Hedonic Housing Prices in Corsica: A hierarchical spatiotemporal approach

WORKSHOP: THEORY AND PRACTICE OF SPDE MODELS AND INLA

LING Yuheng[1]

30 Oct. 2018

[1]PhD student in Economics - University of Corsica - CNRS UMR LISA 6240, France.

Location, location, location

Corse Matin, May 17, 2012

"Une nouvelle exception corse: Les prix de l'immobilier flambent".

Corse Matin, Auguste 28, 2012

"Aussi, que vaut aujourd'hui un appartement dans la cité impériale ? Tout dépend du quartier."

"On language: location, location, location" in The New York Time, June 28, 2009

When asking a real estate professional about the three most important characteristics of a house, the likely answer will be "location, location, location".

Economist's words

Can, Ayse, "Specification and estimation of hedonic housing price models", Regional Science and Urban Economics, sep 1992, 22 (3), 453-474.

- Neighborhood effects

Potential spatial autocorrelation

Economist's words

Can, Ayse, "Specification and estimation of hedonic housing price models", Regional Science and Urban Economics, sep 1992, 22 (3), 453-474.

- Neighborhood effects
- Adjacent effect

Potential spatial autocorrelation

Data

Housing transaction data (collected over time)

Cross section? Panel? Repeated cross section?
Spatiotemporal geostatistical/point-referenced data

Tools

The tools to analyze geo-referenced house transaction data are very limited. *(Dubé and Legros, 2013)*

- Pooling cross-sectional data
- Using a pooled OLS regression *(Palmquist, 2005)*
- Biased coefficients? *(Clark and Linzer, 2015)*

Literature on Corsican property market

Corsican property market studies

Corsican housing market has not been fully explored in literature.

- Spatial inequality, as well as on land-use pressure *(Furt and Tafani, 2014; Kessler and Tafani, 2015; Prunetti et al., 2015)*
- A recent research *(Giannoni et al., 2017)* focuses on the phenomenon that non-local house buyers drive out local house buyers.

A twofold objective

First

We propose a model which can explicitly capture dependences in space and over time simultaneously.

Second

The proposed model is applied to study the Corsican housing market. We intend to investigate the determinants of Corsican apartment prices; in particular, we would like to highlight the impacts of time and space on apartment prices.

Economic cornerstone: Hedonic price theory (HPM)

A New Approach to Consumer Theory

"The good, per se, does not give utility to the consumer; it possesses characteristics, and these characteristics give rise to utility." *(Lancaster, 1966, p134)*

Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition

"A class of differentiated products is completely described by a vector of objectively measured characteristics. Observed product prices and the specific amounts of characteristics associated with each good define a set of implicit prices." *(Rosen, 1976, p34)*

Empirical definition of HPM

Empirical representation of a house price *(Malpezzi, 2008)*

$$P = f\left(S, N, L, C, T, \beta\right) \quad (1)$$

Dealing with Space

Spatial regression models *(Anselin, 1988)*

$$y = \beta W y + X\beta + u \tag{2}$$

$$y = X\beta + \varepsilon \tag{3}$$

$$\varepsilon = \lambda W \varepsilon + u \tag{4}$$

Dealing with Space

Multilevel modeling/hierarchical models *(Raudenbush and Bryk, 2002)*

$$Level1 : y = \Delta\alpha + X\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2) \quad (5)$$

$$Level2 : \alpha = Z\gamma + u, u \sim N(0, \tau^2) \quad (6)$$

- *Goodman and Thibodeau (1998)*
- *Goodman and Thibodeau (2003)*

Special issues on applying HPM

Space and time

Housing transaction data are collected over time.

Tools

The tools to analyze geo-referenced house transaction data are very limited. *(Dubé and Legros, 2013)*

State-of-the-art models dealing with dependences in space and over time

Spatial econometrics and the hedonic pricing model: what about the temporal dimension?

"...the STAR specification outperforms the SAR specification; the STAR specification, with a small good threshold distance value outperforms the OLS specification;" *(Dubé and Legros, 2014, p355)*

Drawbacks

Specification

State-of-the-art models dealing with dependences in space and over time

Hedonic Housing Prices in Paris: An Unbalanced Spatial Lag Pseudo-Panel Model with Nested Random Effects

Baltagi et al. *(2015)* investigate determinants of house prices in Paris over the period 1990-2003.

- Turning repeated cross-sectional data into pseudo-panel data

State-of-the-art models dealing with dependences in space and over time

Hedonic Housing Prices in Paris: An Unbalanced Spatial Lag Pseudo-Panel Model with Nested Random Effects

Baltagi et al. *(2015)* investigate determinants of house prices in Paris over the period 1990-2003.

- Turning repeated cross-sectional data into pseudo-panel data
- N-way nested error component disturbances models *(Baltagi and Chang, 1994)* with a spatial lag term

State-of-the-art models dealing with dependences in space and over time

Hedonic Housing Prices in Paris: An Unbalanced Spatial Lag Pseudo-Panel Model with Nested Random Effects

Baltagi et al. *(2015)* investigate determinants of house prices in Paris over the period 1990-2003.

- Turning repeated cross-sectional data into pseudo-panel data
- N-way nested error component disturbances models *(Baltagi and Chang, 1994)* with a spatial lag term
- Spatial nested random effect model allowing spatial lag effects λ to vary by year.

State-of-the-art models dealing with dependences in space and over time

Hedonic Housing Prices in Paris: An Unbalanced Spatial Lag Pseudo-Panel Model with Nested Random Effects

$$\begin{aligned} y_{taqif} &= \lambda_t \tilde{y}_{taqif} + X_{taqif}\beta + u_{taqif} \quad ; \\ \tilde{y}_{taqif} &= \sum_{a=1}^{N} \sum_{q=1}^{Q_{ta}} \sum_{i=1}^{M_{taq}} \sum_{p=1}^{Ftaqi} w_{taqip} y_{taqip} \quad ; \\ u_{taqif} &= \delta_{ta} + \mu_{taq} + \nu_{taqi} + \varepsilon_{taqif} \end{aligned} \tag{7}$$

Drawbacks

Temporal dependence

Hierarchical spatio-temporal model

- A two-level hierarchical spatio-temporal model *(Banerjee and al. 2014; Cressie and Wikle, 2011; Cameletti and al., 2013)*.

Hierarchical spatio-temporal model

- A two-level hierarchical spatio-temporal model *(Banerjee and al. 2014; Cressie and Wikle, 2011; Cameletti and al., 2013)*.
- $$y\left(s_i, t\right) = z\left(s_i, t\right)\beta + \xi\left(s_i, t\right) + \varepsilon\left(s_i, t\right) \tag{8}$$

Hierarchical spatio-temporal model

- A two-level hierarchical spatio-temporal model *(Banerjee and al. 2014; Cressie and Wikle, 2011; Cameletti and al., 2013)*.
- $$y\left(s_i, t\right) = z\left(s_i, t\right)\beta + \xi\left(s_i, t\right) + \varepsilon\left(s_i, t\right) \quad (8)$$
- $y(s_i, t)$ is a realization of the underlying spatio-temporal process $Y\left(\cdot, \cdot\right)$ representing house prices measured at apartment unit $i = 1, \cdots, d$ located at site s_i and time $t = 1, \cdots, T$.

Hierarchical spatio-temporal model

- A two-level hierarchical spatio-temporal model *(Banerjee and al. 2014; Cressie and Wikle, 2011; Cameletti and al., 2013)*.
- $$y\left(s_i, t\right) = z\left(s_i, t\right)\beta + \xi\left(s_i, t\right) + \varepsilon\left(s_i, t\right) \qquad (8)$$
- $y(s_i, t)$ is a realization of the underlying spatio-temporal process $Y\left(\cdot, \cdot\right)$ representing house prices measured at apartment unit $i = 1, \cdots, d$ located at site s_i and time $t = 1, \cdots, T$.
- $z\left(s_i, t\right)\beta$ represents all covariates referring to fixed effects

Hierarchical spatio-temporal model

- $\xi\,(s_i, t)$ is a so-called spatiotemporal random effects term.

Hierarchical spatio-temporal model

- $\xi(s_i, t)$ is a so-called spatiotemporal random effects term.
- $$\xi(s_i, t) = a\xi(s_i, t-1) + \omega(s_i, t) \tag{9}$$

Hierarchical spatio-temporal model

- $\xi\left(s_i, t\right)$ is a so-called spatiotemporal random effects term.
- $$\xi\left(s_i, t\right)=a \xi\left(s_i, t-1\right)+\omega\left(s_i, t\right) \tag{9}$$
- a is the first-order autoregressive (AR1) coefficient.

Hierarchical spatio-temporal model

- $\xi\left(s_i, t\right)$ is a so-called spatiotemporal random effects term.
- $$\xi\left(s_i, t\right) = a\xi\left(s_i, t-1\right) + \omega\left(s_i, t\right) \quad (9)$$
- a is the first-order autoregressive (AR1) coefficient.
- $\omega\left(s_i, t\right)$ is a time-independent random field (RF).

Hierarchical spatio-temporal model

- $\xi(s_i, t)$ is a so-called spatiotemporal random effects term.
- $$\xi(s_i, t) = a\xi(s_i, t-1) + \omega(s_i, t) \quad (9)$$
- a is the first-order autoregressive (AR1) coefficient.
- $\omega(s_i, t)$ is a time-independent random field (RF).
- $$\omega(s_i, t) \sim N\left(0, \sum = \sigma_\omega^2 \sum\right) \quad (10)$$

Hierarchical spatio-temporal model

- $\xi\left(s_i, t\right)$ is a so-called spatiotemporal random effects term.
- $$\xi\left(s_i, t\right) = a\xi\left(s_i, t-1\right) + \omega\left(s_i, t\right) \tag{9}$$
- a is the first-order autoregressive (AR1) coefficient.
- $\omega\left(s_i, t\right)$ is a time-independent random field (RF).
- $$\omega\left(s_i, t\right) \sim N\left(0, \sum = \sigma_\omega^2 \sum\right) \tag{10}$$
- $$cov\left(\omega\left(s_i, t\right), \omega\left(s_j, t'\right)\right) = \begin{cases} 0 \; if \;\; t \neq t' \\ C_\theta\left(h\right) \;\; if \;\; t = t' \end{cases} \tag{11}$$

where $h = \|s_i - s_j\|$ is the Euclidean distance.

Hierarchical spatio-temporal model

- $\xi\left(s_i, t\right)$ is a so-called spatiotemporal random effects term.
- $$\xi\left(s_i, t\right) = a\xi\left(s_i, t-1\right) + \omega\left(s_i, t\right) \tag{9}$$
- a is the first-order autoregressive (AR1) coefficient.
- $\omega\left(s_i, t\right)$ is a time-independent random field (RF).
- $$\omega\left(s_i, t\right) \sim N\left(0, \sum = \sigma_\omega^2 \sum\right) \tag{10}$$
- $$cov\left(\omega\left(s_i, t\right), \omega\left(s_j, t'\right)\right) = \begin{cases} 0 \; if \;\; t \neq t' \\ C_\theta\left(h\right) \;\; if \;\; t = t' \end{cases} \tag{11}$$

  where $h = \|s_i - s_j\|$ is the Euclidean distance.
- Gaussian white noise

  $$\varepsilon\left(s_i, t\right) \sim N\left(0, \sigma_\varepsilon^2 I_d\right) \tag{12}$$

As a grouped model

ξ represents grouped random effects

- a within group correlation structure $\omega(s_i, t)$

As a grouped model

ξ represents grouped random effects

- a within group correlation structure $\omega(s_i, t)$
- a between group correlation structure measured by a

As a grouped model

ξ represents grouped random effects

- a within group correlation structure $\omega\left(s_i, t\right)$
- a between group correlation structure measured by a
- If $\xi_{s_i,t}$ is the ith element in the domain S in time period t, we have

$$Cov\left(\xi_{s_1,t}, \xi_{s_2,t'}\right) = \sum_{ar1} \otimes \sum_{\omega} \tag{13}$$

Fitting the model

- Matérn correlation function

Fitting the model

- Matérn correlation function
- Gaussian Markov random field (GMRF)

Fitting the model

- Matérn correlation function
- Gaussian Markov random field (GMRF)
- SPDE approach

Fitting the model

- Matérn correlation function
- Gaussian Markov random field (GMRF)
- SPDE approach
- INLA algorithm

Corsica


Genoa
Savona
Bologna
Ravenna
Parc national du Mercantour
La Spezia
Cesena
Nice
Pisa
Florence
Livorno
Cannes
Arezzo
Siena
Toulon
Bastia
Grosseto
Parc Naturel Régional de Corse
Corse
Ajaccio
Porto-Vecchio
Fiumicino
Stintino
Olbia
Sassari
Alghero
Nuoro
Sardegna
Oristano
Tyrrhenian Sea
Map data ©2018 Google, Inst. Geogr. Nacional
lat
lon
44
43
42
41
40
6
8
10
12

Data

"PERVAL" database from "Notaries de France"

- The "PERVAL" database records all type of property transactions in France.

Data

"PERVAL" database from "Notaries de France"

- The "PERVAL" database records all type of property transactions in France.
- High-quality and high-reliability

Data

"PERVAL" database from "Notaries de France"

- The "PERVAL" database records all type of property transactions in France.
- High-quality and high-reliability
- Transaction ID

Data

"PERVAL" database from "Notaries de France"

- The "PERVAL" database records all type of property transactions in France.
- High-quality and high-reliability
- Transaction ID
- Transaction date

Data

"PERVAL" database from "Notaries de France"

- The "PERVAL" database records all type of property transactions in France.
- High-quality and high-reliability
- Transaction ID
- Transaction date
- Transaction price

Data

"PERVAL" database from "Notaries de France"

- The "PERVAL" database records all type of property transactions in France.
- High-quality and high-reliability
- Transaction ID
- Transaction date
- Transaction price
- Characteristics of the property

Data

Our dataset

- The used dataset are extracted from the "PERVAL" database.

Data

Our dataset

- The used dataset are extracted from the "PERVAL" database.
- We are interested in apartment prices.

Data

Our dataset

- The used dataset are extracted from the "PERVAL" database.
- We are interested in apartment prices.
- 7634 observations.

Data

Our dataset

- The used dataset are extracted from the "PERVAL" database.
- We are interested in apartment prices.
- 7634 observations.
- Transactions from 2006 to 2017

Independent variables

Variable	Description/Unit
ROOM	Number of rooms
BATH	Number of bathrooms
GAR	Number of garages
FLOOR	Number of floors
SURF	Living area (square meters)
TYPE	Dummy (=1 if the apartment pertains to this type and 0 otherwise)
SA	Standard apartment (referenced)
DU	Duplex apartment
ST	Studio apartment
CONSTRUCTION PERIOD	Dummy (=1 if the apartment was built during this period and 0 otherwise)
PERIOD A	Time of building 1850-1913 (referenced)
PERIOD B	Time of building 1914-1947
PERIOD C	Time of building 1948-1969

Variable	Description/Unit
PERIOD E	Time of building 1981 / 1991
PERIOD F	Time of building 1992 / 2000
PERIOD G	Time of building 2001 / 2010
PERIOD H	Time of building 2011 / 2020
DBEAD	Distance to the nearest beach (kilometers)
DPuHigSch	Distance to the nearest public high school (kilometers)
DHealFac	Distance to the nearest health facility (kilometers)
DPuHigSch	Distance to the nearest public primary school (kilometers)

Descriptive statistics

Table: Descriptive statistics for hedonic housing prices in Corsica

	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Transaction Price	149467.08	58483.01	57445.76	100000	185347.95	325431.67
log(Transaction Price)	11.84	0.39	10.96	11.55	12.13	12.69
ROOM	2.672	0.967	0	2	3	8
BATHROOM	1.053	0.259	0	1	1	3
PAK	0.795	0.712	0	0	1	8
FLOOR	1.849	1.731	-3*	1	3	12
SURF	59.315	22.191	6	43	73	197
DBEAD	3.782	7.153	0.001	1.040	3.561	52.008
DHealFac	10.421	12.099	0.051	1.636	16.461	72.244
DPuPriSch	1.347	1.698	0.0001	0.469	1.544	39.513
DPuHigSch	9.914	10.689	0.001	1.434	15.809	78.978
SVI	11.653	11.237	0.000	1.503	19.906	47.923

Models

- *Classical linear regression model (M0)*

$$ln\ y\left(s_i,t\right)=z\left(s_i,t\right)\beta+\varepsilon\left(s_i,t\right);\ \varepsilon\left(s_i,t\right)\sim N\left(0,\sigma_\varepsilon^2\right) \quad (14)$$

Models

- *Classical linear regression model (M0)*

$$ln\ y(s_i,t) = z(s_i,t)\beta + \varepsilon(s_i,t)\,;\ \varepsilon(s_i,t) \sim N\left(0,\sigma_\varepsilon^2\right) \quad (14)$$

- *Classical linear regression with space fixed effects (M1)*

$$\begin{aligned} ln\ y(s_i,t) = z(s_i,t)\beta + 112\ municipality\ dummies \\ +\varepsilon(s_i,t) \quad ; \\ \varepsilon(s_i,t) \sim N\left(0,\sigma_\varepsilon^2\right) \end{aligned} \quad (15)$$

Models

- *Classical linear regression model (M0)*

$$ln\ y\left(s_i,t\right)=z\left(s_i,t\right)\beta+\varepsilon\left(s_i,t\right);\ \varepsilon\left(s_i,t\right)\sim N\left(0,\sigma_\varepsilon^2\right) \quad (14)$$

- *Classical linear regression with space fixed effects (M1)*

$$\begin{aligned} ln\ y\left(s_i,t\right)=z\left(s_i,t\right)\beta+112\ municipality\ dummies \\ +\varepsilon\left(s_i,t\right) \quad ; \\ \varepsilon\left(s_i,t\right)\sim N\left(0,\sigma_\varepsilon^2\right) \end{aligned} \quad (15)$$

- *Classical linear regression with space and time fixed effects (M2)*

$$\begin{aligned} ln\ y\left(s_i,t\right)=z\left(s_i,t\right)\beta+112\ municipality\ dummies \\ +48\ quarter\ dummies+\varepsilon\left(s_i,t\right) \quad ; \\ \varepsilon\left(s_i,t\right)\sim N\left(0,\sigma_\varepsilon^2\right) \end{aligned} \quad (16)$$

Fixed effects Models

- Advantages
  - Economic perspective

Fixed effects Models

- Advantages
  - Economic perspective
  - Spatial analysis
- Disadvantages
  - Spatial autocorrelation

Mixed effects Models

- *Hierarchical spatial model (M3)*

$$\begin{aligned} ln\ y(s_i) &= z(s_i)\beta + \xi(s_i) + \varepsilon(s_i) \quad ; \\ \xi(s_i) &= \omega(s_i) \quad ; \\ \varepsilon(s_i) &\sim N\left(0, \sigma_\varepsilon^2\right) \quad ; \\ \omega(s_i) &\sim N\left(0, \sum = \sigma_\omega^2 \sum\right) \end{aligned} \tag{17}$$

Mixed effects Models

- *Hierarchical spatial model (M3)*

$$
\begin{aligned}
ln\ y(s_i) = z(s_i)\beta + \xi(s_i) + \varepsilon(s_i) \quad &; \\
\xi(s_i) = \omega(s_i) \quad &; \\
\varepsilon(s_i) \sim N\left(0, \sigma_\varepsilon^2\right) \quad &; \\
\omega(s_i) \sim N\left(0, \sum = \sigma_\omega^2 \sum\right) \quad & \qquad (17)
\end{aligned}
$$

- *Hierarchical spatiotemporal model: AR1 (M4)*

$$
\begin{aligned}
ln\ y(s_i, t) = z(s_i, t)\beta + \xi(s_i, t) + \varepsilon(s_i, t) \quad &; \\
\xi(s_i, t) = a\xi(s_i, t-1) + \omega(s_i, t) \quad &; \\
\varepsilon(s_i, t) \sim N\left(0, \sigma_\varepsilon^2\right) \quad &; \\
\omega(s_i, t) \sim N\left(0, \sum = \sigma_\omega^2 \sum\right) \quad & \qquad (18)
\end{aligned}
$$

Implementing details

- R-INLA

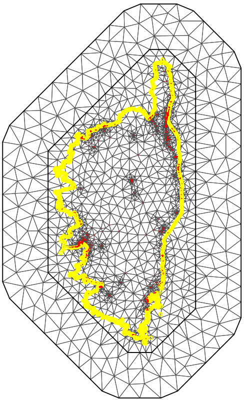
Implementing details

- R-INLA
- Vague prior to hyperparameters

Implementing details

- R-INLA
- Vague prior to hyperparameters
- Mesh (3237 triangles)

Constrained refined Delaunay triangulation



Model selection

Table: Results of DIC

Model	DIC values	Elapsed Time
CLRM	2009.37	6
CLRM+Space fixed effects	-1123.87	6
CLRM+Space and time fixed effects	-1204.59	7
Spatial hierarchical model	-3867.65	43
Spatiotemporal hierarchical model	-4460.54	17287

- M4 is deemed the best model.

Posterior estimates of covariate coefficients

	Model 4		
	mean	0.025 quant	0.975 quant
Intercept	**10.981**	10.886	11.075
ROOM	**0.033**	0.024	0.042
BATHROOM	0.017	-0.001	0.035
GAR	**0.050**	0.041	0.059
FLOOR	**0.019**	0.016	0.022
SURF	**0.010**	0.010	0.011
DU	**0.028**	0.002	0.054
ST	**-0.190**	-0.209	-0.170
PERIOD B	0.000	-0.065	0.065
PERIOD C	-0.006	-0.068	0.056
PERIOD D	0.031	-0.032	0.094
PERIOD E	0.047	-0.016	0.110
PERIOD F	**0.107**	0.038	0.175
PERIOD G	**0.219**	0.154	0.284
PERIOD H	**0.234**	0.169	0.299
DBEAD	**-0.016**	-0.021	-0.011
DHealFac	**-0.005**	-0.008	-0.003
DPuHigSch	0.002	-0.001	0.005
DPuPriSch	0.007	-0.001	0.015

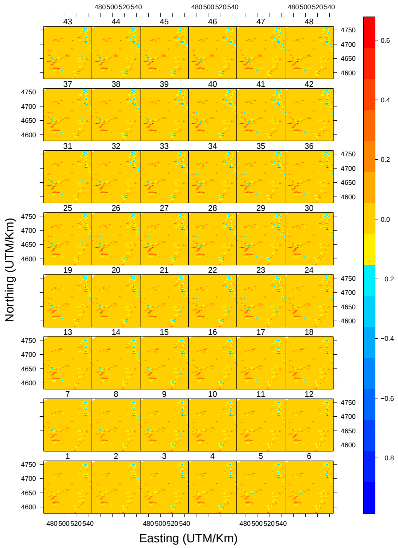
Posterior estimates of the variance parameters

Table: Posterior mean estimates of the variance parameters

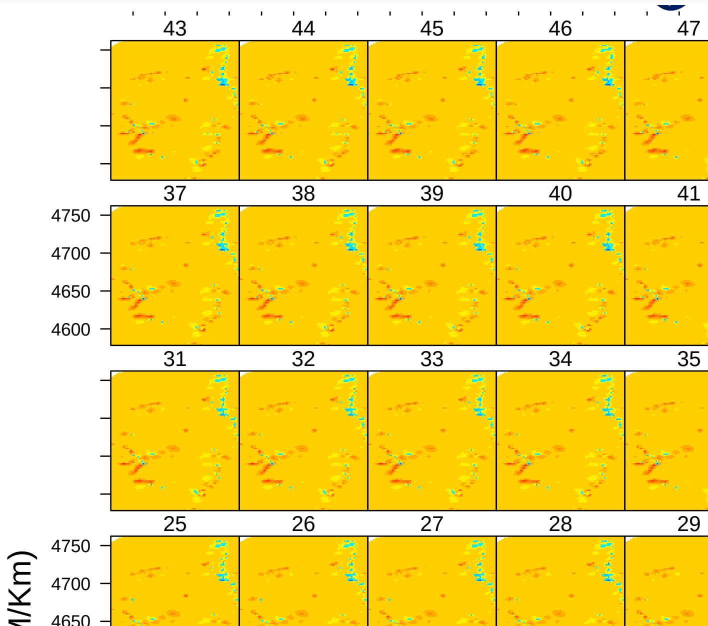
	σ_e^2	σ_w^2	AR1 coef	Range Km
Model0	**0.076**			
Model1	**0.050**			
Model2	**0.049**			
Model3	**0.032**	**0.108**		**1.582**
		(0.090,0.129)		(1.369,1.831)
Model4	**0.028**	**0.106**	**0.990**	**1.503**
		(0.090,0.123)	(0.987,0.993)	(1.289,1.711)

- Main findings

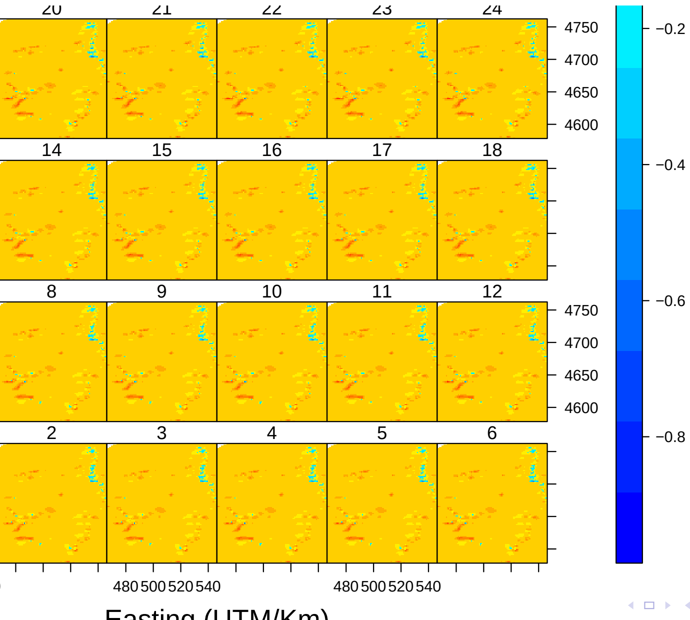
Spatiotemporal random effects visualization



Spatiotemporal random effects visualization



Spatiotemporal random effects visualization



Spatiotemporal random effects

$$ln\ y\left(s_i, t\right) = z\left(s_i, t\right)\beta + \xi\left(s_i, t\right) + \varepsilon\left(s_i, t\right) \quad (19)$$

$$y\left(s_i, t\right) = exp^{z(s_i,t)\beta} \times exp^{\xi(s_i,t)} \times exp^{\varepsilon(s_i,t)} \quad (20)$$

Findings

Locations increase the expected apartment prices up to 82.21%, as well as decrease the expected apartment prices to 55.06%.

Findings

- Several housing structural attributes and accessibility attributes affect apartment prices.

Findings

- Several housing structural attributes and accessibility attributes affect apartment prices.
- It is clear that space and time significantly affect Corsican apartment prices. In particular, locations highly affect apartment prices.

Findings

- Several housing structural attributes and accessibility attributes affect apartment prices.
- It is clear that space and time significantly affect Corsican apartment prices. In particular, locations highly affect apartment prices.
- We can not neglect dependence in space and over time. Hence, fixed effects models are not alternatives to mixed effects models.

Conclusion

- Rather than ad hoc models, hierarchical spatiotemporal models and the INLA-SPDE approach work as a general framework dealing with housing transaction data.

Conclusion

- Rather than ad hoc models, hierarchical spatiotemporal models and the INLA-SPDE approach work as a general framework dealing with housing transaction data.
- It is necessary to incorporate time and space in models when we handle housing transaction data.

Conclusion

- Rather than ad hoc models, hierarchical spatiotemporal models and the INLA-SPDE approach work as a general framework dealing with housing transaction data.
- It is necessary to incorporate time and space in models when we handle housing transaction data.
- The way to gauge time and space effects is also important. Categorical variables in fixed models do not take spatial effects fully into account.

Future studies

- Priors?

Thanks for your attention.