Chapter 12
Modelling Housing Using Multi-dimensional Panel Data

Badi H. Baltagi and Georges Bresson

Abstract This chapter surveys housing models using multi-dimensional panels. While there is a vast literature on housing models using two-dimensional panel data, there are only few papers using multi-dimensional panels. This chapter focuses on housing models, residential mobility and location choice models derived from discrete choice theory utilizing multi-dimensional panels. Examples include nested or hierarchical error components models where a house is located in a street, within a block, within a city, within a county, etc. This chapter introduces some basic concepts of utility functions and discrete choice models used for the hedonic functions and the residential mobility and location choices. Then it surveys some papers on multi-dimensional models of housing hedonic price functions focusing on their estimation methods and their main results. This is followed by a survey of some papers on multi-dimensional models of residential mobility and location choice as well as surveying a few papers on dynamic housing models. It shows that both spatial and temporal dimensions in dynamic systems should be included for hedonic housing models and discrete models of residential location in a multi-dimensional framework. But the inclusion of these multiple dimensions greatly complicates the specification and modeling of such systems. Last, the paper concludes with variational Bayesian approximations which are promising future pathways to potentially overcome many problems in applied modelling of housing and illustrate it using hedonic housing estimation for the city of Paris.

Badi H. Baltagi ✉
Department of Economics and Center for Policy Research, Syracuse University, Syracuse, New York, USA, e-mail: bbaltagi@maxwell.syr.edu

Georges Bresson
Université Paris Panthéon-Assas, Paris, France, e-mail: georges.bresson@u-paris2.fr

12.1 Introduction

This chapter surveys housing models using multi-dimensional panels. For more than a decade, a huge literature within the New Economic Geography has emerged to study the causes of temporal and spatial variations in house prices, residential mobility and location choice. These are major household decisions connected with many activities and travel aspects of households lives. These concepts have been widely researched in various fields including economics, sociology, geography, urban planning, transportation, etc. Location choices and housing investments are inherently dynamic decisions. Moreover, the choice for a household to locate in a given area is a complex decision that is influenced by, among other things, the structural elements of a dwelling as well as the property's spatial relationship to certain amenities. One source of spatial heterogeneity comes from the natural hierarchical and nested structure of the locations of houses, located in a street, within a block, within a city, within a county, within a region, etc. There is a vast literature on such topics mainly using time series and longitudinal (two-dimensional (2D)) data but only few papers using a multi-dimensional (three-dimensional (3D) and more) framework. In this chapter, we will focus on housing models, residential mobility and location choice models derived from discrete choice theory focusing on examples that use multi-dimensional panels.

For example, Baltagi, Fingleton, and Pirotte (2014) focus on the estimation of UK house prices in which spatio-temporal variations in house prices are driven by supply and demand conditions, with spatial effects coming from two distinct sources. One is the direct dependence of house prices in a given locality on house prices in nearby localities. The second source of spatial heterogeneity comes from the presence of hierarchical error components which represent the impact of local (district) effects embedded within wider (county) effects. The panel data includes 353 local authority districts in England over the period 2000-2007. This is done using instrumental variable estimation. Another example is Baltagi, Bresson, and Etienne (2015) who estimate a hedonic housing model based on flats sold in the city of Paris over the period 1990-2003. This is done using maximum likelihood estimation, taking into account the nested structure of the data. Paris is historically divided into 20 *arrondissements*, each divided into four *quartiers* (quarters), which in turn contain between 15 and 169 blocks (*îlot*, in French) per *quartier*.

In Sect. 12.2 we introduce some basic concepts of utility functions and discrete choice models used for the hedonic functions, the residential mobility and location choices. Section 12.3 deals with multi-dimensional models of housing hedonic price functions, their estimation methods and some results. Section 12.4 analyses some multi-dimensional models of residential mobility and location choice. Section 12.5 focuses on multi-dimensional dynamic models of housing models. Section 12.6 highlights variational Bayesian inference, and more specifically mean field variational Bayes approximations to specify and estimate complex multi-dimensional housing models while Section 12.7 concludes.

12.2 Discrete Choice Models and Hedonic Price Functions: A Quick Overview

The pioneering work by Daniel McFadden on location choice is an obvious starting point for a discussion on housing models. One generally considers a household i who chooses to locate in neighborhood j and buy house type k. A standard random utility model (see Holmes and Sieg (2014) for instance) assumes that the indirect utility of household i for location j and house k is given by

$$u_{ijk} = X_j'\beta + Z_k'\gamma + (y_i - p_{jk})\alpha + \varepsilon_{ijk} = f_{ijk}(.) + \varepsilon_{ijk} ,$$

where X_j is a vector of observed characteristics of location j, Z_k is a vector of observed characteristics for house k, y_i is the household income and p_{jk} is the price of housing type k in location j. Each household chooses the neighborhood-housing pair that maximizes utility. Under the assumption that the error terms ε_{ijk} are independent and identically distributed ($i.i.d.$) across i, j and k and follow a type I extreme value distribution, McFadden (1973) (see also McFadden (1974, 1978)), derived the well-known conditional logit choice probabilities:

$$Pr\left[d_{ijk} = 1\right] = \frac{\exp\left(f_{ijk}(.)\right)}{\sum_{j=1}^{J}\sum_{k=1}^{K}\exp\left(f_{ijk}(.)\right)} ,$$

where $d_{ijk} = 1$ if household i has chosen neighborhood j and house type k and zero otherwise. But, the independence of irrelevant alternatives (IIA) property of this model is unattractive. McFadden (1978) proposed the use of a generalized extreme value distribution for the error terms which gives rise to the nested logit model and allows one to relax the assumption that idiosyncratic tastes are independent across locations and houses. But, we need to choose the nesting structure before estimation, mainly if the nested structure is not natural and if we do not have knowledge about the neighborhood structure. One solution is to use random coefficients β_i, γ_i and α_i instead of fixed coefficients β, γ and α. Estimation with random coefficients is challenging and needs the use of simulation-based estimators (SBE) (see Newey and McFadden (1974) or Judd (1998)).

Moreover, Bayesian estimators are also well suited for the estimation of discrete choice models with random coefficients. One application of such model with SBE has been done by Hastings, Kane, and Staiger (2006) who study the effects of open enrollment policies under a particular parent choice mechanism, sorting households among schools within the Mecklenburg Charlotte school district, North Carolina. Bajari and Kahn (2005) used Bayesian methods to study housing demand explaining racial segregation in cities.

Demand estimation has also focused on the role of unobserved neighborhood characteristics or housing quality ζ_j. In that case, the indirect utility function is written as

$$u_{ijk} = X_j'\beta + Z_k'\gamma + (y_i - p_{jk})\alpha + \zeta_j + \varepsilon_{ijk} .$$

Unobserved neighborhood characteristics can be recovered by matching the observed market shares of community j. Then, the remaining parameters can be estimated by a generalized method of moments (GMM) estimator using instrumental variables (IV) to deal with the correlation between housing price p_{jk} and unobserved neighborhood characteristics or housing quality ζ_j. Bayer, Ferreira, and McMillan (2007), using a two-dimensional (2D) panel data, estimate household preferences for school and neighborhood attributes in the presence of sorting. The model embeds a boundary discontinuity design in a heterogeneous residential choice model, addressing the endogeneity of the school and neighborhood characteristics. Their application concerns a restricted-access version of the 1990 U.S. Census, that links detailed characteristics for nearly a quarter of a million households and their houses in the San Francisco Bay Area with their precise residential location. Bayer, McMillan, Murphy, and Timmins (2016), using a three-dimensional panel data (3D), develop a dynamic model of neighborhood choice (see Sect. 12.5). They capture observed and unobserved preference heterogeneity across households and locations of housing transactions in the San Francisco Bay Area from 1994 to 2004.

We turn now to hedonic measures with a strong theoretical grounding (see, for example, Griliches (1971), Rosen (1974), Nelson (1977), Blomquist and Worley (1981), Blomquist and Worley (1982) among others). Also, the use of regression techniques to control for compositional and quality change (Witte, Sumka, and Erekson (1979), Brown and Rosen (1982), Meese and Wallace (1997), to mention a few). The hedonic pricing method is based on the fact that prices of goods (in our case, houses) in a market are affected by their characteristics. This method estimates the value of a commodity based on people's willingness to pay for the commodity as and when its characteristics change. In real estate economics, hedonic pricing is used to adjust for the problems associated with looking for a dwelling that is as heterogeneous as buildings. The hedonic pricing function, which explains the price of a house, will be affected by among other things, the structural characteristics of the house, the characteristics of the neighborhood and the environmental characteristics.

Since the seminal work of Rosen (1974), one generally uses a two-stage procedure for estimating the hedonic price function of the dwelling and for the recovery of marginal willingness to pay functions of heterogeneous individuals for the characteristics of differentiated products. Basically, hedonic models of housing price relate the price (or the logarithm of the price per square meter) to among other things, the characteristics of the dwellings $p_{jk} = f(Z'_k, ...)$. The price gradient associated with this hedonic price function $\partial p_{jk}/\partial Z_{kl}$ denotes the implicit price of the amenity Z_{kl} (number of rooms, quality of air, etc.). The second stage of Rosen's procedure seeks to recover the coefficients of demand (or marginal willingness to pay) and supply (or marginal willingness to accept) functions for the attribute Z_{kl} from the first-order conditions of the equilibrium relationships: $\partial p_{jk}/\partial Z_{kl} = f_d\,(Z_k, B_k)$ for demand and $\partial p_{jk}/\partial Z_{kl} = f_s\,(Z_k, S_k)$ for supply where B_k and S_k represent attributes of the buyer and seller of house k. Bartik (1987) and Epple (1987) have described a source of endogeneity in the second stage of Rosen's procedure that is difficult to overcome without exclusion restriction arguments or the use of IV methods. This has led researchers to avoid the estimation of marginal willingness to pay functions altogether, relying instead on the first-stage hedonic price function and limiting the analysis to the evaluation of marginal changes in amenities (see Gayer, Hamilton, and Viscusi (2000), K. C. Bishop and Timmins (2011) to mention a few).

In some studies, dwellings were assumed to be stratified into blocks or communities j where prices are homogeneous and price trends are roughly parallel. Ideally a model could be estimated in each neighborhood and the elementary geographic zones could be very small sub-markets. In that case, each model is estimated in a particular block, all variables are *de facto* interacted with the block. So, spatial location is not without consequences and hedonic housing price models should incorporate spatial effects. In the econometric literature, spatial effects may result from spatial dependence or from spatial heterogeneity. Spatial dependence means that observations at location j depend on other observations at locations $l \neq j$. Spatial heterogeneity refers to variation in relationships over space and more precisely over every point in space. The distinction is coming from the structure of the dependence which can be related to location and distance, both in a geographic space as well as in a more general economic or social network space (see Anselin (2001), Anselin, LeGallo, and Jayet (2008)).

For spatial effects in real estate, a lot of housing models have been estimated in a 2D framework on panel data with two indexes j and t generally for location and time associated with spatial weight matrices (see for instance Baltagi and Bresson (2011), Bresson and Hsiao (2011), Fingleton (2008), Glaeser (2008), Holly, Pesaran, and Yamagata (2010) to mention a few). But very few models have been developed in a three-dimensional, four-dimensional, or more in a panel data setting. In the next section, we present some of these models and their associated results for these multi-dimensional frameworks.

12.3 Multi-dimensional Models of Housing Hedonic Price Functions: Some Examples

Baltagi et al. (2015) estimate a hedonic housing model based on flats sold in the city of Paris over the period 1990-2003. This is done using maximum likelihood estimation, taking into account the nested structure of the data. Paris is historically divided into 20 *arrondissements*, each divided into four *quartiers* (quarters), which in turn contain between 15 and 169 blocks (*îlot*, in French) per *quartier*. The data set used is an unbalanced pseudo-panel data containing 156,896 transactions. The real estate literature emphasizes the importance of neighborhoods in determining the value of a house or a flat. While one can try and include as many as possible of the neighborhood characteristics in the regression to capture these effects, most attempts may fall short as many neighborhood characteristics are not observed, as in our case. One simple method of capturing the effect of neighbors' prices used by Baltagi et al. (2015) is to estimate a spatial lag regression equation with time-varying coefficients:

$$p_{taqif} = \lambda_t \tilde{p}_{taqif} + Z_{taqif}\beta + \varepsilon_{taqif} \ , |\lambda_t| < 1 \ , \tag{12.1}$$

where $t = 1, ..., T$ for years, $a = 1, ..., N$ for *arrondissements*, $q = 1, ..., Q_{ta}$ for *quartiers*, $i = 1, ..., M_{taq}$ for *îlots* and $f = 1, ..., F_{taqi}$ for flats. p is the transaction price (in logs) for flat f, in *îlot* i nested in *quartier* q, which in turn is nested in *arrondissement* a at time t. Z_{taqif} denotes the vector of K explanatory variables describing the characteristics for this flat (surface in m^2, count data as number of rooms, bedrooms, bathrooms, garage plots, and dummy variables as balcony, whether it is located on a street, boulevard, avenue, or place, period of construction (<1850, 1850-1913, ...,1981-2003), etc). This unbalanced panel is made up of $N = 20$ top-level *arrondissements*, each containing Q_{ta} second-level *quartiers*. The second-level *quartiers* in turn contain M_{taq} third-level *îlots*, which contain the innermost F_{taqi} observations on flats. The number of observations in the higher level groups are $F_{taq} = \sum_{i=1}^{M_{taq}} F_{taqi}$ and $F_{ta} = \sum_{q=1}^{Q_{ta}} F_{taq}$. The total number of observations is $H = \sum_{t=1}^{T} \sum_{a=1}^{N} F_{ta}$. The number of top-level groups is NT, the number of second-level groups is $L = \sum_{t=1}^{T} \sum_{a=1}^{N} Q_{ta}$ and the number of bottom-level groups is $G = \sum_{t=1}^{T} \sum_{a=1}^{N} \sum_{q=1}^{Q_{ta}} M_{taq}$. So, we have a five-dimensional pseudo-panel data structure. The spatial lag coefficient λ_t may be time varying or constant over time and the spatial lag variable $\tilde{p}_{taqif}$ is defined as

$$\tilde{p}_{taqif} = \sum_{a=1}^{N} \sum_{q=1}^{Q_{ta}} \sum_{i=1}^{M_{taq}} \sum_{p=1}^{F_{taqi}} w_{taqip} p_{taqip} \ ,$$

where w_{taqip} denotes the elements of the spatial weights matrices W_t which vary with t. Elements on the diagonal of W_t are set to zero while the off-diagonal elements define the connexion (contiguity or distances) between dwellings. There are at least two reasons why positive spatial correlation may exist. First, dwellings in a neighborhood tend to have similar structural characteristics and second, dwellings in a neighborhood share the same location amenities (see Basu and Thibodeau (1988)). However, many of the price determining factors shared by neighborhoods are difficult to explain explicitly but these “omitted" factors are contained in the neighborhood prices. For each year, Baltagi et al. (2015), using the “Delaunay triangle algorithm", define first-order contiguity matrices W_t for the nearest neighbors (*i.e.*, from 10 to 140 nearest sold flats). Accordingly to the nested structure, the disturbance term is given by

$$\varepsilon_{taqif} = \delta_{ta} + \mu_{taq} + \nu_{taqi} + u_{taqif} \ ,$$

where δ_{ta} is the *arrondissement* effect, μ_{taq} is the *quartier* effect naturally nested in the respective *arrondissement* and ν_{taqi} is the *îlot* effect naturally nested in the respective *quartier*. These could be fixed or random. The remainder disturbance term for the particular flat is random $u_{taqif} \sim iiN(0, \sigma_u^2)$. For the random specification, we assume that $\delta_{ta} \sim iiN(0, \sigma_\delta^2)$, $\mu_{taq} \sim iiN(0, \sigma_\mu^2)$ and $\nu_{taqi} \sim iiN(0, \sigma_\nu^2)$.

Following Antweiler (2001), Baltagi et al. (2015) use block-diagonal matrices of size $(H \times H)$ corresponding in structure to the groups or subgroups they represent. They can be constructed explicitly by using “group membership” matrices consisting of ones and zeros that uniquely assign each of the H observations to one of the G (or L or NT) groups. Let R_ν be such an $(H \times G)$ matrix corresponding to the innermost group level. Then the block-diagonal $(H \times H)$ matrix J_ν can be expressed as the outer product of its membership matrices: $J_\nu = R_\nu R'_\nu$. The inner product $R'_\nu R_\nu$ produces a diagonal matrix $\widetilde{L}_\nu$ of size $(G \times G)$ which contains the number of observations of each group. Similarly, let R_μ be such an $(H \times L)$ matrix corresponding to the second-level groups. Then the block-diagonal $(H \times H)$ matrix J_μ can be expressed as the outer product of its membership matrices: $J_\mu = R_\mu R'_\mu$. Last, let R_δ be such an $(H \times NT)$ matrix corresponding to the top-level groups. Then the block-diagonal $(H \times H)$ matrix J_δ can be expressed as the outer product of its membership matrices: $J_\delta = R_\delta R'_\delta$.

If we pool the observations, the log-likelihood is given by

$$\ln l = -\frac{1}{2} H \ln(2\pi) - \frac{1}{2} \ln|\Omega| + \ln|A| - \frac{1}{2}\varepsilon' \Omega^{-1} \varepsilon ,$$

where

$$\varepsilon = Ay - X\beta , A = I_H - \lambda W ,$$

with $W = diag\,(W_t)$ and $\lambda = diag\,(\lambda_t)$ where W is the block-diagonal spatial weight matrix of size $(H \times H)$. W_t is the spatial weight matrix[1] of size $(F_{ta} \times F_{ta})$ changing at each time period t. λ is the spatial lag matrix of size $(T \times T)$ whose elements λ_t change at each time period t. I_H is an identity matrix of size $(H \times H)$.

The variance-covariance matrix of the disturbance is defined as follows:

$$\Omega = E\left[\varepsilon\varepsilon'\right] = \sigma_u^2 \left[I_H + \rho_\nu J_\nu + \rho_\mu J_\mu + \rho_\delta J_\delta\right] ,$$

with

$$\rho_\delta = \frac{\sigma_\delta^2}{\sigma_u^2} , \rho_\mu = \frac{\sigma_\mu^2}{\sigma_u^2} , \rho_\nu = \frac{\sigma_\nu^2}{\sigma_u^2} .$$

Extending the derivations of Antweiler (2001) to the case of the spatial lag model (12.1), Baltagi et al. (2015) get:

$$\ln l = -\frac{1}{2}\left[H \ln\left(2\pi\sigma_u^2\right) + \sum_{t=1}^{T}\left\{\ln|I_t - \lambda_t W_t| + \sum_{a=1}^{N}\left\{\ln\theta_{ta} + C_{ta} - \frac{\rho_\delta}{\theta_{ta}}\frac{U_{ta}^2}{\sigma_u^2}\right\}\right\}\right] , \tag{12.2}$$

[1] Baltagi et al. (2015) use a block-diagonal weight matrix W of $(156,896 \times 156,896)$ whose smallest sub-block is a weight matrix W_t of $(6,643 \times 6,643)$ for the year 1992 and whose largest sub-block is a weight matrix W_t of $(17,098 \times 17,098)$ for the year 1999.

$$\text{with } C_{ta} = \sum_{q=1}^{Q_{ta}} \left\{ \ln \theta_{taq} + C_{taq} - \frac{\rho_\mu}{\theta_{taq}} \frac{U_{taq}^2}{\sigma_u^2} \right\} ,$$
$$\text{and } C_{taq} = \sum_{i=1}^{M_{taq}} \left\{ \ln \theta_{taqi} + \frac{V_{taqi}}{\sigma_u^2} - \frac{\rho_\nu}{\theta_{taqi}} \frac{U_{taqi}^2}{\sigma_u^2} \right\} ,$$

where I_t is an identity matrix of size $(F_{ta} \times F_{ta})$ and where

$$\begin{cases} \theta_{taqi} = 1 + \rho_\nu F_{taqi} & & V_{taqi} = \sum_{f=1}^{F_{taqi}} \varepsilon_{taqif}^2 , \\ \theta_{taq} = 1 + \rho_\mu \phi_{taq} & \text{with } \phi_{taq} = \left(\sum_{i=1}^{M_{taq}} \frac{F_{taqi}}{\theta_{taqi}} \right) & U_{taqi} = \sum_{f=1}^{F_{taqi}} \varepsilon_{taqif} , \\ \theta_{ta} = 1 + \rho_\delta \phi_{ta} & \text{with } \phi_{ta} = \left(\sum_{q=1}^{Q_{ta}} \frac{\phi_{taq}}{\theta_{taq}} \right) & U_{taq} = \sum_{i=1}^{M_{taq}} \frac{U_{taqi}}{\theta_{taqi}} , \\ & & U_{ta} = \sum_{q=1}^{Q_{ta}} \frac{U_{taq}}{\theta_{taq}} , \end{cases}$$

where $\varepsilon_{taqif} = y_{taqif} - \lambda_t \sum_{a=1}^{N} \sum_{q=1}^{Q_{ta}} \sum_{i=1}^{M_{taq}} \sum_{p=1}^{F_{taqi}} w_{taqip} y_{taqip} - X_{taqif} \beta$. A gradient of this log-likelihood function (12.2) is obtained analytically, but it can also be obtained through numeric approximation. In carrying out this maximization, it is necessary to constrain the optimization such that $|\lambda_t| < 1$, the variance σ_u^2 remains positive, and that the variance ratios ρ_δ, ρ_μ and ρ_ν remain non-negative.

Baltagi et al. (2015) report several ML estimation results. One for the random effects (RE) model ignoring the nested effects, one for the nested RE model ignoring the spatial lag effects and one for the spatial nested RE model.[2] Baltagi et al. (2015) found significant spatial lag effects as well as significant nested random error effects. They emphasize the importance of nested effects in the Paris housing data as well as the spatial lag effects. In fact, they show that the impact of the adjacent neighborhoods becomes relatively small when one takes care of the nested random effects. In addition, due to the unbalanced pseudo-panel aspect of these transactions, they show that one should allow the spatial weight matrix as well as the spatial lag coefficients to vary over time, and that the likelihood ratio tests confirm that they fit the Paris housing data better.

Following LeSage and Pace (2009), Baltagi et al. (2015) compute the marginal effects – which are decomposed into direct, indirect and total marginal effects – and show that the marginal spillover effects due to the neighbors are negligible relative to the direct effects. Moreover, the empirical results show that the marginal effect

[2] For the estimation of nested error component model with unbalanced panel data using simple analysis of variance (ANOVA), maximum likelihood (MLE) and minimum norm quadratic unbiased estimators (MINQUE)-type estimators of the variance components, see Baltagi, Song, and Jung (2001). For Lagrange multiplier testing of nested error component model with unbalanced panel data, see Baltagi, Song, and Jung (2002).

for a specific housing characteristic is lower on average once the nested effects are taken into account.

Baltagi et al. (2014) estimate a nested random effects spatial autoregressive panel data model to explain annual house price variation across 353 local authority districts in England over the period 2000-2007. The nested error components represent the impact of local (district) effects embedded within wider (county) effects. Baltagi et al. (2014) propose new estimators based on the instrumental variable approaches of Kelejian and Prucha (1998) and L. Lee (2003) for the cross-sectional spatial autoregressive model. The estimation methods allow for the endogeneity of the spatial lag variable producing the simultaneous spatial spillover of prices across districts together with the nested random effects in a panel data setting. Monte Carlo results show that these estimators perform well relative to alternative approaches and produce estimates based on real data that are consistent with the theoretical house price model underpinning the reduced form. The empirical results show that there is a significant spatial lag term indicating positive correlation between prices locally and prices in 'nearby' districts and that income within commuting distance has a positive effect, while the stock of housing has a negative effect, on housing price. Also that the nested error components attributable to district and county effects, like the spatial lag term are necessary elements in modeling UK house prices.[3]

From hedonic price functions, we can derive temporal and/or spatial price indexes. This has been done, for instance, by Syed, Hill, and Melser (2008) for the Sydney region. Their data concern 15 regions in Sydney on a quarterly basis from 2001 to 2006 from a data set consisting of 418,877 house sales. As 60% of sales observations are missing for one or more of the core characteristics, they first use multiple-imputation techniques to fill in the gaps in the data set, prior to estimating the hedonic model. In a second stage, they specify and estimate a non-nested three-dimensional hedonic price function. They pool across all the regions and periods in the sample and estimate the region-time specific fixed effects and shadow prices of housing characteristics. This method was first proposed by Aizcorbe and Aten (2004), who refer to it as the time-interaction-country product dummy method.

$$
\begin{aligned}
p_{jth} = \alpha &+ \sum_{\tau=2}^{T} \beta_\tau q_{\tau h} + \sum_{\kappa=2}^{J} \gamma_\kappa r_{\kappa h} + \sum_{\tau=2}^{T}\sum_{\kappa=2}^{J} \delta_{\tau\kappa} b_{\tau\kappa h} \\
&+ \sum_{m=2}^{M_\kappa} \eta_{\kappa m} d_{\kappa m h} + Z_{jth}\theta + \varepsilon_{jth} \,, \\
&\text{for } j = 1, \ldots, J,\ t = 1, \ldots, T \text{ and } h = 1, \ldots, H_{jt} \,,
\end{aligned}
$$

[3] Baltagi and Pirotte (2014) derive the Best Linear Unbiased Predictor (BLUP) for a spatial nested error components panel data model. This predictor is useful for panel data applications that exhibit spatial dependence and a nested hierarchical structure. The predictor allows for unbalancedness in the number of observations in the nested groups. It could be interesting for forecasting average housing prices located in a county nested in a state.

where p is the log of the price of a dwelling h belonging to region-period jt, $q_{\tau h}$ (resp. $r_{\kappa h}$) are dummy variables such that $q_{\tau h} = 1$ (resp. $r_{\kappa h} = 1$) if the observation h is from period t (resp. from region j) and zero otherwise. The dummy variables $b_{\tau\kappa h}$ denote interactions between periods and regions taking the value of 1 if the observation h is from region-period jt and zero otherwise. The postcode dummies are denoted by $d_{\kappa mh}$, where $d_{\kappa mh} = 1$ for observation h's postcode and zero otherwise. Z is a set of quality characteristics including the dwelling type, the number of bedrooms, bathrooms, lot size, etc. Spatial correlation between observations is defined by a spatial autoregressive process on the error term: $\varepsilon_{jth} = \lambda W \varepsilon_{jth} + u_{jth}$ where $u_{jth} \sim N(0, \omega_{jth}\sigma^2)$. The spatial weight matrix W is a contiguity matrix and the variance of u_{jth} is subscripted with jt allowing for heteroskedasticity. The coefficients δ_{jt} measure the region-period specific fixed effects for the logarithms of the price level after controlling for the effects of the attributes of the dwellings. The model is estimated using the maximum likelihood method. The advantage of this region-time-dummy model is that the temporal and regional price indexes are derived directly from the estimated coefficients $\hat{\beta}_t$, $\hat{\gamma}_j$, $\hat{\delta}_{jt}$, $\hat{\eta}_{jm}$ and $\hat{\theta}$. Let $P_{j,t,s}$ the price index for region j in year t and quarter s. Then, the relative prices are given by

$$\frac{P_{j,t,s}}{P_{j,t,1}} = \exp\left(\hat{\beta}_{t,s} + \hat{\delta}_{j,t,s}\right) \text{ for } s = 2, 3, 4\,,$$
$$\text{and } \frac{P_{j,t+1,1}}{P_{j,t,1}} = \exp\left(\hat{\beta}_{t+1,1} + \hat{\delta}_{j,t+1,1}\right)\ .$$

So, it is possible to construct a temporal price index for each region j over the entire time period of the dataset. Results are normalized such that the price index for the initial region (Inner Sydney) is equal to 1 for the first quarter of 2001. One can also construct a spatial price index for each quarter s of a specific year t for the entire set of regions. For a given quarter (t, s), spatial price indexes can be constructed from the estimated coefficients $\hat{\gamma}_j$, $\hat{\delta}_{jt}$, $\hat{\eta}_{jm}$ and $\hat{\theta}$. The starting point is a comparison between a postcode m in region l and a postcode n in region j for a particular dwelling h with amenities vector Z_{ch}. This spatial price index is defined as:

$$P_{lmts,jnts}(Z_{ch}) = \exp\left[\left(\hat{\gamma}_j - \hat{\gamma}_l\right) + \left(\hat{\delta}_{jt} - \hat{\delta}_{lt}\right) + \left(\hat{\eta}_{jn} - \hat{\eta}_{lm}\right)\right] \times \left[\prod_{c=1}^{C} \exp\left[Z_{ch}\left(\hat{\theta}_{jc} - \hat{\theta}_{lc}\right)\right]\right]\,,$$

and the spatial index can be generalized to take into account of all dwellings sold in postcodes lm:

$$P_{lmts,jnts} = \exp\left[\left(\hat{\gamma}_j - \hat{\gamma}_l\right) + \left(\hat{\delta}_{jt} - \hat{\delta}_{lt}\right) + \left(\hat{\eta}_{jn} - \hat{\eta}_{lm}\right)\right] \times \left[\prod_{h=1}^{H_{lmts}} \prod_{c=1}^{C} \exp\left[Z_{ch}\left(\hat{\theta}_{jc} - \hat{\theta}_{lc}\right)\right]\right]^{1/H_{lmts}}\ .$$

This is close to a .

Combining the temporal and spatial indexes allows a price comparison of dwellings between different location-year-quarter triplets. Syed et al. (2008) found that their hedonic house price indexes rose significantly from 2001 to 2003, after which they fell slightly. This finding is consistent with the Australian Bureau of Statistics (ABS) index. But their indexes, however, are less volatile than their ABS counterpart, rising noticeably less in the boom and falling less thereafter. In the spatial dimension, they found large and systematic differences in the price of housing across regions of Sydney. The regional dispersion narrowed during the boom period but appears to have increased again since then.

Several authors have shown that values of complex assets are difficult to accurately quantify and information asymmetry affects asset prices through various channels (see Agarwal and Hauswald (2010), Baker and Wurgler (2007), Carlin, Kogan, and Lowery (2013), Kelly and Ljungqvist (2012) to mention a few). The subprime crisis (poor household mortgage decisions and subsequent foreclosure), the housing market collapse in the US, followed by the financial crisis have revealed that uninformed buyers overpay. The house buying mechanism is a field in which households' ability (or inability) to use market information may have strong effects on housing decisions. It could be done through the choice of mortgage product and through the purchase transaction (see Carlin et al. (2013) and Turnbull and van der Vlist (2015)). House purchases may involve residential mortgages and associated complex financial instruments which have been assigned as a major cause of waves of foreclosures during and after the 2007-2008 financial crisis. Turnbull and van der Vlist (2015) show that buyers who are uninformed in the housing market pay more for houses than buyers who are informed. They use pseudo-panels of repeated sales based on neighborhood census block-level. This data is for 426,021 parcels located in Orange County, Florida, over the period 2000-2012. The authors split fair market value and uninformed buyer effects by first identifying for each of the market sales in the period 2000-2006 which of the units foreclosed in 2007-2012. The future foreclosure dummy FF equals 1 if a market transaction completed in 2000-2006 is followed by a foreclosure in 2007-2012 and equals zero otherwise. Turnbull and van der Vlist (2015) estimate an hedonic price function of the log of market price in first differences on the neighborhood block-level j:

$$p_{itj} - p_{lsj} = \left(Z_{itj} - Z_{lsj}\right)\beta_Z + \left(FF_{ij} - FF_{lj}\right)\beta_{FF} + \varepsilon_{itj} - \varepsilon_{lsj}\,,$$
$$\text{for } t, s = 1, \ldots, T,\ i, l = 1, \ldots, N,\ j = 1, \ldots, J \text{ for all } i \neq j \text{ and } t \neq s\,,$$

where p_{itj} is the log of the price of property i sold at time t located in area j. Z is the vector of relevant house characteristics and amenities and FF is the penalty associated with being foreclosed *ex post* (over 2007-2012). The model of first differences at the neighborhood block-level basically treats sales within the neighborhood block as repeat sales while accounting for observed structural differences. This is a model on pseudo-panels of repeated observations "*à la* Deaton (1985)". This model also allows for clustered errors at the neighborhood block-level j. Results show that buyers who are later foreclosed paid a 2.7% (resp. a 4.6%) premium for properties bought between 2000 and 2006 (resp. between 2005 and 2006). Estimation on different sub-periods also reveal strong correlation between home buyers' house prices and future foreclosures. To check whether effects vary across housing market segments, Turnbull and van der Vlist (2015) estimate quantile regression models. Results show that the effect for the penalty associated with being foreclosed is larger for the lower end of the housing market. Buyers in 2005-2006 who ended up foreclosed paid up to 3.5% above fair market value in the lower end of the housing market while foreclosed owners paid a little over 1% percent more in the higher end of the housing market.

Evaluating non-marginal changes in amenities in the hedonic model requires estimating the underlying hedonic demand or marginal willingness-to-pay (MWTP) function. In general, however, it is necessary to make assumptions that restrict preference heterogeneity. As stated earlier, the usual approach, based on Rosen (1974)'s seminal paper, suffers from a number of econometric problems, especially when the hedonic price function is non-linear in the amenity of interest or when buyers simultaneously choose both the hedonic price and the quantity of that amenity that they will consume. The introduction of a nonlinear budget constraint creates a serious endogeneity problem when using statistical inference to recover the parameters describing the consumer's preferences. And the instrumental-variable approaches generally used have relied on questionable exclusion restrictions. These problems have led most researchers to abandon the estimation of the MWTP function altogether and to use only local MWTP measurements. But, Bajari and Benkard (2005) have demonstrated that this endogeneity problem may be avoided by replacing the statistical inference used in the second stage of Rosen (1974)'s method with a "preference inversion" procedure that inverts the first-order conditions of utility maximization to recover demand at the individual level. However, strict functional-form assumptions, necessary for their inversion procedure, limit their approach. In contrast, K. C. Bishop and Timmins (2018) have shown how to recover the unconditional distribution of linear MWTP functions with a simple and transparent data-driven estimation approach. Constructing a rich 3D panel dataset of property transactions and of buyers covering the six core counties of the San Francisco Bay Area, they recover the full distribution of demand functions for clean air. First, they find that estimating the full demand function, rather than simply recovering a local estimate of marginal willingness to pay, is important. Second, they find that the mean willingness to pay to avoid a 33% increase in ozone is $\$1,021$ and that the median willingness to pay to avoid the increase is \$770. There is considerable heterogeneity as evidenced by the interquartile range of $\$1,106$. Their data-driven estimates of the welfare effects associated with a nonmarginal change in air quality differ substantially from those recovered using the existing approaches to welfare estimation.

12.4 Multi-dimensional Models of Residential Mobility and Location Choice: Some Examples

Residential mobility and location choice are significant household decisions and have been widely researched in various fields including economics, sociology, geography, regional science, urban planning, housing policy, transportation, etc. Decisions of residential mobility and location choice are closely related to the household housing process with a large range of factors that contribute to each choice. Due to the vastness of the literature on such topics, we will focus on few examples of residential mobility and location choice. Readers could read profitably the survey of Dieleman (2001) on residential mobility. Since the seminal works of Rossi (1955) and Alonso (1964), a huge number of research on residential location choice has been published. *"Reasons for moving are divided into those which pertain to the decision to move out of the former home - "pushes" - and those reasons pertaining to the choice among places to move to - "pulls"* (Rossi (1955), p. 8). For instance, push factors may include negative externalities like noise, pollution or crime, changes in housing affordability, dissatisfaction with the current dwelling, changes in household structure, etc. Pull factors often include better access to good quality public services (schools and health care facilities), employment, leisure and recreational opportunities, etc. (see B. H. Y. Lee and Waddell (2010) and Hoang and Wakely (2000) for a review). Our purpose is not to review the main factors of residential mobility and relocation but to summarize few multi-dimensional studies of residential mobility and relocation.

One interesting study has been done by Davies and Pickles (1985) in a multi-dimensional framework. They propose a model that conceptualizes residential mobility as a sequence of choices between staying and moving. Household i will move in time period t if and only if random utility derived from the most-favored alternative dwelling available u_{itb} is larger than the random utility derived from the current dwelling u_{ita}:

$$u_{ita} = V(y_{it}, Z_{ta}) + \varepsilon_{ita} = V_{ita} + \varepsilon_{ita} \text{ with } \varepsilon_{ita} = \mu_{ia} + g(d_{it}) + \nu_{ita},$$
$$u_{itb} = V(y_{it}, Z_{tb}) + \varepsilon_{itb} = V_{itb} + \varepsilon_{itb} \text{ with } \varepsilon_{itb} = \mu_{ib} + h(t) + \nu_{itb},$$

where y_{it} is a vector of observed characteristics of household i at time t, Z_{ta} (resp. Z_{tb}) is a vector of observed characteristics of the current dwelling (resp. the most-favored alternative dwelling available). V_{ita} and V_{itb} are the systematic utilities while ε_{ita} and ε_{itb} are the random components of utilities. These random components are likely to be correlated over time for each household. ε_{ita} is the sum of the unexplained household heterogeneity μ_{ia}, a function $g(d_{it})$ of the duration of stay for household i at time t and a remainder term ν_{ita}, independently distributed over both households and time. For the other random component ε_{itb}, the unexplained household heterogeneity μ_{ib} also applies. Moreover, a time trend $h(t)$ represent fluctuations in market conditions. Davies and Pickles (1985) used a quadratic specification for the duration of stay $g(d_{it}) = \beta_1 d_{it} + \beta_2 d_{it}^2$, and a cubic specification for the housing market function $h(t) = \beta_3 t + \beta_4 t^2 + \beta_5 t^3$.

The likelihood $L(z_{it})$ of the observed sequence of outcomes is the product of the probabilities of the observed choice for each time period:

$$L(z_{it}) = \prod_{t=1}^{T} \{Pr\,[u_{itb} > u_{ita}]\}^{z_{it}} \{1 - Pr\,[u_{itb} > u_{ita}]\}^{1-z_{it}} \ ,$$

$$\text{with} \quad Pr\,[u_{itb} > u_{ita}] = \int_{-V_{it}-\mu_i+g(d_{it})-h(t)}^{\infty} \phi\,(v_{itb} - v_{ita})\, d\,(v_{itb} - v_{ita}) \ ,$$

where $z_{it} = 1$ if household i moves in time period t and zero elsewhere, $V_{it} = V_{itb} - V_{ita}$, $\mu_i = \mu_{ib} - \mu_{ia}$ and $\phi\,(.)$ is the probability density of the difference between the two random components. Assuming that they follow Weibull distributions leads to the following likelihood with a household-specific error term μ_i:

$$L(z_{it}) = \prod_{t=1}^{T} \frac{\exp\,[-V_{it} - \mu_i + g(d_{it}) - h(t)]^{z_{it}}}{1 + \exp\,[-V_{it} - \mu_i + g(d_{it}) - h(t)]} \ .$$

Three problems arise with this likelihood: the integration over the error term distribution is almost analytically intractable; the initial observation complicates the handling of endogenous variables such as duration of stay d_{it} and numerical methods are required for parameter estimation. To overcome these problems, Davies and Pickles (1985) derived an approximation of the likelihood using the generalized Beta-logistic approach developed by Davies (1984).

The panel data is for 887 households participating in the Michigan Panel Study of Income Dynamics over the period 1968-1977. The dependent variable was a residential move within the county or an intercounty move with no change in the head-of-household's job. Among the main explanatory variables were the duration of stay, a room adequacy index (actual rooms / required rooms), an income adequacy index (actual income / needs), the age of the head of household, and the education level. First, they show that the room adequacy index has a U-shaped relationship with residential mobility. Renters have the shortest initial duration status while owners have the longest. But, there is no evidence of a similar U-shaped relationship anticipated for the income adequacy index. Second, they show that changing financial circumstances does not seem to play any role in the life cycle variation in residential mobility in the United States. Moreover, they are not able to show any effect of income surplus on residential mobility. These are unexpected results. Davies and Pickles (1985) argue that these results may be due to the housing market being highly segmented, not just between renting and owner-occupation, but between different types of property and their location. It could be interesting to redo this study on more recent data. It will probably give different conclusions for the last decade which has known troubled financial periods. Davies and Pickles (1985) found a strong negative relationship between age of the head of household and residential mobility. This strong negative relationship is present even when changing space requirements and financial pressures are accounted for. Age of the head of household is the dominant life cycle and acts as a proxy variable for changing needs and financial circumstances through the life cycle.

Several studies have acknowledged the interdependencies between residential choice, job choice, and transportation mode choice. However, most previous studies focused on two dimensions only. Guo, Feng, and Timmermans (2020) jointly consider a multidimensional model of residential choice, job choice, commuting mode choice and assume individuals and households consider different life domains jointly. Based on a 3D dataset (individuals, choice situations and alternatives) collected in the region of Shenyang (China), Guo et al. (2020) specify a mixed multinomial logit model which allows for unobserved heterogeneity in individual preferences. This model was estimated to capture the effects of different residential, job and commuting attributes on multi-dimensional choice, accounting for the panel nature of the data. Guo et al. (2020) find that, housing tenure, size, price, distance to the bus stop, and housing location are important housing characteristics explaining the residential mobility choice process. Second, salary, job type, co-worker relationships and job environment are also significant factors in the job mobility choice process. Guo et al. (2020) show that time-related factors influence commute mode choice and choice of public transportation modes is sensitive to commuting costs while car mode choice is not. The model estimation highlights that people are relatively satisfied with their current situation and do not frequently make changes. And people are less inclined to move house relative to changing job. Finally, Guo et al. (2020) show that both unobserved heterogeneity and demographic characteristics affect the multiple dimensions of choice.

Explaining the factors which determine housing tenure choices is important. For instance, Fu, Zhu, and Ren (2015) estimate multilevel multinomial logistic regressions for housing types to study home ownership in urban China. They base their estimation on a sample of $2,585,480$ households from the 2005 National Population Sample Survey of China and available information for 205 urban areas (prefectures-level data) (see Huang and Clark (2002) for a similar study in China but in a 2D framework). For one household i in prefecture j, the within-prefecture multinomial logistic model for the odds of housing type m is given by

$$\log\left[\frac{Pr\left(\text{housing type}_{mij}\right)}{Pr\left(\text{private rental housing}_{ij}\right)}\right] = \beta_{mj0} + \sum_{k=1}^{K} \beta_{km}\left(Z_{h,kijm} - \overline{Z}_{h,kjm}\right) + \varepsilon_{ij} \, .$$

The $m = 1, .., 5$ housing types refer to owning self-built housing, owning commodity housing, owning affordable housing, owning privatized *danwei* housing and public rental housing. $Z_{h,kijm}$ is the value of household-level covariate k associated with household i in prefecture j for the m-th housing type. $\overline{Z}_{h,kjm}$ is the sample mean of covariate k within prefecture j. The household-level error term ε_{ij} is assumed to be $i.i.N(0, \sigma^2)$. The between-prefecture model for housing types is:

$$\beta_{mj0} = \gamma_{00m} + \sum_{s=1}^{S} \gamma_{0sm} Z_{p,sjm} + \eta_{0jm} ,$$

where $Z_{p,sjm}$ is the prefecture-level covariate s in prefecture j for the m-th housing type and η_{0jm} is the prefecture-level error term, which is assumed to be $i.i.N(0, \sigma_m^2)$.

Using generalized linear mixed model with random effects estimation methods (GLMM), Fu et al. (2015) show, at the household level, that redistributors (*e.g.*, cadres) and supporting clerical staff were more likely to achieve home ownership than manual workers did. Both non-agricultural status and working in state sectors confer benefits in obtaining reform-era housing with heavy subsidies or better qualities. When one takes into account education and earnings, the advantage of redistributors (*e.g.*, cadres) over manual workers in home ownership could be explained by work units. At the prefecture-level, they show that the marketization only reduced the local home ownership of self-built housing, affordable housing and privatized *danwei* housing but not that of commodity housing. In contrast, political and market connections promote all types of home ownership except self-built housing, and have significant positive association with the odds of renting public housing.

In the literature, numerous studies focus on how neighborhoods change in terms of income level, housing values, environment amenities or different racial preferences, etc. Racial and ethnic composition may have effects on neighborhood economic change (see for instance Sykes (2003)). Some studies have examined how neighborhood minority composition is associated with change in neighborhood relative economic status. This is, for instance, the case of the paper of Jun (2016) in a 3D framework. He uses the Neighborhood Change Database (NCDB), which includes the decennial census data across the USA from 1970 to 2000 at the census tract level. The multilevel modeling fits the data structure that a neighborhood is nested in a metropolitan area and allows for answering the research question whether the effect of neighborhood racial/ethnic composition on neighborhood economic change is conditioned by metropolitan-level factors. Jun (2016) shows that both neighborhood percentage Black and Hispanic are negatively related to neighborhood economic gain and are conditioned by metropolitan-level factors. Although this negative effect of neighborhood minority composition has been consistent over the four ten years panel, – the 1970s, 1980s, 1990s, and 2000s – its impact level is lower in the latest panel compared to the early panel. The negative effect of neighborhood minority composition has also declined by the interactions with metropolitan minority composition. In the later panels, metropolitan minority composition turned out to moderate the negative effect of neighborhood minority composition.

We now turn to studies focusing on residential mobility, location choice and the impact of the U.S. housing choice voucher program.[4] Eriksen and Ross (2015)

[4] The Section 8 housing choice voucher program is the federal government's major program for assisting very low-income families, the elderly, and the disabled to afford decent, safe, and sanitary housing in the private market. Since housing assistance is provided on behalf of the family or individual, participants are able to find their own housing, including single-family homes, townhouses and apartments. The participant is free to choose any housing that meets the requirements of the program and is not limited to units located in subsidized housing projects. A family that is issued a

estimate the effect of increasing the supply of housing vouchers on rents using a panel of housing units in the American Housing Survey. Their full-sample is a 3D panel dataset of 8, 388 rental housing units located in the 135 identified MSAs in the public use version of the AHS with a reported rent in 1997 and at least one other year between 1999 and 2003. Eriksen and Ross (2015) find an elasticity which is close to zero. When splitting the sample at arbitrary values of a relevant variable (ratio of the rent of a rental unit in the base year to the US Department of Housing and Urban Development fair market rent for that metropolitan statistical area in the same year), they find statistically significant negative effects for those units which were initially 80% below the 1997 ratio and statistically significant positive elasticities for those units which were between 80% and 120% of that ratio. They do not find that an increase in vouchers affected the overall price of rental housing but do estimate differences in effects based on an individual unit's rent before the voucher expansion. Their results are consistent with voucher recipients renting more expensive units after receiving the subsidy. They also find that the largest price increases were for units near the maximum allowable voucher rent in cities with an inelastic housing supply. Several research studies have shown that expanding the housing choice voucher program has varying price effects on different parts of the rental housing market. While recipients will certainly benefit from the added vouchers, there is ambiguity in the impacts on quality of life for non-recipients in the broader housing market. Excessive crowding in homes, dubbed overcrowding, is one measure of quality of life because overcrowding can lead to adverse outcomes such as stunted child development, adverse mental health, and an increased risk in spreading infectious diseases such as COVID-19.

Kole (2022) makes use of an exogenous increase in the supply of housing vouchers to explore its effects on overcrowding in the general housing market in the U.S. Using the dataset of Eriksen and Ross (2015), estimation of a linear probability model shows that adding vouchers reduces the incidence of overcrowding: a 10% increase in the supply of vouchers reduces the likelihood that a housing unit is overcrowded by 0.081 percentage points. The mechanism behind the effect is shown to correspond with anecdotes about overcrowding: households experiencing misfortune, financial difficulties, or other tenuous circumstances double-up with higher-income households. Hence, a voucher enables the troubled household to move into more suitable living arrangements.

Henderson, Soberon, and Rodriguez-Poo (2022) develop a nonparametric estimator that works for multidimensional fixed effects model, has a closed form solution and can be estimated in a single step. They apply their method on the relationship between the price of rental housing and housing vouchers using the dataset of Eriksen and Ross (2015). As previously stated, an heterogeneous result with respect to the ratio of the rent of a rental unit to the US Department of Housing and Urban Development fair market rent has been observed in the literature (Eriksen and Ross (2015)) using arbitrary splits of the sample. Henderson et al. (2022) avoid these arbitrary splits by adopting a semiparametric approach whereby they obtain an elas-

housing voucher is responsible for finding a suitable housing unit of the family's choice where the owner agrees to rent under the program.

ticity for each rental unit in the sample and confirm existence of both negative and positive impacts of housing vouchers on the price of rental housing. They find that positive elasticities are concentrated in the Western United States and specifically in areas which are more supply inelastic, but overall, negative elasticities are more prominent in this dataset. This suggests that increasing rents for those who do not receive subsidies are likely localized and not predominant in the U.S.

Explaining the residential choices and the residential mobility is not sufficient. It seems important to jointly model the residential mobility and the duration of stay at a location preceding the relocation. A lot of research treated the decision to move as a binary choice decision (move/no-move) and modeled this decision as a function of various factors (see above). Others have used duration models (see Deng, Gabriel, and Nothaft (2003)) to represent the stay at a location between moves, treating the reason for a move as an exogenous variable. An interesting study, done in a multi-dimensional framework by Eluru, Sener, Bhat, Pendyala, and Axhausen (2009), has extended these previous studies in three ways. First, the move decision is treated as an endogenous variable in a multinomial unordered choice modeling framework. Second, the duration of stay is modeled as a grouped choice, supposing that households treat the duration of stay at a residential location in terms of time-period ranges as opposed to exact continuous durations. Third, they consider heterogeneity of exogenous variables using random coefficients in both the equation for the move as well as the equation for the duration of stay preceding a relocation. In sum, Eluru et al. (2009) estimated a joint unordered choice-grouped choice model system with random coefficients.

Let the households be represented by the index $i = 1, \ldots, N$, let the different move reasons (*e.g.*, personal reasons, employment reasons, etc.) be represented by the index $m = 1, \ldots, M$ and let the duration categories (*e.g.*, < 2 years, $2-5$ years, $5-10$ years, etc.) be represented by the index $j = 1, 2, \ldots, J$. The specification of Eluru et al. (2009) allows the possibility of multiple move records per household defined by the index $t = 1, 2, \ldots, T$ as the different moving choice occasions for households i. The system of equations jointly models the reason for move and the duration of stay as follows:

$$\begin{cases} u_{imt} = X_{it}'\beta_{im} + \eta_{im} + \varepsilon_{imt} \,, \\ d_{imt} = j \text{ if } \psi_{m,j-1} < d_{imt}^{*} < \psi_{m,j} \,, \end{cases}$$

$$\text{with } d_{imt}^{*} = X_{it}'\alpha_{im} \pm \eta_{im} + \zeta_{imt} \,.$$

The first equation of the system is associated with the random utility u_{imt} for a household i corresponding to the reason to move m at choice occasion t. The $(Q \times 1)$ vector X_{it} is the vector of attributes associated with household i and its choice environment (*e.g.*, sex, age, employment status, family type, transportation mode to work, etc.) at the t-th choice occasion. The $(Q \times 1)$ random coefficient vector $\beta_{im} = \beta_m + \gamma_{im}$ is the sum of a vector β_m of mean effects of the elements of X_{it} for move reason m and a random vector γ_{im} with its q-th element $(q = 1, \ldots, Q)$ representing unobserved factors specific to household i and his choice environment. η_{im} expresses unobserved individual factors that simultaneously impact the propensity of moving for a certain reason m and the duration of stay. ε_{imt} is an idiosyncratic random error term assumed to be identically and independently standard Gumbel distributed across individuals, move reasons and choice occasions.

The second equation of the system is associated with d^*_{imt}, being the latent (continuous) duration of stay for household i before moving for reason m at the t-th choice occasion. This latent duration is mapped to the grouped duration category d_{imt} by the ψ thresholds (with infinite bounds as in the usual ordered-response modeling framework). d_{imt} is observed only if the end of the duration of stay at a residential location is associated with alternative m. The $(Q \times 1)$ random coefficient vector $\alpha_{im} = \alpha_m + \delta_{im}$ is the sum of the vector α_m of mean effects for category m, and the random vector δ_{im} of unobserved factors specific to household i and his duration of stay. ζ_{imt} is an idiosyncratic random error term, assumed identically and independently logistic distributed across individuals, reasons for move, and choice occasions, with variance λ^2. The elements of the random vectors γ, δ and η are normally distributed: $\gamma_{imq} \sim N(0, \sigma^2_{\gamma_{mq}})$, $\delta_{imq} \sim N(0, \sigma^2_{\delta_{mq}})$ and $\eta_{im} \sim N(0, \sigma^2_{\eta_m})$ for $q = 1, ..., Q$.

Correlation in unobserved individual factors between the reason to move and the duration of stay may be positive or negative, it is indicated by the ± sign in front of η_{im} in the duration category equation. If a positive sign seems logical for the propensity of a move for a given reason m in the first equation, a negative sign in the second equation suggests that unobserved individual factors will decrease the duration of stay preceding such a potential move. In the estimation, Eluru et al. (2009) considered both the positive and negative signs on the η_{im} terms in the second equation of the system. But the negative sign for all m provided statistically superior results. Conditional on γ_{im} and η_{im} for each (and all) m, the probability of a household i choosing to move for reason m at the t-th choice occasion is given by

$$P_{imt} = \frac{\exp\left(X'_{it}\beta_{im} + \eta_{im}\right)}{\sum_{m=1}^{M} \exp\left(X'_{it}\beta_{im} + \eta_{im}\right)} .$$

Conditional on δ_{im} and η_{im}, the probability of a household i choosing to stay for a particular duration category j preceding a move for reason m at the t-th choice occasion is given by

$$R_{imtj} = G\left(\frac{\psi_{m,j} - \left\{X'_{it}\alpha_{im} \pm \eta_{im}\right\}}{\lambda}\right) - G\left(\frac{\psi_{m,j-1} - \left\{X'_{it}\alpha_{im} \pm \eta_{im}\right\}}{\lambda}\right) ,$$

where $G(.)$ is the cumulative distribution of the standard logistic distribution. Let Ω be a vector that include all the parameters β_m, α_m, λ, $\sigma_{\gamma_{mq}}$, $\sigma_{\delta_{mq}}$ and σ_{η_m} for $m = 1, ..., M$ and $q = 1, ..., Q$. Let c_i be a vector stacking the coefficients γ_{im}, δ_{im} and η_{im} across all m for household i. Let Σ be another vector stacking the standard error terms $\sigma_{\gamma_{mq}}$, $\sigma_{\delta_{mq}}$ and σ_{η_m} and let $\Omega_{-\Sigma}$ represent a vector of all parameters except the standard error terms. Then, the unconditional likelihood function for all the households is given by

$$L(\Omega) = \prod_{i=1}^{N} L_i(\Omega) = \prod_{i=1}^{N} \int_{c_i} \{L_i(\Omega_{-\Sigma}|c_i)\}\, d\Phi(c_i|\Sigma) \ , \tag{12.3}$$

$$\text{with } L_i(\Omega_{-\Sigma}|c_i) = \prod_{m=1}^{M} \prod_{t=1}^{T} \prod_{j=1}^{J} \left[P_{imt} R_{imtj}\right]^{D_{imt} E_{ijt}} \ ,$$

where $\Phi(.)$ is the multi-dimensional cumulative normal distribution and $L_i(\Omega_{-\Sigma}|c_i)$ is the likelihood function, for household i and for a given value of $\Omega_{-\Sigma}$ and c_i. D_{imt} (resp. E_{ijt}) is a dummy variable taking a value of 1 if household i chooses to move for reason m (resp. chooses to stay for duration category j) on the t-th choice occasion and 0 otherwise. Equation (12.3) needs the evaluation of a multi-dimensional integral of size equal to the number of rows in c_i. Eluru et al. (2009) apply Quasi-Monte Carlo simulation techniques based on the Halton sequence to approximate this integral in the likelihood function and maximize the logarithm of the resulting simulated likelihood function across individuals with respect to Ω (see Bhat (2001, 2003)). Eluru et al. (2009) use a longitudinal data set of households from a stratified sample of municipalities in the Zurich region of Switzerland over the period 1985-2004. The data set includes 1012 households and 2590 move records. They found that several demographic, socioeconomic, and commute related variables (*e.g.*, age, gender, family reasons, education/employment reasons, accommodation related reasons, surrounding environment related reasons, vicinity to family and friends, etc.) have a significant influence on the reason for move and the duration of stay. In the duration of stay model, Eluru et al. (2009) found that household size creates heterogeneity across the sample of households. They show that people who own dwellings have a lower probability of moving for surrounding vicinity related reasons than those renting their units. Likewise, people who live in smaller homes have higher probabilities of short duration stays probably because they are looking for larger homes. Having a mix of job opportunities located close to residential neighborhoods increases the duration of stay in the dwelling. Reducing commute distances promotes longer durations of stay, etc. Eluru et al. (2009) found that common unobserved factors jointly affect the reason to move and the duration of stay and calls for a joint modeling framework that allows error correlation structures.

Endogeneity (or simultaneity) is a fundamental aspect of modelling housing that should be taken into account both for hedonic housing price functions and for choice models of residential location. This is the object of the next section.

12.5 Multi-dimensional Dynamic Models of Housing Models

In hedonic housing price functions, some explanatory variables, in addition to the dependent variable and its spatial lag, may be endogenous following the simultaneous choice of the house price and of the quantities of attributes. This is particularly true for floor space (see Fingleton and LeGallo (2008) who extended Kelejian and Prucha (1998) feasible generalized spatial two-stage least squares estimator to account for endogenous variables due to system feedback, given an autoregressive or a moving average error process). As for hedonic price functions, endogeneity is expected to occur mainly as a result of the omission of attributes in discrete choice models of residential mobility. In the literature, several methods have been proposed to consider endogeneity. Berry, Levinsohn, and Pakes (1995) proposed a fixed effects procedure, by product and market, to solve market-level endogeneity in the automobile sector. Guevara and Ben-Akiva (2006) applied to residential location choice models the control function method, which is based on the inclusion of an additional variable that controls for the endogeneity problem (see Heckman (1978) and Blundell and Powell (2004)). They applied residential location choice models based on 630 households of renters who had moved to their present location between 1999 and 2001 in Santiago (Chile). The results show that price endogeneity is significant in choice models of residential location and that the control function method is suitable to account for it.

Endogeneity is not limited to the correlation between the dependent variables and attributes (in the equation or omitted) or to the simultaneity of demand and supply, the marginal willingness to pay and the marginal willingness to accept. Location choices and housing investments are fundamentally dynamic decisions over multiple time periods. In the 2D panel data literature, some dynamic models have been applied on real estate topics. For instance, Engle, Lilien, and Watson (1985) used a version of a dynamic multiple-indicator multiple-cause (DYMIMIC) model for an hedonic price model of the resale housing market for a suburb of San Diego, California during the period 1973-1980. The specification of the model features hedonic equations for each house sale and a dynamic equation for the capitalization rate which is taken to be an unobservable time series to be estimated jointly with the unknown parameters. Engle et al. (1985) used maximum likelihood with an EM algorithm based upon Kalman filtering. Some authors have used, in a 2D framework, the dynamic factor models (DFM) and/or large-scale Bayesian vector autoregressive (LBVAR) models to forecast housing price. These models are interesting to study the "ripple effect", *i.e.*, the propagation of shocks to house prices across regions. For instance, Das, Gupta, and Kabundi (2010) forecast regional house price inflation for five metropolitan areas of South Africa using principal components obtained from quarterly macroeconomic time series in the period 1980 to 2006. In the majority of the cases, the dynamic factor model statistically outperforms the vector autoregressive models, using both the classical and the Bayesian treatments. They also considered spatial and non-spatial specifications. Das et al. (2010) indicate that macroeconomic fundamentals in forecasting house price inflation are important. Li and Leatham (2011) investigate moving trends of house prices in 42 metropolitan areas in the United States from the perspective of large-scale models, which are also DFM and LBVAR models. These models accommodate a large panel data comprising 183 monthly series for the U.S. economy, and an in-sample period of 1980 to 2007 are used to forecast one- to twelve-months-ahead house price growth rate over the out-of-sample horizon of 2008 to 2010. Li and Leatham (2011) show that DFM consistently outperforms its LBVAR alternative for forecasting the house price growth rate for the overall U.S. housing market. The forecasting power of DFM does not decrease as the number of forecasted period ahead increases, while LBVAR has its best performance for two-month-ahead forecast and then its forecasting accuracy decays. Beenstock and Felsenstein (2015) using data from 9 regions of Israel over 1987-2010 apply spatial panel cointegration methods for a dynamic model of regional housing markets in which people prefer to live where housing is cheaper and building contractors prefer to build in regions where construction is more profitable. Based on dynamic hedonic price functions, the analysis of nonstationary spatial panel data shows that although housing starts vary directly with profitability as measured by house prices relative to building costs, they vary inversely with profitability in neighboring regions. Beenstock and Felsenstein (2015) show that there is a non negligible spatial substitution in housing construction and this substitution effect suggests that contractors have local building preferences since they regard neighboring regions as close substitutes but not more distant regions. Abate and Anselin (2016) investigate the interactions between house price fluctuations and output growth rate across 373 metropolitan statistical areas in the US over the period 2001-2013. In a panel data context, they use time varying spatial econometric hedonic price functions. They show that the spatial correlation coefficient across metropolitan areas has been increasing over time, indicating an increasing synchronization of house prices across metropolitan statistical areas during the sample period.

Spatio-temporal models of hedonic price functions have been recently proposed to jointly take into account time effects and spatial effects either through multifactor error structure or through specific weight matrices. For instance, Holly et al. (2010) considered the determination of real house prices in a panel made up of 49 US States over 29 years. An error correction model with a cointegrating relationship between real house prices and real incomes is found once they take proper account of both heterogeneity and cross sectional dependence. See also Latif (2015) for a study on the impact of new immigration on housing rent, using Canadian province-level panel data from 1983 to 2010. Latif (2015) uses panel cointegration regressions and panel vector error correction models and shows that immigration flow has a significant positive impact on housing rent both in the short run and in the long run. There are also extensions of the spatial hedonic price functions which use a weight matrix that expresses spatio-temporal relations instead of purely spatial. A general $(N \times N)$ spatio-temporal weight matrix W is obtained by splitting its construction into two separate matrices of the same dimension. The first matrix, S, captures the spatial relations among the N observations and a second matrix, T, expresses the temporal direction of observations. Smith and Wu (2011) have proposed a spatio-temporal weight matrix defined as the Hadamard product between two spatial and temporal distance weight matrices $W = S \odot T = \left[s_{jl}\right] \odot \left[t_{jl}\right]$. It identifies the spatio-temporal neighbors that affect hedonic price determination. The elements s_{jl} indicate the way observation j is spatially connected to observation l. The elements on the diagonal s_{jj} are set to zero while the off-diagonal elements are defined by an inverse distance function: $s_{jl} = d_{jl}^{-\gamma}$ if $d_{jl} < \overline{d}$ and 0 elsewhere where d_{jl} is the geographic distance between locations j and l, $d_{jl} < \overline{d}$ is a critical cut-off and $\gamma \geq 0$. The elements t_{jl} represent the time elapsed between realization of observations j and l. One assumes that observations have been ordered chronologically so the first row of T corresponds to the earliest observation while the last row corresponds to the latest observation.

The elements on the diagonal t_{jj} are set to zero while the off-diagonal elements are defined by an inverse function of the time elapsed between two observations: $t_{jl} = |t_j - t_l|^{-\alpha}$ if $|t_j - t_l| < \bar{t}$ and 1 elsewhere. t_j (resp. t_l) is the time when dwelling j (resp. l) is sold. $\bar{t}$ is a critical cut-off value and α is a penalty parameter to be fixed. Several authors have used spatio-temporal models of hedonic price functions with standard spatial specifications (spatial autoregressive (SAR), spatial error (SEM), spatial Durbin model, etc.) but with different spatio-temporal matrices W. They got better results in terms of estimation and/or forecasting as compared to those obtained with usual purely spatial weight matrices. See for instance, Pace, Barry, Gilley, and Sirmans (2000) for an application on the residential market of Bâton Rouge, Louisiana, during 1984-1992, Liu (2013) for an application of housing in Randstad, The Netherlands, during the years 1997-2007, Nappi-Choulet and Maury (2011) for the residential market of Paris for the years 1995-2005, or Thanos, Dubé, and Legros (2016) for the Aberdeen, Scotland, housing market during 2004-2007, to mention a few. Unfortunately, and to our knowledge, nobody has used these spatio-temporal multifactor error structures or the spatio-temporal weight matrices in a three-dimensional framework. But, it could be a promising development for future research.

The developments in the dynamics of modelling housing is not only focused on hedonic price functions. Some authors have been interested in dynamic versions of discrete models of location choice. Forward-looking behavior in the housing market justify dynamic considerations in a model of location choice. Several authors have underlined the need to use dynamic specifications of modelling housing. For instance, Case, Shiller, and Thompson (2012), using questionnaire surveys for home buyers in four U.S. cities over 2003-2012, have shown that the root causes of the speculative bubble can be seen in their long-term home price expectations, which reached abnormal levels relative to the mortgage rate at the peak of the boom and declined sharply since. The downward turning point around 2005 of the long boom that preceded the crisis was associated with changing public understanding of speculative bubbles. But estimating dynamic discrete models of location choice is a rather challenging and stimulating objective and is technically difficult. Bayer et al. (2016) noted that first, estimation of residential sorting and hedonic equilibrium models needs to match a large sample of households, their characteristics to the location and the features of their housing choices. Second, the high dimensionality of the state space (consisting of current lifetime utilities and neighborhood characteristics) – required to define the evolution of a urban system – leads to the curse of dimensionality which puts a brake on the estimation of an acceptable sized dynamic model of residential location decisions.

Diao, Ma, and Ferreira (2015) propose a real-option based dynamic model to simulate real estate developer behavior. In a three-dimensional framework (property, type of property and time for private residential housing in Singapore during 1995-2012), they extend the standard discrete choice model approach by adding an explicit probabilistic representation of development templates available to developers to take into account both developers' option to hold the land undeveloped and the market volatility of different development types. In their proposed simulation framework, Diao et al. (2015) suppose that a developer making investment decisions for a parcel faces a set of alternative development templates in a market with uncertainty. At each time period, the developer estimates future revenue and construction cost of feasible development templates under planning constraints and related real option values. He chooses the template based on the principle of profit maximization, but only does so if the return of the development template is greater than a threshold level (value of the call option), which is a function of the market volatility of the built property as suggested by the real option theory, otherwise, keeps the *status quo*. The model components in the proposed simulation framework are calibrated with private housing data in Singapore. The results show significant volatility in housing price and construction costs, relevant differences in volatility across housing types, and good fit, in the hedonic model, of market prices and construction costs. This kind of research contributes to the microsimulation literature by proposing an interesting approach which takes into account the dynamic and volatile nature of the real estate market but, unfortunately, this remains a simulation study.

Bayer et al. (2016) have proposed a new approach for estimating a three-dimensional dynamic model of demand for houses and neighborhoods that is computationally tractable. Using a semi-parametric estimation approach, they control for unobserved household and neighborhood heterogeneity. Their model adapts dynamic demand models for durable goods in a housing market context. They treat houses as assets and allow households wealth to evolve endogenously. Households anticipate selling their homes at some point in the future and then consider the expected evolution of neighborhood amenities and housing price when deciding where and when to purchase, or sell, their house. They relax the index sufficiency assumption which is standard in the dynamic demand literature. This assumption helps to deal with the computational challenges posed by the large state space typically arising in models of dynamic demand. Instead of treating the logit inclusive value as a sufficient statistic for predicting future continuation values, Bayer et al. (2016) define the continuation value from predicted future lifetime utilities, which depend on the state space in a flexible manner. Last, they use stable and uniform realtor fees to estimate the marginal utility of consumption without the need for a price instrument. They use the fact that households face a monetary trade-off both in the standard sense of deciding which product (neighborhood) to purchase but also in terms of deciding when to move. They take advantage of the fact that realtor fees during the sample period were quite uniform (6% of the house value) in order to identify the marginal utility of consumption when estimating each resident's move-stay decision. The decision variable, d_{it}, denotes both of the choices made by household i in period t, whether to move and where to move, conditional on deciding to move. If a household decides to move, the decision is denoted $d_{it} = j$, $j = 0, 1, ..., J$ where j indexes neighborhoods, J denotes the total number of neighborhoods in the region and 0 denotes the outside option. The data concern housing transactions in the San Francisco Bay Area from 1994-2004 for more than $220,000$ households and 2398 neighborhoods. We give only some results as the paper is highly technical. But, the model and estimation procedure presented in this paper are very general and can be applied to a broad range of dynamic studies in housing markets. The model uses a two-stage estimator. In the first stage, Bayer et al. (2016) use the household location and the mobility decisions to estimate the value of lifetime expected utility for each neighborhood, time period, and household type as well as an unobservable characteristic that captures a household's preference for sub-regions within the San Francisco Bay Area. In the second stage, they recover fully-flexible estimates of per-period utility and regress them on a set of observable attributes. They use a semi-parametric estimation approach to control for the endogeneity of price in this second stage, utilizing outside information relating to the financial cost of moving to pin down the coefficient on house prices. The results indicate that the downward biases associated with static demand estimation are significant for three important non-marketed amenities: air quality, crime, and neighborhood race. For instance, for a 10% change in each amenity, the static model overestimates the willingness to pay for living in close proximity to neighbors of the same race for low-income households. The static estimation is $1,627.03 whereas the corresponding dynamic estimation is $612.09. For high-income households, the bias runs in the opposite direction and the static model underestimates the willingness to pay by a factor of more than two. The static model always underestimates the willingness to pay for living in close proximity to crimes. For low-income households and for a 10% increase in violent crime, the static estimation is -$291.14 while the corresponding dynamic estimation is -$350.18. This is also true for air pollution.

12.6 From Variational Inference ... to Multi-dimensional Housing Models

The presence of numerous latent variables, omitted variables, the definition of dynamic and spatial structures within multi-dimensional frameworks (3D, 4D or more) and the econometric complexity that results will not make things easier and must move us towards the use of flexible models and methods. Among many others promising future pathways is probably the use of variational Bayesian approximations (see for instance Ormerod and Wand (2010), C. Y. Lee and Wand (2016), Blei, Kucukelbir, and McAuliffe (2017), Baltagi, Bresson, and Etienne (2019), Baltagi, Bresson, and Etienne (2020) or Nolan, Menictas, and Wand (2020) to mention a few). These methods facilitate approximate inference for the parameters in complex statistical models and provide fast, deterministic alternatives to Monte Carlo methods to potentially overcome many problems in applied modelling of housing. While the two-level approaches have been developed over the past decade, the three-level approach was recently proposed by Nolan et al. (2020). Their streamlined variational inference algorithm for three-level random effects models relies on theorems provided by Nolan and Wand (2020) concerning linear system solutions and sub-blocks of matrix inverses for three-level sparse matrix problems which are the basis for their QR-decomposition-based streamlined algorithm. This algorithm allows one to obtain mean field variational Bayes approximate posterior density functions for the parameters in the three-level linear mixed model. We will briefly present their method and apply it to data on housing transaction prices in Paris in 2003 for 20 *arrondissements*, each divided in four *quartiers*. Even though four-level and even higher level situations may be of interest, unfortunately the required theory is not yet in place.

Inference based on MCMC can be very slow for such two-level or three-level linear mixed model and MCMC methods may suffer from poor mixing. Variational Bayesian inference can help in tackling the scalability challenge of big data sets and/or models with large sparse covariance matrices as they use a deterministic optimization approach to approximate the posterior distribution. The parameters of the approximate distribution are chosen to minimize some measure of distance (as the Kullback-Leibler divergence) between the approximation and the posterior. Mean field variational Bayes approximation is analogous to Gibbs sampling for conjugate models (see C. M. Bishop (2006), Ormerod and Wand (2010), Pham, Ormerod, and Wand (2013) and C. Y. Lee and Wand (2016) to mention a few).

Consider a generic Bayesian model with observed vector y and parameter vector θ that is continuous over the parameter space Θ. The Bayes theorem allows one to define the posterior distribution as:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)\, p(\theta)}{p(y)} \text{ with } p(y) = \int_{\Theta} p(\theta, y)\, d\theta$$

Let q be an arbitrary density function over Θ. Then, the logarithm of the marginal likelihood satisfies (see C. M. Bishop (2006), Ormerod and Wand (2010)):

$$\begin{aligned}
\log p(y) &= \log p(y) \int_{\Theta} q(\theta)\, d\theta = \int_{\Theta} q(\theta) \log p(y)\, d\theta \\
&= \int_{\Theta} q(\theta) \log \left\{ \frac{p(\theta, y)/q(\theta)}{p(\theta|y)/q(\theta)} \right\} d\theta \\
&= \int_{\Theta} q(\theta) \log \left\{ \frac{p(\theta, y)}{q(\theta)} \right\} d\theta + \int_{\Theta} q(\theta) \log \left\{ \frac{q(\theta)}{p(\theta|y)} \right\} d\theta \\
&= \log \underline{p}(y, q) + KL(q, p)
\end{aligned}$$

where $KL(q, p)$ is the Kullback-Leibler divergence between $q(\theta)$ and $p(\theta|y)$. Furthermore, $\log \underline{p}(y, q)$ is a lower bound on the marginal log-likelihood. The Kullback-Leibler divergence becomes

$$\begin{aligned}
KL(q, p) &= E_q[\log q(\theta)] - E_q[\log p(\theta|y)] \\
&= E_q[\log q(\theta)] - E_q[\log p(\theta, y)] + \log p(y)
\end{aligned}$$

where the last term, $\log p(y)$, is a constant. The minimization of the Kullback-Leibler divergence is thus equivalent to maximizing the scalar quantity,

$$\log \underline{p}(y, q) = E_q\left[\log \left(\frac{p(\theta, y)}{q(\theta)} \right)\right]$$

which is usually referred as the evidence lower bound (ELBO).
Let $\{\theta_1, ..., \theta_M\}$ be a partition of the parameter vector θ. The MFVB approximates the posterior distribution $p(\theta|y)$ by the product of the q-densities:[5]

$$q(\theta) = \prod_{j=1}^{M} q_j(\theta_j)$$

The optimal q-densities which minimize the Kullback-Leibler divergence are given by

$$q_j^*(\theta_j) \propto \exp\left[E_{q(-\theta_j)}\left\{\log p\left(\theta_j|\text{ rest}\right)\right\}\right], \; j = 1, ..., M \tag{12.4}$$

where $E_{q(-\theta_j)}$ denotes expectation with respect to $\prod_{k\neq j} q_k(\theta_k)$.
rest $\equiv \left\{y, \theta_1, ..., \theta_{j-1}, \theta_{j+1}, ..., \theta_M\right\}$ is the set containing the rest of the random vectors in the model, except θ_j and the distributions $\left(\theta_j|\text{ rest}\right)$ are the full conditionals in the MCMC literature.
The iterative scheme for obtaining the optimal q-densities under product restriction (12.4) is

---

1. Initialize $q_1^*(\theta_1)\ , q_2^*(\theta_2)\ , \cdots , q_M^*(\theta_M)$
2. Cycle through updates:

$$q_1^*(\theta_1) \leftarrow \frac{\exp\left[E_{q(-\theta_1)}\{\log p(y,\theta)\}\right]}{\int \exp\left[E_{q(-\theta_1)}\{\log p(y,\theta)\}\right] d\theta_1}$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$q_M^*(\theta_M) \leftarrow \frac{\exp\left[E_{q(-\theta_M)}\{\log p(y,\theta)\}\right]}{\int \exp\left[E_{q(-\theta_M)}\{\log p(y,\theta)\}\right] d\theta_M}$$

until the increase in $\log \underline{p}(y, q)$ is negligible.

Compared to the minimization of the KL divergence, the maximization of the ELBO is often a more convenient objective of the optimization over the free distributional parameters. Nolan et al. (2020) apply this principle and derive the MFVB approximation of the three-level linear mixed model.
Suppose that the response data vector y is modeled according to a Bayesian version of the Gaussian linear mixed model[6]

$$y|\beta, u, R \sim N(X\beta + Zu, R), \, u|G \sim N(0, G), \beta \sim N(\mu_\beta, \Sigma_\beta)$$

for hyperparameters μ_β and Σ_β and such that β and $u|G$ are independent. X and Z are the fixed effects and random effects design matrices associated with the fixed

[5] This is known as the *mean field restriction*. The term *mean field* originated from physics.
[6] This draws heavily from the results in Nolan et al. (2020).

effects and random effects vectors β and u.[7] Prior specification for the covariance matrices G and R involves auxiliary covariance matrices A_G and A_R with conjugate Inverse G-Wishart distributions (generalization of inverse Wishart distributions)[8] (Wand (2017)).

Full Bayesian inference for the β, G and R and the random effects u involves the posterior density function $p(\beta, u, A_G, A_R, G, R|y)$, but typically is analytically intractable and Markov chain Monte Carlo approaches are required for practical exact inference. Variational approximate inference involves mean field restrictions such as

$$\begin{aligned} p\left(\beta, u, A_G, A_R, G, R|y\right) &\approx q\left(\beta, u, A_G, A_R\right) q\left(G, R\right) \\ &\approx q\left(\beta, u\right) q(A_G) q(A_R) q(G) q(R) \end{aligned} \tag{12.5}$$

Then, the forms and optimal parameters for the q-densities are obtained by minimizing the Kullback-Leibler divergence of the right-hand side of (12.5) from its left-hand side. The optimal q-density parameters are interdependent and a coordinate ascent algorithm (see Ormerod and Wand (2010)) is used to obtain their solution. The optimal q-density for (β, u), denoted by $q^*(\beta, u)$, is a multivariate Normal density function with mean vector $\mu_{q(\beta,u)}$ and covariance matrix $\Sigma_{q(\beta,u)}$. The coordinate ascent algorithm is such that they are updated according to (see C. Y. Lee and Wand (2016), Nolan et al. (2020))[9]

$$\Sigma_{q(\beta,u)} \leftarrow \left\{ C' E_q(R^{-1}) C + \begin{bmatrix} \Sigma_\beta^{-1} & 0 \\ 0 & E_q(G^{-1}) \end{bmatrix} \right\}$$

$$\mu_{q(\beta,u)} \leftarrow \Sigma_{q(\beta,u)} C' E_q(R^{-1}) \left(y + \begin{bmatrix} \Sigma_\beta^{-1} \mu_\beta \\ 0 \end{bmatrix} \right)$$

where $E_q(G^{-1})$ and $E_q(R^{-1})$ are the q-density expectations of G^{-1} and R^{-1} and $C = [X, Z]$.

Let us now turn our attention to the mean field variational Bayes (MFVB) approximation of the three-level linear mixed model proposed by Nolan et al. (2020).

$$y_{ij}|\beta, u_i^{L1}, u_{ij}^{L2}, \sigma^2 \sim N\left(X_{ij}\beta + Z_{ij}^{L1} u_i^{L1} + Z_{ij}^{L2} u_{ij}^{L2}, \sigma_\varepsilon^2 I\right)$$

where $1 \le i \le m$ is the index of the outer group (the 20 *arrondissements* of Paris) and $1 \le j \le n_i$ is the index of the inner group (the 4 *quartiers* of each *arrondissement* of Paris in our particular example). Inside each inner group, we have o_{ij} observations

---

[7] This terminology is different from what the panel data literature dubs as "fixed" and "random" effects.

[8] See Appendix.

[9] Very often, $R = \sigma_\varepsilon^2 I$ where I is an identity matrix and where $(\sigma_\varepsilon^2 | a_{\sigma_\varepsilon^2})$ follows an inverse χ^2 and $a_{\sigma_\varepsilon^2} (\equiv A_R)$ also follows an inverse χ^2 distribution.

for each variable (o_{ij} housing transaction prices in our example). Of course, we can have unbalanced inner groups with, in each, unbalanced observations for each variable. y_{ij} is a $\left(o_{ij} \times 1\right)$ vector. X_{ij} is an $(o_{ij} \times p)$ matrix of covariates, Z_{ij}^{L1} (resp. Z_{ij}^{L2}) is an $(o_{ij} \times q_1)$ (resp. $(o_{ij} \times q_2)$) block-diagonal matrix of the X_{ij} submatrices.[10] Thus, we have a three level random-intercept-random-slopes model with the first level (outer group), the second level (inner group) and the third level being the set of observations in each inner group.
The full Bayesian model (with priors on parameters and hyperparameters) is given by (see Nolan et al. (2020)):

$$y_{ij} \mid \beta, u_i^{L1}, u_{ij}^{L2}, \sigma^2 \sim N\left(X_{ij}\beta + Z_{ij}^{L1}u_i^{L1} + Z_{ij}^{L2}u_{ij}^{L2}, \sigma_\varepsilon^2 I\right)$$

with

$$\begin{cases} \begin{pmatrix} u_i^{L1} \\ u_{ij}^{L2} \end{pmatrix} | \Sigma^{L1}, \Sigma^{L2} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma^{L1} & 0 \\ 0 & \Sigma^{L2} \end{pmatrix}\right) \\ \beta \sim N\left(\mu_\beta, \Sigma_\beta\right), \sigma_\varepsilon^2 | a_{\sigma_\varepsilon^2} \sim \text{Inverse-}\chi^2(\nu_{\sigma_\varepsilon^2}, 1/a_{\sigma_\varepsilon^2}) \\ a_{\sigma_\varepsilon^2} \sim \text{Inverse-}\chi^2(1, 1/(\nu_{\sigma_\varepsilon^2} s_{\sigma_\varepsilon^2}^2)) \\ \Sigma^{L1} | A_{\Sigma^{L1}} \sim \text{Inverse-G-Wishart}(G_{full}, \nu_{\Sigma^{L1}} + 2q_1 - 2, (A_{\Sigma^{L1}})^{-1}) \\ A_{\Sigma^{L1}} \sim\sim \text{Inverse-G-Wishart}(G_{diag}, 1, \left\{\nu_{\Sigma^{L1}} diag(s_{\Sigma^{L1},1}^2, \cdots, s_{\Sigma^{L1},q_1}^2)\right\}^{-1}) \\ \Sigma^{L2} | A_{\Sigma^{L2}} \sim \text{Inverse-G-Wishart}(G_{full}, \nu_{\Sigma^{L2}} + 2q_2 - 2, (A_{\Sigma^{L2}})^{-1}) \\ A_{\Sigma^{L2}} \sim\sim \text{Inverse-G-Wishart}(G_{diag}, 1, \left\{\nu_{\Sigma^{L2}} diag(s_{\Sigma^{L2},1}^2, \cdots, s_{\Sigma^{L2},q_2}^2)\right\}^{-1}) \end{cases}$$

and $\nu_{\Sigma^{L1}} > 0$, $\nu_{\Sigma^{L2}} > 0$, $s_{\Sigma^{L1},1}^2, \cdots, s_{\Sigma^{L1},q_1}^2 > 0$ and $s_{\Sigma^{L2},1}^2, \cdots, s_{\Sigma^{L2},q_2}^2 > 0$. The minimal mean field restriction needed for a tractable variational inference algorithm is

$$p(\beta, u, a_{\sigma_\varepsilon^2}, A_{\Sigma^{L1}}, A_{\Sigma^{L2}}, \sigma_\varepsilon^2, \Sigma^{L1}, \Sigma^{L2}|y) \approx q(\beta, u, a_{\sigma_\varepsilon^2}, A_{\Sigma^{L1}}, A_{\Sigma^{L2}})q(\sigma_\varepsilon^2, \Sigma^{L1}, \Sigma^{L2})$$

The optimal q-density functions for the parameters of interest being as follows:

$$\begin{cases} q^*(\beta, u) \text{ has a } N(\mu_{q(\beta,u)}, \Sigma_{q(\beta,u)}) \text{ density function} \\ q^*(\sigma_\varepsilon^2) \text{ has an Inverse-}\chi^2(\xi_{q(\sigma_\varepsilon^2)}, \lambda q(\sigma_\varepsilon^2)) \text{ density function} \\ q^*(\Sigma^{L1}) \text{ has an Inverse-G-Wishart}(G_{full}, \xi_{q(\Sigma^{L1})}, \Lambda_{q(\Sigma^{L1})}) \text{ density function} \\ q^*(\Sigma^{L2}) \text{ has an Inverse-G-Wishart}(G_{full}, \xi_{q(\Sigma^{L2})}, \Lambda_{q(\Sigma^{L2})}) \text{ density function} \end{cases}$$

[10] If $q_1 \neq q_2$, it means that some random effects of level 1 or level 2 concern only some covariables. If $p \neq q_1$ (with $q_1 < p$) or $p \neq q_2$ (with $q_2 < p$), it also means that some covariables have only fixed effects. In our example, we will assume that $p = q_1 = q_2$.

and Appendix B.2 of Nolan et al. (2020) provides expressions for the q-densities for naïve mean field variational Bayesian inference for the parameters (see also C. Y. Lee and Wand (2016)). But, due to the nature of the random effects design matrices, Z^{L1} and Z^{L2} are large sparse matrices. Indeed, X, Z^{L1} and Z^{L2} are respectively $(N \times p)$, $(N \times N_{g_1})$ and $(N \times N_{g_2})$ matrices with $N = \sum_{i=1}^{m} \sum_{j=1}^{n_i} o_{ij}$, $N_{g_1} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} q_1$ and $N_{g_2} = \sum_{i=1}^{m} \sum_{j=1}^{n_i} q_2$. The total number of elements in $[X, Z^{L1}, Z^{L2}]$ is $N(p+N_{g_1}+N_{g_2})$. The number of non-zero elements in $[X, Z^{L1}, Z^{L2}]$ is $N(p+q_1+q_2)$. And so by difference, the number of zeros elements in $[X, Z^{L1}, Z^{L2}]$ is $N(N_{g_1}+N_{g_2}-q_1-q_2)$. If we go back to our example of $m = 20$ *arrondissements* and $(n_1 = \cdots = n_m = 4)$ *quartiers*, assuming a model with only one explanatory variable and one intercept $(p = q_1 = q_2 = 2)$ and assuming 100 transactions per *quartier* $(o_{ij} = 100, \forall i, j)$, then the total number of elements in $[X, Z^{L1}, Z^{L2}]$ is $2,576 \times 10^3$ and the ratio of zero elements is 98.136%. Nolan et al. (2020) and Nolan and Wand (2020) have proposed streamlined variational Bayes approximations with specific solutions to multilevel sparse matrix problems using only the non-zero elements of $[X, Z^{L1}, Z^{L2}]$ (in our case using only 1.864% of all the elements). They show that the following relatively small sub-blocks of $\Sigma_{q(\beta,u)}$ are required for variational inference concerning σ_ε^2, Σ^{L1} and Σ^{L2}:

$$\begin{gathered}\Sigma_{q(\beta)}, \Sigma_{q(u_i^{L1})}, \Sigma_{q(u_{ij}^{L2})}, E_q\left[\left(\beta - \mu_{q(\beta)}\right)\left(u_i^{L1} - \mu_{q(u_i^{L1})}\right)'\right] \\ E_q\left[\left(\beta - \mu_{q(\beta)}\right)\left(u_{ij}^{L2} - \mu_{q(u_{ij}^{L2})}\right)'\right] \text{ and } E_q\left[\left(u_i^{L1} - \mu_{q(u_i^{L1})}\right)\left(u_{ij}^{L2} - \mu_{q(u_{ij}^{L2})}\right)'\right]\end{gathered} \tag{12.6}$$

and the MFVB updates of $\mu_{q(\beta)}$ and each of the sub-blocks of $\Sigma_{q(\beta,u)}$ in (12.6) are expressible as a three-level sparse matrix least squares problem. Nolan et al. (2020) propose a QR-decomposition-based streamlined algorithm for obtaining MFVB approximate posterior density functions for the parameters in the three-level linear mixed model.[11]

As an illustration, we apply Nolan et al. (2020)'s method on a subset of the housing transaction prices data of Baltagi et al. (2015). We use only the year 2003 and the associated transaction prices of the 10,983 sold flats. As shown in Figure 12.1, Paris is divided into $m = 20$ *arrondissements*, each of which is divided into $n_i = n = 4$ *quartiers* (denoted from 1 to 80 on Figure 12.1).[12]

To fit the logarithm of the housing price (in euros per square meter), we use 6 covariates: the surface of the flat (in sq.m), the number of bathrooms, the garage plots, the number of maid's rooms, the floor level of the flat and the presence (or

[11] For a balanced model, with $n_i = n$ and $o_{ij} = o$, Nolan et al. (2020) show that the order of magnitudes of the number of floating point operations for naïve MFVB and QR-decomposition-based streamlined MFVB algorithms are $O(m^3 n^3 o)$ and $O(mno)$ and therefore the streamlined MFVB can be 78 times to 2,300 times faster than the naïve MFVB depending on the size of m. The R code of this QR-decomposition-based streamlined algorithm is available at http://matt-wand.utsacademics.info/statsPapers.html.

[12] However, for the writing of the three-level model and for the Figures, we re-number the *quartiers* from 1 to 4. Thus, the *quartiers* (5,6,7,8) of the 2nd *arrondissement* are noted (1,2,3,4), ... and the *quartiers* (77, 78, 79, 80) of the 20th *arrondissement* are also noted (1,2,3,4).

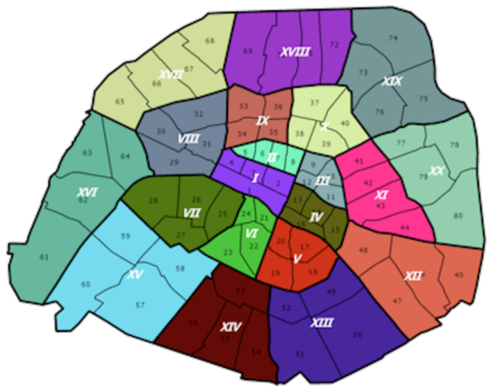

Fig. 12.1: Administrative division of Paris.

not) of a balcony.

$$
\begin{aligned}
y_{ij} &= \sum_{k=1}^{p=7} \left(\beta_k + u_{i,k}^{L1} + u_{ij,k}^{L2}\right) X_{ij,k} + \varepsilon_{ij} \\
&\quad 1 \le i \le m = 20,\ 1 \le j \le n_i (\equiv n) = 4,\ \forall i
\end{aligned}
$$

where $\left(\beta_1 + u_{i,1}^{L1} + u_{ij,1}^{L2}\right)$ is the random intercept. Results for the fixed effects estimates ($\mu_{q(\beta)}$) and their standard errors (from $\Sigma_{q(\beta)}$) are given in Table 12.1. Coefficients are significantly different from zero at 5% level and this model fits well the data (the R^2 is close to one). Estimate of σ_ε (= 0.259) is small as compared to the mean (= 8.194) of the dependent variable. But more interestingly, while the estimated variance of the fixed effect of the intercept ($\Sigma_{q(\beta_1)}$) is small, the mean of the estimated variances of the random effects of the outer group ($(1/m)\sum_{i=1}^{m} \Sigma_{q(u_{i,1}^{L1})}$) is larger while the mean of the estimated variances of the random effects of the inner group is smaller ($(1/(m.n))\sum_{i=1}^{m}\sum_{j=1}^{n} \Sigma_{q(u_{ij,1}^{L2})}$). Thus, there is indeed more heterogeneity between *arrondissements* than between *quartiers* within the same *arrondissement*.

One advantage of this flexible model is that it provides marginal effects $\left(\beta_k + u_{i,k}^{L1} + u_{ij,k}^{L2}\right)$ for a specific covariable X_k, that take into account multi-dimensional random effects. Figure 12.2 gives the nonparametric densities of the marginal effect for each of the covariates, both for the outer group $\left(\beta_k + u_{i,k}^{L1}\right)$ and for the inner group $\left(\beta_k + u_{i,k}^{L1} + u_{ij,k}^{L2}\right)$. The density curves of the marginal effects for the outer group tend to be smoother and more centered around the fixed effect value (β_k) than those of the marginal effects of the inner group. These curves high-

	$\mu_{q(\beta)}$	s.e $\left(\Sigma_{q(\beta)}\right)$	t_value	Pr(>|t|)
intercept	8.02490	0.02778	288.91765	0.0
surface	0.00122	0.00048	2.54670	0.01089
bathrooms	0.05808	0.02171	2.67596	0.00746
garage plots	0.04377	0.02358	1.85599	0.06348
maid rooms	0.08867	0.02807	3.15910	0.00159
floor levels	0.01214	0.00616	1.97215	0.04862
balcony	0.15424	0.02996	5.14748	0.0
	statistics			
N	10,983			
σ_ε^2	0.06704			
R^2	0.99901			
$\log \underline{p}\,(y,q)$	-1277.57706			
$\Sigma_{q(\beta_1)}$	0.00077			
$(1/m)\sum_{i=1}^{m}\Sigma_{q(u_{i,1}^{L1})}$	0.00773			
$(1/(m.n))\sum_{i=1}^{m}\sum_{j=1}^{n}\Sigma_{q(u_{ij,1}^{L2})}$	0.00127			

Table 12.1: MFVB approximation of the fixed effects parameters of the three-level linear mixed model.

light the differences in heterogeneity of marginal effects at the *arrondissement* and *quartier* levels. Thus, an increase of one unit in the number of bathrooms in the flat increases the price per square meter by an average of 5.8% but with a relatively high variability between 3.5% and 7.5%. Similarly, the presence or absence of a balcony in the flat increases the price per square meter by an average of 15.4% but with a relatively high variability between 13.5% and 17%. But it is especially the 3D histograms of marginal effects that reveal the heterogeneity of such effects across *arrondissements* and *quartiers* (see Figure 12.3). These figures clearly show the differences even between *quartiers* in the same *arrondissement*. These three-level random effects models and the associated streamlined MFVB approximation can be useful tools to estimate multi-dimensional housing prices models. In order to gain even more flexibility, we can hope that in the future four-level or even five-level random effects models will be developed in order to induce, for instance, a nested structure as *year*, *arrondissement*, *quartier*, *block* and *flats* in order to be able to estimate precise multi-dimensional panels.

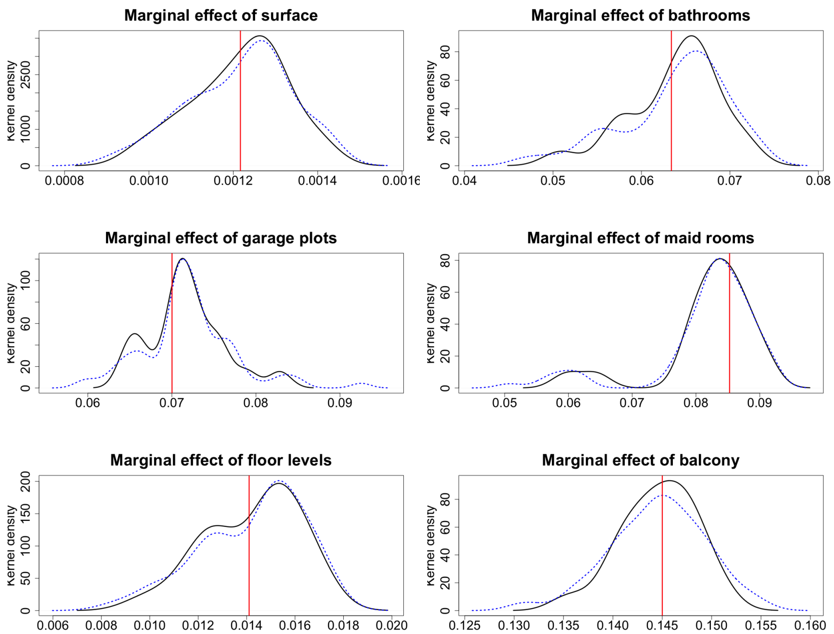

Fig. 12.2: MFVB approximation of densities of marginal effects. Vertical line: fixed effect β. Solid line: random effect - *arrondissement* $\beta + u^{(L1)}$. Dashed line: random effect - *arrondissement* + *quartier* $\beta + u^{(L1)} + u^{(L2)}$.

12.7 Conclusion

The development of modelling housing in multi-dimensional frameworks (3D, 4D or more) is still in its infancy. As compared to the huge literature in a 2D framework, which explains why there are relatively few multi-dimensional housing studies. The limitation comes from the availability of the data and the complexity of the methods relative to time series or longitudinal dimensions. The previous papers show that both spatial and temporal dimensions in dynamic systems should be included for hedonic housing models and discrete models of residential location in a three-dimensional framework. But the inclusion of these multiple dimensions greatly complicates the specification and modeling of such systems. Extending models with unobserved neighborhood characteristics to deal with the endogenous neighborhood characteristics or introducing rationing in housing markets (see Geyer and Sieg (2013)) is not trivial.

Part of the attractiveness of a neighborhood may be driven by the characteristics of neighbors (for instance, higher-income households attract higher-income households while lower-income households repel higher-income households). As Kuminoff, Smith, and Timmins (2013) said *“households “sort" across neighbor-*

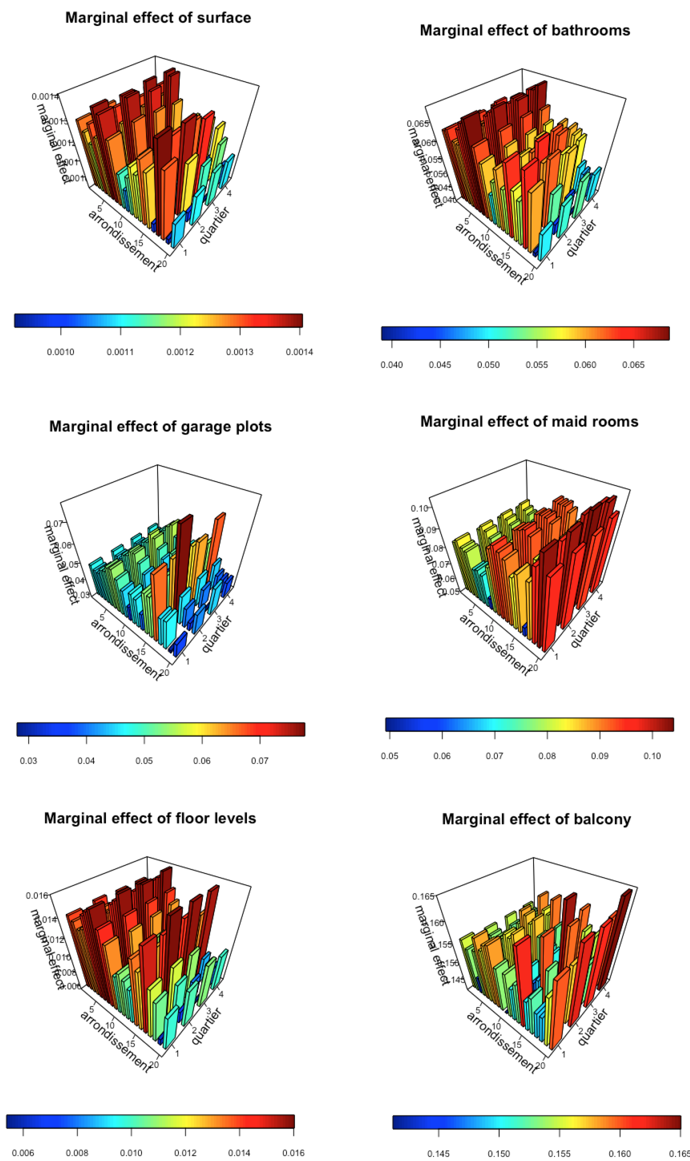

Fig. 12.3: 3D histograms of marginal effects.

hoods according to their wealth and their preferences for public goods, social characteristics, and commuting opportunities ... These "equilibrium sorting" models use the properties of market equilibria, together with information on household behavior, to infer structural parameters that characterize preference heterogeneity. These results can be used to develop theoretically consistent predictions for the welfare implications of future policy changes. Analysis is not confined to marginal effects or a partial equilibrium setting. Nor is it limited to prices and quantities... These capabilities are just beginning to be understood and used in applied research" (p. 1007). Since three decades, econometric methods have made significant progress and considerably grew to eliminate non credible assumptions such as homogenous preferences and exogenous amenities. But now, in a 2D framework, the structural estimators still rely on parametric assumptions for utility functions, on specific statistical distributions (log-normal, Type I extreme value, generalized extreme value, etc.) used to capture sources of unobserved heterogeneity and some strong assumptions to eliminate potential sources of market frictions. As suggested by Kuminoff et al. (2013), one approach could be to refine the current estimators through the lens of the econometric literature on partial identification (see Manski (2007)) which views economic models as sets of assumptions, some of which are plausible and some of which are "esoteric" (according to Tamer (2010)'s expression) and are needed only to complete a model. One of the key advantages of such approach is that it could characterize the potential sensitivity of outcomes to the least credible assumptions. Last, we have shown that variational Bayesian approximations are promising future pathways to potentially overcome many problems in applied modelling of housing, hoping that four-level and five-level random effects models will be available soon.

Appendix: Inverse G-Wishart distributions

Inverse G-Wishart distributions are a generalization of inverse Wishart distributions. They correspond to the matrix inverses of random matrices that have a G-Wishart distribution (see Atay-Kayis and Massam (2005), Nolan et al. (2020) or Maestrini and Wand (2021)). For any positive integer d, let G be an undirected graph with d nodes labeled $1, \cdots, d$ and set E consisting of sets of pairs of nodes that are connected by an edge. We say that the symmetric $(d \times d)$ matrix M respects G if $M_{ij} = 0$ for all $\{i, j\} \notin E$. A $(d \times d)$ random matrix X has an Inverse G-Wishart distribution with graph G and parameters $\xi > 0$ and symmetric $(d \times d)$ matrix Λ, written $X \sim$ Inverse-G-Wishart(G, ξ, λ), if and only if the density function of X satisfies

$$p(X) \propto |X|^{-(\xi+2)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}(\Lambda X^{-1})\right\}$$

over arguments X such that X is symmetric and positive definite and X^{-1} respects G. Two important special cases are $G = G_{full}$($\equiv$ totally connected d-node graph) for which the Inverse G-Wishart distribution coincides with the ordinary inverse Wishart distribution, and $G = G_{diag}$($\equiv$ totally disconnected d-node graph), for which the Inverse G-Wishart distribution coincides with a product of independent inverse Chi-Squared random variables. The subscripts of G_{full} and G_{diag} reflect the fact that X^{-1} is a full matrix and X^{-1} is a diagonal matrix in each special case. The $G = G_{full}$ case corresponds to the ordinary inverse Wishart distribution. If $d = 1$, the graph $G = G_{full} = G_{diag}$ and the inverse G-Wishart distribution reduces to the Inverse Chi-Squared distribution.

References

Abate, G. D., & Anselin, L. (2016). House price fluctuations and the business cycle dynamics. *Aarhus University, Dept. of Economics and Business Economics, Working Paper 2016-6*.

Agarwal, S., & Hauswald, R. (2010). Distance and private information in lending. *The Review of Financial Studies*, *23*, 2757–2788.

Aizcorbe, A., & Aten, B. (2004). An approach to pooled time and space comparisons. *SSHRC Conference on Index Number Theory and the Measurement of Prices and Productivity, Vancouver*.

Alonso, W. (1964). *Location and Land Use*. Harvard University Press.

Anselin, L. (2001). Spatial econometrics. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics* (pp. 310–330). Blackwell.

Anselin, L., LeGallo, J., & Jayet, H. (2008). Spatial panel econometrics. In L. Mátyás & P. Sevestre (Eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice* (pp. 625–660). Kluwer Academic Publishers.

Antweiler, W. (2001). Nested random effects estimation in unbalanced panel data. *Journal of Econometrics*, *101*, 295–313.

Atay-Kayis, A., & Massam, H. (2005). A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika*, *92*(2), 317–335.

Bajari, P., & Benkard, C. L. (2005). Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach. *Journal of Political Economy*, *113*(6), 1239–1276.

Bajari, P., & Kahn, M. (2005). Estimating housing demand with an application to explaining racial segregation in cities. *Journal of Business and Economic Statistics*, *23*, 20–33.

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, *21*, 129–151.

Baltagi, B. H., & Bresson, G. (2011). Maximum likelihood estimation and Lagrange multiplier tests for panel seemingly unrelated regressions with spatial lag and spatial errors: an application to hedonic housing prices in Paris. *Journal of Urban Economics*, *69*, 24–42.

Baltagi, B. H., Bresson, G., & Etienne, J.-M. (2015). Hedonic housing prices in Paris: an unbalanced spatial lag pseudo-panel model with nested random effects. *Journal of Applied Econometrics*, *30*, 509–528.

Baltagi, B. H., Bresson, G., & Etienne, J.-M. (2019). Carbon dioxide emissions and economic activities: A mean field variational Bayes semiparametric panel data model with random coefficients. *Annals of Economics and Statistics*, *134*, 43–77.

Baltagi, B. H., Bresson, G., & Etienne, J.-M. (2020). Growth empirics: a Bayesian semiparametric model with random coefficients for a panel of OECD countries. In M. Pesaran, T. Li, & D. Terrell (Eds.), *Advances in Econometrics: Essays in Honor of Cheng Hsiao* (Vol. 41, pp. 217–253). Emerald Group Publishing Limited.

Baltagi, B. H., Fingleton, B., & Pirotte, A. (2014). Spatial lag models with nested random effects: an IV procedure with application to English house prices. *Journal of Urban Economics*, *80*, 76–86.

Baltagi, B. H., & Pirotte, A. (2014). Prediction in a spatial nested error components panel data model. *International Journal of Forecasting*, *30*, 407–414.

Baltagi, B. H., Song, S., & Jung, B. (2001). The unbalanced nested error component regression model. *Journal of Econometrics*, *101*, 357–381.

Baltagi, B. H., Song, S., & Jung, B. (2002). Simple LM tests for the unbalanced nested error component regression model. *Econometric Reviews*, *21*, 167–187.

Bartik, T. (1987). The Estimation of demand parameters in hedonic price models. *Journal of Political Economy*, *95*, 81–88.

Basu, S., & Thibodeau, T. (1988). Analysis of spatial autocorrelation in house prices. *Journal of Real Estate Finance and Economics*, *17*, 61–85.

Bayer, P., Ferreira, F., & McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, *115*, 588–638.

Bayer, P., McMillan, R., Murphy, A., & Timmins, C. (2016). A dynamic model of demand for houses and neighborhoods. *Econometrica*, *84*, 893–942.

Beenstock, M., & Felsenstein, D. (2015). Estimating spatial spillover in housing construction with nonstationary panel data. *Journal of Housing Economics*, *28*, 1–144.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*, 841–889.

Bhat, C. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B*, *35*, 677–693.

Bhat, C. (2003). Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B*, *37*, 837–855.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Bishop, K. C., & Timmins, C. (2011). Hedonic prices and implicit markets: estimating marginal willingness to pay for differentiated products without instrumental variables. *NBER working paper 17611*.

Bishop, K. C., & Timmins, C. (2018). Using panel data to easily estimate hedonic demand functions. *Journal of the Association of Environmental and Resource*

Economists, *5*(3), 517–543.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877.

Blomquist, G., & Worley, L. (1981). Hedonic prices, demand for urban housing amenities and benefit estimates. *Journal of Urban Economics*, *9*, 212–221.

Blomquist, G., & Worley, L. (1982). Specifying the demand for housing characteristics: the exogeneity issue. In D. Diamond & G. Tolley (Eds.), *The Economics of Urban Amenities* (pp. 89–102). Academic Press.

Blundell, R., & Powell, J. (2004). Endogeneity in semi-parametric binary response models. *Review of Economic Studies*, *71*, 655–679.

Bresson, G., & Hsiao, C. (2011). A functional connectivity approach for modeling cross-sectional dependence with an application to the estimation of hedonic housing prices in Paris. *Advances in Statistical Analysis*, *95*, 501–529.

Brown, J., & Rosen, H. (1982). On the estimation of structural hedonic price models. *Econometrica*, *50*, 765–768.

Carlin, B., Kogan, S., & Lowery, R. (2013). Trading complex assets. *The Journal of Finance*, *58*, 1937–1960.

Case, K., Shiller, R., & Thompson, A. (2012). What have they been thinking? Home buyer behavior in hot and cold markets. *NBER Working Paper No. 18400*.

Das, S., Gupta, R., & Kabundi, A. (2010). Dynamic factor models and vector autoregressive models. *Journal of Forecasting*, *30*, 288–302.

Davies, R. B. (1984). A generalized Beta-logistic model for longitudinal data with an application to residential mobility. *Environment and Planning, A*, *16*, 1375–86.

Davies, R. B., & Pickles, A. (1985). A panel study of life-cycle effects in residential mobility. *Geographical Analysis*, *17*, 199–216.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, *30*, 109–126.

Deng, Y., Gabriel, S. A., & Nothaft, F. E. (2003). Duration of residence in the rental housing market. *The Journal of Real Estate Finance and Economics*, *26*, 267–285.

Diao, M., Ma, X., & Ferreira, J. (2015). A real-option based dynamic model to simulate real estate developer behavior. *14th International Conference on Computers in Urban Planning and Urban Management (CUPUM), Cambridge, MA USA*.

Dieleman, F. (2001). Modelling residential mobility: a review of recent trends in research. *Journal of Housing and the Built Environment*, *16*, 249–265.

Eluru, N., Sener, I., Bhat, C., Pendyala, R., & Axhausen, K. (2009). Understanding residential mobility: a joint model of the reason for residential relocation and stay duration. *Transportation Research Board 88th Annual Meeting, Washington, DC*.

Engle, R. F., Lilien, D. M., & Watson, M. (1985). A DYMIMIC model of housing price determination. *Journal of Econometrics*, *28*, 307–326.

Epple, D. (1987). Hedonic prices and implicit markets: estimating demand and supply functions for differentiated products. *Journal of Political Economy*, *95*, 59–80.

Eriksen, M. D., & Ross, A. (2015). Housing vouchers and the price of rental housing. *American Economic Journal: Economic Policy*, *7*(3), 154–76.

Fingleton, B. (2008). Housing supply, housing demand and affordability. *Urban Studies*, *45*, 1545–1563.

Fingleton, B., & LeGallo, J. (2008). Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties. *Papers in Regional Science*, *87*, 319–339.

Fu, Q., Zhu, Y., & Ren, Q. (2015). The downside of marketization: A multilevel analysis of housing tenure and types in reform-era urban China. *Social Science Research*, *49*, 126–140.

Gayer, T., Hamilton, J., & Viscusi, W. (2000). Private values of risk tradeoffs at superfund sites: housing market evidence on learning about risk. *Review of Economics and Statistics*, *82*, 439–451.

Geyer, J., & Sieg, H. (2013). Estimating an model of excess demand for public housing. *Quantitative Economics*, *4*, 483–513.

Glaeser, E. (2008). *Agglomeration and Spatial Equilibrium*. Oxford University Press.

Griliches, Z. (1971). *Price Indexes and Quality Change, Studies in New Methods of Measurement*. Harvard University Press.

Guevara, C. A., & Ben-Akiva, M. (2006). Endogeneity in residential location choice models. *Transportation Research Record*, *1977*, 60–66.

Guo, J., Feng, T., & Timmermans, H. J. (2020). Co-dependent workplace, residence and commuting mode choice: Results of a multi-dimensional mixed logit model with panel effects. *Cities*, *96*(102448), 1–9.

Hastings, J., Kane, T., & Staiger, D. (2006). Parental preferences and school competition: Evidence from a public school choice program. *NBER Working Papers 11805*.

Heckman, J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, *46*, 931–959.

Henderson, D. J., Soberon, A., & Rodriguez-Poo, J. M. (2022). Nonparametric multidimensional fixed effects panel data models. *Econometric Reviews*, *41*(3), 321–358.

Hoang, H., & Wakely, P. (2000). Status, quality and the other trade-off: towards a new theory of urban residential location. *Urban Studies*, *37*, 7–35.

Holly, S., Pesaran, M. H., & Yamagata, T. (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics*, *158*, 160–173.

Holmes, T., & Sieg, H. (2014). Structural estimation in urban economics. In G. Duranton, J. V. Henderson, & W. C. Stran (Eds.), *Handbook of Regional and Urban Economics, vol.5* (pp. 69–114). Elsevier, North-Holland.

Huang, Y., & Clark, W. A. V. (2002). Housing tenure choice in transitional urban China: a multilevel analysis. *Urban Studies*, *39*, 7–32.

Judd, K. (1998). *Numerical Methods in Economics*. MIT Press.

Jun, H.-J. (2016). The effect of racial and ethnic composition on neighborhood economic change: a multilevel and longitudinal look. *Housing Studies*, *31*, 102-125.

Kelejian, H., & Prucha, I. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, *17*, 99–121.

Kelly, B., & Ljungqvist, A. (2012). Testing asymmetric-information asset pricing models. *The Review of Financial Studies*, *25*, 1366–1413.

Kole, K. (2022). Housing vouchers reduce residential crowding. *Journal of Housing Economics*, *55*, 101822.

Kuminoff, N. V., Smith, V. K., & Timmins, C. (2013). The new economics of equilibrium sorting and policy evaluation using housing markets. *Journal of Economic Literature*, *54*, 1007–1062.

Latif, E. (2015). Immigration and housing rents in Canada: a panel data analysis. *Economic Issues*, *20*, 91–108.

Lee, B. H. Y., & Waddell, P. (2010). Residential mobility and location choice: a nested logit model with sampling of alternatives. *Transportation*, *37*, 587–601.

Lee, C. Y., & Wand, M. P. (2016). Streamlined mean field variational Bayes for longitudinal and multilevel data analysis. *Biometrical Journal*, *4*, 868-895.

Lee, L. (2003). Best spatial two-stage least square estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews*, *22*, 307–335.

LeSage, J., & Pace, R. (2009). *Introduction to Spatial Econometrics*. CRC Press/Taylor & Francis.

Li, Y., & Leatham, D. J. (2011). Forecasting housing prices: dynamic factor model versus LBVAR model. *Agricultural & Applied Economics Association's 2011 AAEA and NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania*.

Liu, X. (2013). Spatial and temporal dependence in house price prediction. *Journal of Real Estate Finance and Economics*, *47*, 341–369.

Maestrini, L., & Wand, M. P. (2021). The inverse G-Wishart distribution and variational message passing. *Australian & New Zealand Journal of Statistics*, *63*(3), 517–541.

Manski, C. (2007). *Identification for Prediction and Decision*. Harvard University Press.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105–142). Academic Press.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, *3*, 303–328.

McFadden, D. (1978). Modelling the choice of residential location. In A. Karlqvist, L. Lundqvist, F. Snickars, & J. Weibull (Eds.), *Spatial Interaction Theory and Planning Models* (pp. 75–96). North Holland:.

Meese, R., & Wallace, N. E. (1997). The construction of residential housing price indices: a comparison of repeat-sales, hedonic-regression and hybrid approaches. *Journal of Real Estate Finance and Economics*, *14*, 51–73.

Nappi-Choulet, I., & Maury, T. (2011). A spatial and temporal autoregressive local estimation for the Paris housing market. *Journal of Regional Science*, *51*, 732–750.

Nelson, J. (1977). Residential choice, hedonic prices, and the demand for urban air quality. *Journal of Urban Economics*, *5*, 357–369.

Newey, W. K., & McFadden, D. (1974). Large sample estimation and hypothesis testing. In R. F. Engle & D. McFadden (Eds.), *Handbook of Econometrics* (Vol. 4, pp. 2111–2245). Elsevier, North-Holland.

Nolan, T. H., Menictas, M., & Wand, M. P. (2020). Streamlined computing for variational inference with higher level random effects. *Journal of Machine Learning Research*, *21*, 1–62.

Nolan, T. H., & Wand, M. P. (2020). Streamlined solutions to multilevel sparse matrix problems. *The ANZIAM Journal*, *62*(1), 18–41.

Ormerod, J., & Wand, M. (2010). Explaining variational approximations. *The American Statistician*, *64*, 140–153.

Pace, P., Barry, R., Gilley, O., & Sirmans, C. (2000). A method for spatial-temporal forecasting with an application to real estate prices. *International Journal of Forecasting*, *16*, 229–246.

Pham, T. H., Ormerod, J. T., & Wand, M. P. (2013). Mean field variational Bayesian inference for nonparametric regression with measurement error. *Computational Statistics & Data Analysis*, *68*, 375–387.

Rosen, S. (1974). Hedonic prices and implicit markets. *Journal of Political Economy*, *82*, 34–55.

Rossi, P. (1955). *Why Families Move: a Study in the Social Psychology of Urban Residential Mobility*. The Free Press, Glencoe, Illinois.

Smith, T., & Wu, P. (2011). A spatio-temporal model of housing prices based on individual sales transactions over time. *Journal of Geographical Systems*, *11*, 333–355.

Syed, I., Hill, R. J., & Melser, D. (2008). Flexible spatial and temporal hedonic price indexes for housing in the presence of missing data. *The University of New South Wales, School of Economics, Discussion Paper: 2008/14*.

Sykes, L. L. (2003). Income rich and asset poor: a multilevel analysis of racial and ethnic differences in housing values among baby boomers. *Population Research and Policy Review*, *22*, 1–20.

Tamer, E. (2010). Partial identification in econometrics. *Annual Review of Economics*, *2*, 167–195.

Thanos, S., Dubé, J., & Legros, D. (2016). Putting time into space: the temporal coherence of spatial applications in the housing market. *Regional Science and Urban Economics*, *58*, 78–88.

Turnbull, G. K., & van der Vlist, A. J. (2015). Uninformed house buyers and foreclosures. *University of Central Florida, Department of Finance, Working Paper 1503*.

Wand, M. P. (2017). Fast approximate inference for arbitrarily large semiparametric regression models via message passing. *Journal of the American Statistical*

Association, *112*(517), 137–168.

Witte, A., Sumka, H., & Erekson, H. (1979). An estimate of a structural hedonic price model of the housing market: an application of Rosen's theory of implicit markets. *Econometrica*, *47*, 1151–1173.

Index