.01
	(0.02)	(0.03)	(0.03)	(0.03)
Controls: Basic		X	X	X
Controls: Basic+Math Note+Health			X	X
Controls: All Controls				X
Observations	10,853	8,177	8,004	7,363

Std.errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Analyzing transition from unemployment in Wave IV to employment in Wave V, we find that Neuroticism is the only personal trait that is significant and it has positive sign (Table 4.11). Those individuals with higher scores in Neuroticism are more likely to become employed in Wave V, than other individuals. The way Neuroticism influence the probability of being employed or unemployed might seem contradictory, as we find that neurotic people tend to loose their job more often, but at the same time, in line with the literature, they tend to find their job faster. According toFrey and Stutzer (2002) , neurotic people have a tendency to quick action, that once being unemployed helps them to find a new place faster. However, employees prefer more stable workers on a long-term basis, as “quick actions” sometimes can lead to wrong decisions on a work place.(Le Gallo, 2000)

Table 4.11.: Transition from Unemployment to Employment

	(1)	**(2)**	**(3)**	**(4)**
Conscientiousness	-0.02 (0.04)	-0.05 (0.04)	-0.02 (0.04)	-0.03 (0.05)
Extroversion	-0.01 (0.03)	0.01 (0.04)	-0.01 (0.04)	0.01 (0.04)
Agreeableness	-0.07 (0.04)	0.02 (0.05)	0.01 (0.06)	0.04 (0.06)
Neuroticism	0.14*** (0.04)	0.13*** (0.04)	0.12*** (0.04)	0.09* (0.05)
Openness	0.07* (0.04)	0.08 (0.05)	0.06 (0.05)	0.06 (0.06)
Controls: Basic		X	X	X
Controls: Basic+Math Note+Health			X	X
Controls: All Controls				X
Observations	8,718	6,642	6,520	6,040

Std.errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Being in an unemployment spell

When we run a simple Probit regression taking as a dependent variable a dummy that is equal to 1 if the person has ever been in an unemployment spell and 0 otherwise, we get the same results as before. When including only Big Five traits as the explanatory variables, Conscientiousness and Agreeableness have a significant negative effect on the probability of having at least one unemployment spell(probability of being fired or leave a job at least once), while other personal traits have significant positive effect. The results are robust when including other controls for all the personal characteristics, except for the Agreeableness.(Table 4.12).

Table 4.12.: One or more unemployment spells (Probit)

	(1)	**(2)**	**(3)**	**(4)**
Conscientiousness	-0.12^{***} (0.02)	-0.11^{***} (0.02)	-0.09^{***} (0.02)	-0.09^{***} (0.02)
Extroversion	0.07^{***} (0.02)	0.05^{***} (0.02)	0.05^{***} (0.02)	0.06^{***} (0.02)
Agreeableness	-0.17^{***} (0.02)	-0.03 (0.03)	-0.02 (0.03)	-0.03 (0.03)
Neuroticism	0.16^{***} (0.02)	0.15^{***} (0.02)	0.13^{***} (0.02)	0.12^{***} (0.02)
Openness	0.13^{***} (0.0)	0.12^{***} (0.03)	0.11^{***} (0.03)	0.12^{***} (0.03)
Controls: Basic		X	X	X
Controls: Basic+Math Note+Health			X	X
Controls: All Controls				X
Observations	10,562	9,727	9,530	9,160

Std.errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Number of unemployment spells

The results of the next step of the analysis reflected in Table4.13 state that each of the Big Five traits has a significant impact on the number of unemployment spells(dependent variable). As one can see, Conscientiousness and Agreeableness have a significant negative effect, while other traits have a significant positive effect on the number of times the person has been unemployed. However, we notice that Agreeableness looses its significance when other control variables are added.

Table 4.13.: Number of unemployment spells (Ordered Probit)

	(1)	**(2)**	**(3)**	**(4)**
Conscientiousness	-0.12^{***} (0.02)	-0.10^{***} (0.02)	-0.08^{***} (0.02)	-0.08^{***} (0.02)
Extroversion	0.06^{***} (0.02)	0.04^{**} (0.02)	0.04^{**} (0.02)	0.05^{***} (0.19)
Agreeableness	-0.18^{***} (0.02)	-0.03 (0.02)	-0.03 (0.03)	-0.04 (0.03)
Neuroticism	0.15^{***} (0.02)	0.14^{***} (0.02)	0.12^{***} (0.02)	0.10^{***} (0.21)
Openness	0.13^{***} (0.02)	0.12^{***} (0.02)	0.11^{***} (0.02)	0.11^{***} (0.02)
Controls: Basic		X	X	X
Controls: Basic+Math Note+Health			X	X
Controls: All Controls				X
Observations	10,562	9,727	9,530	9,160

Std.errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

To confirm the results obtained, we use a Poisson regression. We use a Poisson distribution following Cameron et al., 2009 to control for a violation of the distribution assumption that the variance equals the mean. Table 4.14 shows that the results we previously obtained are robust, however, in this case, Agreeableness does not loose its significance, even when all the control variables are added to the analysis. Apart from the results on the Big Five characteristics the general results on the analysis of the unemployment spells show us that education and race have significant negative effect on the number of unemployment spells, as long as gender, health level and and ADHD status have positive significant effect. That means that, for example, those that reported their health as poor are more likely to be fired(have more unemployment spells).

Table 4.14.: Number of unemployment spells (Poisson)

	(1)	**(2)**	**(3)**	**(4)**
Conscientiousness	-0.14^{***} (0.02)	-0.11^{***} (0.02)	-0.08^{***} (0.02)	-0.08^{***} (0.02)
Extroversion	0.10^{***} (0.02)	0.08^{***} (0.02)	0.07^{***} (0.02)	0.08^{***} (0.02)
Agreeableness	-0.38^{***} (0.02)	-0.12^{***} (0.03)	-0.11^{***} (0.03)	-0.12^{***} (0.03)
Neuroticism	0.20^{***} (0.02)	0.21^{***} (0.02)	0.17^{***} (0.02)	0.14^{***} (0.02)
Openness	0.15^{***} (0.02)	0.09^{***} (0.03)	0.06^{**} (0.03)	0.06^{**} (0.03)
Controls: Basic		X	X	X
Controls: Basic+Math Note+Health			X	X
Controls: All Controls				X
Observations	10,562	9,727	9,530	9,160

Std.errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Our analysis shows that the best predictor of the unemployment status is the level of Neuroticism of an individual. The more neurotic is the person, the less is the probability of being employed in general. Apart from that, Conscientiousness has a significant negative effect on the probability of having at least one unemployment spell(probability of being fired or leave a job at least once) and on the number of an unemployment spells. Thus, more conscious people tend to have less unemployment spells and have lower probability of being fired or loose the job. Other personal traits, such as Neuroticism, Openness and Extroversion have a significant positive effect on the number of unemployment spells. It can be easily explained by psychological meaning of every trait. More neurotic people tend to loose the job more often, due to the bad decisions they can perform. Moreover, Neuroticism is connected to Openness, as neurotic people, as well as people that are opened to new experience tend to change their place of work more often. Finally, extroverted people take a decision to change their place of work or sometimes even job, more often than others.

From the detailed results presented in the Appendix, one can see that for transition variables the general trend is as follows. First, education has positive effect when the transition to employment is concerned and negative in transitions to unemployment. The opposite result keeps for health variable. Another important factor in analyzing the unemployment status is the personal income of the individual. Thus, the higher is the personal income of the person, the lower is the number of unemployment spells and the lower the probability of becoming unemployed. The subjective estimation of proper health as poor increases the probability of transition to unemployment. Other results cannot be presented as robust ones, as they differ depending on the transition variables, that can be easily explained by psychological reasons.

4.6. Conclusions and Next Steps

Unemployment is an event that can be suffered by most people. The psychological consequences of unemployment have been researched extensively. However, previous research into unemployment has not been investigated in relation to individual differences. Our main goal in this analysis has been to examine the role of personality traits in determining the success of unemployed workers in finding a job and their success in the labour market, measured by a number of unemployment spells. We can show that personality traits are major determinants of job transitions and unemployment status. Not all five dimensions of the Big Five contribute, however, to explaining observed individual unemployment status in the same way. Neuroticism and Openness significantly decrease the probability of being employed and raise the probability of having a higher number of unemployment spells. Curiously, in earlier studies Extroversion and Agreeableness were identified as the traits that raise unemployment risk (Cuesta and Budria-Rodriguez (2012)). Additionally, these traits were identified as negative determinants of earnings (Heineck (2011)).

Moreover, it was found that Extroversion had a significant effect on the number of unemployment spells, while Uysal and Pohlmeier (2011) found that Extroversion and Agreeableness revealed no explanatory power. The relevance of personality traits in explaining individual unemployment status is also confirmed by focusing on the effects across different occupations and sectors. Our results contribute to the discussion on individual heterogeneity in unemployment by showing that the differences in personalities are able to explain parts of individual differences in employment history.

This indicates that appropriate screening of the unemployed by assessing the personality of the individual and eventually offering appropriate interventions (e.g. training of self-regulatory skills, McCrae and Lockenhoff (2010)) may improve their success in the labour market. Governments typically focus on labour market institutions and observed individual-related characteristics, especially education, to deal with high levels of unemployment. The results of this note warn that the effectiveness of such policies may differ importantly among individuals with different unobserved personality characteristics.

Future research should focus on the channels through which personality traits affect the chances of finding a job. Empirical estimates of such a structural approach would help to evaluate the role of specific training programs (e.g. application training) for the unemployed.

5. Conclusions and Further Research

Taking into account the topics dealt with in the analysis, the part on conclusions and further research is divided into two parts. Part I is dedicated to the general conclusions and the plan of further research on spatial trends and spatial econometric structures. Part II is based on the conclusions and possible further research on personal characteristics and unemployment analysis based on the Add Health data.

5.1. Part I

The focus in the chapters based on spatial trends and spatial econometric structures has been on cross-sectional spatial data modelization. The analysis contributes, on one hand, to the use of specification tests in order to assess the robustness of the results, once it is recognized that our ability to accurately model spatial data is very limited. On the other hand, it also contributes to the debate around the W weight matrix in formulating spatial econometrics models. There are several challenges when econometrically modelling economic relations for which space can have a relevant role. The combination of missing correlated over space variables along with many sources of potential nonlinearities has led scholars to consider that conventional parametric spatial models are not necessarily the best modelling strategy. We firstly argue that it is critical to use specifications tests in order to validate results. This explains why we propose (and develop on) (i) the use of a nonparametric test for spatial structure of unknown (either linear or nonlinear) form along with other available tests, and (ii) the use of geographically restricted semiparametric models that assume that the true model is unknown. To put it differently, we argue that, supported by powerful statistical tools, semiparametric models and estimators could be used before going through conventional spatial lag models. The first chapter based on the spatial analysis has studied hedonic price functions and estimates because housing price formation has been a central object for the implementation of policy decisions and therefore model specification may be particularly important. The three studied cites are independent from each other. Regarding specifications issues, the main conclusion is the following: spatial dependence structure (potentially of unknown form) can be controlled by restricted semiparametric models that do not use W matrix specification, however there might be spatial data relationships for which spatial structure can only be controlled at the cost of assuming a W matrix of connections. Given the potential consequences of a wrong specification, those models that avoid a W matrix are preferred. However, if the modeller requires the use of a weight matrix (as might happen), according to our results, the use of robust specification tests is recommended when choosing the best parametric model. As a result, this Chapter makes it obvious that there are some types of datasets where spatial dependence can be controlled using simple models, that include longitude, latitude or some combinations of both. The fact that the new approach with delta-models produced better results in some datasets than in others, opened the way to Chapter 3.

The general contribution of the second part of the analysis of spatial models is the exploitation of the analysis techniques and how they perform in different environments. The analysis is based on the delta models and the general technique presented in Chapter 1. These techniques are applied to the datasets that present different characteristics and have different economic behaviour. This approach lets us highlight the main differences between the datasets where delta models perform better than classic spatial models or vice versa. The results of the analysis performed could not give us a proper answer on what characteristics are crucial in the dataset to distinguish and apply a proper type of methodology. This leads naturally to several lines of further research.

From a broarder standpoint, this thesis has contributed to the spatial econometric literature around open concerns for the scholars, in particular to the literature on spatial dependence bias, on functional form bias and on spatial heterogeneity bias. Nevertheless, other concerns remain open for further research, namely, the potential interaction between geoadditive terms and covariates of particular interest for researchers. From the applied side, the first part of the study has centered on housing prices, however there are other well studied cross-sectional datasets (along with their economic relationships) for which space is a natural source of variation and explanation that has kept the attention of geographical modellers (industry location and knowledge spillovers, among others). And the second part of the research dealt with possible differences in the methodology applied, depending on the inherent characteristics of the data.

Furthermore, taking into account that we could not define the main necessary characteristics in our dataset, we might be able to define the proper environment for every type of information using another extension of the spatial models. This extension can include the spatiotemporal modelization of the processes. Thus, the main idea is to symbolize the spatial process with a finite set of symbols that is capable of controlling the spatial structure of the process

$$X_s \to \sigma_s \in \Gamma$$

where X_s is a spatial process that is observed at the location s andσ_s is a symbol associated to the spatial process X_s that collects information on locations and their neighbours. We might want to select k subsets from $\Gamma : W_1, ..., W_k = \mathit{\Gamma}$ by the way that $|W_1| < |W_2| < \ldots < |W_k|$ and analyse with a measure of information (or degree of disorder) each of the W_i

$$h(W_i) = -\sum_{\sigma \in W_i} p(\sigma) log(p(\sigma)) - \text{Shannon entropy of } W_i$$

This allows us to determine the dominant part (stochastic or deterministic). If the sequence $h(W_i)$ grows with i (numerical derivative), the stochastic part dominates over the deterministic. If there is j such that the sequence $h(W_i)$ $(i > j)$ does not increase, then the spatial process is structured (little disorder) and therefore the deterministic part dominates the spatial process analysed.

That is the main reason why it might be useful to apply the natural extension to Spatio-Temporal processes X_{ts}. The key lies in the process of symbolization. It might control both the temporal and spatial structure of the process.

The process of the spatiotemporal representation can be structured as:

$$X_{ts} \to \delta_{ts} = (\pi_t, \sigma_s) \in \mathit{\Pi} \times \mathit{\Gamma}$$

- $X_{ts}$is a a spaciotemporal process observed at time t at location s
- π_t fixed at the location s, is the pattern of order m of the consecutive values $X_{ts}, X_{t+1s}, ..., X_{t+m-1,s}$, i.e. the permutation that orders the m values from smallest to largest.
- σ_s fixed at the moment t, is a symbol associated to X_{ts} that collects information on locations and their neighbours as if it were a spatial process.

The general model can be divided into two parts: $h(\Pi, \Gamma) = h(\Pi|\Gamma) + h(\Gamma)$, where $h(\Pi|\Gamma)$ is a temporal entropy and $h(\Gamma)$ is a spatial entropy. The benefits of this modelling is that the independence of space and time can be contrasted. Moreover, we might be able to contrast the existence of a spaciotemporal trend (using the same procedure as before) and evaluate whether the trend or dominant deterministic part detected comes from time, space or both.

5.2. Part II

This part of the thesis aims to face the problem of the unemployment that might be directly connected with the certain psychological deterministic part of the human personality. Our contribution is based on a wider research of the topic, as the previous analysis has not been viewed in relation to individual differences. The main goal was to study the relations between the Big-Five Personality traits and different outcomes on the labor market, such as the success of unemployed workers seeking a job and their success in the labor market, measured by a number of unemployment spells. We have detected that personality traits are major determinants of job transitions and unemployment status.

However, different characteristics have different explanatory power and different type of influence on the observed individual unemployment status. The results found in this analysis create similar patterns to the analysis performed previously, thus, Extroversion had a significant effect on some of the unemployment characteristics studied. What is more, Neuroticism and Openness significantly decrease the probability of being employed and raise the probability of having a higher number of unemployment spells.

Our results suggest important links between personality traits and unemployment, and this is substantiated by the robustness analyses performed, where we included other personal characteristics, such as health information, math grades, etc. Statistically significant and economically meaningful relations remain even when we control for possible effects across different occupations and sectors, even though some sector specific differences (more extroverted people work in sales and social sectors) exist. Our results contribute to the discussion on individual heterogeneity in unemployment by showing that the differences in personalities are able to explain parts of individual differences in employment history.

An important direction for future work in this area would be to analyse the development of and possible changes in personality traits, with particular attention to labour situations, external circumstances, and potentially unobserved characteristics. Future research should also focus on the channels through which personality traits affect the chances of finding a job. Our results suggest that there is a potential for policies that exploit the association between positive non-cognitive traits and achievements in the labour market—outcomes that are important determinants of long-term living standards. The results found warn that the effectiveness of such policies may differ importantly among individuals with different unobserved personality characteristics. Empirical estimates of such a structural approach would help to evaluate the role of specific training programs (e.g. application training) for the unemployed.

Bibliography

Almlund, M., Duckworth, A. L., Heckman, J. J., and Kautz, T. D. (2011). Personality psychology and economics. Working Paper 16822, National Bureau of Economic Research.

Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical analysis*, 20(1):1–17.

Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27(3):247–267.

Anselin, L., Bera, A. K., Florax, R., and Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26(1):77–104.

Baltagi, B. (2005). *Econometric analysis of panel data*. Wiley, Chichester [u.a.], 3. ed. edition.

Barrick, M. R. and Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, 44(1):1–26.

Barrick, M. R., Mount, M. K., and Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and assessment*, 9(1-2):9–30.

Basile, R., DurbÃ¡n, M., MÃnguez, R., MarÃa Montero, J., and Mur, J. (2014). Modeling regional economic dynamics: Spatial dependence, spatial heterogeneity and nonlinearities. *Journal of Economic Dynamics and Control*, 48:229–245.

Basile, R. and Minguez, R. (2018). *Advances in Spatial Econometrics: Parametric vs. Semiparametric Spatial Autoregressive Models*, pages 81–106.

Bhattacharjee, A. and Holly, S. (2013). Understanding Interactions in Social Networks and Committees. *Spatial Economic Analysis*, 8(1):23–53.

Bivand, R., Pebesma, E., and Gomez Rubio, V. (2013). *Applied Spatial Data Analysis with R*, pages 59–82.

Borghans, L., Duckworth, A. L., Heckman, J. J., and Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of human Resources*, 43(4):972–1059.

Brauns, H. and Steinmann, S. (1999). Educational reform in france, west-germany and the united kingdom: updating the casmin educational classification. *Zuma Nachrichten*, 23(44):7–44.

Cameron, A. C., Trivedi, P. K., et al. (2009). *Microeconometrics using stata*, volume 5. Stata press College Station, TX.

Chambers, J. M. (2014). Object-oriented programming, functional programming and r. *Statistical Science*, 29(2).

Ciccarelli, C. and Elhorst, J. (2018). A dynamic spatial econometric diffusion model with common factors: The rise and spread of cigarette consumption in Italy. *Regional Science and Urban Economics*, 72(C):131–142.

Cobb-Clark, D. and Schurer, S. (2012). The stability of big-five personality traits. *Economics Letters*, 115(1):11–15.

Coenen, J., Borghans, L., and Diris, R. (2021). Personality traits, preferences and educational choices: A focus on stem. *Journal of Economic Psychology*, 84:102361.

Corr, P. J. and Matthews, G. (2009). *The Cambridge handbook of personality psychology.* Cambridge University Press Cambridge, UK:.

Cuesta, M. and Budria-Rodriguez, S. (2012). Overeducation dynamics and personality. *Education Economics*, 20:1–24.

De Fruyt, F. and Mervielde, I. (1999). Riasec types and big five traits as predictors of employment status and nature of employment. *Personnel psychology*, 52(3):701–727.

Ditzen, J. (2018). Estimating dynamic common-correlated effects in stata. *The Stata Journal*, 18(3):585–617.

Donnellan, M. B., Oswald, F. L., Baird, B. M., and Lucas, R. E. (2006). The mini-ipip scales: tiny-yet-effective measures of the big five factors of personality. *Psychological assessment*, 18(2):192.

Dubin, R. A. (1992). Spatial autocorrelation and neighborhood quality. *Regional science and urban economics*, 22(3):433–452.

Egan, M., Daly, M., Delaney, L., Boyce, C. J., and Wood, A. M. (2017). Adolescent conscientiousness predicts lower lifetime unemployment. *Journal of Applied psychology*, 102(4):700.

Eilers, P., Marx, B., and Durban, M. (2015). Twenty years of p-splines. *SORT (Statistics and Operations Research Transactions)*, 39:149–186.

Elhorst, J. P. (2013). *Spatial Econometrics*. Springer-Verlag Berlin Heidelberg.

Engelhardt, C. (2017). Unemployment and personality: Are conscientiousness and agreeableness related to employability? Technical report, Hannover Economic Papers (HEP).

Florax, R., Folmer, H., and Rey, S. (2003). Specification searches in spatial econometrics: The relevance of hendry's methodology. *Regional Science and Urban Economics*, 33:557–579.

Frey, B. and Stutzer, A. (2002). What can economists learn from happiness research? *J. Econ. Lit.*, 40.

Garcia-Cordoba, J., Matilla-Garcia, M., and Ruiz Marin, M. (2019). A test for deterministic dynamics in spatial processes. *Spatial Economic Analysis*, pages 1–17.

Geniaux, G. and Martinetti, D. (2018). A new method for dealing simultaneously with spatial autocorrelation and spatial heterogeneity in regression models. *Regional Science and Urban Economics*, 72:74–85. New Advances in Spatial Econometrics: Interactions Matter.

Geron, A. (2017). *Hands-on machine learning with Scikit-Learn and TensorFlow : concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, Sebastopol, CA.

Gibbons, S. and Overman, H. G. (2012). Mostly pointless spatial econometrics? *Journal of Regional Science*, 52(2):172–191.

Heineck, G. (2011). Does it pay to be nice? personality and earnings in the united kingdom. *ILR Review*, 64(5):1020–1038.

Ismail, N. A. (2011). Graduates' characteristics and unemployment: A study among malaysian graduates. *International journal of business and social science*, 2(16).

J.B, T. (1995). Economic context and the health effects of unemployment. *J Health Soc Behav.*, 36(3):213–229.

Judge, T. A., Higgins, C. A., Thoresen, C. J., and Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel psychology*, 52(3):621–652.

Kettunen, J. (1997). Education and unemployment duration. *Economics of education review*, 16(2):163–170.

Le Gallo, J. (2000). Econometrie spatiale 1 -autocorrelation spatiale. *LATEC, Laboratoire d'Analyse et des Techniques EConomiques, CNRS UMR 5118, Universite de Bourgogne, LATEC - Document de travail - Economie (1991-2003).*

Le Gallo, J. (2002). Econometrie spatiale : l autocorrelation spatiale dans les modeles de regression lineaire.

LeSage, J. and Pace, R. (2009). *Introduction to Spatial Econometrics*. Statistics: A Series of Textbooks and Monographs. CRC Press.

LeSage, J. P. (2005). *Encyclopedia of Social Measurement*, chapter Spatial Econometrics, pages 613–619. Elsevier.

Lucas, R. E., C. A. E. G. Y. . D. E. (2004). Unemployment alters the set point for life satisfaction. *Psychological Science*, 15(1):8â13.

Mankiw, N., Romer, D., and Weil, D. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 107(May):407–437.

Mayer, G. (2010). The trend in long-term unemployment and characteristics of workers unemployed for more than 99 weeks. *Congressional Research Service*, (7):2–27.

McCrae, R. and Lockenhoff, C. (2010). *SelfâRegulation and the FiveâFactor Model of Personality Traits*, pages 145 – 168.

McCrae, R. R. and Costa, P. T. (1985). Updating norman's" adequacy taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of personality and social psychology*, 49(3):710.

McMillen, D. (2012). Perspectives on spatial econometrcis: linear smoothing with structured models. *Journal of Regional Science*, 52(2):192–209.

McMillen, D. (2015). *Package McSpatial*.

Montero, J., Minguez Salido, R., and Durban, M. (2012). Sar models with nonparametric spatial trends. a p-spline approach. *Estadistica Espanyola.*, 54:89–111.

Mount, M. K., B. M. R. . S. G. L. (1998). Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance,*, 11:145â165.

Mur, J., Herrera, M., and Ruiz, M. (2011). Selecting the w matrix. parametric vs nonparametric approaches.

Nandi, A. and Nicoletti, C. (2014). Explaining personality pay gaps in the uk. *Applied Economics*, 46(26):3131–3150.

Nunez, I. and Livanos, I. (2010). Higher education and unemployment in europe: An analysis of the academic subject and national effects. *Higher Education*, 59.

Pace, R. K. and Gilley, O. W. (1997). Using the spatial configuration of the data to improve estimation. *The Journal of Real Estate Finance and Economics*, 14(3):333–340.

Pebesma, E. and Bivand, R. (2005). Classes and methods for spatial data in r. *R News*, 5.

Pesaran, M. H. (2015). *Time Series and Panel Data Econometrics*. Number 9780198759980 in OUP Catalogue. Oxford University Press.

Prevoo, T. and ter Weel, B. (2015). The importance of early conscientiousness for socio-economic outcomes: Evidence from the british cohort study. *Oxford Economic Papers*, 67(4):918–948.

Ritter, J. A. and Taylor, L. J. (2011). Racial disparity in unemployment. *The Review of Economics and Statistics*, 93(1):30–42.

Theodossiou, I. and Zangelidis, A. (2009). The social gradient in health: The effect of absolute income and subjective social status assessment on the individual's health in europe. *Economics & Human Biology*, 7(2):229–237.

Uysal, S. D. and Pohlmeier, W. (2011). Unemployment duration and personality. *Journal of economic psychology*, 32(6):980–992.

Viinikainen, J. and Kokko, K. (2012). Personality traits and unemployment: Evidence from longitudinal data. *Journal of Economic Psychology*, 33(6):1204–1222.

Z, S. ((2017)). Spatial Trend Analysis. In Sen, Z., editor, *Innovative Trend Methodologies in Science and Engineering*, pages 227–280. Springer International Publishing, Cham.

A. Appendix

Table A.1.: Results for Baltimore Housing Prices

	OLS	SAR	SEM	SLX	SAC	SDM	SDEM	GNS	Spline	C-spline	F-spline	MA	MB	MC	MD	M E	MF	M G
Constant	7.16	-0.91	6.95	8.78	-0.05	13.05	3.18	16.66”		-17690.00	-48.30	2m	2m’	785.57	2.41	3.46	-3.84	5.98
Nroom	0.29	0.53	0.75	0.27	0.08	0.29	0.22	0.47	0.95	0.33	0.29	0.21	0.21	0.15	0.19	0.17	0.17	0.16
Dwell	6.30”	5.74”	6.96”’	6.33”	4.46”	6.35”’	7.52”’	6.72”’	7.13”’	6.79”	6.95”	6.09”	6.12”	5.51”	5.03’	5.28”	5.30”	5.26”
Nbath	6.08”’	5.37”’	6.57”’	5.57”’	4.08”	5.59”’	6.87”’	6.72”’	6.19”’	5.82”	5.89”	5.71”’	5.75”’	5.67”’	4.97”	5.32”’	5.34”’	5.31”’
Patio	9.40”’	6.92”’	7.55”’	5.25’	7.39”’	5.21”’	9.81”’	5.63”	6.34”	8.16”	7.94”	7.94”’	7.98”’	9.35”’	8.64”’	8.98”’	9.02”’	8.94”’
Firepl	10.69”’	8.21”’	8.95”’	8.93”’	8.46”’	8.88”’	9.48”’	7.59”’	7.58”’	8.80”’	8.66”’	8.92”’	8.95”’	10.05”’	9.55”’	9.86”’	9.89”’	9.83”’
AC	8.20”’	7.15”’	7.34”’	7.18”’	7.12”’	7.17”’	7.14”’	7.25”’	6.85”’	6.91”	6.93”	6.26”’	6.25”’	6.29”	6.50”’	6.35”	6.41”	6.28”
Bment	3.81”’	3.78”’	3.59”’	3.36”’	3.77”’	3.36”’	3.34”’	3.02”’	3.37”’	3.36”	3.41”’	3.71”’	3.69”’	3.39”’	3.74”’	3.52”’	3.53”’	3.51”’
Nstor	-4.36	-5.08’	-3.93	-2.41	-5.82”	-2.44	-1.20	-3.13	-3.49	-4.19	-4.08	-4.63’	-4.60’	-4.73	-5.23’	-4.99’	-4.98’	-5.01’
Gar	5.53”’	5.65”’	5.23”’	5.74”’	5.71”’	5.74”’	4.58”’	4.86”’	4.33”	4.93”	4.94”	5.15”’	5.14”’	5.52”’	5.76”’	5.61”’	5.62”’	5.60”’
Age	0.00	0.04	-0.01	0.01	0.06	0.00	0.03	0.01	0.01	-0.02	-0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Citcou	12.46”’	10.12”’	12.43”’	10.87”’	9.08”’	10.86”’	13.42”’	10.32”’	11.52”’	16.55”’	16.83”’	10.72”’	10.74”’	13.10”’	12.67”’	13.18”’	13.17”’	13.18”’
Lotsz	0.04”	0.02	0.03	0.02	0.02	0.02	0.04”’	0.01	0.03	0.04’	0.04’	0.03’	0.03’	0.03”	0.03’	0.03’	0.03’	0.03’
Sqft	0.37’	0.38’	0.26	0.25	0.47”	0.25	0.08	0.18	0.19	0.29	0.28	0.31	0.31	0.31	0.37	0.34	0.34	0.34
WNroom				0.02		0.02	-2.07”’	-4.37”										
WDwell				-3.29		-3.27	-1.30	-5.39										
WNbath				-1.55		-1.70	17.85”’	-11.20”’										
WPatio				-7.21’		-7.28’	14.31”’	2.92										
WFirepl				10.96’		10.75’	31.64”’	9.62’										
WAC				17.45”’		17.02”’	26.05”’	-1.10										
WBment				5.36		5.19	16.69”	-0.82										
WNstor				0.35		0.27	8.19”’	4.81										
WGar				7.69		7.59	0.71	0.22										
WAge				2.24		2.07	9.73”	-0.02										
WCitcou				0.04		0.04	0.11	-6.87										
WLotsz				-1.71		-1.71	26.68”’	0.07’										
WSqft				0.10”		0.10”	0.14”’	0.48										
rho		0.27”’			0.36”’	0.00		0.45”’										
lambda			0.31”’		-0.34		0.96”’	-0.71”’										
Log-Lik	-926.53	-817.7957	-751.92	-926.53	-816.07	-802.20	-778.74	-797.55				-908.65	-909.11	-920.57	-920.22	-921.09	-921.39	-920.79
Moran(pv)	0.02	0.09	0.02	0.46	0.39	0.46	0.36	0.92	0.02	0.32	0.34	0.25	0.29	0.02	0.01	0.02	0.02	0.02
dh-test(pv)	0.02	0.05	0.12	0.01	0.12	0.12	0.13	0.12	0.04	0.02	0.01	0.02	0.02	0.03	0.02	0.04	0.04	0.04
AIC	1.88	1.67	1.53	1.84	1.66	1.51	1.66	1.65		5.1	5.09	1.86	1.85	1.86	1.87	1.87	1.87	1.87
	OLS	SAR	SEM	SLX	SAC	SDM	SDEM	GNS	Spline	C-spline	F-spline	MA	MB	MC	MD	ME	M F	MG

Table A.2.: Results for Boston Housing Prices

	OLS	SAR	SEM	SLX	SAC	SDM	SDEM	GNS	Spline	C-spline	F-spline	Model A	Model B	Model C	Model D	Model E	Model F	Model G
Constant	36.46”’	19.25”’	30.11”’	37.77”’	28.09”’	23.94”’	27.49”’	25.74”’		704400.00”	-274.30	46.69”’	46.78”’	34.64”’	47.61”’	47.57”’	-435.14	-122.45
Crime	-0.11”’	-0.09”’	-0.13”’	-0.12”’	-0.13”’	-0.12”’	-0.13”’	-0.15”’	-0.15”’	-0.12”’	-0.12”’	-0.13”’	-0.13”’	-0.11”’	-0.13”’	-0.13”’	-0.11	-0.11”’
Zoning	0.05”’	0.04”’	0.04”’	0.03	0.04”’	0.03”	0.04”	0.03”	0.04”	0.04”	0.04”	0.03”	0.03”	0.04”’	0.03”	0.03”	0.05”’	0.05”’
Industry	0.02	0.02	-0.01	-0.01	-0.01	-0.05	-0.03	-0.02	-0.01	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.01	0.01
Charlesr	0.02”’	1.29	-0.35	-0.54	-0.27	-1.31	-0.45	-0.1	1.33	2.42”	2.15’	2.68”’	2.69”’	2.54”’	2.69”’	2.67”’	2.60”’	2.59”’
Noxsq	-17.77”’	-13.06”’	-18.58”’	-29.57”’	-17.80”’	-17.96”’	-18.72”’	-19.23”’	-23.05”’	-20.18”’	-19.33”’	-21.80”’	-21.83”’	-16.01”’	-22.92”’	-23.09”’	-15.90”’	-15.90”’
Rooms2	3.81”’	3.68”’	4.23”’	4.06”’	4.24”’	4.33”’	4.37”’	4.66”’	3.82”’	3.62”’	3.58”’	3.74”’	3.74”’	3.78”’	3.71”’	3.71”’	3.78”’	3.78”’
Houseage	0.01”’	0.01	-0.02’	-0.03’	-0.02	-0.03”	-0.03”	-0.03”	-0.03’	-0.02	-0.02	-0.02	-0.02	0.00	-0.02	-0.02	0.01	0.01
Distance	-1.48”’	-1.24”’	-1.62”’	-1.88”’	-1.62”’	-1.51”’	-1.59”’	-1.61”’	-2.25”’	-2.37”’	-2.45”’	-3.22”’	-3.24”’	-1.33”’	-1.28”’	-3.11”’	-1.38”’	-1.38”’
Access	0.31”’	0.26”’	0.32”’	0.35”’	0.32”’	0.29”’	0.31”’	0.31”’	0.31”’	0.36”’	0.35”’	0.36”’	0.36”’	0.33”’	0.33”’	0.34”’	0.31”’	0.31”’
Taxrate	-0.01”’	-0.01”’	-0.01”’	-0.01”’	-0.01”’	-0.01”’	-0.01”’	-0.01”’	-0.01”’	-0.01”’	-0.01”’	-0.01”’	-0.01”’	-0.01”’	-0.01”’	-0.01”’	-0.01”’	-0.01”’
Pupil/Teacher	-0.95”’	-0.63”’	-0.64”’	-0.61”’	-0.64”’	0.44”’	-0.58”’	-0.57”’	-0.74”’	-0.99”’	-0.95”’	-1.02”’	-1.02”’	-0.92”’	-1.03”’	-1.04”’	-0.91”’	-0.91”’
Blackpop	0.01”’	0.01”’	0.01”’	0.01”’	0.01”’	0.01”’	0.01”’	0.01”’	0.01”’	0.01”’	0.01”’	0.01”’	0.01”’	0.01”’	0.01”’	0.01”’	0.01”’	0.01”’
Lowclass	-0.52”’	-0.40”’	-0.44”’	-0.45”’	-0.43”’	-0.41”’	-0.41”’	-0.39”’	-0.52”’	-0.53”’	-0.53”’	-0.53”’	-0.54”’	-0.53”’	-0.53”’	-0.53”’	-0.53”’	-0.53”’
WCrime				0.14”		0.1	-0.05	-0.17’										
WZoning				0.02		-0.01	0.01	0.02										
WIndustry				-0.04		0.03	0.05	0.05										
WCharlesr				7.91”’		5.11”’	7.09”’	7.28”’										
WNoxsq				11.65		5.79	-3.96	-11.02										
WRooms2				-0.76		-3.60”’	0.53	2.89’										
WHouseage				0.07”’		0.06”	0.04	0.02										
WDistance				0.72		0.94’	0.12	-0.47										
WAccess				-0.25’		-0.23”	-0.06	0.07										
WTaxrate				0.01		0.01”	0.01	0.01										
WPupil/Teacher				0.50”		-0.13	-0.26	-0.25										
WBlackpop				-0.01		-0.01”	-0.01	0.01										
WLowclass				-0.18’		0.11	-0.1	-0.19										
rho		0.31”’			0.06	0.59”’		-0.31										
lambda			0.68”’		0.64”’		0.64”’	0.81										
Log-lik	-1751.8	-1467.43	-1256.67	-1729.1	-1431.09	-1420.53	-1420.98	-1420.03				-1725.4	-1725.9	-1741.3	-1728.3	-1729.0	-1742.6	-1742.5
Moran(pv	0	0	0	0	0.46	0	0	0.54	0	0	0	0	0	0	0	0	0	0
dh-test(pv)	0.04	0.10	0.12	0.05	0.01	0.07	0.09	0.01	0.09	0.03	0.03	0.03	0.03	0.05	0.05	0.04	0.04	0.04
AIC	3.52	2.96	2.54	3.47	2.89	2.53	2.88	2.88		3.1	3.09	3.49	3.49	3.52	3.52	3.52	3.52	3.52
	OLS	SAR	SEM	SLX	SAC	SDM	SDEM	GNS	Spline	C-spline	F-spline	Model A	Model B	Model C	Model D	Model E	Model F	Model G

Table A.3.: Results for Toledo Housing Prices

	OLS	SAR	SEM	SLX	SAC	SDM	SDEM	GNS
Constant	8105.30	8441.69	8410.68	-2831.41	8564.30	-4299.56	-7219.56	-3694.60
Neighbor	-2758.37	-2842.90	-2821.70	-1098.85	-2867.48	-895.53	-708.78	-882.77
Lot sqft	0.27	0.29	0.29	0.15	0.30	0.11	0.19	0.20
Total sqft	7.28	6.96	7.15	4.73	6.91	5.05	4.29	5.99
Family room	30926.93"'	30859.95"'	30546.77"'	21272.60"	30718.27"'	21876.78"'	20990.69"	19292.71"
Rec room	-1608.90	-1860.70	-1832.99	-4541.93	-1944.34	-4104.28	-4953.70	5143.33
Air cond	-37981.82"'	-37797.45"'	-37228.27"'	-22340.99	-37519.17"'	-23296.76	-19710.70	-18300.67
Baths	16449.63"'	16578.64"'	16604.19"'	18759.07"'	16616.57"'	18497.00"'	18988.56"'	18549.71"'
Condition	8430.70'	8766.38"	8190.45'	8004.52'	8674.14"	8672.42"	8044.16"	7720.24'
Garage cond	-10996.77"'	-10976.26"'	-10895.37"'	-10979.44"'	-10936.93"'	-11240.48"'	-11136.00"'	-10911.83"'
Age	-143.77	-149.68	-144.95	-223.14"	-150.07	-225.93"	-235.34"'	-244.49"'
Wneighbor				3455.48		3092.83	2922.37	5347.99
Wlotsqft				-1.72'		-1.73'	-1.72"	-1.47"
Wtotalsqft				5.63		8.30	7.68	-2.88
Wfam room				20387.20		24679.64	17493.16	15391.37
Wrecroom				14095.51		14814.49	15395.43'	15748.86"
Wac				-52454.49		-59448.07	-50695.33'	-30990.27
Wbath				-4416.34		-2325.33	-5640.27	-3662.30
Wcond				12393.72		15397.47	13039.72	7687.00
Wgarcond				-3100.97		-5199.50	-2807.53	-3202.35
Wage				-85.77		-120.96	-89.60	-18.34
rho		-0.01			-0.02	-0.18		0.14
lambda			0.03		0.01		-0.26	-0.40
Log-lik	-1033.10	-989.26	-955.34	-1018.90	-989.25	-944.65	-978.24	-977.73
Moran(pv)	0.57	0.64	0.57	0.47	0.63	0.43	0.37	0.80
dh-test(pv)	0.02	0.03	0.05	0.11	0.10	0.11	0.03	0.02
AIC	2.09	2.09	1.93	2.08	2.00	1.93	2.00	2.00
	OLS	SAR	SEM	SLX	SAC	SDM	SDEM	GNS

Table A.4.: Results for Toledo Housing Prices(cont.)

	Spline	C-spline	F-spline	Model A	Model B	Model C	Model D	Model E	Model F	Model G
Constant		3.299e+08	3.445e+06'	1252522880	-685092529.5	1090229886	1821513.36"	1834964.09"	3664887.02"	1224987.63"
Neighbor	-3696.39	-1348.00	-1162.00	-2165.26	-1886.25	-2935.71	-2583.57	-2377.45	-2377.25	-2377.64
Lot sqft	0.61	0.55	0.54	0.56	0.55	0.39	0.38	0.38	0.38	0.38
Total sqft	0.47	3.95	3.86	2.67	2.06	4.78	4.65	4.78	4.78	4.78
Family room	23652.30"	29010.00"	28620.00"	25736.93"'	25529.68"'	30432.49"'	30407.07"'	30322.9"'	30324.16"'	30321.64"'
Rec room	-6587.14	-4645.00	-5318.00	-5686.75"	-5248.25"	-2898.33'	-2573.88'	-2599.84'	-2599.08'	-2600.6'
Air cond	-24095.51	-27840.00	-25680.00	-22861.58	-23047.23	-34972.25"	-35151.44"	-34790.75"	-34792.74"	-34788.76"
Baths	20972.00"'	18410.00"'	18650.00"'	19030.72"'	19253.35"'	17702.32"'	17747.41"'	17722.46"'	17721.84"'	17723.07"'
Condition	4063.32	5434.00	4823.00	5442.27	5182.25	7984.58'	7694.87	7471.27	7469.58	7472.95
Garage cond	-8194.50"'	-9691.00"'	-9479.00"'	-9686.32"'	-9809.15"'	-10572.82"'	-10628.51"'	-10584.73"'	-10584.11"'	-10585.34"'
Age	-215.12'	-257.00'	-265.50'	-261.17"	-263.42"	-175.77'	-181.21'	-188.06'	-188.07'	-188.05'
Wneighbor										
Wlotsqft										
Wtotalsqft										
Wfam room										
Wrecroom										
Wac										
Wbath										
Wcond										
Wgarcond										
Wage										
rho										
lambda										
Log-lik				-1024.4	-1025.2	-1028.6	-1029	-1029.4	-1029.4	1029.4
Moran(pv)	0.57	0.51	0.53	0.89	0.78	0.79	0.77	0.88	0.88	0.88
dh-test(pv)	0.05	0.01	0.01	0.01	0.01	0.02	0.02	0.02	0.02	0.02
AIC		17.6	17.6	2.08	2.08	2.09	2.09	2.08	2.08	2.08
	Spline	C-spline	F-spline	Model A	Model B	Model C	Model D	Model E	Model F	Model G

Table A.5.: LR and LM tests. Baltimore.

Comparation of models	Degrees of freedom	Critical value	Result
OLS vs SLX	13	22.36	62.76
OLS vs SAR	1	3.84	217.47
OLS vs SEM	1	3.84	349.22
SAR vs SAC	1	3.84	31.19
SEM vs SAC	1	3.84	-100.57
SLX vs SDM	1	3.84	185.89
SEM vs SDM	13	22.36	-100.57
SAR vs SDM	13	22.36	31.19
SLX vs SDEM	1	3.84	232.81
SAR vs SDEM	13	22.36	78.10
SEM vs SDEM	13	22.36	-53.65

SEM vs SAR	Lag	Error
Non-robust LM	21.88	3.71
p-value	0.00	0.00
Robust LM	19.01	0.83
p-value	0.05	0.36

Table A.6.: LR and LM tests. Boston.

Comparation of models	Degrees of freedom	Critical value	Result
OLS vs SLX	13	22.36	45.40
OLS vs SAR	1	3.84	568.74
OLS vs SEM	1	3.84	990.26
SAR vs SAC	1	3.84	72.68
SEM vs SAC	1	3.84	-348.84
SLX vs SDM	1	3.84	617.14
SEM vs SDM	13	22.36	-327.72
SAR vs SDM	13	22.36	93.80
SLX vs SDEM	1	3.84	616.24
SAR vs SDEM	13	22.36	92.90
SEM vs SDEM	13	22.36	-328.62

SEM vs SAR	Lag	Error
Non-robust LM	63.59	150.60
p-value	0.00	0.00
Robust LM	2.31	89.32
p-value	0.13	0.00

Table A.7.: LR and LM tests. Toledo.

Comparation of models	Degrees of freedom	Critical value	Result
OLS vs SLX	10	18.31	28.40
OLS vs SAR	1	3.84	87.68
OLS vs SEM	1	3.84	155.52
SAR vs SAC	1	3.84	89.22
SEM vs SAC	1	3.84	21.38
SLX vs SDM	1	3.84	148.50
SEM vs SDM	10	18.31	21.38
SAR vs SDM	10	18.31	89.22
SLX vs SDEM	1	3.84	81.32
SAR vs SDEM	10	18.31	22.04
SEM vs SDEM	10	18.31	-45.80

SEM vs SAR	Lag	Error
Non-robust LM	4.52	1.38
p-value	0.03	0.24
Robust LM	6.77	3.63
p-value	0.01	0.06

Table A.8.: Results for NUTS2 Dataset

	GNS	OLS	MA	MB	MC	MD	ME	MF	MG	MH	SAC	SAR	SDEM	SDM	SEM	SLX
Constant	0,080”’	0,093	0,007”	0,08	0,000”	0,09”’	-0,09”’	0,077”	0.088”	0.049	0,096”’	0,058”	0,094”’	0,074”	0,096”’	0,090”’
Human capital	-0,003	0,000	-0,001	-0,001	0,000	-0,001	0,000	0,000	0,001	-0,002	-0,003	-0,002	-0,002	-0,003	-0,003	-0,002
GDP	0,002’	-0,001	0,000	0,000	0,000	0,000	-0,001	0,000	-0,001	0,001	0,002”	0,001	0,001	0,002	0,002”	0,001
Population Growth	0,015’	0,000	0,008	0,006	-0,002	0,004	0,001	-0,003	-0,002	0,015	0,019”	0,009	0,015’	0,014	0,019	0,015
Physical Capital	0,021”’	0,035	0,033”’	0,033”’	0,033”’	0,033”’	0,036”’	0,034”’	0,036”’	0,030”’	0,022”’	0,020”’	0,023”’	0,021”’	0,023”’	0,021”’
WHuman capital	0,010’												0,009	0,010		0,010
WGDP	-0,007”’												-0,007”	-0,007”		-0,009”’
WPopulation Growth	-0,019”												-0,029”’	-0,016”		-0,042”’
WPhysical Capital	-0,002												0,010	-0,004		0,023”’
rho	0,591”’										-0,115	0,660”’		0,662”’		
lambda	0,123										0,842”’		0,696”’		0,794”’	
Moran(p-value)	0,72	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,68	0,00	0,00	0,00	0,00	0,00
dh-test(p-value)	0,02	0,07	0,29	0,31	0,13	0,14	0,10	0,13	0,10	0,11	0,02	0,10	0,10	0,10	0,15	0,07
	GNS	OLS	MA	MB	MC	MD	ME	MF	MG	MH	SAC	SAR	SDEM	SDM	SEM	SLX

Table A.9.: Results for GEcon Dataset

	GNS	OLS	MA	MB	MC	MD	ME	MF	MG	MH	SAC	SAR	SDEM	SEM	SLX
Constant	3,252”’	3,805”’	0,000’	0,000	0,000”	0,000”’	0,001”	0,019”’	0,000	0,000	4,199”’	1,156”	3,084”’	3,144”’	3,350”’
Distance to coast (km)	139,548”’	-471,761	-516,009	-423,227	-507,959	-590,273	-462,758	-462,758	-471,904	-1211,060	107,525”’	-237,028”’	-100,677”’	55,290”’	-546,409
Distance to coast (km)	-0,004”	-0,001	0,000	-0,001	-0,001	0,000	-0,001	-0,001	-0,001	0,001	-0,004”	-0,001”’	-0,005”	-0,003”’	-0,004
Elevation (km)	0,001	-0,001’	-0,002”’	-0,002”’	-0,002”’	-0,002”’	-0,001	-0,002”’	-0,001’	-0,003”’	0,001	0,000	0,002	0,000	0,002
Dist. to mn lake (km)	0,000”’	0,000”’	0,000	0,000”	0,000	0,000	0,000”’	0,000”	0,000”	0,000”	0,000	0,000	0,000”’	0,000	0,000’
Dist. to mn river (km)	0,000’	0,000	0,000	0,000”’	0,000	0,000”’	0,000	0,000”’	0,000	0,000	0,000	0,000	0,000	0,000	0,000’
Dist. to ice-free ocean (km)	-0,140”’	0,472	0,516	0,423	0,508	0,590	0,463	0,463	0,472	1,211	-0,108”’	0,237”’	0,101”’	-0,055”’	0,546
Dist. to navigable river (km)	-0,001”	-0,001”’	-0,001”	-0,001”	-0,001”’	-0,001”	0,000	-0,001”’	-0,001”’	0,000	-0,001”’	-0,001”’	-0,001”	-0,001”’	-0,001’
Veg. category	0,019	0,056”’	0,063”’	0,062”’	0,061”’	0,061”’	0,056”’	0,060”’	0,056”’	0,062”’	0,018	0,060”’	0,016	0,034	0,019
Grid cell population, 2000	0,000”’	0,000”’	0,000”’	0,000”’	0,000”’	0,000”’	0,000”’	0,000”’	0,000”’	0,000”’	0,000”’	0,000”’	0,000”’	0,000”’	0,000”’
Avg precipitation, prior data	-0,001”	0,001”	0,000	0,000	0,000’	0,000	0,001”’	0,000’	0,001”	0,000	-0,001”	0,000”	-0,001”	0,000	-0,001”
Soil category	0,003	0,004”	0,002	0,002	0,002	0,002	0,004’	0,002	0,003	0,002	0,003	0,004”	0,002	0,003	0,003
Avg temperature, prior data	-0,200”’	-0,189”’	-0,168”’	-0,163”’	-0,181”’	-0,158”’	-0,086	-0,174”’	-0,191”’	0,006	-0,204”’	-0,144”’	-0,201”’	-0,196”’	-0,189”’
WDistance to coast (km)	405,584”’												-899,864”’		-5019,592
WDistance to coast (km)	0,005												0,005		0,004
WElevation (km)	-0,004												-0,003’		-0,004”
WDist. to mn lake (km)	0,000”’												0,000”’		0,000”’
WDist. to mn river (km)	0,000												0,000		0,000
WDist. to ice-free ocean (km)	-0,406”’												0,900”’		5,020
WDist. to navigable river (km)	0,000												0,000		0,000
WVeg. category	0,045												0,020		0,018
WGrid cell population, 2000	0,000”’												0,000”’		0,000”’
WAvg precipitation, prior data	0,002”												0,002”’		0,002”’
WSoil category	0,003												-0,001		-0,002
WAvg temperature, prior data	-0,036												0,053		0,040
rho	-0,500”’										-0,586”’	0,436”’			
lambda	0,790”’										0,816”’		0,542”’	0,579”’	
Moran(p-value)	0,874	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,818	0,000	0,000	0,000	0,000
dh-test(p-value)	0,351	0,420	0,288	0,313	0,421	0,416	0,420	0,423	0,420	0,426	0,362	0,421	0,400	0,421	0,398
	GNS	OLS	MA	MB	MC	MD	ME	MF	MG	MH	SAC	SAR	SDEM	SEM	SLX

Table A.10.: Results for California Prices

	GNS	OLS	MA	MB	MC	MD	ME	MF	MG	MH
Constant	30949,965”’	-46139,647”’	22230,990”’	-1869,001”’	679,201”’	2130,442”’	-569,229”’	-42509,737”’	-10,066”’	2,587
Housing age	1159,251”’	1882,121”’	1170,281”’	1162,020”’	1152,106”’	1172,236”’	1161,584”’	1157,900”’	1170,883”’	1141,883”’
Total room	-10,042”’	-19,733”’	-7,251”’	-7,839”’	-8,126”’	-8,067”’	-8,403”’	-8,250”’	-8,642”’	-7,071”’
Bed. number	75,731”’	100,944”’	94,186”’	122,837”’	113,182”’	117,242”’	111,883”’	113,821”’	110,074”’	89,731”’
Population	-31,378”’	-35,319”’	-37,755”’	-37,988”’	-38,705”’	-37,416”’	-38,641”’	-38,386”’	-38,840”’	-38,025”’
Households	79,585”’	124,803”’	63,447”’	35,087”’	48,605”’	40,836”’	51,321”’	47,701”’	55,186”’	67,840”’
Median inc	36734,281”’	47748,381”’	39869,022”’	40269,322”’	40230,435”’	40327,951”’	40354,844”’	40297,522”’	40465,702”’	39785,866”’
WHousing age	1461,697”’									
WTot. room	-40,181”’									
WBed. number	211,435”’									
WPop	-27,032”’									
WHouseholds	100,585”’									
WMed.inc	32553,167”’									
rho	-0,041									
lambda	0,653”’									
Moran(p-value)	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
dh-test(p-value)	0,038	0,104	0,392	0,391	0,106	0,099	0,105	0,105	0,105	0,093
	GNS	OLS	MA	MB	MC	MD	ME	MF	MG	MH

Table A.11.: Results for California Prices

	SAC	SAR	SDEM	SDM	SEM	SLX
Constant	30653,921"'	-52610,582"'	31896,967"'	4849,920"'	31284,487"'	32398,866"'
Housing age	1212,721"'	2039,695"'	1150,475"'	1603,491"'	1204,226"'	1149,584"'
Total room	-7,288"'	-21,241"'	-9,873"'	-11,070"'	-7,189"'	-9,949"'
Bed. number	53,326"'	100,486"'	74,811"'	76,703"'	53,163"'	75,520"'
Population	-30,274"'	-32,916"'	-31,187"'	-29,152"'	-30,266"'	-30,679"'
Households	84,744"'	121,697"'	78,914"'	81,540"'	84,384"'	76,440"'
Median inc	36141,689"'	44088,888"'	36593,661"'	38385,684"'	36119,892"'	36562,699"'
WHousing age			1369,102"'	-6,659"'		1146,347"'
WTot. room			-39,404"'	-24,608"'		-42,333"'
WBed. number			206,787"'	0,660"'		216,508"'
WPop			-24,921"'	26,302"'		-11,429"'
WHouseholds			93,811"'	-17728,949"'		51,617"'
WMed.inc			30470,984"'	30472,984"'		31436,363"'
rho	0,010	0,151"'		0,649"'		
lambda	0,714"'		0,632"'		0,718"'	
Moran(p-value)	0,000	0,000	0,000	0,000	0,000	0,000
dh-test(p-value)	0,053	0,101	0,098	0,113	0,147	0,096
	SAC	SAR	SDEM	SDM	SEM	SLX

Table A.12.: Results for AirBnb Prices Texas

	GNS	OLS	MA	MB	MC	MD	ME	MF	MG	MH	SAC	SAR	SDEM	SDM	SEM	SLX
Constant	-116,113"'	-112,410"'	-7,713"'	4,459"'	-6,926"'	-3,881"'	-0,257"'	-15,380"'	-0,006"'	-0,888"	-79,915"'	-130,487"'	-123,379"'	-117,650"'	-111,394"'	-126,985"'
Bedrooms	168,827"'	176,220"'	172,503"'	173,150"'	172,879"'	172,911"'	173,408"'	173,463"'	173,361"'	172,535"'	156,443"'	169,828"'	172,850"'	171,385"'	173,776"'	173,642"'
Wbedrooms	136,085"'												14,971"'	-18,838"'		16,271"'
rho	-0,870"'										-0,747"'	0,206"'		0,252"'		
lambda	0,782"'										0,796"'		0,249"'		0,254"'	
Moran(p-value)	0,304	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
dh-test(p-value)	0,252	0,361	0,392	0,391	0,390	0,391	0,388	0,387	0,388	0,392	0,212	0,302	0,309	0,337	0,362	0,306
	GNS	OLS	MA	MB	MC	MD	ME	MF	MG	MH	SAC	SAR	SDEM	SDM	SEM	SLX

Table A.13.: 20-Item Mini-IPIP Questionnaire

How to create Big Five variables?
Each measure of those personality inventories was constructed from four questions. The respondents rated each statement about their personality using a five-level Likert scale ranging from strongly agree (1 point) to strongly disagree (5 points). For example, the items used to assess Conscientiousness include "I get chores done right away". (Reverse coded, strongly disagree (5 points) means low Conscientiousness.) "I often forget to put things back in their proper place". "I like order". (Reverse coded.) "I make a mess of things". The items for Neuroticism include "I have frequent mood swings". (Reverse coded.) "I am relaxed most of the time". "I get upset easily". (Reverse coded.) "I seldom feel blue". The constructed measures range from 1 to 5, with the higher scores representing higher levels of the personality traits.

Item	Factor	Text	Original Item Code(Add Health)
1	E	Am the life of the party.	H4PE1
2	A	Sympathize with others' feelings	H4PE2
3	C	Get chores done right away.	H4PE3
4	N	Have frequent mood swings.	H4PE4
5	I	Have a vivid imagination.	H4PE5
6	E	Don't talk a lot. (R)	H4PE9
7	A	Am not interested in other people's problems. (R)	H4PE10
8	C	Often forget to put things back in their proper place. (R)	H4PE11
9	N	Am relaxed most of the time. (R)	H4PE12
10	I	Am not interested in abstract ideas. (R)	H4PE13
11	E	Talk to a lot of different people at parties.	H4PE17
12	A	Feel others' emotions.	H4PE18
13	C	Like order.	H4PE19
14	N	Get upset easily.	H4PE20
15	I	Have difficulty understanding abstract ideas. (R)	H4PE21
16	E	Keep in the background. (R)	H4PE25
17	A	Am not really interested in others. (R)	H4PE26
18	C	Make a mess of things. (R)	H4PE27
19	N	Seldom feel blue. (R)	H4PE28
20	I	Do not have a good imagination. (R)	H4PE29

Table A.15.: Descriptive Statistics (Explanatory variables)

	Wave I	Wave IV	Wave V
Background			
Health==excellent/very good(%)	–	58.99	52.64
Health==good(%)	–	32.27	33.58
Helath==fair/poor(%)	–	8.74	13.78
Depression (%)	-	15.54	25.13
ADHD ever (%)	–	95.49	–
Math: A(%)	26.08	–	–
Math: B(%)	29.79	–	–
Math: below B(%)	38.03	–	–
Math: not taken(%)	0.61	–	–
Math: not ABCD(%)	5.50	–	–
Demographic controls			
Female(%)	57.56	–	–
White(%)	66.18	–	–
Black(%)	20.90	–	–
Hispanic(%)	14.53	–	–
Asian(%)	6.57	–	–
Other race(%)	8.28	–	–
Age	16.04	29.05	37.94
	(1.78)	(1.75)	(1.89)
Ever married by Wave(%)	0.55	51.29	69.82
Urban(%)	33.37	–	–
Siblings	–	2.91	2.79
	–	(2.35)	(2.09)
Education			
Less than high school(%)	–	6.26	4.25
High school graduate(%)	–	57.90	27.89
College graduate(%)	–	35.84	67.86
Personal income			
Personal income: <$5,000(%)	–	11.66	9.92
Personal income: $5,001-$9,999(%)	–	4.84	4.23
Personal income: $10,000-$14,999(%)	–	6.08	3.97
Personal income: $15,000-$19,999(%)	–	6.41	3.79
Personal income: $20,000-$24,999(%)	–	8.16	5.47
Personal income: $25,000-$29,999(%)	–	9.21	5.22
Personal income: $30,000-$39,999(%)	–	18.64	11.34
Personal income: $40,000-$49,999(%)	–	13.09	11.57
Personal income: $50,000-$74,999(%)	–	15.59	20.44
Personal income: $75,000-$99,999(%)	–	3.84	10.78
Personal income: $100,000-$149,999(%)	–	1.56	8.57
Personal income: $150,000-$199,999(%)	–	0.34	2.23
Personal income:>$200,000(%)	–	0.57	2.46
Observations	10914	10914	10914

Table A.16.: Transition from unemployment to unemployment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	**PerCh**	**Basic**	**Basicmath**	**Add**	**Parents**	**Full**	**Big5mean**	**Interactions**
ConscientiousnessD							0.14	0.23
							(0.12)	(0.14)
ExtraversionD							0.04	0.16
							(0.12)	(0.14)
AgreeablenessD							0.06	0.11
							(0.12)	(0.15)
NeuroticismD							0.07	0.06
							(0.12)	(0.15)
OpennessD							0.02	0.08
							(0.12)	(0.15)
EducationW5		-0.40"'	-0.38"'	-0.35"'	-0.36"'	-0.21"	-0.22"	-0.26"
		(0.07)	(0.07)	(0.07)	(0.08)	(0.09)	(0.09)	(0.12)
UrbanW4		-0.09	-0.08	-0.07	-0.01	0.05	0.05	0.07
		(0.10)	(0.10)	(0.10)	(0.11)	(0.12)	(0.12)	(0.12)
MarriedW5		-0.40"'	-0.41"'	-0.40"'	-0.50"'	-0.39"'	-0.39"'	-0.55"'
		(0.09)	(0.09)	(0.10)	(0.10)	(0.12)	(0.12)	(0.17)
Gender		0.23"	0.24"	0.26"	0.25"	0.48"'	0.47"'	2.03'
		(0.10)	(0.10)	(0.10)	(0.11)	(0.13)	(0.12)	(1.14)
Age		-0.01	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
		(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Race		-0.09	-0.08	-0.06	-0.01	-0.02	-0.02	-0.24
		(0.10)	(0.10)	(0.10)	(0.11)	(0.12)	(0.12)	(0.18)
Math			0.08'	0.07'	0.08"	0.07	0.08	0.12'
			(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.07)
Depression				0.09	0.09	0.09	0.10	-0.01
				(0.10)	(0.10)	(0.12)	(0.12)	(0.16)
ADHD				-0.17	-0.10	-0.25	-0.19	-3.51
				(0.24)	(0.24)	(0.27)	(0.27)	(126.4)
Health				0.12'	0.12'	-0.02	-0.01	0.02
				(0.06)	(0.07)	(0.07)	(0.07)	(0.08)
Siblings					0.02	0.01	0.01	0.02
					(0.02)	(0.02)	(0.02)	(0.02)
Personal Income						-0.20"'	-0.20"'	-0.18"'
						(0.02)	(0.02)	(0.04)

Standard errors in parentheses
"' p<0.01, " p<0.05, ' p<0.1

Table A.17.: Transition from unemployment to unemployment(cont.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	PerCh	Basic	Basicmath	Add	Parents	Full	Big5mean	Interactions
Male×Married								0.33
								(0.24)
Male×EducationW5								0.07
								(0.18)
Male×Depression								0.22
								(0.23)
Male×ADHD								3.35
								(126.4)
Male×Conscientiousness								-0.07
								(0.16)
Male×Extraversion								-0.19
								(0.13)
Male×Agreeableness								-0.06
								(0.16)
Male×Neuroticism								0.01
								(0.15)
Male×Openness								-0.09
								(0.16)
Male×Race								0.47’
								(0.24)
Male×Personal income								-0.05
								(0.05)
Male×Health								-0.23’
								(0.13)
Male×Math								-0.09
								(0.10)
Conscientiousness	-0.09	-0.07	-0.06	-0.05	-0.03	-0.01		
	(0.06)	(0.07)	(0.07)	(0.07)	(0.08)	(0.09)		
Extraversion	0.01	0.03	0.04	0.05	-0.01	0.01		
	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)	(0.08)		
Agreeableness	-0.16”	0.05	0.07	0.07	0.06	0.06		
	(0.06)	(0.08)	(0.09)	(0.09)	(0.09)	(0.10)		
Neuroticism	0.24”’	0.22”’	0.21”’	0.19”’	0.14’	0.09		
	(0.05)	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)		
Openness	-0.05	-0.04	-0.05	-0.05	-0.02	0.01		
	(0.06)	(0.08)	(0.08)	(0.08)	(0.09)	(0.10)		
Constant	-1.94”’	-1.34	-1.25	-1.42	-1.49	-1.00	-0.64	-0.60
	(0.39)	(1.04)	(1.07)	(1.09)	(1.16)	(1.30)	(1.17)	(1.22)
Observations	10,853	8,177	8,016	8,004	7,454	7,363	7,400	7,363

Standard errors in parentheses
”’ p<0.01, ” p<0.05, ’ p<0.1

Table A.18.: Transition from unemployment to unemployment with interactions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	**PerCh**	**basic**	**basicmath**	**Add**	**Parents**	**Full**	**big5mean**
Conscientiousness	-0.09	-0.06	-0.06	-0.05	-0.04	0.01	
	(0.08)	(0.07)	(0.07)	(0.07)	(0.08)	(0.09)	
Extraversion	0.08	0.03	0.04	0.04	-0.01	0.01	
	(0.08)	(0.06)	(0.06)	(0.06)	(0.07)	(0.08)	
Agreeableness	-0.10	0.06	0.08	0.08	0.07	0.08	
	(0.10)	(0.08)	(0.09)	(0.09)	(0.09)	(0.11)	
Neuroticism	0.28”’	0.22”’	0.22”’	0.21”’	0.16”	0.10	
	(0.08)	(0.07)	(0.07)	(0.07)	(0.08)	(0.09)	
Openness	-0.11	-0.04	-0.05	-0.06	-0.03	-0.01	
	(0.10)	(0.08)	(0.08)	(0.08)	(0.09)	(0.10)	
EducationW5		-0.43”’	-0.41”’	-0.38”’	-0.35”’	-0.21”	-0.23”’
		(0.10)	(0.10)	(0.10)	(0.11)	(0.09)	(0.09)
UrbanW4		-0.09	-0.08	-0.07	-0.01	0.08	0.04
		(0.10)	(0.10)	(0.10)	(0.11)	(0.12)	(0.12)
MarriedW5		-0.39”’	-0.40”’	-0.39”’	-0.50”’	-0.52”’	-0.40”’
		(0.09)	(0.10)	(0.10)	(0.10)	(0.17)	(0.12)
Male	0.51	-0.04	0.08	0.27	0.87’	0.97”	1.75’
	(0.80)	(0.32)	(0.41)	(0.47)	(0.52)	(0.45)	(1.03)
Age		-0.01	-0.02	-0.03	-0.02	-0.02	-0.02
		(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Race		-0.26’	-0.26’	-0.25’	-0.02	-0.22	-0.03
		(0.14)	(0.14)	(0.15)	(0.11)	(0.18)	(0.12)
Math			0.10’	0.10’	0.13”	0.12’	0.07
			(0.05)	(0.05)	(0.06)	(0.07)	(0.05)
Depression				-0.04	-0.01	-0.03	0.10
				(0.14)	(0.15)	(0.16)	(0.12)
ADHD				-3.45	-3.04	-3.49	-0.23
				(182.4)	(74.97)	(127.8)	(0.281)
Health5				0.15”	0.14”	0.01	-0.01
				(0.07)	(0.07)	(0.08)	(0.07)

Standard errors in parentheses
”’ p<0.01, ” p<0.05, ’ p<0.1

Table A.19.: Transition from unemployment to unemployment with interactions(Cont.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	PerCh	basic	basicmath	Add	Parents	Full	big5mean
Male×Depression				0.24	0.19	0.25	
				(0.19)	(0.20)	(0.23)	
Male×ADHD				3.38	3.07	3.30	
				(182.4)	(74.98)	(127.8)	
Male×Race		0.32'	0.32'	0.34'		0.41'	
		(0.19)	(0.19)	(0.20)		(0.24)	
Male×Health				-0.17	-0.17	-0.19	
				(0.11)	(0.11)	(0.13)	
Male×Math			-0.04	-0.03	-0.08	-0.09	
			(0.08)	(0.08)	(0.08)	(0.10)	
Male×EducationW5		0.06	0.05	0.03	-0.02		
		(0.13)	(0.14)	(0.14)	(0.15)		
Male×Conscientiousness	0.03						-0.11
	(0.11)						(0.15)
Male×Extraversion	-0.14						-0.18
	(0.11)						(0.13)
Male×Agreeableness	-0.01						-0.03
	(0.13)						(0.16)
Male×Neuroticism	-0.02						0.01
	(0.11)						(0.15)
Male×Openness	0.04						-0.07
	(0.13)						(0.15)
Siblings					0.05'	0.04	0.01
					(0.03)	(0.03)	(0.02)
Male×Sibings					-0.07	-0.07	
					(0.04)	(0.05)	
Personal Income						-0.18"'	-0.20"'
						(0.04)	(0.02)
Male×Personla Income5						-0.05	
						(0.05)	
Male×Married						0.30	
						(0.24)	
ConscientiousnessD							0.21
							(0.14)
ExtraversionD							0.15
							(0.14)
AgreeablenessD							0.08
							(0.14)
NeuroticismD							0.06
							(0.14)
OpennessD							0.07
							(0.14)
Constant	-2.41"'	-1.27	-1.21	-1.29	-1.72	-1.25	-0.72
	(0.58)	(1.06)	(1.09)	(1.11)	(1.21)	(1.36)	(1.18)
Observations	10,853	8,177	8,016	8,004	7,454	7,363	7,363

Standard errors in parentheses
"' $p<0.01$, " $p<0.05$, ' $p<0.1$

Table A.20.: Transition from unemployment to employment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	PerCh	basic	basicmath	Add	Parents	Full	big5mean	interact
ConscientiousnessD							-0.01	0.01
							(0.06)	(0.07)
ExtraversionD							0.02	0.04
							(0.06)	(0.07)
AgreeablenessD							0.01	-0.02
							(0.06)	(0.08)
NeuroticismD							0.08	0.05
							(0.06)	(0.08)
OpennessD							0.03	0.01
							(0.06)	(0.08)
EducationW5		-0.20"'	-0.20"'	-0.19"'	-0.20"'	-0.08	-0.06	-0.01
		(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.08)
UrbanW4		-0.11'	-0.11'	-0.11'	-0.10	-0.08	-0.08	-0.07
		(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
MarriedW5		-0.23"'	-0.23"'	-0.22"'	-0.24"'	-0.15"	-0.16"	-0.17'
		(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.09)
Male		0.08	0.08	0.06	0.08	0.18"'	0.16"	-0.01
		(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.66)
Age		0.01	0.01	0.01	0.01	0.01	0.01	0.01
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Race		-0.22"'	-0.20"'	-0.22"'	-0.21"'	-0.18"'	-0.19"'	-0.23"'
		(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.09)
Math			0.04	0.04	0.04	0.02	0.02	0.05
			(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Depression				-0.01	-0.01	-0.01	0.01	0.02
				(0.06)	(0.06)	(0.06)	(0.06)	(0.09)
ADHD				0.41"'	0.42"'	0.38"'	0.41"'	0.66"'
				(0.12)	(0.12)	(0.13)	(0.13)	(0.21)
Health5				0.03	0.02	-0.03	-0.03	-0.05
				(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Sibings					0.03'	0.02	0.02	0.02
					(0.01)	(0.01)	(0.01)	(0.01)
Personal Income						-0.10"'	-0.10"'	-0.11"'
						(0.01)	(0.01)	(0.02)

Standard errors in parentheses
"' p<0.01, " p<0.05, ' p<0.1

Table A.21.: Transition from unemployment to employment(Cont.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	PerCh	basic	basicmath	Add	Parents	Full	big5mean	interact
MaleMarried								0.02
								(0.13)
MaleEducationW5								-0.12
								(0.11)
MaleDepression								-0.05
								(0.13)
MaleADHD								-0.39
								(0.26)
Malecons								-0.02
								(0.08)
Maleextra								-0.03
								(0.07)
Maleagree								0.05
								(0.09)
Maleneuro								0.03
								(0.08)
Maleopen								0.06
								(0.09)
MaleRace								0.11
								(0.13)
Maleinc5								0.03
								(0.02)
MaleHealth								0.08
								(0.07)
Malemath								-0.06
								(0.05)
Conscientiousness	-0.02	-0.05	-0.04	-0.02	-0.04	-0.03		
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)		
Extraversion	-0.01	0.01	0.01	-0.01	0.01	0.01		
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)		
Agreeableness	-0.07	0.02	0.02	0.01	0.05	0.04		
	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)		
Neuroticism	0.14"'	0.13"'	0.13"'	0.12"'	0.12"'	0.09'		
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)		
Openness	0.07'	0.08	0.07	0.06	0.05	0.06		
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)		
Constant	-2.02"'	-1.88"'	-1.80"'	-1.92"'	-1.86"'	-1.69"	-1.16'	-1.26'
	(0.26)	(0.66)	(0.67)	(0.68)	(0.70)	(0.72)	(0.63)	(0.66)
Observations	8,718	6,642	6,529	6,520	6,085	6,040	6,068	6,040

Standard errors in parentheses
"' p<0.01, " p<0.05, ' p<0.1

Table A.22.: Transition from unemployment to employment with interactions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	PerCh	basic	basicmath	Add	Parents	Full	big5mean
Conscientiousness	-0.03	-0.05	-0.03	-0.01	-0.03	-0.02	
	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	
Extraversion	0.01	0.01	-0.01	-0.01	0.01	0.01	
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	
Agreeableness	-0.08	0.02	0.02	0.02	0.05	0.04	
	(0.07)	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)	
Neuroticism	0.13”’	0.13”’	0.13”’	0.11”	0.12”’	0.08’	
	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	
Openness	0.06	0.07	0.07	0.07	0.06	0.06	
	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	
EducationW5		-0.18”	-0.16”	-0.16”	-0.17”	-0.08	-0.07
		(0.07)	(0.07)	(0.08)	(0.08)	(0.06)	(0.06)
UrbanW4		-0.11’	-0.11’	-0.11’	-0.09	-0.07	-0.08
		(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
MarriedW5		-0.23”’	-0.23”’	-0.23”’	-0.25”’	-0.16’	-0.16”
		(0.06)	(0.06)	(0.06)	(0.06)	(0.09)	(0.06)
Male	-0.32	0.11	0.35	0.19	0.26	-0.02	-0.02
	(0.53)	(0.26)	(0.31)	(0.34)	(0.36)	(0.26)	(0.57)
Age		0.01	0.01	0.01	0.01	0.01	0.01
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Race		-0.28”’	-0.28”’	-0.30”’	-0.21”’	-0.25”’	-0.18”’
		(0.08)	(0.08)	(0.08)	(0.06)	(0.09)	(0.06)
Math			0.07”	0.07”	0.06’	0.05	0.02
			(0.03)	(0.03)	(0.03)	(0.04)	(0.03)
Depression				0.04	0.03	0.01	-0.01
				(0.08)	(0.08)	(0.09)	(0.06)
ADHD				0.58”’	0.62”’	0.62”’	0.40”’
				(0.20)	(0.20)	(0.21)	(0.13)
Health5				0.02	0.01	-0.05	-0.03
				(0.04)	(0.04)	(0.05)	(0.04)
Observations	8,718	6,642	6,529	6,520	6,085	6,040	6,040

Standard errors in parentheses
”’ p<0.01, ” p<0.05, ’ p<0.1

Table A.23.: Transition from unemployment to employment with interactions(Cont.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	PerCh	basic	basicmath	Add	Parents	Full	big5mean
Male×Depression				-0.09	-0.07	-0.05	
				(0.12)	(0.12)	(0.13)	
Male×ADHD				-0.28	-0.31	-0.37	
				(0.25)	(0.25)	(0.26)	
Male×Race		0.14	0.14	0.15		0.16	
		(0.11	(0.11)	(0.12)		(0.13)	
Male×Health				0.10	0.08	0.08	
				(0.06)	(0.07)	(0.07)	
Male×Math			-0.06	-0.06	-0.06	-0.05	
			(0.05)	(0.05)	(0.05)	(0.05)	
Male×EducationW5		-0.04	-0.07	-0.06	-0.05		
		(0.10)	(0.10)	(0.10)	(0.10)		
Male×Conscientiousness	0.03						-0.02
	(0.07)						(0.08)
Male×Extraversion	-0.03						-0.03
	(0.07)						(0.07)
Male×Agreeableness	0.05						0.04
	(0.09)						(0.09)
Male×Neuroticism	0.04						0.02
	(0.07)						(0.08)
Male×Openness	0.01						0.04
	(0.09)						(0.09)
Siblings					0.02	0.01	0.02
					(0.02)	(0.02)	(0.01)
Male×Siblings					0.01	0.02	
					(0.03)	(0.03)	
Personal Income						-0.11”’	-0.10”’
						(0.01)	(0.01)
Male×Personal Income5						0.02	
						(0.02)	
Male×Married						0.01	
						(0.13)	
ConscientiousnessD							-0.01
							(0.07)
ExtraversionD							0.04
							(0.07)
AgreeablenessD							-0.02
							(0.07)
NeuroticismD							0.07
							(0.07)
OpennessD							0.01
							(0.07)
Constant	-1.92”’	-1.90”’	-1.91”’	-2.04”’	-1.98”’	-1.62”	-1.17’
	(0.37)	(0.67)	(0.68)	(0.70)	(0.72)	(0.73)	(0.64)
Observations	8,718	6,642	6,529	6,520	6,085	6,040	6,040

Standard errors in parentheses
”’ p<0.01, ” p<0.05, ’ p<0.1

Table A.24.: Transition from employment to unemployment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	PerCh	basic	basicmath	Add	Fam	Full	big5mean	interact
ConscientiousnessD							0.07	0.02
							(0.07)	(0.08)
ExtraversionD							-0.08	-0.16’
							(0.07)	(0.08)
AgreeablenessD							-0.03	-0.04
							(0.07)	(0.08)
NeuroticismD							-0.11	-0.10
							(0.07)	(0.08)
OpennessD							-0.04	0.01
							(0.07)	(0.08)
EducationW5		-0.30”’	-0.30”’	-0.24”’	-0.23”’	-0.02	-0.01	-0.02
		(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.09)
UrbanW4		0.02	0.01	0.01	-0.03	0.02	0.03	0.03
		(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
MarriedW5		-0.36”’	-0.35”’	-0.33”’	-0.30”’	-0.16”	-0.15”	-0.08
		(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.10)
Male		0.01	0.02	0.03	-0.03	0.17”	0.17”	0.36
		(0.06)	(0.06)	(0.07)	(0.07)	(0.08)	(0.07)	(0.76)
Age		0.01	0.01	0.01	0.01	0.03	0.02	0.02
		(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Race		-0.13”	-0.13”	-0.11’	-0.10	-0.06	-0.06	-0.11
		(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.10)
Math			-0.01	-0.01	0.01	-0.02	-0.02	0.01
			(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
Depression				0.03	0.02	0.02	0.01	0.04
				(0.06)	(0.06)	(0.07)	(0.07)	(0.09)
ADHD				0.09	0.07	0.02	-0.01	0.22
				(0.15)	(0.16)	(0.17)	(0.17)	(0.27)
Health5				0.26”’	0.24”’	0.15”’	0.16”’	0.14”’
				(0.04)	(0.04)	(0.05)	(0.05)	(0.05)
Siblings					0.01	0.01	0.01	0.01
					(0.01)	(0.02)	(0.02)	(0.02)
Personal Income						-0.18”’	-0.17”’	-0.16”’
						(0.01)	(0.01)	(0.02)
Observations	8,718	6,642	6,529	6,520	6,085	6,040	6,068	6,040

Standard errors in parentheses
”’ p<0.01, ” p<0.05, ’ p<0.1

Table A.25.: Transition from employment to unemployment(Cont.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	PerCh	basic	basicmath	Add	Fam	Full	big5mean	interact
Male×Married								-0.13
								(0.14)
Male×EducationW5								0.02
								(0.12)
Male×Depression								-0.01
								(0.14)
Male×ADHD								-0.31
								(0.36)
Male×Conscientiousness								0.09
								(0.10)
Male×Extraversion								0.12
								(0.08)
Male×Agreeableness								-0.01
								(0.11)
Male×Neuroticism								-0.07
								(0.10)
Male×Openness								-0.13
								(0.11)
Male×Race								0.12
								(0.14)
Male×Personal Income5								-0.04’
								(0.02)
Male×Health								0.05
								(0.08)
Male×Math								-0.06
								(0.06)
Conscientiousness	-0.01	-0.01	0.01	0.05	0.08	0.07		
	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)		
Extraversion	-0.03	-0.03	-0.03	-0.02	-0.04	-0.03		
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)		
Agreeableness	-0.13”’	-0.06	-0.06	-0.05	-0.04	-0.02		
	(0.04)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)		
Neuroticism	0.11”’	0.04	0.05	0.01	0.01	-0.07		
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)		
Openness	-0.03	0.01	-0.01	-0.02	-0.02	-0.02		
	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)		
Constant	-1.28”’	-0.97	-1.02	-1.52”	-1.64”	-1.56’	-1.67”	-1.70”
	(0.26)	(0.67)	(0.68)	(0.70)	(0.73)	(0.80)	(0.71)	(0.73)
Observations	8,718	6,642	6,529	6,520	6,085	6,040	6,068	6,040

Standard errors in parentheses
”’ p<0.01, ” p<0.05, ’ p<0.1

Table A.26.: Transition from employment to unemployment with interactions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	PerCh	basic	basicmath	Add	Fam	Full	big5mean
Conscientiousness	-0.01	-0.01	0.01	0.05	0.08’	0.08	
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	
Extraversion	-0.01	-0.03	-0.03	-0.02	-0.04	-0.03	
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	
Agreeableness	-0.17”	-0.06	-0.06	-0.05	-0.04	-0.01	
	(0.07)	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)	
Neuroticism	0.11”	0.04	0.05	0.01	0.01	-0.07	
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	
Openness	-0.03	0.01	-0.01	-0.02	-0.02	-0.03	
	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	
EducationW5		-0.29”’	-0.27”’	-0.21”’	-0.22”’	-0.02	-0.01
		(0.07)	(0.07)	(0.07)	(0.08)	(0.06)	(0.06)
UrbanW4		0.02	0.01	0.01	-0.03	0.03	0.02
		(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)
MarriedW5		-0.36”’	-0.35”’	-0.33”’	-0.30”’	-0.09	-0.16”
		(0.06)	(0.06)	(0.06)	(0.06)	(0.10)	(0.07)
Male	-0.11	0.04	0.29	0.37	0.13	0.42	0.10
	(0.53)	(0.25)	(0.31)	(0.34)	(0.37)	(0.29)	(0.66)
Age		0.01	0.01	0.01	0.01	0.02	0.02
		(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Race		-0.13	-0.12	-0.10	-0.10	-0.12	-0.06
		(0.08)	(0.08)	(0.08)	(0.07)	(0.10)	(0.07)
Math			0.02	0.02	0.03	0.01	-0.02
			(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
Depression				0.06	0.07	0.04	0.02
				(0.08)	(0.08)	(0.09)	(0.07)
ADHD				0.18	0.25	0.23	0.02
				(0.25)	(0.25)	(0.27)	(0.17)
Health5				0.26”’	0.24”’	0.15”’	0.15”’
				(0.04)	(0.04)	(0.05)	(0.05)
Observations	8,718	6,642	6,529	6,520	6,085	6,040	6,040

Standard errors in parentheses
”’ p<0.01, ” p<0.05, ’ p<0.1

Table A.27.: Transition from employment to unemployment with interactions(Cont.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	PerCh	basic	basicmath	Add	Fam	Full	big5mean
Male×Depression				-0.06	-0.13	-0.04	
				(0.12)	(0.13)	(0.14)	
Male×ADHD				-0.13	-0.26	-0.34	
				(0.31)	(0.32)	(0.35)	
Male×Race		0.01	-0.02	-0.02		0.14	
		(0.11)	(0.12)	(0.12)		(0.14)	
Male×Health				-0.01	0.03	0.03	
				(0.07)	(0.07)	(0.08)	
Male×Math			-0.05	-0.05	-0.05	-0.06	
			(0.05)	(0.05)	(0.05)	(0.06)	
Male×EducationW5		-0.01	-0.05	-0.06	-0.03		
		(0.09)	(0.10)	(0.10)	(0.11)		
Male×Conscientiousness	-0.01						0.09
	(0.07)						(0.10)
Male×Extraversion	-0.04						0.11
	(0.07)						(0.08)
Male×Agreeableness	0.10						-0.02
	(0.09)						(0.10)
Male×Neuroticism	0.01						-0.05
	(0.07)						(0.10)
Male×Openness	-0.02						-0.12
	(0.09)						(0.10)
Siblings					0.01	-0.01	0.01
					(0.02)	(0.02)	(0.02)
Male×Siblings					0.02	0.02	
					(0.03)	(0.03)	
Personal Income						-0.16”’	-0.18”’
						(0.02)	(0.01)
Male×Personal Income5						-0.03	
						(0.02)	
Male×Married						-0.11	
						(0.14)	
ConscientiousnessD							0.02
							(0.08)
ExtraversionD							-0.15’
							(0.08)
AgreeablenessD							-0.03
							(0.08)
NeuroticismD							-0.10
							(0.08)
OpennessD							0.01
							(0.08)
Constant	-1.26”’	-0.98	-1.13	-1.66”	-1.71”	-1.67”	-1.57”
	(0.37)	(0.68)	(0.70)	(0.71)	(0.75)	(0.81)	(0.72)
Observations	8,718	6,642	6,529	6,520	6,085	6,040	6,040

Standard errors in parentheses
”’ $p<0.01$, ” $p<0.05$, ’ $p<0.1$

Table A.28.: Transition from employment to employment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	PerCh	basic	basicmath	Add	Parents	Full	big5mean	interact
ConscientiousnessD							-0.05	-0.03
							(0.04)	(0.04)
ExtraversionD							0.06’	0.10”
							(0.04)	(0.04)
AgreeablenessD							0.01	0.02
							(0.04)	(0.04)
NeuroticismD							-0.01	0.01
							(0.04)	(0.04)
OpennessD							-0.02	-0.02
							(0.04)	(0.04)
EducationW5		0.36”’	0.35”’	0.30”’	0.29”’	0.07”	0.06’	0.17”’
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
UrbanW4		0.02	0.03	0.02	0.02	0.02	0.02	0.02
		(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
MarriedW5		0.09”’	0.09”	0.04	0.04	-0.12”’	-0.12”’	-0.27”’
		(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Male		0.36”’	0.35”’	0.37”’	0.36”’	0.10”	0.12”’	0.37
		(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.40)
Age		0.01	0.01	0.01	0.02’	0.01	0.01	0.01
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Race		-0.03	-0.03	-0.04	-0.07’	-0.05	-0.05	-0.10”
		(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)
Math			-0.02’	-0.02	-0.02	0.01	0.01	-0.01
			(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Depression				-0.05	-0.05	-0.05	-0.05	-0.06
				(0.03)	(0.03)	(0.04)	(0.04)	(0.05)
ADHD				-0.35”’	-0.36”’	-0.30”’	-0.30”’	-0.47”’
				(0.07)	(0.08)	(0.08)	(0.08)	(0.13)
Health5				-0.24”’	-0.23”’	-0.10”’	-0.10”’	-0.11”’
				(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Sibings					-0.03”’	-0.02”	-0.02”	-0.02”
					(0.01)	(0.01)	(0.01)	(0.01)
Personal Income						0.19”’	0.19”’	0.18”’
						(0.01)	(0.01)	(0.01)
Observations	10,853	8,177	8,016	8,004	7,454	7,363	7,400	7,363

Standard errors in parentheses
”’ $p<0.01$, ” $p<0.05$, ’ $p<0.1$

Table A.29.: Transition from employment to employment(Cont.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	PerCh	basic	basicmath	Add	Parents	Full	big5mean	interact
MaleMarried								0.36"'
								(0.08)
MaleEducationW5								-0.23"'
								(0.07)
MaleDepression								-0.01
								(0.08)
MaleADHD								0.28'
								(0.17)
Malecons								-0.02
								(0.05)
Maleextra								-0.06
								(0.04)
Maleagree								0.03
								(0.06)
Maleneuro								-0.02
								(0.05)
Maleopen								-0.01
								(0.06)
MaleRace								0.14'
								(0.07)
Maleinc5								0.02
								(0.01)
MaleHealth								0.01
								(0.05)
Malemath								0.05
								(0.03)
Conscientiousness	0.03'	0.04	0.03	-0.01	-0.02	-0.03		
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)		
Extraversion	0.05"'	0.05"	0.05"	0.05"	0.06"'	0.03		
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)		
Agreeableness	-0.03	-0.02	-0.02	-0.03	-0.04	-0.02		
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)		
Neuroticism	-0.18"'	-0.10"'	-0.10"'	-0.06"	-0.06"	-0.02		
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)		
Openness	0.06"'	-0.02	-0.03	-0.02	-0.01	-0.01		
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)		
Constant	0.69"'	-0.49	-0.53	-0.09	-0.06	-0.38	-0.54	-0.65'
	(0.14)	(0.36)	(0.36)	(0.37)	(0.38)	(0.41)	(0.37)	(0.38)
Observations	10,853	8,177	8,016	8,004	7,454	7,363	7,400	7,363

Standard errors in parentheses
"' p<0.01, " p<0.05, ' p<0.1

Table A.30.: Transition from employment to employment with interactions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	PerCh	basic	basicmath	Add	Parents	Full	big5mean
Conscientiousness	0.05”	0.04’	0.03	-0.01	-0.02	-0.03	
	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	
Extraversion	0.05”	0.05”	0.05”	0.05”	0.06”’	0.03	
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Agreeableness	0.03	-0.01	-0.01	-0.02	-0.04	-0.01	
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Neuroticism	-0.12”’	-0.10”’	-0.10”’	-0.06”	-0.06”	-0.02	
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	
Openness	0.06”	-0.03	-0.03	-0.02	-0.01	-0.01	
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
EducationW5		0.43”’	0.41”’	0.36”’	0.36”’	0.07”	0.07”
		(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)
UrbanW4		0.02	0.03	0.02	0.02	0.02	0.02
		(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
MarriedW5		0.10”’	0.09”’	0.05	0.04	-0.27”’	-0.12”’
		(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)
Male	0.76”’	0.65”’	0.54”’	0.60”’	0.83”’	-0.45”’	0.56
	(0.29)	(0.15)	(0.17)	(0.19)	(0.21)	(0.16)	(0.34)
Age		0.01	0.01	0.01	0.02’	0.01	0.01
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Race		-0.13”’	-0.13”’	-0.15”’	-0.07’	-0.09’	-0.05
		(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)
Math			-0.03’	-0.02	-0.03	-0.02	0.01
			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Depression				-0.06	-0.05	-0.06	-0.06
				(0.04)	(0.04)	(0.05)	(0.04)
ADHD				-0.39”’	-0.44”’	-0.46”’	-0.30”’
				(0.11)	(0.12)	(0.13)	(0.08)
Health5				-0.24”’	-0.22”’	-0.11”’	-0.10”’
				(0.02)	(0.02)	(0.03)	(0.03)
Observations	10,853	8,177	8,016	8,004	7,454	7,363	7,363

Standard errors in parentheses
”’ p<0.01, ” p<0.05, ’ p<0.1

Table A.31.: Transition from employment to employment with interactions(Cont.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	PerCh	basic	basicmath	Add	Parents	Full	big5mean
Male×Depression				0.02	0.01	0.01	
				(0.07)	(0.07)	(0.08)	
Male×ADHD				0.03	0.14	0.27	
				(0.15)	(0.16)	(0.17)	
Male×Race		0.26”’	0.25”’	0.27”’		0.12’	
		(0.06)	(0.06)	(0.07)		(0.08)	
Male×Health				-0.04	-0.04	0.03	
				(0.04)	(0.04)	(0.04)	
Male×Math			0.02	0.02	0.02	0.06’	
			(0.03)	(0.03)	(0.03)	(0.03)	
Male×EducationW5		-0.17”’	-0.14”’	-0.15”’	-0.16”’		
		(0.05)	(0.06)	(0.06)	(0.06)		
Male×Conscientiousness	-0.01						-0.04
	(0.04)						(0.05)
Male×Extraversion	0.02						-0.05
	(0.04)						(0.04)
Male×Agreeableness	0.03						0.02
	(0.05)						(0.05)
Male×Neuroticism	-0.07						-0.02
	(0.04)						(0.05)
Male×Openness	-0.11”						-0.05
	(0.05)						(0.05)
Siblings					-0.03”’	-0.02’	-0.02”
					(0.01)	(0.01)	(0.01)
Male×Siblings					-0.02	-0.01	
					(0.02)	(0.02)	
Personal Income						0.18”’	0.19”’
						(0.01)	(0.01)
Male×Personal Income5						0.01	
						(0.01)	
Male×Married						0.36”’	
						(0.08)	
ConscientiousnessD							-0.03
							(0.04)
ExtraversionD							0.09”
							(0.04)
AgreeablenessD							0.01
							(0.04)
NeuroticismD							0.01
							(0.04)
OpennessD							-0.01
							(0.04)
Constant	0.10	-0.63’	-0.64’	-0.19	-0.26	-0.26	-0.60
	(0.18)	(0.36)	(0.37)	(0.38)	(0.39)	(0.42)	(0.39)
Observations	10,853	8,177	8,016	8,004	7,454	7,363	7,363

Standard errors in parentheses
”’ p<0.01, ” p<0.05, ’ p<0.1

Table A.32.: Employed in WaveIV

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	PerCh	basic	basicmath	Add	Parents	Full	big5mean
Conscientiousness	0.06’	0.04	0.03	0.01	0.01	-0.02	
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	
Extraversion	0.01	0.03	0.03	0.03	0.02	-0.02	
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Agreeableness	0.10”’	-0.01	-0.01	-0.01	-0.01	–0.01	
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	
Neuroticism	-0.21”’	-0.17”’	-0.18”’	-0.14”’	-0.14”’	-0.11”’	
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	
Openness	-0.01	-0.05	-0.06	-0.05	-0.05	-0.04	
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	
EducationW4		0.44”’	0.42”’	0.39”’	0.39”’	0.17”’	0.17”’
		(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
UrbanW4		0.09’	0.10’	0.10’	0.09’	0.12”	0.12”
		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Married		0.29”’	0.29”’	0.29”’	0.29”’	0.21”’	0.21”’
		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Male		-0.05	-0.05	-0.05	-0.06	-0.23”’	-0.22”’
		(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
Age		-0.01	0.01	-0.01	-0.01	-0.03’	-0.03’
		(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Race		0.17”’	0.18”’	0.18”’	0.16”’	0.16”’	0.17”’
		(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Math			-0.05”	-0.05”	-0.05”	-0.02	-0.02
			(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Depression				-0.01	-0.01	0.03	0.02
				(0.05)	(0.05)	(0.05)	(0.05)
ADHD				-0.25”	-0.26”’	-0.19’	-0.20’
				(0.10)	(0.10)	(0.11)	(0.11)
Health				-0.15”’	-0.15”’	-0.13”’	-0.14”’
				(0.03)	(0.03)	(0.04)	(0.04)
Sibings					-0.02’	-0.01	-0.01
					(0.01)	(0.01)	(0.01)
Personal Income						0.17”’	0.17”’
						(0.01)	(0.01)
ConscientiousnessD							-0.06
							(0.05)
ExtraversionD							-0.05
							(0.05)
AgreeablenessD							0.01
							(0.05)
NeuroticismD							-0.10’
							(0.05)
OpennessD							0.01
							(0.05)
Constant	1.55”’	1.00”	0.99”	1.34”’	1.50”’	1.85”’	1.33”’
	(0.22)	(0.45)	(0.46)	(0.47)	(0.48)	(0.52)	(0.44)
Observations	9,442	8,704	8,543	8,539	8,530	8,202	8,233

Table A.33.: Employed in Wave IV with interactions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	PerChInt	basicint	basicmathint	Addint	Parentsint	Fullint	big5meanint
Conscientiousness	0.08’	0.05	0.03	0.01	0.01	-0.02	
	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	
Extraversion	-0.02	0.03	0.03	0.03	0.02	-0.02	
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	
Agreeableness	0.10’	-0.01	-0.01	-0.01	-0.01	-0.01	
	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	
Neuroticism	-0.19”’	-0.17”’	-0.17”’	-0.14”’	-0.14”’	-0.11”’	
	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	
Openness	0.01	-0.05	-0.05	-0.05	-0.05	-0.03	
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	
EducationW4		0.47”’	0.45”’	0.41”’	0.41”’	0.18”’	0.17”’
		(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)
UrbanW4		0.09’	0.10’	0.09’	0.09’	0.12”	0.12”
		(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
Married		0.20”’	0.29”’	0.29”’	0.29”’	0.19”’	0.22”’
		(0.06)	(0.05)	(0.05)	(0.05)	(0.07)	(0.05)
Male	0.42	0.01	0.15	0.08	0.08	-0.20	0.06
	(0.45)	(0.18)	(0.23)	(0.27)	(0.28)	(0.22)	(0.48)
Age		-0.01	0.01	-0.01	-0.01	-0.03’	-0.03’
		(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Race		0.19”’	0.20”’	0.19”’	0.16”’	0.24”’	0.16”’
		(0.06)	(0.06)	(0.06)	(0.05)	(0.07)	(0.05)
Observations	9,442	8,704	8,543	8,539	8,530	8,202	8,202

Standard errors in parentheses
”’ p<0.01, ” p<0.05, ’ p<0.1

Table A.34.: Employed in Wave IV with interactions(Cont.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	PerChInt	basicint	basicmathint	Addint	Parentsint	Fullint	big5meanint
Male×Race		-0.04	-0.04	-0.04		-0.18’	
		(0.09)	(0.09)	(0.09)		(0.10)	
Male×EducationW4		-0.05	-0.05	-0.05	-0.06		
		(0.08)	(0.08)	(0.09)	(0.09)		
Male×Married		0.19”				0.04	
		(0.09)				(0.10)	
Male×Conscientiousness	-0.05						0.01
	(0.06)						(0.07)
Male×Extraversion	0.05						0.06
	(0.06)						(0.06)
Male×Agreeableness	-0.05						-0.03
	(0.08)						(0.08)
Male×Neuroticism	-0.08						-0.13’
	(0.06)						(0.07)
Male×Openness	-0.02						-0.02
	(0.07)						(0.07)
Math			-0.04	-0.04	-0.04	-0.01	-0.02
			(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Male×Math			-0.03	-0.02	-0.02	-0.04	
			(0.04)	(0.04)	(0.04)	(0.04)	
Depression				0.03	0.03	0.06	0.03
				(0.06)	(0.06)	(0.07)	(0.05)
ADHD				-0.31’	-0.31’	-0.22	-0.20’
				(0.16)	(0.16)	(0.18)	(0.11)
Health				-0.17”’	-0.17”’	-0.18”’	-0.14”’
				(0.05)	(0.05)	(0.05)	(0.04)
Male×Depression				-0.09	-0.08	-0.07	
				(0.09)	(0.09)	(0.10)	
Male×ADHD				0.10	0.08	0.06	
				(0.20)	(0.20)	(0.22)	
Male×Health				0.05	0.05	0.10	
				(0.07)	(0.07)	(0.07)	
Constant	1.44”’	0.97”	0.90’	1.28”’	1.42”’	1.81”’	1.33”’
	(0.31)	(0.46)	(0.47)	(0.49)	(0.49)	(0.53)	(0.45)
Observations	9,442	8,704	8,543	8,539	8,530	8,202	8,202

Standard errors in parentheses
”’ p<0.01, ” p<0.05, ’ p<0.1

Table A.35.: Employed in Wave IV with interactions(Cont.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	PerChInt	basicint	basicmathint	Addint	Parentsint	Fullint	big5meanint
Siblings					-0.01	0.01	-0.01
					(0.01)	(0.01)	(0.01)
MaleSiblings					-0.01	-0.02	
					(0.02)	(0.02)	
Personal Income						0.16"'	0.17"'
						(0.01)	(0.01)
Male×Personal Income						0.01	
						(0.02)	
ConscientiousnessD							-0.06
							(0.06)
ExtraversionD							-0.08
							(0.06)
AgreeablenessD							0.02
							(0.06)
NeuroticismD							-0.04
							(0.06)
OpennessD							0.01
							(0.06)
Constant	1.44"'	0.97"	0.90'	1.28"'	1.42"'	1.81"'	1.33"'
	(0.31)	(0.46)	(0.47)	(0.49)	(0.49)	(0.53)	(0.45)
Observations	9,442	8,704	8,543	8,539	8,530	8,202	8,202

Standard errors in parentheses
"' p<0.01, " p<0.05, ' p<0.1

Table A.36.: Employed in Wave V

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	PerCh	basic	basicmath	Add	Parents	Full	big5mean	interact
Conscientiousness	0.03	0.02	0.01	-0.04	-0.06	-0.04		
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)		
Extraversion	0.03	0.03	0.03	0.02	0.04	0.01		
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)		
Agreeableness	0.13”’	0.02	0.01	0.01	0.01	-0.05		
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)		
Neuroticism	-0.20”’	-0.13”’	-0.13”’	-0.08”	-0.07’	-0.01		
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)		
Openness	0.04	-0.01	-0.01	0.01	0.01	0.02		
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)		
EducationW5		0.41”’	0.39”’	0.34”’	0.32”’	0.11”	0.09	0.09
		(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
UrbanW4		0.03	0.03	0.02	0.03	-0.06	-0.04	-0.04
		(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
MarriedW5		0.39”’	0.39”’	0.36”’	0.36”’	0.20”’	0.20”’	0.20”’
		(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
Male		0.03	0.02	0.01	0.04	-0.26”’	-0.21”’	-0.21”’
		(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
Age		-0.01	-0.01	0.01	0.01	-0.02	-0.02	-0.02
		(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Race		0.13”	0.12”	0.10’	0.07	0.06	0.05	0.05
		(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Math			-0.02	-0.01	-0.03	0.02	0.01	0.01
			(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Depression				-0.05	-0.04	-0.04	-0.03	-0.03
				(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
ADHD				-0.15	-0.17	-0.03	-0.07	-0.07
				(0.12)	(0.13)	(0.14)	(0.14)	(0.14)
Health5				-0.29”’	-0.28”’	-0.20”’	-0.20”’	-0.20”’
				(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Sibings					-0.02”		-0.01	-0.01
					(0.01)		(0.01)	(0.01)
Personal Income						0.20”’	0.20”’	0.20”’
						(0.01)	(0.01)	(0.01)
ConscientiousnessD							-0.10	-0.10
							(0.06)	(0.06)
ExtraversionD							0.08	0.08
							(0.06)	(0.06)
AgreeablenessD							-0.01	-0.01
							(0.07)	(0.07)
NeuroticismD							0.02	0.02
							(0.06)	(0.06)
OpennessD							0.01	0.01
							(0.06)	(0.06)
Constant	1.25”’	0.60	0.59	1.16’	1.26”	1.18’	1.07’	1.07’
	(0.22)	(0.57)	(0.58)	(0.60)	(0.63)	(0.68)	(0.62)	(0.62)
Observations	9,679	7,291	7,155	7,143	6,669	7,080	6,648	6,648

Standard errors in parentheses
”’ $p<0.01$, ” $p<0.05$, ’ $p<0.1$

Table A.37.: Employed in Wave V with interactions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Perch	basic	basicmath	Add	Parents	Full	big5mean	big5mean
Conscientiousness	0.01	0.02	0.01	-0.03	-0.06	-0.06		
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)		
Extraversion	0.01	0.03	0.03	0.02	0.04	0.03		
	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)		
Agreeableness	0.18”’	0.02	0.01	0.01	0.01	-0.04		
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)		
Neuroticism	-0.22”’	-0.13”’	-0.13”’	-0.09”	-0.08’	-0.01		
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)		
Openness	0.03	-0.01	-0.01	0.01	0.01	0.01		
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)		
EducationW5		0.45”’	0.41”’	0.36”’	0.33”’	0.10’	0.09’	0.09’
		(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)
UrbanW4		0.03	0.03	0.02	0.03	-0.05	-0.03	-0.03
		(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
MarriedW5		0.38”’	0.39”’	0.36”’	0.36”’	0.13	0.21”’	0.21”’
		(0.05)	(0.05)	(0.05)	(0.05)	(0.08)	(0.06)	(0.06)
Male	-0.14	0.24	0.06	-0.04	-0.09	-0.76”’	-0.65	-0.65
	(0.45)	(0.21)	(0.26)	(0.29)	(0.31)	(0.26)	(0.59)	(0.59)
Age		-0.01	-0.01	0.01	0.01	-0.01	-0.01	-0.01
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Race		0.14”	0.12’	0.10	0.08	0.11	0.04	0.04
		(0.07)	(0.07)	(0.07)	(0.06)	(0.08)	(0.06)	(0.06)
Math			-0.03	-0.03	-0.05	-0.02	0.01	0.01
			(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Depression				-0.03	-0.04	-0.01	-0.03	-0.03
				(0.07)	(0.07)	(0.08)	(0.06)	(0.06)
ADHD				-0.10	-0.18	-0.12	-0.08	-0.08
				(0.20)	(0.20)	(0.22)	(0.14)	(0.14)
Health5				-0.30”’	-0.29”’	-0.20”’	-0.19”’	-0.19”’
				(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Observations	9,679	7,291	7,155	7,143	6,669	6,615	6,615	6,615

Standard errors in parentheses
”’ p<0.01, ” p<0.05, ’ p<0.1

Table A.38.: Employed in Wave V with interactions(Cont.)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Perch	basic	basicmath	Add	Parents	Full	big5mean	big5mean
Male×Depression				-0.03	0.02	-0.06		
				(0.10)	(0.11)	(0.13)		
Male×ADHD				-0.07	0.01	0.11		
				(0.25)	(0.26)	(0.29)		
Male×Race		-0.03	-0.01	0.01		-0.15		
		(0.10)	(0.10)	(0.10)		(0.13)		
Male×Health				0.06	0.02	0.04		
				(0.06)	(0.06)	(0.07)		
Male×Math			0.03	0.03	0.04	0.06		
			(0.04)	(0.04	(0.05)	(0.05)		
Male×EducationW5		-0.08	-0.05	-0.04	-0.03			
		(0.08)	(0.08)	(0.08)	(0.09)			
Male×Conscientiousness	0.03						0.02	0.02
	(0.06)						(0.08)	(0.08)
Male×Extraversion	0.07						-0.01	-0.01
	(0.06)						(0.07)	(0.07)
Male×Agreeableness	-0.11						-0.06	-0.06
	(0.08)						(0.09)	(0.09)
Male×Neuroticism	0.04						0.07	0.07
	(0.06)						(0.08)	(0.08)
Male×Openness	0.02						0.12	0.12
	(0.07)						(0.09)	(0.09)
Siblings					-0.03”	-0.02	-0.01	-0.01
					(0.02)	(0.02)	(0.01)	(0.01)
Male×Siblings					0.02	0.02		
					(0.02)	(0.03)		
Personal Income						0.18”’	0.20”’	0.20”’
						(0.01)	(0.01)	(0.01)
Male×Personal Income						0.05”		
						(0.02)		
Male×Married						0.14		
						(0.13)		
ConscientiousnessD							-0.10	-0.10
							(0.07)	(0.07)
ExtraversionD							0.09	0.09
							(0.07)	(0.07)
AgreeablenessD							0.02	0.02
							(0.07)	(0.07)
NeuroticismD							0.01	0.01
							(0.07)	(0.07)
OpennessD							-0.04	-0.04
							(0.07)	(0.07)
Constant	1.29”’	0.51	0.56	1.12’	1.31”	1.44”	1.03	1.03
	(0.30)	(0.58)	(0.59)	(0.61)	(0.64)	(0.71)	(0.63)	(0.63)
Observations	9,679	7,291	7,155	7,143	6,669	6,615	6,615	6,615

Standard errors in parentheses
”’ p<0.01, ” p<0.05, ’ p<0.1

Table A.39.: Number of unemployment spells(Ordered Probit)

Agreeableness	-0.19"'	-0.03	-0.03	-0.03
	(1)	(3)	(5)	(7)
VARIABLES	PerCh	basic	basicmath	Add
Conscientiousness	-0.12"'	-0.10"'	-0.10"'	-0.08"'
	(0.02)	(0.02)	(0.02)	(0.02)
Extraversion	0.06"'	0.04"	0.04"	0.04"
	(0.02)	(0.02)	(0.02)	(0.02)
	(0.02)	(0.02)	(0.02)	(0.02)
Neuroticism	0.15"'	0.14"'	0.15"'	0.12"'
	(0.02)	(0.02)	(0.02)	(0.02)
Openness	0.13"'	0.12"'	0.11"'	0.11"'
	(0.02)	(0.02)	(0.02)	(0.02)
EducationW4		-0.39"'	-0.39"'	-0.36"'
		(0.02)	(0.02)	(0.03)
UrbanW4		-0.03	-0.03	-0.03
		(0.03)	(0.03)	(0.03)
Married		-0.25"'	-0.25"'	-0.25"'
		(0.03)	(0.03)	(0.03)
Male		0.31"'	0.31"'	0.31"'
		(0.03)	(0.03)	(0.03)
Age		-0.01	-0.01	-0.01
		(0.01)	(0.01)	(0.01)
Race		-0.12"'	-0.11"'	-0.12"'
		(0.03)	(0.03)	(0.03)
Math			0.03"	0.03"
			(0.01)	(0.01)
Depression				0.08"'
				(0.03)
ADHD				0.30"'
				(0.06)
Health				0.09"'
				(0.02)
Observations	10,562	9,727	9,536	9,530

Standard errors in parentheses
"' p<0.01, " p<0.05, ' p<0.1

Table A.40.: Number of unemployment spells(Ordered Probit)(Cont.)

	(1)	(3)	(5)	(7)
VARIABLES	PerCh	basic	basicmath	Add
Constant cut20	3.33”’			
	(0.19)			
Constant cut1	0.46”’	0.04	-0.01	0.15
	(0.13)	(0.27)	(0.27)	(0.28)
Constant cut2	1.25”’	0.88”’	0.84”’	0.99”’
	(0.13)	(0.27)	(0.27)	(0.28)
Constant cut3	1.77”’	1.44”’	1.40”’	1.56”’
	(0.13)	(0.27)	(0.27)	(0.28)
Constant cut4	2.06”’	1.73”’	1.69”’	1.85”’
	(0.13)	(0.27)	(0.27)	(0.28)
Constant cut5	2.23”’	1.92”’	1.87”’	2.03”’
	(0.13)	(0.27)	(0.27)	(0.28)
Constant cut6	2.42”’	2.13”’	2.09”’	2.25”’
	(0.14)	(0.27)	(0.27)	(0.28)
Constant cut7	2.54”’	2.24”’	2.20”’	2.36”’
	(0.14)	(0.27)	(0.28)	(0.28)
Constant cut8	2.64”’	2.31”’	2.27”’	2.43”’
	(0.14)	(0.27)	(0.27)	(0.28)
Constant cut9	2.71”’	2.39”’	2.35”’	2.51”’
	(0.14)	(0.27)	(0.28)	(0.28)
Constant cut10	2.76”’	2.45”’	2.42”’	2.58”’
	(0.14)	(0.27)	(0.28)	(0.28)
Constant cut11	2.92”’	2.61”’	2.59”’	2.75”’
	(0.15)	(0.28)	(0.28)	(0.29)
Constant cut12	2.96”’	2.66”’	2.64”’	2.81”’
	(0.15)	(0.28)	(0.29)	(0.29)
Constant cut13	2.99”’	2.69”’	2.67”’	2.84”’
	(0.15)	(0.28)	(0.29)	(0.29)
Constant cut14	3.04”’	2.72”’	2.70”’	2.87”’
	(0.16)	(0.28)	(0.29)	(0.29)
Constant cut15	3.07”’	2.75”’	2.74”’	2.90”’
	(0.16)	(0.28)	(0.29)	(0.30)
Constant cut16	3.13”’	2.82”’	2.82”’	2.99”’
	(0.17)	(0.29)	(0.30)	(0.30)
Constant cut17	3.17”’	2.87”’	2.87”’	3.04”’
	(0.17)	(0.29)	(0.30)	(0.31)
Constant cut18	3.22”’	2.92”’	2.94”’	3.11”’
	(0.18)	(0.30)	(0.31)	(0.31)
Constant cut19	3.27”’	2.98”’	3.01”’	3.19”’
	(0.18)	(0.30)	(0.31)	(0.32)
Observations	10,562	9,727	9,536	9,530

Standard errors in parentheses
”’ p<0.01, ” p<0.05, ’ p<0.1

Table A.41.: Number of unemployment spells(Poisson regression)

	(1)	(3)	(5)	(7)	(9)	(11)	(13)
VARIABLES	PerCh	basic	basicmath	Add	Parents	Full	big5mean
Conscientiousness	-0.14”’	-0.11”’	-0.12”’	-0.08”’	-0.08”’	-0.08”’	
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Extraversion	0.09”’	0.08”’	0.07”’	0.07”’	0.07”’	0.08”’	
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
Agreeableness	-0.38”’	-0.12”’	-0.11”’	-0.11”’	-0.11”’	-0.12”’	
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Neuroticism	0.202”’	0.209”’	0.227”’	0.165”’	0.166”’	0.144”’	
	(0.0200)	(0.0220)	(0.0223)	(0.0229)	(0.0229)	(0.0236)	
Openness	0.151”’	0.0925”’	0.0728”’	0.0616”	0.0629”	0.0565”	
	(0.0239)	(0.0263)	(0.0266)	(0.0265)	(0.0265)	(0.0273)	
EducationW4		-0.482”’	-0.475”’	-0.424”’	-0.420”’	-0.343”’	-0.346”’
		(0.0264)	(0.0275)	(0.0280)	(0.0282)	(0.0300)	(0.0299)
UrbanW4		-0.0375	-0.0274	-0.0299	-0.0270	0.00762	0.0189
		(0.0314)	(0.0319)	(0.0320)	(0.0320)	(0.0328)	(0.0327)
Married		-0.341”’	-0.363”’	-0.358”’	-0.361”’	-0.342”’	-0.346”’
		(0.0306)	(0.0311)	(0.0311)	(0.0311)	(0.0321)	(0.0320)
Male		0.581”’	0.573”’	0.590”’	0.589”’	0.668”’	0.683”’
		(0.0321)	(0.0326)	(0.0330)	(0.0330)	(0.0349)	(0.0341)
Age		0.0286”’	0.0202”	0.0146	0.0154’	0.0261”’	0.0232”
		(0.00848)	(0.00899)	(0.00904)	(0.00905)	(0.00931)	(0.00930)
Race		-0.215”’	-0.178”’	-0.184”’	-0.179”’	-0.160”’	-0.148”’
		(0.0301)	(0.0307)	(0.0311)	(0.0316)	(0.0324)	(0.0323)
Math			0.0276”	0.0249’	0.0243’	0.0180	0.0207
			(0.0131)	(0.0131)	(0.0132)	(0.0135)	(0.0134)
Depression				0.207”’	0.202”’	0.197”’	0.206”’
				(0.0308)	(0.0309)	(0.0317)	(0.0315)
ADHD				0.379”’	0.381”’	0.340”’	0.385”’
				(0.0575)	(0.0575)	(0.0585)	(0.0578)
Health				0.157”’	0.157”’	0.157”’	0.167”’
				(0.0223)	(0.0223)	(0.0230)	(0.0227)
Sibings					0.00649	0.00143	0.000820
					(0.00613)	(0.00639)	(0.00638)
Personal Income						-0.0668”’	-0.0675”’
						(0.00586)	(0.00584)
ConscientiousnessD							-0.0292
							(0.0309)
ExtraversionD							0.0809”
							(0.0314)
AgreeablenessD							-0.121”’
							(0.0328)
NeuroticismD							0.196”’
							(0.0322)
OpennessD							0.0165
							(0.0323)
Constant	-0.155	-0.775”’	-0.589’	-0.842”’	-0.897”’	-0.928”’	-0.842”’
	(0.145)	(0.296)	(0.303)	(0.307)	(0.308)	(0.315)	(0.272)
Observations	10,562	9,727	9,536	9,530	9,522	9,160	9,195

Standard errors in parentheses
”’ p<0.01, ” p<0.05, ’ p<0.1

Table A.42.: Unemployment spell existence

	(1)	(3)	(5)	(7)	(9)	(11)	(13)
VARIABLES	PerCh	basic	basicmath	Add	Parents	Full	big5mean
Conscientiousness	-0.123”’	-0.109”’	-0.110”’	-0.0901”’	-0.0921”’	-0.0858”’	
	(0.0194)	(0.0209)	(0.0211)	(0.0214)	(0.0214)	(0.0219)	
Extraversion	0.0691”’	0.0529”’	0.0516”’	0.0519”’	0.0499”’	0.0606”’	
	(0.0178)	(0.0190)	(0.0193)	(0.0193)	(0.0193)	(0.0199)	
Agreeableness	-0.179”’	-0.0310	-0.0234	-0.0238	-0.0209	-0.0328	
	(0.0233)	(0.0262)	(0.0264)	(0.0265)	(0.0265)	(0.0272)	
Neuroticism	0.165”’	0.152”’	0.156”’	0.127”’	0.127”’	0.116”’	
	(0.0191)	(0.0211)	(0.0214)	(0.0220)	(0.0220)	(0.0226)	
Openness	0.126”’	0.120”’	0.117”’	0.111”’	0.112”’	0.118”’	
	(0.0226)	(0.0249)	(0.0251)	(0.0252)	(0.0252)	(0.0260)	
EducationW4		-0.397”’	-0.388”’	-0.363”’	-0.352”’	-0.294”’	-0.289”’
		(0.0255)	(0.0265)	(0.0270)	(0.0272)	(0.0290)	(0.0288)
UrbanW4		-0.0199	-0.0272	-0.0268	-0.0232	-0.0129	-0.00486
		(0.0298)	(0.0302)	(0.0303)	(0.0303)	(0.0312)	(0.0311)
Married		-0.261”’	-0.260”’	-0.257”’	-0.259”’	-0.255”’	-0.262”’
		(0.0287)	(0.0290)	(0.0291)	(0.0291)	(0.0299)	(0.0298)
Male		0.257”’	0.259”’	0.260”’	0.263”’	0.328”’	0.328”’
		(0.0304)	(0.0307)	(0.0310)	(0.0311)	(0.0329)	(0.0321)
Age		-0.00508	-0.0107	-0.0119	-0.0123	0.000796	-0.000725
		(0.00806)	(0.00849)	(0.00856)	(0.00857)	(0.00886)	(0.00884)
Race		-0.133”’	-0.124”’	-0.126”’	-0.111”’	-0.107”’	-0.101”’
		(0.0287)	(0.0290)	(0.0293)	(0.0298)	(0.0305)	(0.0304)
Math			0.0327”’	0.0307”	0.0306”	0.0220’	0.0232’
			(0.0124)	(0.0124)	(0.0125)	(0.0127)	(0.0127)
Depression				0.0600”	0.0549’	0.0464	0.0524’
				(0.0295)	(0.0296)	(0.0303)	(0.0302)
ADHD				0.284”’	0.284”’	0.266”’	0.291”’
				(0.0665)	(0.0665)	(0.0678)	(0.0671)
Health				0.0878”’	0.0857”’	0.0799”’	0.0832”’
				(0.0223)	(0.0223)	(0.0230)	(0.0228)
Sibings					0.0191”’	0.0141”	0.0135”
					(0.00600)	(0.00624)	(0.00621)
persinc						-0.0509”’	-0.0518”’
						(0.00568)	(0.00565)
ConscientiousnessD							-0.130”’
							(0.0291)
ExtraversionD							0.0848”’
							(0.0296)
AgreeablenessD							-0.0416
							(0.0310)
NeuroticismD							0.149”’
							(0.0303)
OpennessD							0.0885”’
							(0.0304)
Constant	-0.523”’	0.108	0.150	-0.0138	-0.0882	-0.294	0.164
	(0.138)	(0.283)	(0.288)	(0.293)	(0.294)	(0.302)	(0.259)
Observations	10,562	9,727	9,536	9,530	9,522	9,160	9,195

Standard errors in parentheses
”’ p<0.01, ” p<0.05, ’ p<0.1