The Effects of Geographical Indications and Natural Conditions on Wineyard Sale Prices

Jean-Sauveur Ay[a] and Julie Le Gallo[b]

September 8, 2013

Abstract

It is common knowledge that the taste of a wine depends on Natural Conditions (NCs) where grapes have grown, and that Geographical Indications (GIs), by grouping together similar NCs, also provide some information about the taste of a wine. Both GIs and NCs potentially add value to wineyards. However, disentangling their relative contributions is complex because of their spatial nesting. In this paper, we propose an estimation method that allows evaluating these contributions. This method takes into account some potential unobserved NCs and the resulting endogeneity of GIs in the hedonic equation. Using original data about wineyard sales from Burgundy (France), we find, in accordance with previous studies, that GIs are a more important source of wineyard price variation than NCs. However, taking into account the possibility of spatially omitted biophysical variables implies at least more than a doubling of the explained part from NCs (from 8% to 17%, where the GIs' parts fall from 51% to 37%). Taking into account the endogenity of GIs also sharply decreases their economic importance. From a naive per-hectare premium of €1.32 million for the most famous GIs (*Grand Cru*), the estimate of our prefered model is about €0.35 million, still highly significant nervertheless.

Keywords: Geographical indications ; wineyard prices ; endogeneity ; ordered models
J.E.L. Codes: C25, C26, C51, R33

(a): Corresponding author. CNRS, UMR 7204 CERSP F-75005 PARIS (principal affiliation) and INRA, UMR 1041 CESAER F-21000 DIJON (associated researcher). jsay.site@gmail.com.
(b): CRESE, Université de Franche Comté F-25000 BESANCON. jlegallo@univ-fcomte.fr.

Preliminary results, please do not quote. *The authors acknowledge the data providers, which allowed this work to be done. Wineyard prices are obtained from SCAFR – Terre d'Europe with the help of Jean Cavailhès and Mohamed Hilal (INRA Dijon). Spatial delineations of Geographical Indications come from the Institut National de l'Origine et de la Qualité (INAO) and we are grateful to Catherine Burrier (INAO) and Cécile Détang-Dessandre (INRA) to have obtained the collaboration. Data about soil quality are obtained with the friendly help of Jean-Marc Brayer and Pierre Curmi (AgroSup Dijon). Climate variables are obtained through the Observatoire du Développement Rural, we thank Eric Cahuzac (INRA) and its colleagues for working on these variables and making them available.*

1 Introduction

It is generally acknowledged that the taste of a wine depends on Natural Conditions (NCs) where grapes have grown and that consumers rarely taste a wine before purchasing. As a mean to differentiate wines produced in similar NCs and to provide this simplified, objective information to potential consumers, the relevance of Geographical Indications (GIs) is obvious. What is less obvious is the relative contributions of these two nested characteristics – NCs and GIs – in the final value of wines. This decomposition is nevertheless central because, in addition to the value added from information availability (Akerlof 1970 ; Nelson 1970 ; Menapace and Moschini 2012), GIs are also potential sources of undeserved rents for producers (Mussa and Rosen 1978 ; Besanko et al., 1987 ; Mérel and Sexton 2012). The economic outcomes of GIs are a combination of vertuous informational content and surplus extraction by limiting supply and artificially segmenting wine markets, making their recognition conflicting in trade negociations (Josling, 2006).

A large number of empirical papers is concerned with the determinants of wine value. Based on data about wine prices, it appears that producer's reputation (Combris et al. 1997 ; Ali and Nauges 2007), technology (Gergaud and Ginsburgh, 2008) and expert's opinion (Ali et al. 2008 ; Dubois and Nauges 2010) are important explanatory variables. However, uninformed consumers also use bottle price as a signal of quality, making the causal interpretation more delicate (Nerlove 1995 ; Costanigro et al. 2007 ; Schnabel and Storchmann 2010). Wine quality is also signaled through GIs that are found to provide positive premiums (Ashenfelter et al. 1995 ; Combris et al. 2000 ; Carew and Florkowski 2010) increasing with wine prices (Costanigro et al., 2010). Even if they are more scarce, some papers study the effects of NCs such as year to year climate variations (Lecocq and Visser 2006 ; Ashenfelter 2008) or land characteristics and exposure (Gergaud and Ginsburgh, 2008). Analysing the effects of NCs on bottle prices is complicated by the necessity to match precisely the wines to the NCs where grapes grown. Consequently, the two closest papers to this one (Ashenfelter and Storchmann 2010 ; Cross et al., 2011) prefer using wineyard sale prices to identify the value of what we call NCs. The results of these two last papers are contrasted, using respectively data from the Mosel Valley (Germany) and the Willamette Valley (OR, United States). The first finds a strong effect of NCs through solar radiation index and the second does not find any significant effect, with or without controlling by GIs. Both find a positive effect of GIs on wineyard sale prices, up to \$7,000 per-acre for Cross et al. (2011) (Table 2, p. 155).

Starting with a reduced equation from the hedonic theory applied to farmland (Palmquist, 1989), the present methodology semiparametrically estimates the effects of NCs through B-splines (Anglin and Gencay 1996 ; Bao and Wan 2004 ; Li and Racine 2007). It integrates potentially measurement errors and omitted biophysical variables that describes NCs, a recurrent weakness of previous studies (Oczkowski, 2001). However, allowing for this possibility requires to model GIs as functions of NCs in a first step. Although the exogeneity of biophysical variables is indisputable, GIs are spatially designated to picture similar (observed and unobserved) NCs, making them endogenous. In this context, our empirical strategy is twofold. (i) NCs are inherently a continous spatial signal. Therefore, we argue that semiparametric Spatial Trend Surfaces (STSs) from geographical coordinates may act as proxies of the unobservable biophysical variables (Kammann and Wand 2003 ; Fik et al. 2003 ; McMillen 2010). (ii) Spatially discontinous GIs are designated on the basis of historical considerations sometimes orthogonal to the NCs of wineyards (favoritism, personal influences or false believes about biological mecanisms, according to Stanziani 2004 and Norman 2010). Therefore, the endogeneity of GIs is controlled by using a more than two century old administrative subdivision (the *communes*) as multinomial instrumental variable. We evaluate their appropriatness for the identification both through spatial error autocorrelation (Anselin 1988, used as a test for misspecification by McMillen, 2003 and Le Gallo and Fingleton 2012) and through some adaptations of the IV post-regression tests to our particular case.

We apply our proposed methodology on a quasi-exhaustive database about wineyard sales 1992–2008 in two wine regions within the French administrative region of Burgundy. These two regions of wine production (*Côte de Beaune*, CDB, and *Côte de Nuits*, CDN) probably group the most expensive wineyards of the world, as they contain 37 wines in the 50 most expensive.[1] Within these regions, GIs actually consist in a common hierarchy of wineyard plots (and resulting wines) according to a nationally wide scheme, the *Appellations d'Origine Controlée (AOCs)*. First legally established in 1855 for Burgundy under another scheme, these delineations initially come from land classification schemes by the monks from *Cîteaux* around the Xth century (Stanziani 2004 ; Norman 2010). Actual GIs divide symetrically the two wine regions in *Grands Crus* (GCRUs), *Premiers Crus* (PCRUs), *Villages* (VILLs), *Bourgognes Régionaux* (BOURs) and *Grands Ordinaires* (BGORs), listed from the most famous to the less.[2] These wineyards from Burgundy constitute a well-shaped application for at least three reasons:

1. GIs are delimited very finely (at the plot scale) and wineyard sale prices can be perfectly matched with this information. Moreover, the presence of GIs' names on the labels of wine bottles is highly regulated through mandatory information and font sizes as examples.

2. CDB and CDN produce overwhelmingly *terroir* wines, implying a high homogenity in terms of both grape variety, technology and wine making process (Norman, 2010). Two grape varieties represent more than 95% of the acreages (*chardonnay* and *pinot noir*).

3. Wine production and wineyard classification have a long history. This long-run temporal predetermination of GIs provides some current variations of GIs orthogonal to NCs. What was probably the result from lobbying two century ago is today arbitrary, hence exogenous.

The first point allows us to model GIs without errors-in-variables and to have high variations of GIs at a fine scale, i.e., for locally similar NCs. The second point reinforces the use of the hedonic theory applied to farmland that substitutes the producers by the landowners as the agent from which the value is infered: Burgundy wine production presents less unobserved heterogeneity than other potential French wine region. The third point allows us to identify the structural (causal) effect of GIs even if some NCs, correlated with GIs, are omitted from the regression functions. *Contributions*.

The outline of the paper is as follows. The section 2 presents the empirical issues relative to our study: omitted variables, endogeneity of GIs' designations. The data are presented in section 3. The results are reported in section 4. Finally, section 5 concludes.

[1] http://www.wine-searcher.com/most-expensive-wines (last accessed: September 8, 2013)

[2] These five GIs are a grouping of the approximatively 800 official *AOCs* that the Burgundy counts, just for its wines. This classification in five items (distinctively and systematically reported on the labels of wine bottles) provides an information about an objective quality, which is not the case for the numerous within *AOCs*. See http://www.vins-bourgogne.fr/gallery_files/site/12881/13118/18581.pdf (last accessed: September 8, 2013).

2 Empirical Issues

2.1 Separating GIs from NCs within an hedonic model

Because land is a non reproducible fixed asset, the Ricardian principle is that its price capitalizes the comparative advantages provided to the final user. Therefore land price is a convenient and well-used metric to recover the values associated to a large variety of spatialized attributes that matter economically (soil quality, legislative constrains, climate, ecosystem services, agricultural subsidies, see Ay and Latruffe 2013). Integrated within the hedonic framework, this moves the focus of valuation from the output market to the landowners though the assumption of competitive land markets. In our case, this helpfully allows to neglect the complex effects of producer's wine portfolio, technology, reputation and ability, and the effects of consumer's information, preference and habits to concentrate the modeling effort on the willingness-to-pay of potential landowners with the highest bids.

By buying a wineyard plot i, a buyer acquires both the effects of NCs on grape quality and the right to put the associated GIs on the wine that comes from this plot. Buyers and sellers are assumed to have a perfect information about both GIs and NCs, so the values of these attributes are totally capitalized in the observed price per hectare p_i. The econometrician only observes a part of the biophysical variables representing the NCs (B_i), and the vector of GIs noted D_i. The other biophysical variables W_i are not observable but matter for observed land prices.

$$p_i = \alpha + \boldsymbol{\beta} B_i^\top + \boldsymbol{\gamma} W_i^\top + \boldsymbol{\delta} D_i^\top + \varepsilon_i \tag{1}$$

The row vectors $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are the marginal prices and ε_i the residual with the usual properties. This is the classical hedonic framework, which, without loss of generality, we assume to be linear in paramaters for mathematical convenience. This equation admits non linear effects of variables, the general hedonic case according to Ekeland et al. (2004). In order to simplify notations, we consider the endogenous and the exogenous variables as nets of other price determinants that are neither GIs nor NCs: price inflation, buyer/seller characteristics or urban influence as examples. According to the Frisch-Waugh-Lovell (FWL) theorem, it is always possible to orthogonalize the variables of a regression model and analyse their effects separatly (Davidson and MacKinnon, 2004).

Coupled with the zero conditional mean assumption $E(\varepsilon_i \mid B_i, D_i, W_i) = 0$, equation (1) describes a structural relationship. Therefore, the Total Sum of Squares (SST) of wineyard prices (their variance) can be decomposed in the sum of the Explained Sum of Squares (SSE) and the Residual Sum of Squares (SSR) according to the classical formula: $SST = SSE + SSR$. The assumption of structural residuals involves a null covariance between the explanatory variables and the residuals, ensuring a uniquely-defined additive decomposition between the explained and the unexplained part of price variations. The respective contributions of GIs and NCs are nested in SSE and are clearly indistinguishable at this point. Opening the black box of the explained part of price variations leads to an additive decomposition based on price's conditional expectations evaluated at the sample averages of the different variable sets.

$$SSE = \sum\nolimits_i \left(\widehat{p}_i^B - \overline{p}\right)^2 + \sum\nolimits_i \left(\widehat{p}_i^D - \overline{p}\right)^2 + \sum\nolimits_i \left(\widehat{p}_i^W - \overline{p}\right)^2 + 2 \cdot \widehat{\Omega} \tag{2}$$

$$\begin{aligned} \text{with} \quad \widehat{p}_i^B &\equiv E(p_i \mid B_i, \overline{D}, \overline{W}) = \widehat{\alpha} + \widehat{\boldsymbol{\beta}} B_i^\top + \widehat{\boldsymbol{\delta}} \overline{D}^\top + \widehat{\boldsymbol{\gamma}} \overline{W}^\top && (3) \\ \widehat{p}_i^D &\equiv E(p_i \mid \overline{B}, D_i, \overline{W}) = \widehat{\alpha} + \widehat{\boldsymbol{\beta}} \overline{B}^\top + \widehat{\boldsymbol{\delta}} D_i^\top + \widehat{\boldsymbol{\gamma}} \overline{W}^\top && (4) \\ \widehat{p}_i^W &\equiv E(p_i \mid \overline{B}, \overline{D}, W_i) = \widehat{\alpha} + \widehat{\boldsymbol{\beta}} \overline{B}^\top + \widehat{\boldsymbol{\delta}} \overline{D}^\top + \widehat{\boldsymbol{\gamma}} W_i^\top && (5) \end{aligned}$$

$$\text{and} \quad \widehat{\Omega} = \widehat{\beta}\widehat{\delta} \cdot \text{cov}(B_i, D_i) + \widehat{\beta}\widehat{\gamma} \cdot \text{cov}(B_i, W_i) + \widehat{\delta}\widehat{\gamma} \cdot \text{cov}(D_i, W_i) \tag{6}$$

The decomposition (2) is non unique but has the interest – relatively to alternative formulations, see Appendix XX – to treat symetrically both sets of variables. The interaction effect Ω is not strictly allocable to a particular set of variables. It represents the loss of precision in the price decomposition due to correlations between wineyard attributes. It is interesting to note a parallel of this interaction effect with the problem of covariate imbalance in the treatment effect litterature (Rosenbaum and Rubin 1984 ; Imbens 2000). The respective contributions of each set of variables are more precisely estimated when covariates are perfectly balanced, in which case $\widehat{\Omega} = 0$.

If we are only interested in the joint effect of biophysical variables $B_i \cup W_i$ (noted SSB) to describe the effects of NCs, we can bind the SSE from B and W as well as the interaction between these two subsets of variables.

$$SSB \equiv \sum\nolimits_i \left(\widehat{p}_i^B - \overline{p}\right)^2 + \sum\nolimits_i \left(\widehat{p}_i^W - \overline{p}\right)^2 + \sum\nolimits_i \left(\widehat{p}_i^B - \overline{p}\right)\left(\widehat{p}_i^W - \overline{p}\right) \tag{7}$$

Only looking for SSB allows us to neglect the correlations between biophysical variables, letting our methodology free of assumptions about them. Nevertheless, the correlations with D_i stay of interest, as the direct part from the sum of squares of GIs which that are the same as in (2): $SSD = \sum_i \left(\widehat{p}_i^D - \overline{p}\right)^2$. So, the decomposition of interest is about the uniquely defined $SSE = SST - SSR$ and consists in indentifying the sum of SSB, SSD and two weighted covariances – between (B_i, D_i) and (W_i, D_i) – that we note respectively Ω_{BD} and Ω_{WD}.

2.2 Omitted variable bias and proxy solution

By considering W_i as unobservable, this framework is sufficiently general to include the possibility that B_i contains error-in-variables,[3] another usual problem in analysing the effects of biophysical variables on land prices. Therefore, the following "naive" estimation of the structural equation (1) would imply biased parameters (Wooldridge, 2002).

$$p_i = \alpha^o + \boldsymbol{\beta}^o B_i^\top + \boldsymbol{\delta}^o D_i^\top + \varepsilon_i^o \tag{8}$$

In the (likely) case of positive correlations between W_i and both B_i and D_i, the coefficients $\boldsymbol{\beta}^o$ and $\boldsymbol{\delta}^o$ are upward biased as the estimated residuals. The effects of W_i are allocated between the three last terms of the Right Hand Side of (8). In the resulting decomposition $SST = SSB^o + SSD^o + SSR^o + 2(\Omega_{BD}^o + \Omega_{WB}^o)$, SSB^o is downward biased relatively to SSB and the two others sums of square are upward biased. Note that, contrary to SSB^o and SSR^o, the bias in SSD^o only comes from the bias in the parameters $\boldsymbol{\beta}^o$ and $\boldsymbol{\delta}^o$.

[3] Because the correlations between B_i and W_i are not constrained, one can consider the difference between measured and true values of B_i as an additional column of W_i.

One solution to recover the structural decomposition consists in including a supplementary term (a proxy) in the regression function. Because of the strong spatial patterns displayed by observed and unobserved biophysical variables (both from litterature and evidences from our data, see section 3), we use a Spatial Trend Surfaces (STSs) of geographical coordinates as this additional term. This approach – also called GeoAdditive modeling – allows to control some unobserved local effects in the regressions (Kammann and Wand 2003 ; Fik et al. 2003 ; McMillen 2010). In our case, the STS ℓ substitutes with errors the effect of W_i.

$$\boldsymbol{\gamma} W_i = \ell(x_i, y_i) + \eta_i \tag{9}$$

The variables x_i and y_i represent respectively the longitude and the latitude of the plot i. The term η_i is a residual component, uncorrelated with the STS by construction. Because of the zero conditional mean assumption on the structural model, we also know that these proxy residuals are uncorrelated with the structural main residuals ε_i. The resulting empirical model from the proxy solution induces a specific set of estimators indexed r.

$$p_i = \alpha^r + \boldsymbol{\beta}^r B_i^\top + \boldsymbol{\delta}^r D_i^\top + \ell^r(x_i, y_i) + \varepsilon_i^r \tag{10}$$

It is generally acknowledged that the proxy variable solution provides a reliable estimation of the coefficients, but under strong assumptions. Wooldridge (2002, Chapter 4, pp. 63–64) provides the sufficient conditions to trust a proxy variable solution, that we adapt to our case.

$$\text{cov}(x_i, \varepsilon_i) \; = \text{cov}(y_i, \varepsilon_i) \; = 0 \tag{11}$$
$$\text{cov}(x_i, B_i) = \text{cov}(y_i, B_i) = 0 \tag{12}$$
$$\text{cov}(x_i, D_i) = \text{cov}(y_i, D_i) = 0 \tag{13}$$

The first condition (11) implies that the proxy variables do not have their own effect on the outcome (i.e., independently from W_i). This is rarely controversial in classical approaches (according to Wooldridge, 2002) but can be problematic here as it is possible that location impacts wineyard prices even if W_i is accounted for. Geographical coordinates have probably some proper effects: neighborhood effects, external economies, local land scarcity as exemples. For the price decompositions, the main implication of this condition is to know if the effects of $\ell^r(x_i, y_i)$ have to be included in the SSB^r or in the SSR^r. If (11) is verified, the effects of the proxies clearly have to be included in SSB^r. If not, the effects have to be included according to their correlations with B_i (observable) and W_i (unobservable). Under general settings, including the proxy effects in SSB^r (i.e., considering (11) as verified) provides a upper bound and including only the correlation with B_i a lower bound of the true SSB. We compute both.

The second condition (12) is necessary for the good allocation of effects between the STS and the observed biophysical variables. It is not fundamental here because both effects could be grouped in NCs if the condition (11) is verified: this condition is auxiliary to the first. However, to obtain reliable decomposition, we substitute a less usual condition which comes from the fact that the effects of W_i are directly of interest here, contrary to the classical proxy framework which is mainly concerned on the coefficients of the true variables included in the model.

$$\mathrm{cov}(\varepsilon_i^r, W_i) = 0 \tag{14}$$

This condition can be called "completeness" as it implies that not any effects of NCs are in the residuals from the proxy specification. Even if less conventional, this assumption could be tested easily in our case by knowing the amount of spatial autocorrelation in ε_i^r. As such, standard spatial autocorrelation test on residuals from the proxy model (Moran's **I** for example, Anselin, 1988) allows us to evaluate the presence of omitted NCs in the proxy equation.

The last condition (13) is more complicated, as it drives the decomposition between our two main effects of interest: GIs and NCs. Without it, GIs are endogenous in the proxy regression and contrary to the failure of (11) we cannot estimate an interval of credible values for SSB and SSD without it. Relaxing this condition implies a non-trivial shift of the proxy solution that we adress in the following subsection. Just keep in mind that the proxy solution needs this condition to be reliable. Finally, a unbiased proxy estimation – conditions (11), (12), (13) verified – assures that $\varepsilon_i^r = \varepsilon_i + \eta_i$. Consequently, SSR^r overestimate SSR and, even if the parameters are without bias, a reliable estimation of SSB through SSB^r can not be obtained without (14).

2.3 Endogeneity of GIs' designations

2.3.1 History behind GIs and the two stage models

It is useful to divide GIs' designation choices in two additive parts: one that is based on NCs and another that is not. The first part is a function of both observed and unobserved NCs, observability still being defined from the econometrician point of view. For the practionners and with experience, NCs are more precisely known and this knowledge is used in GIs' designations. The second part contains the deviations from this underlying NCs-based first part, in particular due to the previously mentioned historical conditions. We assume that this second part is composed of a deterministic term $\boldsymbol{\psi} Z_i^\top$ and a random one ξ_i. Because the GIs of interest have an ordinal structure (providing a quality classification of the resulting wines as in Ashenfelter and Storchmann 2010), we model the designation choices through a latent variable.

$$d_i^* = \boldsymbol{\theta} B_i^\top + \boldsymbol{\lambda} W_i^\top + \boldsymbol{\psi} Z_i^\top + \xi_i \tag{15}$$

This variable d_i^* is the sum of the NCs-based part (the two first terms of the RHS), the "historical" part (the third term) and a residual (the last term). The first part does not represent truly what is usually called the potential of a wineyard in terms of wine quality, an appreciation that has to be done on the wine market (by the consumers in particular). Even based on "objective" NCs, this part represents the subjective quality potential in the minds of peoples that delineated the GIs, through the coefficients $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$. The matrix Z_i typically contains historical conditions that have influenced GIs' designations: favoritism, personal influences or false believes about biological mecanisms. Because these latter variables are not observables, we use a more than two century old administrative subdivision that is known to be the local scale of political influence in French history. In Burgundy in particular, *communes* existed even before the 1789 French revolution and were the scale at which designation choices were made (Norman, 2010).

Intuitively, these deviations from NCs in GIs' designations provide a mean to separate the effects of GIs on wineyard prices from unobserved NCs. With (15) and $\boldsymbol{\lambda} \neq 0$, the unobserved biophysical variables are clearly present in D_i making (13) implausible. In effect, we have $D_i = \left[d_{1i} \cdots d_{Ji}\right]$ with J the number of GIs and with each component that comes from an indicator function.

$$d_{ij} = \mathbb{1}\left[\mu_{j-1} \leqslant d_i^* \leqslant \mu_j\right] \tag{16}$$

This structural form of GIs' designation choices, (15) and (16), leads to an ordered qualitative model. The vector $\boldsymbol{\mu} = \left[\mu_0 \cdots \mu_J\right]$ groups the thresholds of the latent variable to be estimated. By convention, we put $\mu_0 = -\infty$ and $\mu_J = +\infty$. Coupling these two last equations with (1), we obtain a triangular system with an ordered structure of the J binary endogenous explanatory variables which first appears in Vella (1993). From a structural perspective, this does not imply important changes in the estimation process as the structural mean independence assumption stay verified (Lahiri and Schmidt 1978). However, when coupled with the fact that W_i is unobserved, the proxy solution is no longer usable due to the implausibility of (13). We can nevertheless write this first step of GIs designations with the proxy solution.

$$d_i^* = \theta^r B_i^\top + \psi^r Z_i^\top + \kappa^r(x_i, y_i) + \xi_i^r \tag{17}$$

This equation contain another STS, $\kappa^r(x_i, y_i) \equiv \boldsymbol{\lambda} W_i - \eta_i$, which implies that $\xi_i^r = \eta_i + \xi_i$. This means that we can also evaluate the completeness condition (14) at this stage, through the spatial autocorrelation of ξ_i^r. But in all cases, the errors ε_i^r and ξ_i^r both contains η_i, so are correlated. We propose two solutions face to this problem for the reliability of both estimated coefficients and decompositions.

2.3.2 Using Control Functions (CF)

Because ξ_i^r is contained in D_i, our element of interest is now the correlation between the observed GIs and the unobserved biophysical variables that are not taken into account by the STS κ^r, this means $\operatorname{cov}(D_i, \eta_i)$. The main implication of this endogeneity problem is that price variations from GIs are overestimated in the proxy model. The Control Function (CF) approach assumes that the errors ε_i^r and ξ_i^r are distributed according to a bivariate normal, using the result that the expectation of a marginal distribution conditionally to the another can be written analytically (Heckman 1979 ; Vella 1993 ; Newey et al. 1999).

$$E\widehat{(\eta_i \mid d_i^*)} = E\widehat{(\xi_i^r \mid d_i^*)} = \sum\nolimits_j d_{ij} \times \frac{\phi(\widehat{\mu}_{j-1} - \widehat{d}_i^*) - \phi(\widehat{\mu}_j - \widehat{d}_i^*)}{\Phi(\widehat{\mu}_{j-1} - \widehat{d}_i^*) - \Phi(\widehat{\mu}_j - \widehat{d}_i^*)} \tag{18}$$

Gourieroux et al. (1987) and Chesher and Irish (1987) call this term the generalized residuals, in a similar context than the previous latent variable framework with the assumption of gaussian residuals. Controlling for endogeneity is then easily effectued by putting the conditional expectation as an additional covariate in what is now the second step of estimation.

$$p_i = \alpha^c + \boldsymbol{\beta}^c B_i^\top + \boldsymbol{\delta}^c D_i^\top + \ell^c(x_i, y_i) + \rho^c E\widehat{(\xi_i \mid d_i^*)} + \varepsilon_i^c \tag{19}$$

The conditional expectation is estimated by a first step ordered probit model with the dummies *communes* as the instruments Z_i. The coefficient ρ^c is the covariance between ε_i and ξ_i divided by the variance of ξ_i. According to our omitted variable framework and to the fact that the two error terms are functions of η_i, we expect this coefficient to be positive. Decomposition of price variations with control functions: *to be done*.

2.3.3 Using Instrumental Variables (IV)

We also suggest to estimate the structural parameters in equation (1) by controlling omitted variable bias and endogeneity of GIs' designations by using an Instrumental Variable (IV) approach, that does not necessitate the normal assumptions basing the CF approach. An instrument should be sufficiently correlated with the endogenous variable and uncorrelated with the error terms of (1). In our setting, recall that the system of equations (1), (15) and (16) is triangular, where the estimation of (15) and (16) can be considered as the first sep of a Two Stage Least Squares (2SLS) approach to IV. In (15), variables contained in the matrix Z_i play the role of instruments, i.e., the variables that verify the exclusion condition. For the reasons that are detailed above, these instruments are the *communes* dummies. Because they were constituted 200 years ago, they do not affect directly the price of a bottle and hence can be considered as valid instruments. However, several problems arise in order to implement IV/2SLS in our particular setting. One possibility would be to use directly a standard IV estimator on (1). In other words, (1) is estimated using IV with all exogenous biophysical variables and the set of *communes* in our sample as instruments for the dummies D_i. However, proceeding in this way implies that we overlook that the ordered nature of the dummies D_i that appear in (1) so that there relative effects might not be estimated correctly.

Alternatively, we can use a 2SLS approach: the estimated d_i^* from (15) with an ordered qualitative estimator are directly plugged into (10). However, this method correspond to forbidden regressions (Angrist and Pischke 2008, p. 191–192) as the first stage is nonlinear. To overcome this problem, we propose to adapt the procedure described in Angrist and Pischke (2008) for one endogenous dummy variable. This means running an intermediate linear probability step by regressing the dummies D_i on all the exogenous variables (including the STS) and the predictions of the latent variable from the ordered qualitative models as an instrument: $D_i = \alpha^l + \boldsymbol{\beta}^l B_i^\top + \ell^l(x_i, y_i) + \tau^l \widehat{d_i^*} + \varepsilon_i^l$. The statistical significance of the coefficient τ^ℓ can be used to estimate the relevance of the instruments in explaining the endogenous covariates. Finally, we plug the predictions from this intermediate step $\widehat{D}_i$ in the structural equation.

$$p_i = \alpha^v + \boldsymbol{\beta}^v B_i^\top + \ell^v(x_i, y_i) + \boldsymbol{\delta}^v \widehat{D}_i^\top + \varepsilon_i^v \tag{20}$$

This procedure allows using conventional IV in our particular case of J endogenous dummy variables corresponding to an ordered qualitative variable. However, its main drawback is that is uses the nonlinearity of the first stage as an additional source of identifying information. In order to control this potential caveat, we estimate (15) using different nonlinear estimators for ordered dependent variables.

Decomposition of price variations in IV: *to be done*. nonparametric IV models (Das, 2005)

3 Data

3.1 Sample

Our proposed methodology is applied to the most famous wineyards of Burgundy, which is a French region. Burgundy is geographically cut in two main spatial delineations of interest. The first is an already mentioned administrative delineation that exists at the national scale: the *communes*, which are kind of municipalities. The second describes spatial units known as homogenous in terms of geology and soils, according to a regional soil survey (*Référentiel Pédologique de Bourgogne*, Chrétien, 2000). This survey is effectued by soil scientists, independently from the current research. The left panel of Figure 1 colors the four wine regions of Burgundy, from which we keep only *Côtes de Beaune* (CDB) and *Côtes de Nuits* (CDN). Both *Hautes Côtes de Beaune* and *Hautes Côtes de Nuits* are other wine regions that have their own GIs and do not share the same structure than CDB and CDN in terms of *Villages* (VILLs), *Premiers Crus* (PCRUs) and *Grands Crus* (GCRUs). Morever, GIs from *Hautes Côtes* are younger and spatially segregated from the other, that make them hardly comparable with the CDB and CDN.

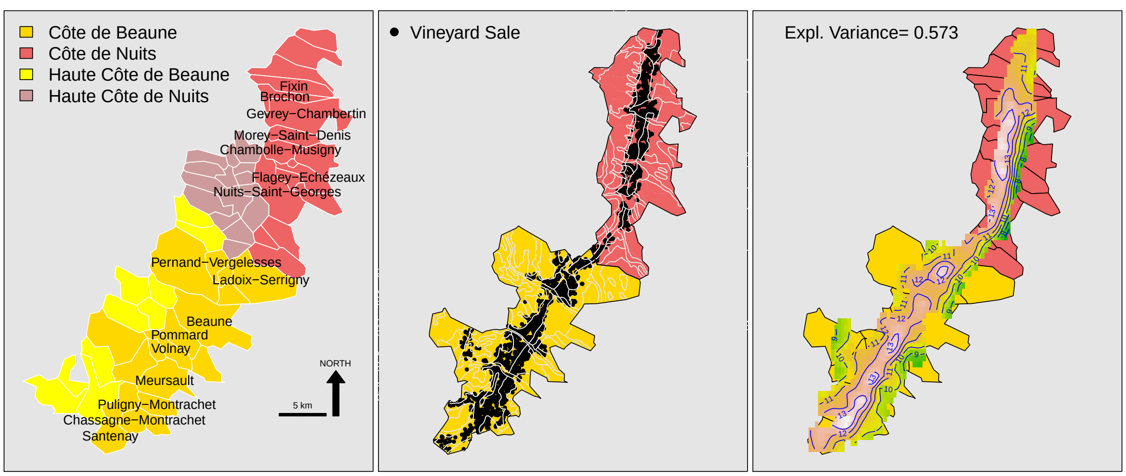


Figure 1: Wine regions and *communes* (left panel), soil units and wineyard sales (middle panel), spatial smoothing of the logarithm of sale prices (in deflated euro/ha, right panel)

The left panel shows the delineation of the *communes* (superimposed with the delination of wine regions) from which only the names of *communes* with well balanced GIs appear (see Appendix A.1 for the details). Northern regions have principally the *pinot noir* as a wine variety to make red wines and southern regions have principally *chardonnay* to make white wines. Some exceptions exist, maybe the more typical are the *communes* of *Pommard* and *Volnay* which mainly produce red wines although being in CDB. The delination of soil units is more biophysically oriented, so the spatial polygons are much more irregular. The middle panel of Figure 1 shows that only few soil units concentrate the essential of wineyard sales. The presence of wine production (and, consequently, wineyard sales) is strongly explained by the geological and soil attributes and sales are principally located at the middle of each *commune* on the East–West gradient. The right panel of Figure 1 displays the spatial distribution of the logarithm of per-ha prices, globally between $\exp(8) \approx 3,000$

and $\exp(14) \approx 1,200,000$ in deflated € 2008. The two wine regions are comparable in average prices and present a core(s)/ periphery structure. CDB (at the South) presents four cores, located at the center from the East–West gradient and regularly along the South–North gradient. CDN (at the North) displays only one core, at the middle of the South–North gradient but shifted at the West.

3.2 Variables

The variables of our final database come from five main sources. The first concerns wineyard sales from the *Société Centrale d'Aménagement du Foncier Rural* (SCAFR), the French regulatory institution of the farmland market. This database normally contains all land sales operated between 1992 and 2008, for which the productive orientation (*nature cadastrale*) is wineyard. This database contains principally the observed price for each sale, the acreages, the identifiants of plots and some qualitative information: date of the sale, socio-demographics of seller and buyer, land tenure and the presence/absence of building. The second database from the *Institut National de l'Origine et de la Qualité* (INAO) reports the spatial delineations of GIs, which allow to know exactly the GI of each parcel of each sale. The third set of variable comes from a Digital Elevation Model that computes for each plot the elevation, slope and exposition. Because the resolution of this DEM is 50 meters, we can consider these biophysical variables as perfectly observed. This is not the case for climate and soil quality variables that come from the two last databases at more aggregated scales: respectively the *communes* and the soil units.

Table 1: Frequencies and proportions of wineyard sales for each GIs within each wine region

	BGOR	BOUR	VILL	PCRU	GCRU	Sum
CDB	177	463	924	393	32	1989
(%)	(5.49)	(14.36)	(28.65)	(12.19)	(0.99)	(61.68)
CDN	124	207	739	126	40	1236
(%)	(3.84)	(6.42)	(22.91)	(3.91)	(1.24)	(38.32)
Sum	301	670	1663	519	72	3225
(%)	(9.33)	(20.78)	(51.56)	(16.10)	(2.23)	(100.00)

With a pooled sample of 3,225 wineyard plots sold, CDB and CDN contain respectively 1,989 (61.7%) and 1,236 (38.3%) observations. The frequencies of GIs are rather similar within wine regions: the GI *Village* (VILL) represents the highest number of sales for both regions, followed by *Bourgogne* (BOUR), *Premier Cru* (PCRU), *Grand Ordinaire* (BGOR) and finally *Grand Cru* (GCRU). From the left to the right of the table, GIs are ordered from the less famous to the most. Sales frequencies are not monotonically related to reputation, the highest numbers of sales correspond to the intermediate GIs.

Our sample of 2,978 wineyard plots represents 1,476 sales. Note that 907 sales are about only one plot and the sale with the highest number of parcels counts 73 parcels. All wineyard plots in one sale have the same reported par-hectare price. From Figure 2, the natural logarithm of per-ha prices is more variable between GIs than between wine regions. This result both implies that there exists

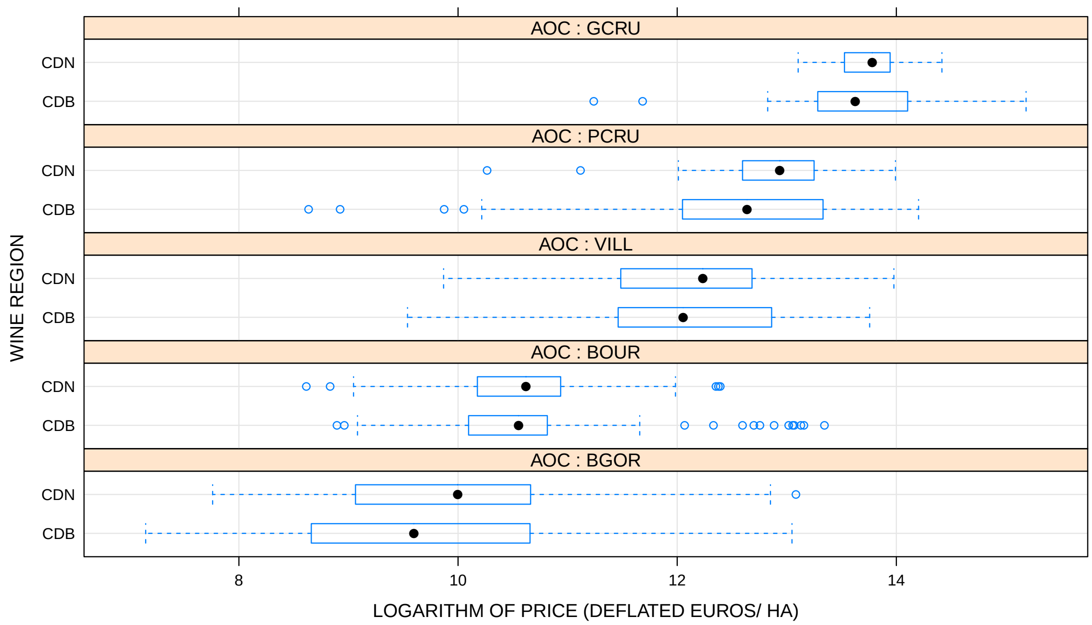


Figure 2: Within wine regions price distributions for each Geographical Indication, from the less famous (bottom) to the most (top)

strong variations of vineyard prices at small scales (within wine regions) and that GIs' effects have globally the same magnitudes (between wine regions). Moreover, the medians are approximatively linear from the BGOR of CDB to the GCRU of CDN, from the bottom to the top of Figure 2. This means that the median price of one ha of GCRU is approximatively two times higher that one hectare (ha) of PCRU, itself two times more expensive than a VILL and so on. This Figure also shows the presence of some potential outliers in terms of price per-hectare, conditionally to the GIs. We report the econometric results with the outliers but we verify systematically that the results are not too sensitive to their presence.

3.3 Summary statistics

The following Table 2 reports the usual summary statistics for continous variables contained in the final samples. Wineyard plots from Burgundy are very small (compared with other agricultural land uses and other wine regions in France). Average acreages correspond respectively to about .56 and .46 acres respectively for the CDB and CDN. The acreages are more variables for the CDB with a Standard Deviation near than three times higher.

The spatial coordinates (centered and reduced from the pooled sample) clearly show the spatial position of the two wine regions that appears at the left panel of Figure 1. If we cut the whole region in four regular rectangles, CDB is at the southern west (low longitudes and latitudes) and CDN is at the northern east (high longitudes and latitudes). We compute the distances from wineyards to the centers of *communes*, which group in general the service available to the population. This variable is used as a proxy of urban influence. The two wine regions share the same proximity of their

Table 2: Summary statistics for continous variables, separated sample for each wine region

	Côte de Beaune (N= 1,846)				*Côte de Nuits (N= 1,132)*			
Variables	Mean	St. Dev.	Min	Max	Mean	St. Dev.	Min	Max
Price (1,000 euro/ha)	252.83	296.10	1.27	3,929.03	247.07	262.00	2.34	1,822.79
Surface (ha)	0.22	0.41	0.001	13.37	0.14	0.14	0.001	1.14
Longitude (scaled)	−0.70	0.51	−1.92	0.48	1.15	0.19	0.47	1.50
Latitude (scaled)	−0.70	0.42	−1.54	0.13	1.15	0.44	0.06	2.01
Distance to Center (km)	1.19	0.57	0.08	3.67	1.00	0.51	0.07	2.90
Elevation (100 m)	2.66	0.48	2.12	4.76	2.69	0.26	2.17	3.79
Slope (degree)	4.67	4.81	0.00	23.61	3.14	3.37	0.40	19.65
Temperature (Celsus)	11.16	0.26	10.92	11.49	11.10	0.18	10.92	11.515
Precipitations (mm)	809.94	17.76	789.37	842.16	816.27	27.01	749.96	841.33
Solar Radiation (Joules)	984.93	17.86	952.00	1,000.52	972.09	16.00	953.08	1,003.61
Humidity (mm)	940.57	5.47	931.05	944.90	936.71	3.64	930.73	943.98
Wind (km)	30.25	2.48	27.31	33.41	30.48	2.25	28.90	35.69
Snow (cm)	26.14	7.10	17.05	39.95	31.25	6.04	22.39	39.34
Water Holding Capacity (mm)	91.08	36.91	0	153	73.32	40.51	0	188
Soil Depth (cm)	52.96	19.61	0	93	47.96	19.74	0	80
Stone Rate (permil)	13.55	14.76	0	50	23.71	16.28	0	85
Silt Rate (percent)	49.93	7.41	0	60	46.21	12.28	0	66
Sand Rate (percent)	15.04	6.43	0	33	15.41	5.59	0	40
Clay Rate (percent)	34.70	4.63	0	55	35.98	9.31	0	45

vineyards to these centers. The within distributions of all biophysical variables are rather similar between the two wine regions. Nevretheless, we can note that elevation appears as more variable for the CDB and that slope is higher on average. Despite a North–South difference in locations, average temperatures do not really differ between the regions. Aggregate variables (the 12 last rows) present smaller coefficients of variations that the others perfectly observed variables. It is particular true for climate variables compared with elevation and slope.

In a unreported analysis, we find high correlations between the numerous biophysical variables that are relevant to explain both GIs' designation and wineyard prices. Hence, we operate two Principal Component Analysis (PCA) to reduce the dimension of the covariates and to decrease problems of multicolinearity in the econometric models. Correlated covariate are separated between climate (rows 8–14) and soil (rows 15–20) variables that are plugged in two PCAs. We keep the two principal axis of both. Each of them explains respectively 87 and 72 percents of the empirical variances of the initial variables (see Appendix A.1 for the details).

4 Results

4.1 Naive and proxy hedonic models

4.1.1 Marginal significance of variables

We first report the results from naive hedonic models that ignore the possibility of omitted biophysical variables and the endogeneity of GIs' designations. To take into account the strong nonlinearities in the effects of biophysical variables on prices, we use both semiparametric B-Splines[4] and polynomial terms. The splines fit better the data in general but the results are less easily interpretable. As it is more usual to use polynomial terms in econometrics, we provide the results from the two methods. We estimate six specifications for each of the three samples (pooled, CDB and CDN), the spline orders are chosen according to a forward selection and the polynomials are limited to the second order to keep the interpretations simple. The first two specifications only contain the variables about NCs, only observed biophysical variables for (I) and with STS of geographical coordinates for (II). The specifications (III) and (IV) only contain the GIs' dummies, respectively with and without STS. The two last specifications (V) and (VI) include both GIs and NCs, with and without STS. The results obtained from splines models are reported in the following Table 3. The details of the coefficients from polynomials models are reported in Table 10 of the Appendix A.3.

In Table 3, the values reported are the increases in terms of SSE that follow the introduction of all the spline terms of each variable, indicating the marginal significances for each of them (Fox and Weisberg, 2010). These numbers are close to the usual Fisher statistics for the joint nullity of the spline coefficients associated to each variable (the "order" columns also represent the degrees of freedom). Knowing that SST is 2,636 for the pooled sample, the R-squared can be computed from the last row of the top panel: $R^2 = 1 - SSR/SST$. They are respectively .31, .53, .52, .65, .60 and .67 for the specification (I)–(VI). The SST for the CDB and the CDN samples are respectively 1,766 and 871, so the R-squared ranges are .40–.67 and .47–.75 for all the specifications. For specifications (I), elevation appears as the most important biophysical variable, even if for the CDN sample WET is equally strong. The effects of slope is strong for the pooled sample but not for the others. Looking at the specification (II), the inclusion of STS sharply decreases the coefficients associated to biophysical variables and increase the R-squared. STS are individually highly significants such that specifications (II) fit the data better than specifications (III) that include only GIs. However, GIs have strong effects too where included alone in (III). For the three samples, the SSE just stemming from GIs are close to the SSR which means that the partial R-squared for the all four dummies is about 1/2. Including STS simultaneously with GIs – specification (IV) – implies a division by two of the marginal effects of GIs. This results still holds for the specifications with both biophysical variables and GIs: between (V) and (VI). By comparing the individual significances of biophysical variables (elevation in particular) between (I) and (V), we see that including GIs strongly decrease their effects. The reverse is also true, albeit at a lesser extend when we include observed biophysical variables in addition to GIs: by comparing specifications (III) and (V). The same evidences are found on the more classical polynomial regressions of order

[4] http://cran.r-project.org/web/packages/crs/vignettes/spline_primer.pdf (last accessed: September 8, 2013).

Table 3: Marginal Explained Sums of Squares for each variables and samples, models with splines

		Pooled Sample					
Variable	Order	(I)	(II)	(III)	(IV)	(V)	(VI)
YEAR	1	6.97	8.74	12.76	7.53	5.51	7.21
GIs	4			1272.07	526.46	741.13	379.81
DEPTH	3	5.60	0.24			6.99	1.69
ELEV.	4	228.58	7.00			17.77	4.87
HEAT	3	23.14	21.15			37.17	13.88
SLOPE	3	102.91	14.19			8.80	3.76
ROUGH	3	36.39	20.47			10.99	0.37
WET	3	94.32	24.04			52.58	4.12
X	5		146.83		93.29		34.79
XY^a	X		340.44		152.87		97.35
Y	5		148.66		120.95		41.60
EXPO.	3	28.33	19.86			1.84	0.83
DIST.	3	29.43	36.36	11.99	15.80	13.06	18.21
Residuals	1	1806.82	1248.48	1257.35	912.70	1065.69	868.67

		Cote de Beaune						*Cote de Nuits*					
Variable	Order	(I)	(II)	(III)	(IV)	(V)	(VI)	(I)	(II)	(III)	(IV)	(V)	(VI)
YEAR	1	2.12	4.18	14.38	5.37	5.06	4.82	9.37	7.98	3.25	2.80	3.08	2.43
GIs	4			836.05	303.57	403.47	245.57			443.82	108.19	189.21	85.39
DEPTH	3	16.32	3.79			7.15	3.29	1.31	0.82			3.52	0.56
ELEV.	4	204.95	4.78			8.40	3.31	7.79	1.48			1.95	2.95
HEAT	3	36.86	8.06			39.79	3.80	52.03	1.90			33.98	1.02
SLOPE	3	17.91	5.88			4.29	1.38	8.15	2.20			4.42	4.81
ROUGH	3	39.22	8.06			22.15	2.60	52.35	8.84			5.50	1.17
WET	3	85.71	39.72			73.76	25.61	58.20	15.32			32.61	6.27
X	5		63.09		72.44		22.45		46.78		20.85		11.47
XY^a	X		182.20		94.98		54.69		57.38		40.63		31.85
Y	5		48.60		57.92		22.84		25.00		68.34		10.62
EXPO.	3	33.66	28.35			0.51	0.50	7.82	3.10			5.72	2.48
DIST.	3	26.37	20.78	17.16	15.1A	13.09	12.32	9.59	0.94	0.05	0.43	1.41	0.07
Residuals	1	1067.62	830.14	845.52	625.68	664.16	584.58	463.41	303.52	394.32	236.88	274.20	218.12

Notes: (a) The XY rows are for the interaction between the splines of each geographical coordinate, to increase the flexibility of STS two, reported in Table 10 of the Appendix A.3.

4.1.2 Marginal effects

The marginal significances of GIs as a whole are highly variable according to the different specifications (with and without biophysical variables and with and without STS) but some features remain relatively robust between samples. Because the effects of GIs are modeled through dummies variables, the following Table 4 reports the marginal effects of GIs' dummies[5] and the standard errors estimated by the delta method. For spline functions, the marginal effects are displayed by plotting conditional predictions with all other variables at their samples means (Fox and Hong, 2009). These effect plots are reported in Figure 6, Appendix A.2 for each of the three samples.

Table 4: GIs' marginal effects and standard errors from spline models, GI ref.: *Grand Ordinaire*

	Bourgogne		*Village*		*Premier Cru*		*Grand Cru*	
	effect	s.e.	effect	s.e.	effect	s.e.	effect	s.e.
POOL (III)	92.80	24.00	722.60	100.30	1441.80	197.90	4310.00	695.60
POOL (IV)	54.90	21.70	605.30	104.90	829.70	157.20	3030.70	591.20
POOL (V)	68.30	21.90	685.80	101.10	1273.10	198.20	4747.70	817.10
POOL (VI)	48.50	21.60	526.90	94.90	726.10	141.20	2535.80	517.80
CDB (III)	120.50	38.20	881.30	167.10	1595.60	297.10	4083.50	1124.30
CDB (IV)	87.30	34.60	772.70	481.70	1049.80	1238.10	3496.40	3256.00
CDB (V)	77.40	32.50	810.60	164.90	1137.00	244.70	4447.80	1317.70
CDB (VI)	102.30	57.10	813.20	1079.50	1103.20	4784.40	3565.10	8959.00
CDN (III)	64.20	29.30	579.80	117.90	1496.30	303.20	3955.10	769.70
CDN (IV)	14.80	40.40	355.70	196.50	510.30	286.90	1234.30	748.00
CDN (V)	58.60	33.70	605.00	170.30	1088.40	321.10	3030.10	798.30
CDN (VI)	29.10	35.10	365.40	153.20	543.50	241.30	1347.20	573.60

The marginal effects of GIs on wineyard prices are high. Relatively to the GI *Grand Ordinaire*, a wineyard designated *Bourgogne* is 92.8% ($\widehat{\sigma} = 24$) more expensive according to specification (III) on the pooled sample. This number falls to 48.5% ($\widehat{\sigma} = 21.6$) when omitted biophysical variables are taken into account through the proxy STS solution. When evaluated at the average price of a *Grand Ordinaire* wineyard (€ 51,968), these premiums correspond respectively to € 48,227 and € 25,205. Recall that Cross et al. (2011) found $7,000 for the biggest premium in Oregon. For the more famous GIs, the premiums are incommensurate: respectively € 273,823, € 377,345 and € 1,317,825 for *Village*, *Premier Cru* and *Grand Cru* according to the conservative specification (VI) on the pooled sample. These magnitudes are similar within each wine regions CDB and CDN, even if certain differences appear. For such high numbers, the associated uncertainty is high as the standard errors are quite important but the premiums are often significant two-by-two, i.e., between adjacent GIs on the reputation scale. Including STS implies an important decrease of the premiums of the *Premier Cru*, in particular relatively to the premiums of *Village* (it is clear for the pooled

[5] We use semi-log models, the marginal effects of dummies are $100 \times [\exp(\delta - \sigma_\delta/2) - 1]$, see Kennedy (1981).

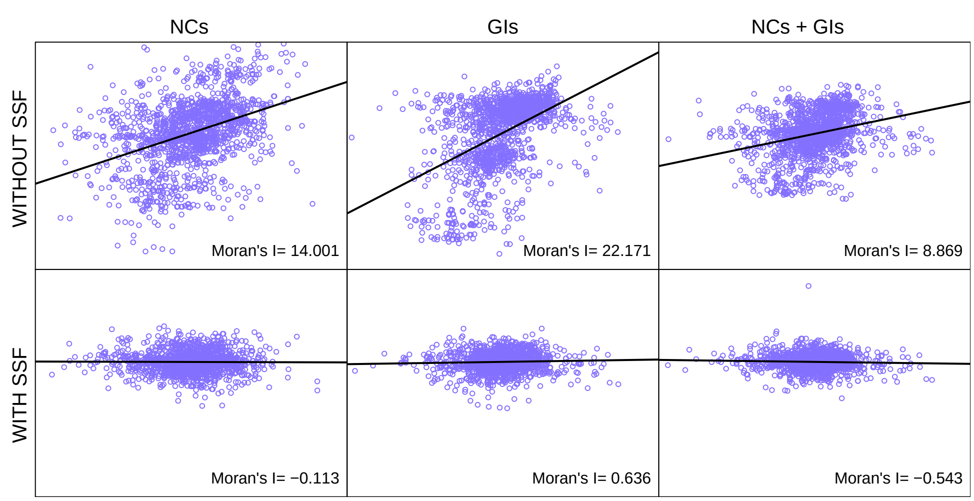
sample between specifications (III) and (IV), and (V) and (VI) but also for the others samples). Note that the specifications with STS on the CDB sample – (IV) and (VI) – present high standard errors for the effects of famous the GIs (*Village*, *Premier Cru* and *Grand Cru*).

From Figure 6 of the Appendix A.2 displays the marginal effects of elevation and slope for the pooled sample, the CDB and CDN. They are reported with and without controlling for GIs, specifications (I) and (V). Between the two specifications, the magnitudes of the effects of this biophysical variables decrease: the curves are more flattened in (V). In general, the effects of elevation are shaped as an inverted U and the effects of slope are concave. For the pooled sample, the elevation with the higher wine prices is of about 240m, For the CDB and the CDN the values are about 280 and 260. The slope effects are highly increasing at low values and become non significant from 10 degrees. Globally the effects from the pooled sample are closest to those from CDB, which can be understood by the higher number of observations. The hedonic marginal prices of these biophysical attributes are the derivative of the effect curves. Elevation and slope have positive or negative values, depending at the level of evaluation.

4.1.3 Spatial autocorrelation

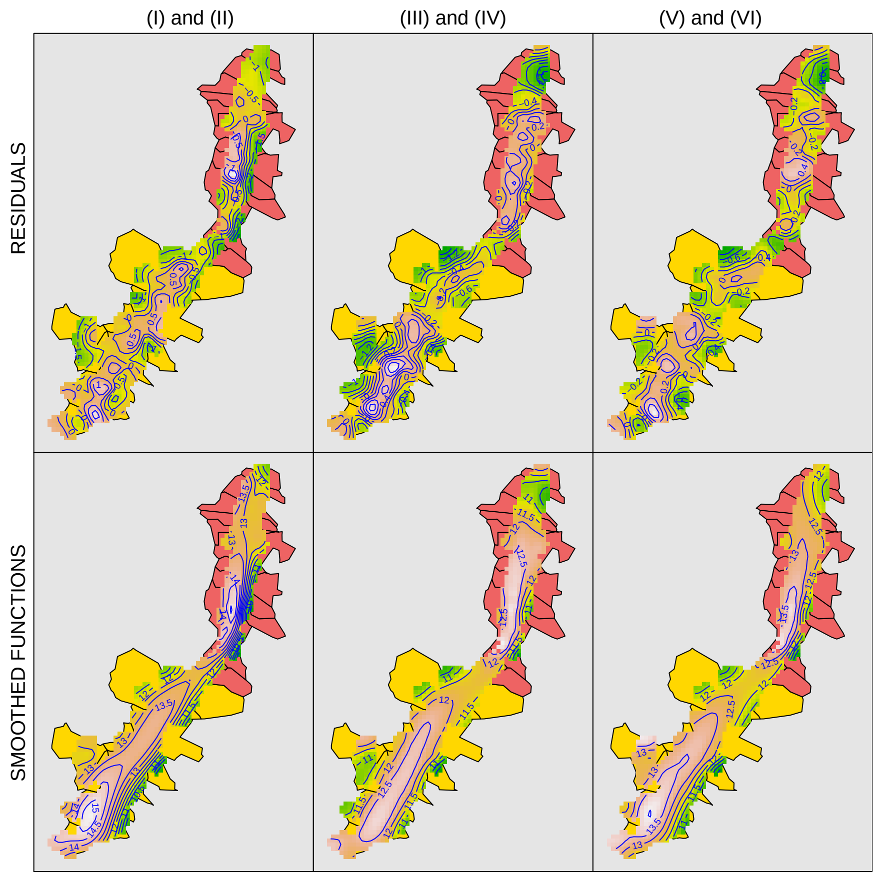
As we argue in subsection 2.1, one way to evaluate the relevance of the proxy solution (equation 14) is to test for spatial autocorrelation of the errors from the hedonic equation. The following Figure 3 displays the Moran's plots for the specifications (I)–(VI) on the pooled sample, for the North-West to the South-East. These plots present on the x-axis the estimated errors for each sale i and on the y-axis the average errors of these neighbors, weighted according to their distances to i. The spatial weight matrix is a gaussian spatial kernel with a bandwith chosen in order to have at least one neighbor for each observation.

Figure 3: Moran plots for spatial autocorrelation of residuals, pooled sample



This Figure clearly shows the interest of STS to control for the spatial autocorrelation of the residuals, and this is true for each specification. Not any spatial effect stay in the residuals after controlling by STS. This involves that the proxy STS solution is sufficient to take into account any omitted spatial biophysical variable. More formally, a bootstrap inference of the Moran's **I** statistics indicate that we reject the spatial independance of errors for the models without STS but we cannot reject it for the models with STS (95% for (II) and (VI), 90% for (IV)). To investigate deeper the spatial structure of the residuals and the proposed solution throught STS, the following Figure 4 displays three STS estimated on the residuals from the specifications (I), (III) and (V) at the top and the three estimated STS $\widehat{\ell^r}(x_i, y_i)$ for the specifications (II), (IV) and (VI) at the bottom.

Figure 4: Maps of estimated residuals from models without STS – (I), (III) and (V), top panel – and the estimated STSs when there are in models: (II), (IV) and (VI), bottom panel



Residuals on the three map at the top look closely to the spatial distribution of price as it appears in Figure 1 but some differences exist in details. Errors are more important in the CDN for the specification (I) with only NCs and the errors are more important in the CDB in the specification (III) with only GIs. In specification (V) the errors are equally variable between the two wine regions. The three STS at the bottom globally displays the same spatial patterns, even if some differences also appears in details. The STS are less marked from the left to the right, according to what Table 3 presents. CDN presents a high East-West gradient in the specification (II), which decreases (but stays) in the specifications (IV) and (VI).

4.1.4 Decompositions

If we turn to the wineyard price decompositions between the GIs (SSD) and the NCs (SSB) for the specifications (V) and (VI) that contain both GIs and NCs, contrasting evidences appear. They are summarized in the following Table 5. For both specifications, GIs appear as a more important source of wineyard price variation than NCs. However, taking into account the possibility of spatially omitted biophysical variables for the pooled sample implies more than a doubling of the explained part from NCs, from 7.62% to 16.56% between (V) and (VI). The naive hedonic model (and, therefore, previous analysis from other papers) underestimates the share of price variations that comes from NCs and overestimate the variations from GIs. Both with and without taking into account the possibility of omitted biophysical variables, CDB has smaller proportions of explained variations due to NCs than CDN, an interesting result to describe the differences between the two wine regions. It is also interesting to note that the interaction effects are increasing when adding a STS to the regression functions. Hence, controlling for omitted variables is done at the cost of a loss of precision in the decompositions.

Table 5: Decompositions of price variations according to naive models (V) and proxy solution (VI)

Sample	Spec.	SSB	SSD	SSR	Ω
POOL	(V)	7.62	50.97	40.42	0.99
POOL	(VI)	16.56	37.63	32.95	12.86
CDB	(V)	11.20	47.08	37.61	4.11
CDB	(VI)	15.75	42.20	33.11	8.94
CDN	(V)	15.07	49.14	31.50	4.29
CDN	(VI)	24.33	31.51	25.06	19.10

4.2 Ordered models of GIs' designations

In this section, we report the results from the first step of ordered qualitative models on GIs' designations. Because these models of designation have an interest of they own, we estimate four specifications both with logistic and probit errors for a total of eight estimations. All the specifications contain the perfectly observed biophysical variables (elevation, slope and exposition), what distinguishes the four specifications is the following:

- (i) : only *commune* and soil unit dummy variables in addition
- (ii) : only climate and soil biophysical variables in addition
- (iii) : (i) with STS
- (iv) : (ii) with STS

We cannot include simultaneously the biophysical variables and the dummies because of perfect colinearity (recall that continous biohysical variables come from aggregate sources at the *commune* and soil units scales) so these specifications are complementary. Because the *communes* dummies are our instruments, we only use the specifications (i) and (iii) with the probit errors as the first step in our CF and IV approaches. The other estimations are displayed to show that the processes underlying GI's designations are closely related to the effects of the previous hedonic models. The effects of biophysical variables on GIs' designation probabilities are reported in Figure 7, Appendix A.2. (according to the methodology of Fox and Hong 2009). We also see important nonlinearities of the effects of biophysical variables. For the elevation effect, the maximum probabilities for the most famous GIs (*Grand Cru*, *premier Cru* and *Village*) are about 250m for the pooled sample. This is the same value that we found for the hedonic model, making explicit the nesting of NCs and GIs. The average probabilities at these elevations are about 70% and 20% for respectively the *Village* and the *Premier Cru*. The maximum probabilities for the less famous GIs are at the top of the distribution: above 400m. Low elevations also present high probabilities of designations both for *Grand Ordinaire* and *Bourgogne*. For the slope effects, the maximum probabilities for famous GIs are about 8–9 deggres if we neglect the corner effects for the high values of slope (only 15 vineyard plots have a slope greater than 20: .05% of the sample). GIs with low reputation are clearly located on small slopes.

For CDB, the effects of elevation and slope are really close to what we found on the pooled sample. The maximum probabilities for famous GIs are for elevations of about 275m and these are more marked than for the pooled sample.For CDN, the results are well contrasted to the other samples. The elevation is globally, positively related with the probability of a famous GI. The slope effect is U shapped and the other biophysical variables have they own (very significant) effects. Raw coefficients from polynomial estimations (instead of splines) are available in the Table 11 at the Appendix A.3. To evaluate the capacity of these ordered models to represent GIs' designation, the following Table 6 reports both the frequencies between observed and predicted GIs (top panel) and the percent of good predictions in terms of GIs (bottom panel). Predictions from ordered qualitative models are computed by maximum probabilities.

Table 6: Frequencies between observed and predicted GIs (rows 1–5) plus percent of good predictions (rows 6–8)

		Pooled Sample				Separated Samples			
	OBS	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)
BGOR	285	168	160	197	196	246	178	212	214
BOUR	603	528	463	595	500	501	511	593	543
VILL	1539	1979	2087	1733	1868	1854	1952	1703	1763
PCRU	483	303	268	440	413	376	336	439	441
GCRU	68	0	0	13	1	1	1	31	17
ALL	100	63.9	57.69	70.72	64.07	68.57	62.46	72.77	69.11
CDB	100	66.58	54.66	70.8	61.7	69.23	57.96	72.21	66.14
CBN	100	59.54	62.63	70.58	67.93	67.49	69.79	73.67	73.94

In general, ordered models overestimate the number of the intermediate GIs (*Village*) and underestimate the GIs at the extremes. Note that the most famous GI (*Grand Cru*) is only well predicted when the ordered models are estimated on separated samples. The differences are decreasing when we include STS, between specifications (i)–(ii) and (iii)–(iv). The ordered models perform well in predicting the GIs at the plot scale: more than 70% of correct predictions for the best specification (iii). It is important because with five items for the ordered endogenous variable, a random allocation produces 20% of correct predictions. For all samples, the specification with splines and discrete variables (iii) perform the best, so is used in the CF and IV. These last results clearly show the relevance of our instrument *communes* to explain the GIs: the first of the condition to be a good instrument. It appears by comparing the specifications (iii) and (iv).

4.3 Control Functions

The raw results from the control function approach are available in Table 12 of the Appendix A.3. The specification (VII) is without STS and the (VIII) with STS so the generalized residuals used as CFs are respectively from ordered models (i) and (iii). These results have to be compared to those from specifications (V) and (VI) that do not take into account the endogeneity of GIs (the raw results from (VI) are reported in the Table 10). The R-squared are comparable with and without CF, as the shapes of effects of biophysical variables (so we do not report them). As expected, the coefficients associated to the CFs at the bottom of the table are positive and highly significant (except for the specification (VII) on the sample CDN, the last column). Biggest induced changes are relative to the effects of GIs. Just by including CF, the coefficient associated to *Bourgogne* becomes not significant, the coefficient of *Premier Cru* a little under the coefficient of *Village* and the coefficient of *Grand Cru* diminishes from 3.3 in specification (VI) to 2.1 with in CF of specification (VIII).

The following table reports the marginal effects of GIs, with their associated Standard Errors. This results have to be compared with those of Table 4 to see the effects of CF.

Table 7: GIs' marginal effects and standard errors from control function models, ref.: *Grand Ordinaire*

	Bourgogne		*Village*		*Premier Cru*		*Grand Cru*	
	effect	s.e.	effect	s.e.	effect	s.e.	effect	s.e.
POOL (VII)	33.19	21.26	372.89	106.63	477.91	182.52	1068.07	484.51
POOL (VIII)	26.90	20.90	302.33	93.05	289.09	125.92	679.36	332.31
CDB (VII)	54.51	35.10	519.64	192.13	619.30	302.60	1449.98	924.95
CDB (VIII)	45.82	34.20	403.03	158.28	390.01	211.93	1126.64	761.63
CDN (VII)	3.27	24.06	124.40	75.54	145.67	119.94	328.01	245.06
CDN (VIII)	10.93	26.64	213.96	114.06	223.03	168.34	435.46	330.62

There are strong decreases of the marginal effects of GIs. The *Bourgogne* is 28.8% more expensive than the reference modality *Grand Ordinaire* (instead of 48.5), the premium is around € 14,000 instead of € 25,200. By comparing marginal effect of GIs betwen specifications (VIII) and (VI) for the more famous GIs, we finds premiums of respectively € 157,099.3, € 150,239.5 and € 353,070.6 instead of € 273,823, € 377,345 and € 1,317,825. The decreases of GIs' premiums are from − 40% to − 75%, the most famous GIs are relatively more impacted.

4.4 Instrumental Variables

Table 8: R squared and Student's t for intermediate IV

	BGOR	BOUR	VILL	PCRU	GCRU
POOLED (IX)	0.381	0.229	0.114	0.307	0.186
	-22.092	-10.195	2.611	15.900	15.060
POOLED (X)	0.429	0.298	0.197	0.358	0.234
	-15.973	-10.289	1.901	13.909	12.354
CDB (IX)	0.398	0.295	0.096	0.374	0.196
	-16.109	-8.974	1.665	13.275	10.444
CDB (X)	0.449	0.377	0.275	0.541	0.367
	-16.756	-6.790	1.218	13.010	9.012
CDN (IX)	0.506	0.244	0.230	0.225	0.251
	-13.297	-3.701	0.773	6.764	11.420
CDN (X)	0.577	0.341	0.358	0.311	0.375
	-10.374	-4.941	0.889	6.899	9.469

Table 9: GIs' Premiums from Instrumental Variables

	Bourgogne		*Village*		*Premier Cru*		*Grand Cru*	
	effect	s.e.	effect	s.e.	effect	s.e.	effect	s.e.
POOL (IX)	773.26	449.17	1404.86	583.90	4661.77	2047.65	11440.03	8496.62
POOL (X)	193.39	137.00	989.73	427.10	1792.16	951.11	4055.66	2388.12
CDB (IX)	43.76	145.29	962.57	764.37	1749.07	1470.26	2039.24	2380.22
CDB (X)	-48.20	61.42	216.87	256.48	47.43	156.57	13232.04	15501.95
CDN (IX)	30.45	111.85	103.99	143.94	165.98	349.12	8530.99	8468.65
CDN (X)	-75.77	18.85	-35.73	42.23	-95.73	5.34	86.15	186.51

5 Conclusion

What the Imbens 2000, Zhao et al 2013 bring in the presence of omitted variables? When the treatment is based on unobservables they are not good.

Even based on the classical hedonic theory, our methodology makes three main distinctions that are necessary for our question:

- Use semiparametric splines to take into account nonlinear effect of biophysical variables on wineyard prices.

- Use Spatial Trend Surface to control for unobservable biophysical variables and error-in-variables.
- Use Control Functions and Instrumental Variables to control for the endogeneity of Geographical Indications, using the historical determinants of the designation choices.

References

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The quarterly journal of economics* : 488–500.

Ali, H. H., S. Lecocq and M. Visser (2008). The impact of gurus: Parker grades and en primeur wine prices. *The Economic Journal* 118: F158–F173.

Ali, H. H. and C. Nauges (2007). The pricing of experience goods: the example of en primeur wine. *American Journal of Agricultural Economics* 89: 91–103.

Anglin, P. M. and R. Gencay (1996). Semiparametric estimation of a hedonic price function. *Journal of Applied Econometrics* 11: 633–648.

Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Anselin, L. (1988). *Spatial econometrics: methods and models*, *4*. Springer.

Ashenfelter, O. (2008). Predicting the quality and prices of bordeaux wine*. *The Economic Journal* 118: F174–F184.

Ashenfelter, O., D. Ashmore and R. Lalonde (1995). Bordeaux wine vintage quality and the weather. *Chance* 8: 7–14.

Ashenfelter, O. and K. Storchmann (2010). Using hedonic models of solar radiation and weather to assess the economic effect of climate change: the case of mosel valley vineyards. *The Review of Economics and Statistics* 92: 333–349.

Ay, J.-S. and L. Latruffe (2013). The empirical content of the present value model: A survey of the instrumental uses of farmland prices. *Factor Markets Working Papers* .

Bao, H. X. and A. T. Wan (2004). On the use of spline smoothing in estimating hedonic housing price models: empirical evidence using hong kong data. *Real estate economics* 32: 487–507.

Besanko, D., S. Donnenfeld and L. J. White (1987). Monopoly and quality distortion: effects and remedies. *The Quarterly Journal of Economics* 102: 743–767.

Carew, R. and W. J. Florkowski (2010). The importance of geographic wine appellations: hedonic pricing of burgundy wines in the british columbia wine market. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 58: 93–108.

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

Combris, P., S. Lecocq and M. Visser (1997). Estimation of a hedonic price equation for bordeaux wine: does quality matter? *The Economic Journal* 107: 390–402.

Combris, P., S. Lecocq and M. Visser (2000). Estimation of a hedonic price equation for burgundy wine. *Applied Economics* 32: 961–967.

Costanigro, M., J. J. McCluskey and C. Goemans (2010). The economics of nested names: name specificity, reputations, and price premia. *American Journal of Agricultural Economics* 92: 1339–1350.

Costanigro, M., J. J. McCluskey and R. C. Mittelhammer (2007). Segmenting the wine market based on price: hedonic regression when different prices mean different products. *Journal of Agricultural Economics* 58: 454–466.

Cross, R., A. J. Plantinga and R. N. Stavins (2011). What is the value of terroir? *American Economic Review* 101: 152.

Das, M. (2005). Instrumental variables estimators of nonparametric models with discrete endogenous regressors. *Journal of Econometrics* 124: 335–361.

Davidson, R. and J. G. MacKinnon (2004). *Econometric theory and methods*. Oxford University Press New York.

Dubois, P. and C. Nauges (2010). Identifying the effect of unobserved quality and expert reviews in the pricing of experience goods: empirical application on bordeaux wine. *International Journal of Industrial Organization* 28: 205–212.

Ekeland, I., J. J. Heckman and L. Nesheim (2004). Identification and estimation of hedonic models. *Journal of Political Economy* 112: 60.

Fik, T. J., D. C. Ling and G. F. Mulligan (2003). Modeling spatial variation in housing prices: a variable interaction approach. *Real Estate Economics* 31: 623–646.

Fox, J. and J. Hong (2009). Effect displays in r for multinomial and proportional-odds logit models: Extensions to the effects package. *Journal of Statistical Software* 32: 1–24.

Fox, J. and S. Weisberg (2010). *An R companion to applied regression*. Sage.

Gergaud, O. and V. Ginsburgh (2008). Natural endowments, production technologies and the quality of wines in bordeaux. does terroir matter? *The Economic Journal* 118: F142–F157.

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

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society* : 153–161.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika* 87: 706–710.

Josling, T. (2006). The war on terroir: geographical indications as a transatlantic trade conflict. *Journal of Agricultural Economics* 57: 337–363.

Kammann, E. and M. P. Wand (2003). Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 52: 1–18.

Kennedy, P. E. (1981). Estimation with correctly interpreted dummy variables in semilogarithmic equations. *American Economic Review* 71: 801.

Le Gallo, J. and B. Fingleton (2012). Measurement errors in a spatial context. *Regional Science and Urban Economics* 42: 114–125.

Lecocq, S. and M. Visser (2006). Spatial variations in weather conditions and wine prices in bordeaux. *Journal of Wine Economics* 1: 114–124.

Li, Q. and J. S. Racine (2007). *Nonparametric econometrics: Theory and practice*. Princeton University Press.

McMillen, D. P. (2003). Spatial autocorrelation or model misspecification? *International Regional Science Review* 26: 208–217.

McMillen, D. P. (2010). Issues in spatial data analysis. *Journal of Regional Science* 50: 119–141.

Menapace, L. and G. Moschini (2012). Quality certification by geographical indications, trademarks and firm reputation. *European Review of Agricultural Economics* 39: 539–566.

Mérel, P. and R. J. Sexton (2012). Will geographical indications supply excessive quality? *European Review of Agricultural Economics* 39: 567–587.

Mussa, M. and S. Rosen (1978). Monopoly and product quality. *Journal of Economic Theory* 18: 301–317.

Nelson, P. (1970). Information and consumer behavior. *The Journal of Political Economy* 78: 311–329.

Nerlove, M. (1995). Hedonic price functions and the measurement of preferences: The case of swedish wine consumers. *European economic review* 39: 1697–1716.

Newey, W. K., J. L. Powell and F. Vella (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica* 67: 565–603.

Norman, R. (2010). *Grand Cru: The Great Wines of Burgundy Through the Perspective of Its Finest Vineyards*. Kyle Cathie.

Oczkowski, E. (2001). Hedonic wine price functions and measurement error. *Economic record* 77: 374–382.

Palmquist, R. B. (1989). Land as a differentiated factor of production: A hedonic model and its implications for welfare measurement. *Land economics* 65: 23–28.

Rosenbaum, P. R. and D. B. Rubin (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association* 79: 516–524.

Schnabel, H. and K. Storchmann (2010). Prices as quality signals: evidence from the wine market. *Journal of Agricultural & Food Industrial Organization* 8.

Stanziani, A. (2004). Wine reputation and quality controls: the origin of the aocs in 19th century france. *European Journal of Law and Economics* 18: 149–167.

Vella, F. (1993). A simple estimator for simultaneous models with censored endogenous regressors. *International Economic Review* 34: 441–57.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. The MIT press.

A Appendix

A.1 Variable recoding

We recode the *communes* by only keeping a dummy for *communes* with strictly more than three GIs (on five possible). This is done to keep the variables well distributed with each GI (the qualitative endogenous variable of the first steps). From the 35 initial, 16 *communes* are finally selected to have their proper fixed effects, see their names in Figure 1. The same is done for soil units. From the 22 initial soil units, seven are finally selected to have strictly more than three GIs within. These two spatial delineations are also used to match some exogenous variables. *Communes* are used to obtain the average climate 1970–2000 and soil units are used to obtain land quality variables.

Figure 5 displays the initial climate and soil variables in the spaces of the principal axis. It allows us to interpret these main axis in terms of the initial variables and facilitates the interpretations of the results.

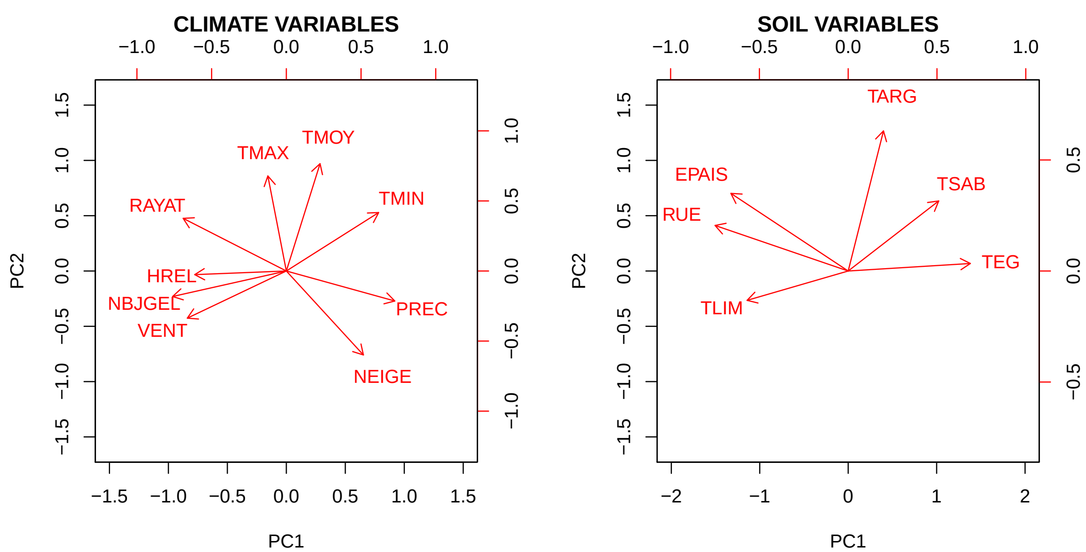
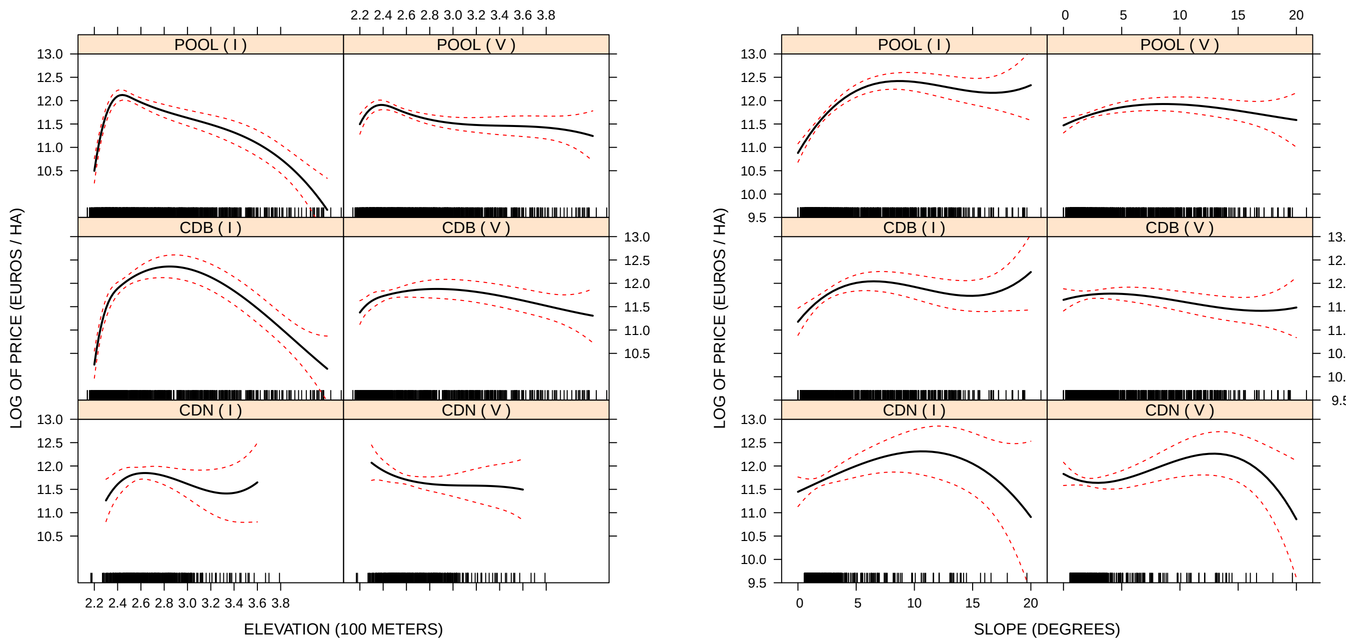


Figure 5: Principal Component Analysis for initial climate and soil variables

For climate variables (left panel), high values for the first axis correspond to high values for cumulative precipitations (`PREC`), quantity of snow (`NEIGE`), and small values of solar radiation (`RAYAT`), humidity (`HREL`) and wind (`VENT`). We call this axis a proxy for a **wet** climate. The second axis is clearly a proxy of a **heat** climate: high values mean high temperatures (`TMOY`, `TMAX` and `TMIN`). For soil variables, the first axis represent the opposite of a classical view of fertility with high values meaning the presence of stones (`TEG`), and low values the presence of high water holding capacity (`RUE`) and silt (`TLIM`). We do not take the opposite of this axis to be included in the regression because for wine production, unfertile soils are in general prefered. We call this variable an index of **rough** soils. For the second axis from soil variables, high values means high values in terms of clay (`TARG`), thickness (`EPAIS`), water holding capacity, and sand (`TSAB`). So, it seems natural to call this dimension a proxy for soil **depth**.

A.2 Spline effects on prices

Figure 6: Marginal B-spline effects of elevation (left) and slope (right) from pooled sample, Cote de Beaune and Cote de Nuits.



A.3 Regression tables

Figure 7: Predicted probabilities of GIs from pooled sample, logistic spline model with continous variables

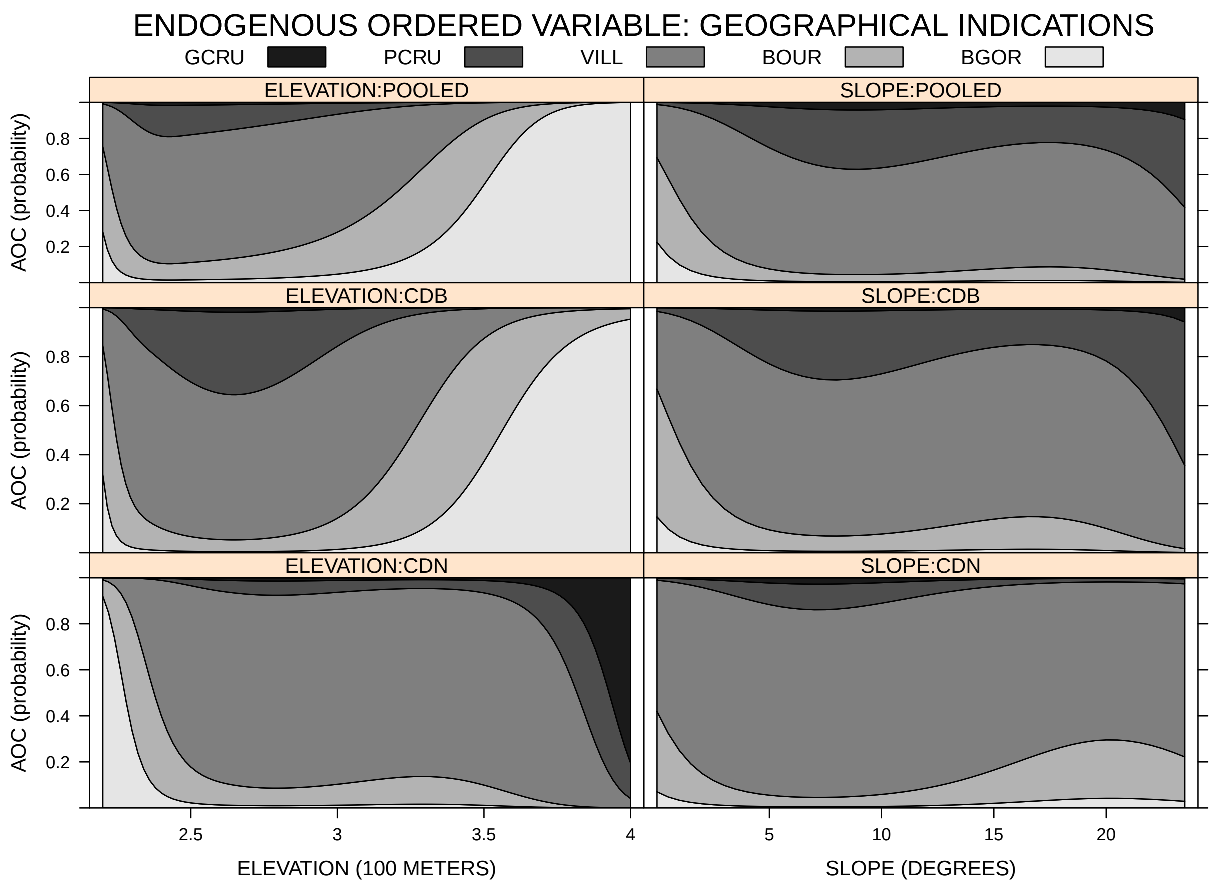


Table 10: Regression results from polynomial models of wineyard prices

	Pooled Sample			Cote de Beaune			Cote de Nuits		
	(I)	(III)	(VI)	(I)	(III)	(VI)	(I)	(III)	(VI)
Elevation	7.934***		−2.413	15.309***		−1.950	−1.486		−7.390
	(1.351)		(1.916)	(1.777)		(17.170)	(3.247)		(6.532)
Elevation^2	−1.387***		0.370	−2.497***		0.277	0.126		1.184
	(0.223)		(0.316)	(0.289)		(2.137)	(0.553)		(1.121)
Slope	0.215***		0.073**	0.111**		0.041	0.186***		0.029
	(0.037)		(0.030)	(0.046)		(15.457)	(0.055)		(0.078)
Slope^2	−0.010***		−0.004**	−0.008***		−0.003	−0.008**		−0.002
	(0.002)		(0.002)	(0.003)		(0.738)	(0.003)		(0.006)
North	−0.671***		−0.087	−1.108***		−0.114	−0.026		−0.005
	(0.157)		(0.108)	(0.170)		(37.753)	(0.232)		(0.443)
South	−0.134**		−0.038	−0.148*		−0.001	−0.227*		−0.136
	(0.068)		(0.054)	(0.083)		(0.082)	(0.117)		(0.098)
West	−0.165		0.021	−0.347**		0.068	0.500		0.211
	(0.172)		(0.116)	(0.171)		(47.455)	(0.395)		(0.342)
Wet	−0.073***		−0.053**	0.093		−0.427	0.184***		−0.105
	(0.022)		(0.026)	(0.068)		(63.293)	(0.052)		(0.094)
Wet^2	−0.113***		−0.043**	0.218**		−0.543	−0.113***		−0.088*
	(0.015)		(0.021)	(0.100)		(88.017)	(0.018)		(0.046)
Heat	0.022		0.094**	0.459***		−0.509	0.299***		−0.076
	(0.033)		(0.048)	(0.143)		(120.490)	(0.067)		(0.092)
Heat^2	−0.033**		−0.060**	0.012		−0.194	−0.232***		−0.012
	(0.017)		(0.025)	(0.027)		(15.881)	(0.051)		(0.094)
Rough	0.037		0.016	−0.142***		0.022	0.323***		0.047
	(0.026)		(0.023)	(0.030)		(1.223)	(0.050)		(0.062)
Rough^2	−0.047**		−0.0001	−0.101***		0.025	−0.071***		0.001
	(0.018)		(0.012)	(0.027)		(4.564)	(0.021)		(0.031)
Depth	0.060		0.029	0.096		0.115	−0.009		−0.028
	(0.044)		(0.036)	(0.099)		(15.640)	(0.059)		(0.069)
Depth^2	0.006		0.018*	0.099***		0.036	−0.029*		−0.002
	(0.020)		(0.010)	(0.024)		(2.399)	(0.017)		(0.020)
AOCfBOUR		0.657***	0.405***		0.791***	0.651		0.496***	0.185
		(0.125)	(0.140)		(0.173)	(10.342)		(0.178)	(0.394)
AOCfVILL		2.107***	1.834***		2.284***	2.148		1.917***	1.435***
		(0.122)	(0.149)		(0.170)	(5.475)		(0.173)	(0.495)
AOCfPCRU		2.736***	2.089***		2.831***	2.431		2.770***	1.789***
		(0.128)	(0.174)		(0.175)	(33.653)		(0.190)	(0.546)
AOCfGCRU		3.786***	3.294***		3.734***	3.594		3.703***	2.523***
		(0.158)	(0.197)		(0.269)	(19.263)		(0.190)	(0.599)
Observations	1,476	1,476	1,478	940	940	941	536	536	537
R^2	0.237	0.523	0.667	0.299	0.521	0.667	0.406	0.547	0.741
Adjusted R^2	0.227	0.521	0.654	0.286	0.518	0.646	0.386	0.541	0.710

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Robust standard errors (sandwich), specifications (VI) for each sample contain Spatial Trend Surfaces of geographical coordinates.

Table 11: Regression Results from Ordered Probit Models on GIs Designations

	Pooled Sample			Cote de Beaune			Cote de Nuits		
	(ii)	(iii)	(iv)	(ii)	(iii)	(iv)	(ii)	(iii)	(iv)
Elevation	12.702***	15.539***	8.490***	19.353***	21.141***	2.875**	15.920***	14.162***	13.234***
	(0.793)	(0.931)	(1.048)	(1.107)	(1.199)	(1.232)	(2.756)	(2.842)	(0.251)
Elevation^2	−2.188***	−2.606***	−1.401***	−3.237***	−3.505***	−0.508**	−2.612***	−2.104***	−1.442***
	(0.129)	(0.152)	(0.173)	(0.177)	(0.192)	(0.200)	(0.486)	(0.496)	(0.097)
Slope	0.242***	0.186***	0.267***	0.155***	0.118***	0.284***	0.209***	0.180***	−0.198***
	(0.020)	(0.023)	(0.024)	(0.027)	(0.029)	(0.030)	(0.049)	(0.051)	(0.057)
Slope^2	−0.010***	−0.008***	−0.013***	−0.007***	−0.006***	−0.013***	−0.013***	−0.010***	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)
North	−1.275***	−1.232***	−1.555***	−1.800***	−1.720***	−2.191***	0.142	−0.023	−0.478**
	(0.095)	(0.100)	(0.109)	(0.118)	(0.123)	(0.131)	(0.203)	(0.196)	(0.210)
South	−0.009	−0.052	−0.068	−0.150***	−0.152**	0.092	0.069	0.020	−0.060
	(0.046)	(0.049)	(0.052)	(0.057)	(0.060)	(0.064)	(0.091)	(0.093)	(0.101)
West	−0.112	−0.241**	0.082	−0.367***	−0.380***	0.095	0.373	0.453	−0.039
	(0.107)	(0.114)	(0.125)	(0.119)	(0.124)	(0.137)	(0.333)	(0.340)	(0.363)
Wet	0.025			0.189***			0.018		
	(0.015)			(0.045)			(0.049)		
Wet^2	−0.023*			0.182***			−0.095***		
	(0.012)			(0.070)			(0.017)		
Heat	0.112***			0.319***			0.292***		
	(0.023)			(0.099)			(0.066)		
Heat^2	−0.039***			−0.033			−0.097*		
	(0.011)			(0.020)			(0.055)		
Rough	0.126***			−0.045**			0.452***		
	(0.016)			(0.022)			(0.034)		
Rough^2	−0.068***			−0.068***			−0.104***		
	(0.010)			(0.019)			(0.014)		
Depth	−0.070**			−0.166***			−0.025		
	(0.029)			(0.056)			(0.048)		
Depth^2	−0.047***			0.026			−0.072***		
	(0.008)			(0.017)			(0.012)		
BGOR|BOUR	16.235***	21.478***	36.739***	26.734***	28.987***	9.112***	20.973***	22.072***	24.357***
	(1.178)	(1.347)	(1.514)	(1.719)	(1.729)	(1.777)	(3.913)	(4.001)	(0.267)
BOUR|VILL	17.285***	22.681***	38.217***	27.985***	30.402***	10.990***	22.159***	23.145***	25.693***
	(1.182)	(1.352)	(1.521)	(1.728)	(1.739)	(1.790)	(3.918)	(4.005)	(0.277)
VILL|PCRU	19.197***	24.944***	40.999***	29.801***	32.577***	13.678***	24.837***	25.976***	29.907***
	(1.190)	(1.364)	(1.534)	(1.740)	(1.757)	(1.809)	(3.936)	(4.022)	(0.284)
PCRU|GCRU	20.529***	26.462***	42.701***	31.563***	34.560***	15.806***	25.700***	26.961***	31.097***
	(1.192)	(1.368)	(1.542)	(1.748)	(1.770)	(1.819)	(3.937)	(4.025)	(0.292)
Observations	2,978	2,978	2,978	1,846	1,846	1,846	1,132	1,132	1,132

Note: *p<0.1; **p<0.05; ***p<0.01
Specifications (iv) for each sample contains Spatial Trend Surfaces of geographical coordinates. The last eight rows report the threshold vector $\boldsymbol{\mu}$.

Table 12: Regression results from control function models of wineyard prices, with polynomial

	Pooled Sample		Cote de Beaune		Cote de Nuits	
	(VII)	(VIII)	(VII)	(VIII)	(VII)	(VIII)
Elevation	2.537**	1.329	6.708***	1.806	−4.393	−4.012
	(1.277)	(1.862)	(1.984)	(2.375)	(3.229)	(5.476)
Elevation^2	−0.485**	−0.236	−1.100***	−0.347	0.619	0.583
	(0.214)	(0.303)	(0.327)	(0.388)	(0.559)	(0.884)
Slope	0.117***	0.111***	0.055	0.105**	0.153***	0.036
	(0.033)	(0.030)	(0.040)	(0.042)	(0.051)	(0.055)
Slope^2	−0.006***	−0.006***	−0.004**	−0.006***	−0.007**	−0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
North	−0.218	−0.343**	−0.359*	−0.359*	0.002	−0.081
	(0.143)	(0.145)	(0.196)	(0.200)	(0.201)	(0.179)
South	−0.115**	−0.087*	−0.100	−0.020	−0.166*	−0.162*
	(0.051)	(0.049)	(0.064)	(0.062)	(0.090)	(0.088)
West	−0.161	−0.100	−0.225*	−0.112	0.572*	0.209
	(0.119)	(0.118)	(0.132)	(0.133)	(0.323)	(0.322)
Wet	−0.074***	−0.011	−0.009	−0.342***	0.185***	−0.066
	(0.016)	(0.021)	(0.058)	(0.109)	(0.044)	(0.053)
Wet^2	−0.100***	−0.082***	0.074	−0.496***	−0.096***	−0.049**
	(0.011)	(0.017)	(0.083)	(0.146)	(0.014)	(0.023)
Heat	−0.026	0.070*	0.208*	−0.473**	0.245***	−0.013
	(0.025)	(0.041)	(0.120)	(0.215)	(0.061)	(0.071)
Heat^2	0.002	−0.119***	0.042**	−0.234***	−0.225***	0.002
	(0.013)	(0.023)	(0.020)	(0.038)	(0.048)	(0.064)
Rough	−0.030	0.034	−0.139***	−0.014	0.215***	0.108**
	(0.020)	(0.023)	(0.024)	(0.030)	(0.049)	(0.054)
Rough^2	−0.021	−0.027**	−0.085***	−0.044*	−0.042**	−0.009
	(0.015)	(0.013)	(0.023)	(0.025)	(0.018)	(0.018)
Depth	0.133***	0.052	0.181**	0.114	0.033	−0.045
	(0.034)	(0.033)	(0.082)	(0.092)	(0.042)	(0.046)
Depth^2	0.033***	0.035***	0.071***	0.052**	−0.006	0.002
	(0.009)	(0.009)	(0.022)	(0.022)	(0.013)	(0.013)
AOCfBOUR	0.287*	0.253	0.435*	0.377	0.032	0.172
	(0.160)	(0.166)	(0.227)	(0.235)	(0.233)	(0.228)
AOCfVILL	1.554***	1.424***	1.824***	1.615***	0.808**	1.277***
	(0.225)	(0.235)	(0.310)	(0.315)	(0.337)	(0.342)
AOCfPCRU	1.754***	1.416***	1.973***	1.589***	0.899*	1.356***
	(0.316)	(0.326)	(0.421)	(0.433)	(0.488)	(0.495)
AOCfGCRU	2.458***	2.128***	2.741***	2.507***	1.454**	1.911***
	(0.415)	(0.433)	(0.597)	(0.621)	(0.573)	(0.586)
POC.R	0.312***					
	(0.083)					
POH.R		0.290***				
		(0.085)				
PSH.R				0.252**		0.131
				(0.112)		(0.117)
PSC.R			0.252**		0.465***	
			(0.104)		(0.114)	
Observations	1,475	1,474	938	938	536	536
R^2	0.569	0.640	0.577	0.638	0.651	0.716
Adjusted R^2	0.563	0.631	0.567	0.624	0.638	0.696

Note: *p<0.1; **p<0.05; ***p<0.01
Robust standard errors (sandwich), specifications (VIII) for each sample contain Spatial Trend Surfaces of geographical coordinates.

Table 13: Regression results from IV

	Pooled Sample		Cote de Beaune		Cote de Nuits	
	(IX)	(X)	(IX)	(X)	(IX)	(X)
Elevation	−1.969	−2.568	3.960	−3.096	−5.170	1.139
	(1.684)	(1.776)	(2.463)	(2.725)	(3.388)	(6.062)
$Elevation^2$	0.315	0.392	−0.622	0.414	0.791	−0.234
	(0.286)	(0.291)	(0.415)	(0.453)	(0.581)	(0.962)
Slope	0.038	0.073*	−0.003	0.130**	0.100	0.215**
	(0.042)	(0.039)	(0.050)	(0.053)	(0.102)	(0.085)
$Slope^2$	−0.002	−0.004*	−0.001	−0.007***	−0.005	−0.014***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.005)	(0.005)
North	0.313	0.088	0.068	−0.234	−0.121	−0.329
	(0.194)	(0.179)	(0.248)	(0.250)	(0.274)	(0.245)
South	−0.176***	−0.074	−0.040	0.132*	−0.208*	−0.215**
	(0.067)	(0.060)	(0.085)	(0.080)	(0.120)	(0.105)
West	−0.064	0.025	−0.106	0.035	0.589	−0.184
	(0.166)	(0.147)	(0.173)	(0.164)	(0.372)	(0.399)
Wet	−0.037*	−0.019	−0.066	−0.662***	0.281***	0.033
	(0.021)	(0.027)	(0.071)	(0.165)	(0.088)	(0.075)
Wet^2	−0.101***	−0.060***	0.034	−0.802***	−0.094***	−0.162***
	(0.015)	(0.023)	(0.098)	(0.214)	(0.022)	(0.038)
Heat	−0.002	0.122**	0.192	−0.911***	0.345***	−0.100
	(0.032)	(0.055)	(0.144)	(0.320)	(0.079)	(0.103)
$Heat^2$	−0.021	−0.102***	0.011	−0.271***	−0.252***	−0.172**
	(0.017)	(0.026)	(0.025)	(0.057)	(0.082)	(0.082)
Rough	0.002	−0.003	−0.195***	0.001	0.226**	0.168*
	(0.028)	(0.030)	(0.036)	(0.043)	(0.095)	(0.095)
$Rough^2$	−0.014	−0.020	−0.098***	−0.057*	−0.029	−0.018
	(0.020)	(0.016)	(0.029)	(0.030)	(0.027)	(0.023)
Depth	0.098**	0.048	0.186*	0.223*	−0.031	−0.007
	(0.044)	(0.042)	(0.101)	(0.114)	(0.062)	(0.060)
$Depth^2$	0.027	0.040***	0.114***	0.079***	−0.010	0.005
	(0.017)	(0.011)	(0.044)	(0.027)	(0.019)	(0.018)
IV.BOUR	2.167***	1.076**	0.363	−0.658	0.266	−1.417*
	(0.514)	(0.467)	(1.011)	(1.186)	(0.857)	(0.778)
IV.VILL	2.711***	2.389***	2.363***	1.153	0.713	−0.442
	(0.388)	(0.392)	(0.719)	(0.809)	(0.706)	(0.657)
IV.PCRU	3.863***	2.940***	2.917***	0.388	0.978	−3.153**
	(0.430)	(0.503)	(0.795)	(1.062)	(1.313)	(1.249)
IV.GCRU	4.748***	3.727***	3.063***	4.893***	4.458***	0.621
	(0.736)	(0.575)	(1.113)	(1.163)	(0.981)	(1.002)
Observations	1,441	1,441	911	911	530	530
R^2	0.269	0.452	0.332	0.463	0.440	0.616
Adjusted R^2	0.260	0.438	0.318	0.442	0.419	0.590

Note: *p<0.1; **p<0.05; ***p<0.01
(1), (4) and (7) for continous biophysical variables. Other columns for fixed effects, spatial functions for (3), (6) and (9)