AJAE online appendix for

The Informational Content of Geographical Indications

Jean-Sauveur Ay*

March 23, 2020

The material contained herein is supplementary to the article named in the title and published in the *American Journal of Agricultural Economics*

Contents

*Umr Cesaer: Agrosup, Inrae, Université Bourgogne Franche-Comté, 21000 Dijon, France.

Details about causal inference

The causal lobbying effect is identified once the spatial heterogeneity is controlled by smoothing geographical coordinates. The sufficient level is expected to be reached once the residuals from auxiliary regressions that do not include *commune* effects are not correlated between *communes*. Using residuals for specification purposes has a long history in econometrics, complemented by generalized residuals for non-linear outcomes (Pagan and Hall, 1983; Gourieroux et al., 1987; Chesher and Irish, 1987). We use the surrogate residuals recently defined by Liu and Zhang (2018).

Define a surrogate variable $S \mid X, y \sim \lambda\left[B(X)^\top \beta - \alpha_y \mid y \right]$ that follows a truncated logistic distribution conditionally on y. The principle of using the observed values of y to estimate the residuals is shared by generalized residuals, the originality of the surrogate approach is to randomly draw the residuals rather than computing them analytically. This allows the estimation of their full distribution instead of only their first two moments (Liu and Zhang, 2018). We estimated the residuals of auxiliary models from N random draws of the surrogate variable S_i with:

$$R_i = S_i - \mathbb{E}(S_i) = S_i + \alpha_{y_i} - B(X_i)^\top \beta, \tag{1}$$

and we regress them on *commune* dummies. By increasing the complexity of $B(X_i)$ through increasing spline base dimensions of the smooth functions of geographical coordinates, the joint significance of *commune* dummies decreases as the unobserved spatial patterns are increasingly accounted for. Failing to reject the null hypothesis of a Fisher test of joint significance is expected to indicate that the sufficient complexity is attained by the auxiliary model. Hence, we estimate a full OGAM of GI designation with *commune* fixed effects and obtained level of spatial smoothing. In the absence of residual effects correlated between *commune*, the estimated ordinal superiority measures are expected to be causal. The F-statistics are bootstrapped to take into account the additional uncertainty attributable to the random draws used in the computation of surrogate residuals.

Table A1: Descriptive Statistics

Table 1: **Descriptive Statistics for the Main Variables used in the Econometric Analysis**

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Acreage [1000 m^2]	59113	0.002	0.003	0.000	0.001	0.002	0.177
Elevation [1000 m]	59113	0.286	0.056	0.210	0.241	0.319	0.505
Slope [degree]	59113	5.772	5.478	0.000	1.556	8.747	36.970
Solar radiation [millions J]	59113	1.060	0.049	0.581	1.048	1.076	1.230
Longitude [degree]	59113	4.837	0.104	4.665	4.740	4.955	5.003
Latitude [degree]	59113	47.060	0.110	46.900	46.980	47.170	47.300
Actual GI [*Coteaux*]	59113	0.164	0.370	0	0	0	1
Actual GI [*Régional*]	59113	0.229	0.420	0	0	0	1
Actual GI [*Village*]	59113	0.428	0.495	0	0	1	1
Actual GI [*Premier Cru*]	59113	0.147	0.354	0	0	0	1
Actual GI [*Grand Cru*]	59113	0.032	0.177	0	0	0	1
1936 GI [*Régional*]	59113	0.565	0.496	0	0	1	1
1936 GI [*Village*]	59113	0.407	0.491	0	0	1	1
1936 GI [*Grand Cru*]	59113	0.027	0.163	0	0	0	1
Aspect [0 – 45]	59113	0.046	0.210	0	0	0	1
Aspect [45 – 90]	59113	0.186	0.389	0	0	0	1
Aspect [90 – 135]	59113	0.362	0.481	0	0	1	1
Aspect [135 – 180]	59113	0.212	0.409	0	0	0	1
Aspect [180 – 225]	59113	0.100	0.300	0	0	0	1
Aspect [225 – 270]	59113	0.044	0.206	0	0	0	1
Aspect [270 – 315]	59113	0.030	0.170	0	0	0	1
Aspect [315 – 360]	59113	0.021	0.142	0	0	0	1

Notes: Topographic variables were computed by a Geographical Information System from a Digital Elevation Model with 5 m resolution. Longitude and latitude variables correspond to the center of each vineyard plot. Current and 1936 GIs are ordered dummy variables. Exposition variable is discretized in the 8 Aspect variables according to the degree quadrants reported in brakets.

Table A2: Significance tests for 1936 GIs

Table 2: **Joint Variable Significance Tests for 1936 GIs**

Variable	(0)	(I)	(II)	(III)	(IV)	(V)
Elevation	982.42**	1 196.2**	197.72**	144.79**	265.02**	253.01**
	[2]	[8.826]	[7.628]	[8.232]	[8.659]	[7.42]
Slope	409.2**	478.13**	466.46**	297.06**	190.45**	169.07**
	[2]	[8.754]	[8.729]	[8.743]	[8.774]	[7.493]
Solar Radiation	859.1**	208.81**	139.42**	99.245**	87.676**	142.83**
	[2]	[8.04]	[1.082]	[8.114]	[7.419]	[7.425]
Spatial Coords	5 814.5**	6 760**	14 559**	17 285**	18 979**	20 906**
	[15]	[48.73]	[97.95]	[147.1]	[194.3]	[235.3]
Pedology	4 099.2**	2 820.6**	898.79**	599.37**	537.03**	539.28**
	[13]	[12]	[12]	[12]	[12]	[12]
Geology	982.42**	1 047**	692.13**	710.2**	585.81**	509.32**
	[14]	[14]	[14]	[14]	[14]	[14]
Exposition	287.18**	177.45**	131.87**	58.532**	43.002**	64.03**
	[7]	[7]	[7]	[7]	[7]	[7]
Commune	8 600.1**	3 720.9**	2 639.2**	2 177.2**	1 831.7**	1 264.7**
	[25]	[25]	[25]	[25]	[25]	[25]
Nb Observ.	50 000	50 000	50 000	50 000	50 000	50 000
McFadden R^2	44.63	49.68	61.32	66.06	69.82	72.36
Pc good pred.	81.86	83.74	87.88	89.84	91.35	92.21
Akaike IC	45	41.21	31.82	28.09	25.12	23.12
Surrogate F	92.72	8.45	5.4	3.43	2.75	2.03

Notes: $^{**}p < 0.001$ for significance tests associated to the chi-square statistics, effective degrees of freedom are inside brackets. Column (0) corresponds to an ordered logit model with quadratic effects for elevation, slope and solar radiation (df= 2) with a full interaction between third-orders polynomials for longitude and latitude (df= $3 + 3 + 3 \times 3 = 15$) and with 13, 14, 7 and 25 dummy variables for pedology, geology, exposition, and *communes* fixed effects, respectively. Models (I) to (V) are OGAMs with elevation, slope and solar radiation additively specified with a maximum of 9 edf, shrinked endogenously by a quadratic penalization. Spatial coordinates are specified in increasing order of complexity with the maximum edf of 50, 100, 150, 200, and 250. The last row reports the bootstraped F-statistics for the joint nullity of *commune* effects on residuals from auxiliary regressions without *commune* dummies.

Figure A1: GIs within *communes*

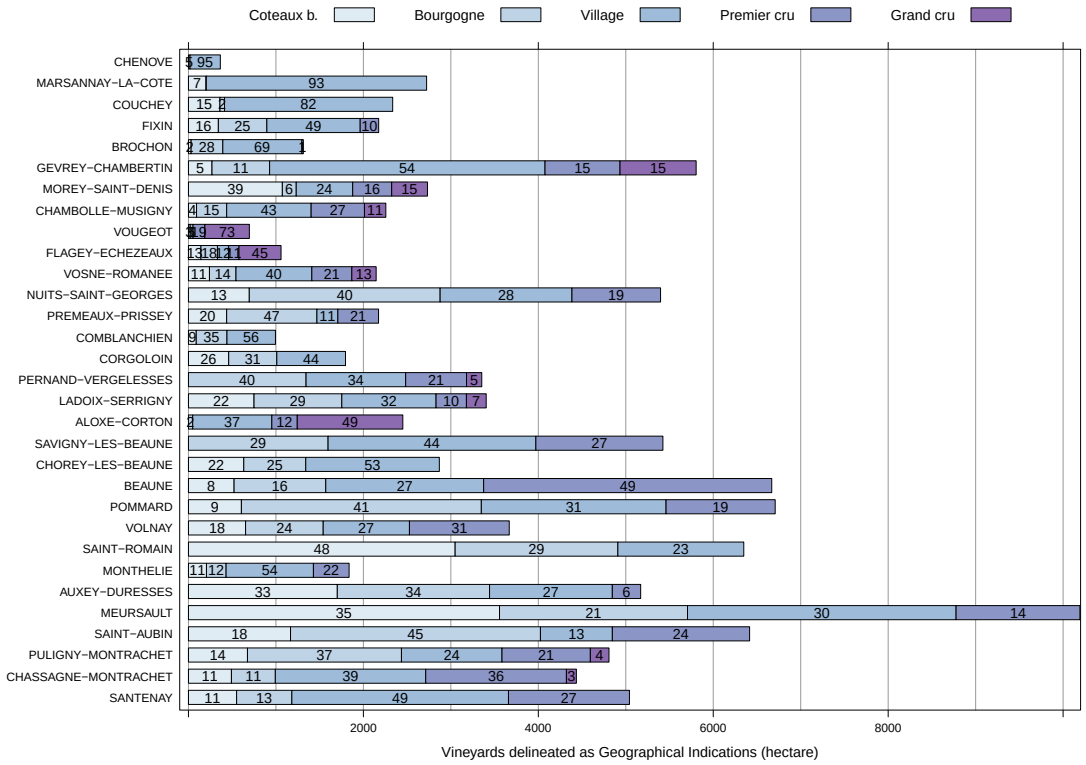

Figure 1: **The Distribution of GI Levels within each *Commune***

Notes: For each administrative *commune* on the y-axis, the bar represents the vineyard area (in hectare, on the x-axis) designated in each vertical level of GIs. The numbers reported within the bar is the percentage that each GI item represents in the total vineyard area of each *commune*.

Figure A2: Effects of topographic variables

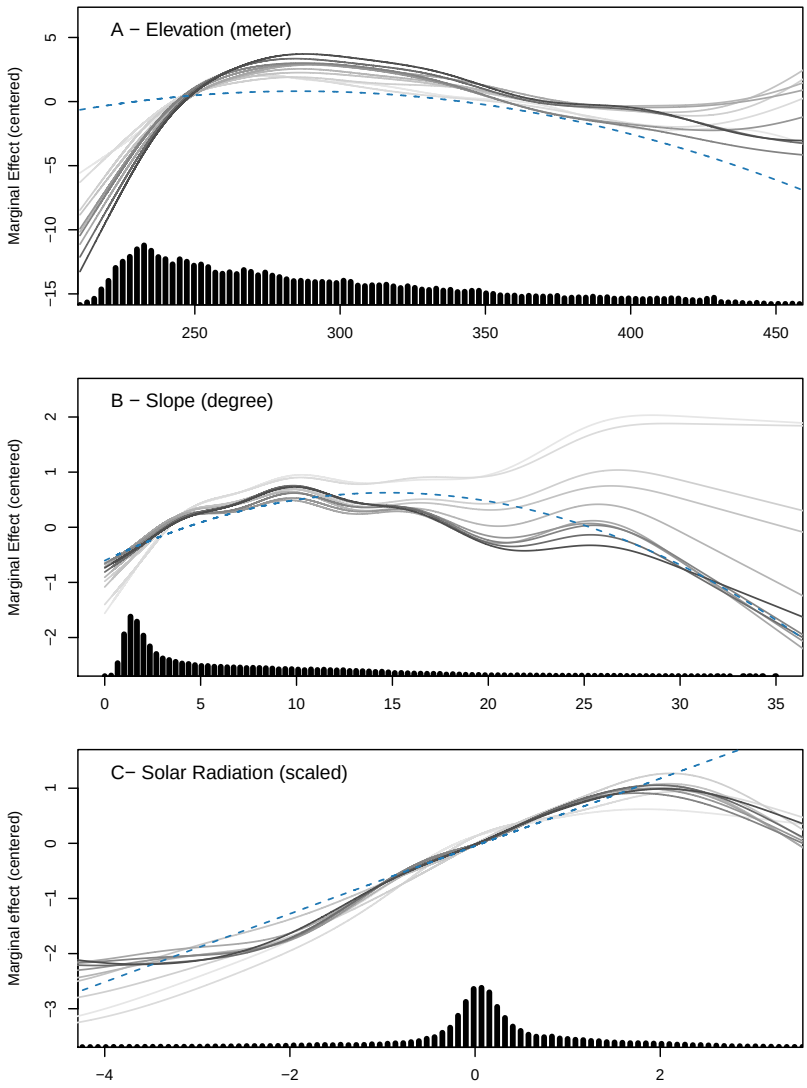

Figure 2: **Nonlinear Effects of Topographic Variables on Current GI Designations**

Notes: Dotted lines represent the quadratic effects from model (0) in Table 1 of the main text, centered at zero with other explanatory variables at their sample means. Continuous lines represent the centered effects from 10 OGAMs with increasing darkening for increasing effective degrees of freedom for spatial smoothing. Models (I) to (V) in the main Table 1 are a subset of these OGAMs with maximum effective degrees of freedom distributed between 100 and 1000. The histograms at the bottom of each plot represent the marginal distributions of each explanatory variable.

Figure A3: Effects of geographic coordinates

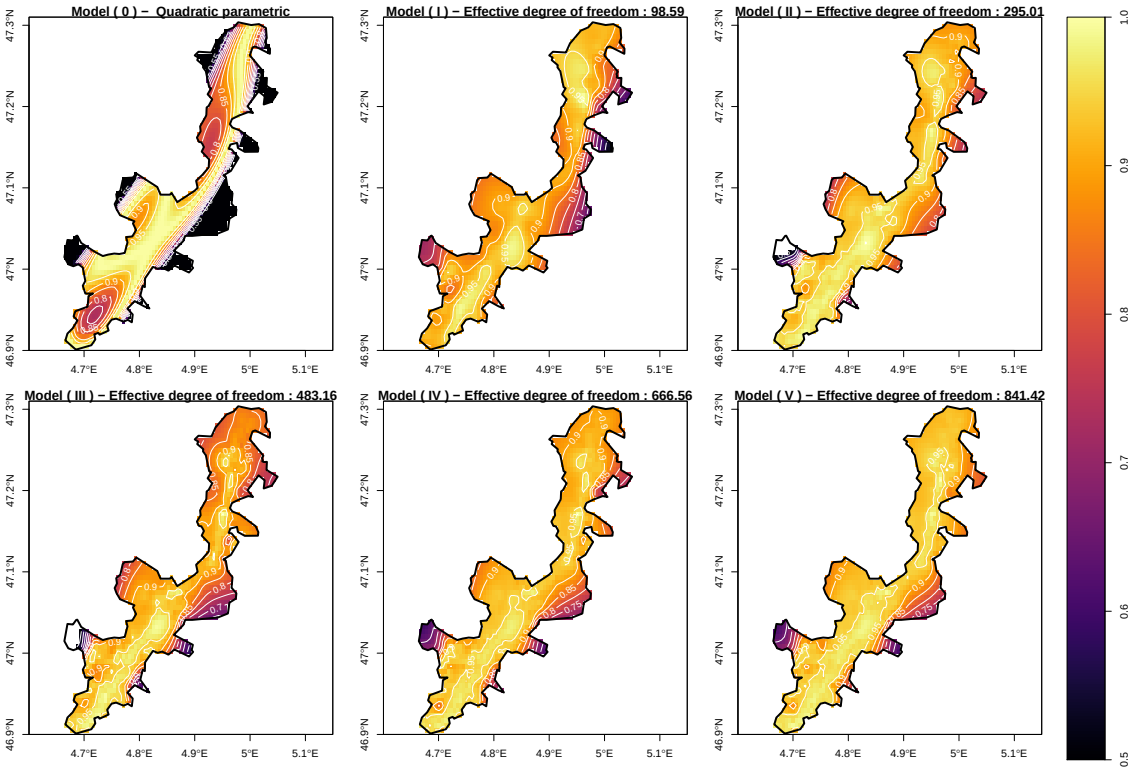

Figure 3: **Smoothed Functions of Geographic Coordinates from Current GI Designations**

Notes: The smooth surfaces are predicted only from geographic coordinates, other explanatory variables are fixed at their sample means. Predictions of the latent vineyard quality index are normalized to be inside the unit interval. Each Panel corresponds to a model reported in Table 1 of the main text, from model (0) at the top-left to model (V) at the bottom-right. The effective degrees of freedom for the smooth functions are reported at the top of each plot.

Figure A4: Causal evidences from residuals

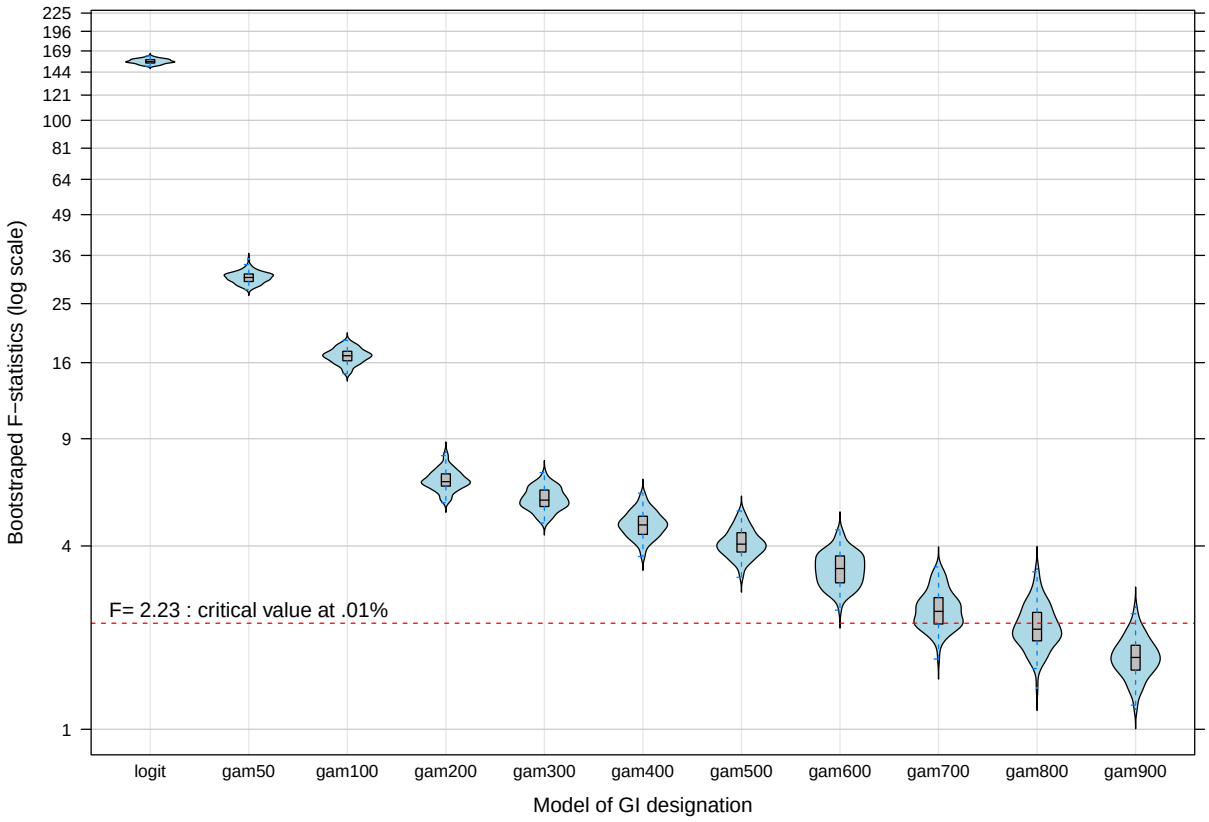

Figure 4: **Joint Significance of *commune* dummies on surrogate residuals**

Notes: For each model on the x-axis (with increasing level of spatial smoothing), the Figure reports the distribution of the bootstrapped F-statistics about the joint significance of *commune* dummies on surrogate residuals (log scale).

Figure A5: Correlation of ordinal superiority

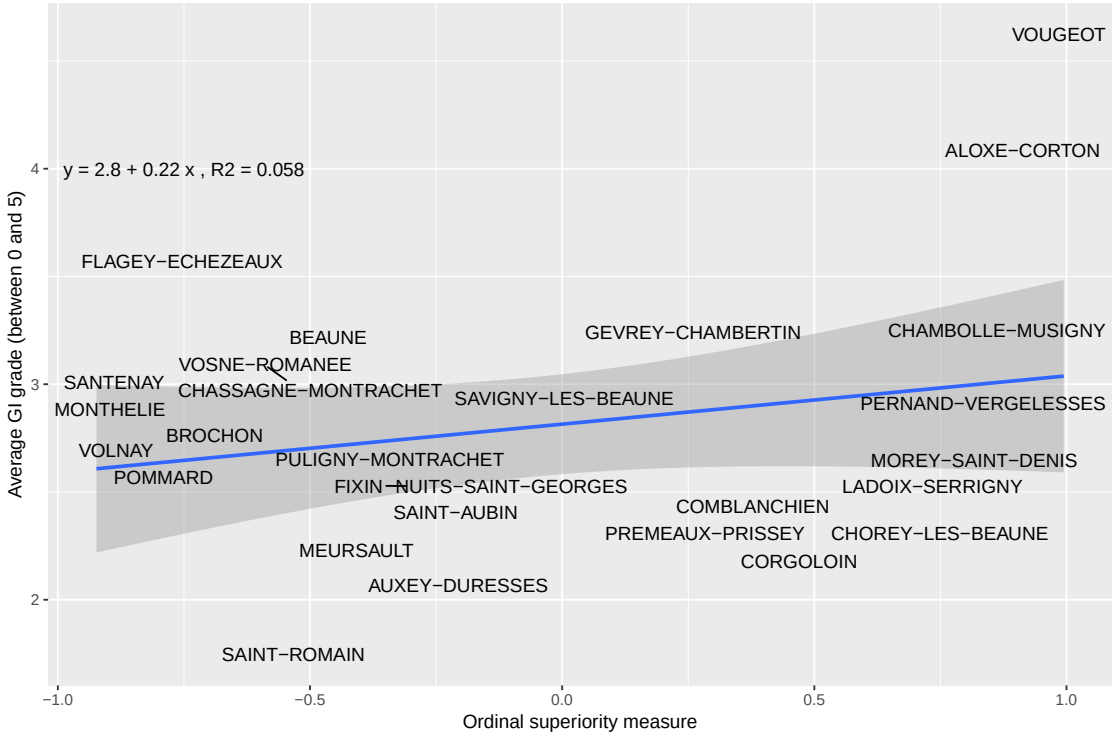

Figure 5: **Relation between Average GI Level and Ordinal Superiority Measures**

Notes: The ordinal superiority measures on the x-axis are those of Figure 2 in the main text. The average GI level for each *commune* is the area-weighted average of GIs vertical levels, coded from 1 to 5 from the worst *Coteaux bourguignons* to the best *Grands Crus*. The nullity of the slope cannot be rejected at 1% ($t = 1.27$).

Figure A6: Topographic variables for 1936 GIs

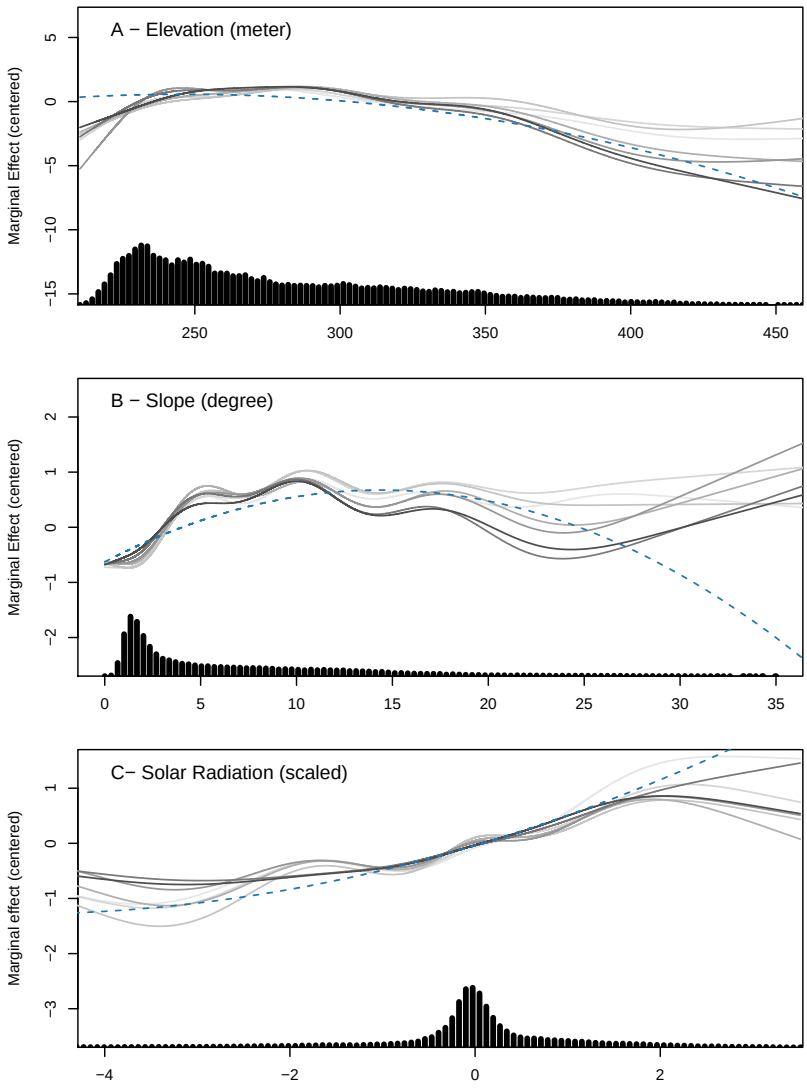

Figure 6: **Nonlinear Effects of Topographic Variables on 1936 GI Designations**

Notes: Dotted lines represent the quadratic effects from model (0) in Table 2 of this appendix, centered at zero with all other explanatory variables at their sample means. Continuous lines represent the centered effects from 10 OGAMs with increasing darkening for increasing effective degrees of freedom for spatial smoothing terms. Models (I) to (V) in Table 2 are a subset of these OGAMs with maximum effective degrees of freedom uniformly distributed between 50 and 350. The histograms at the bottom of each plot represent the marginal distributions of the explanatory variables.

Figure A7: Ordinal superiority for 1936 GIs

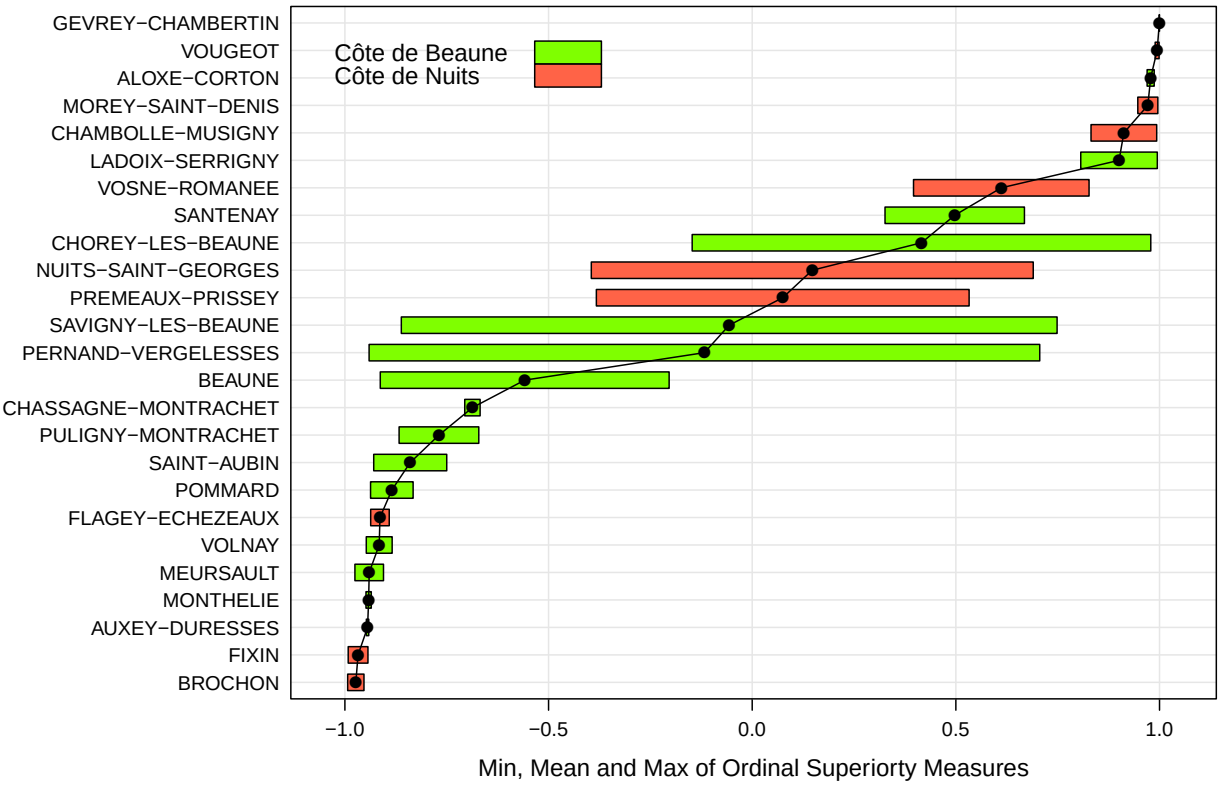

Figure 7: **Ordinal Superiority Measures for 1936 GI designation scheme**

Notes: For a given *commune* on the y-axis, ordinal superiority measures are computed as the difference between the estimated fixed effect μ_c and the average fixed effect $\overline{\mu}$ of all *commune* according to: $\Delta_c = 2 \times \Lambda[(\mu_c - \overline{\mu})/\sqrt{2}] - 1$. The horizontal bars represent the range of measures according to the OGAMs with 150, 200 and 250 maximum edf for the effects of spatial coordinates. Black dots represent the average of these measures. Relatively privilegied *communes* appear at the top of the y-axis, whereas relatively disadvantged *communes* appear at the bottom.

References

Chesher, A. and Irish, M. (1987). Residual analysis in the grouped and censored normal linear model. *Journal of Econometrics* 34: 33–61.

Gourieroux, C., Monfort, A., Renault, E. and Trognon, A. (1987). Generalised residuals. *Journal of Econometrics* 34: 5–32.

Liu, D. and Zhang, H. (2018). Residuals and diagnostics for ordinal regression models: A surrogate approach. *Journal of the American Statistical Association* : 1–10.

Pagan, A. R. and Hall, A. D. (1983). Diagnostic tests as residual analysis. *Econometric Reviews* 2: 159–218.