Draft June 1, 2009.
Not for distribution.
Please do not cite.

Bayesian Inference in Large Hierarchical Spatial Data Structures

Cristina Salvioni[a], Donald J. Lacombe[b] and Garth Holloway[c]

Abstract

In processing discrete-choice spatially referenced data evolving from large hierarchical structures, investigators typically encounter three imposing difficulties. One difficulty is the evaluation of a large-scale determinant resulting from the Jacobian of the transformation from the error structure to the latent variables used to simulate the discrete choices. A second difficulty results from an imposing inversion of an $N \times N$ matrix resulting from the need to simulate conditionally truncated draws. And, third, where investigator choice is involved, there is the over-arching need to evaluate and identify sets of preferred modelling substructures. When N is large, these problems often impart formidable computational demands on investigators, raising scope for the search for alternatives to conventional implementation. This paper demonstrates how each of the difficulties can be mitigated by exploiting hierarchical substructures embedded within the sampling framework. We demonstrate the procedures within a rich set of data sampled from 13000 professional (large economic size) farm households in some 4,000 municipalities throughout Italy collected during the 2005 production year. The empirical investigation explores the factors affecting the adoption of organic production and certification, and the depiction of so-called 'neighbourhood effects.' An important phenomenon affecting adoption decision-making is the decision of a "similar" farm household, where similarity is measured in terms of "location"/municipality. The hierarchical spatial model estimated in this paper allows the investigator to analyze the differences in the neighbourhood effects among regions that can be caused by the different infrastructures and the different social capital endowments available within regions.

Keywords: Markov chain Monte Carlo methods, Gibbs sampling, data augmentation, hierarchical spatial probit.

[a] DASTA- Faculty of Economics, University of Chieti-Pescara, Italy, email: salvioni@unich.it; [b] Departments of Agricultural and Resource Economics and Economics, Research Associate Professor, Regional Research Institute, West Virginia University, email: lacombe@ohio.edu; [c] Department of Agricultural and Food Economics, University of Reading, United Kingdom, phone +44 118 378 6775, fax +44 118 975 64 67, email: garth.holloway@reading.ac.uk.

Bayesian Inference in Large Hierarchical Spatial Data Structures

In processing discrete-choice spatially referenced data evolving from large hierarchical structures, investigators typically encounter three imposing difficulties. One difficulty is the evaluation of a large-scale determinant resulting from the Jacobian of the transformation from the error structure to the latent variables used to simulate the discrete choices. A second difficulty results from an imposing inversion of an N×N matrix resulting from the need to simulate conditionally truncated draws. And, third, where investigator choice is involved, there is the over-arching need to evaluate and identify sets of preferred modelling substructures. When N is large, these problems often impart formidable computational demands on investigators, raising scope for the search for alternatives to conventional implementation. This paper demonstrates how each of the difficulties can be mitigated by exploiting hierarchical substructures embedded within the sampling framework. We demonstrate the procedures within a rich set of data from a sample of 13000 professional (large economic size) farm households in some 4,000 municipalities throughout Italy collected during the 2005 production year. The empirical investigation explores the factors affecting the adoption of organic production and certification, and compares and contrasts alternative representations of the so-called 'neighbourhood effects.' An important phenomenon affecting adoption decision-making is the decision of a "similar" farm household, where similarity is measured in terms of "location"/municipality. The hierarchical spatial model estimated in this paper allows the investigator to analyze the differences in the neighbourhood effects among regions that can be caused by the different infrastructures and the different social capital endowment available within regions.

Section two reviews the origins of hierarchical modeling in Bayesian inference; explores its use in agriculture and related fields; and reviews prior use in spatial econometric settings. Section three presents a background to the Italian empirical setting. Section four presents notation, section five outlines the theoretical framework and section six presents findings of our preliminary application to the Italian data. Conclusions are offered in section seven.

Hierarchical Models, Hierarchical Modeling and Their Origins

The hierarchical, normal, linear model has been an important component for applied Bayesian statisticians since its development by Lindley and Smith (Lindley and Smith, 1972). In short, the hierarchical, normal, linear model is capable of drawing nuanced empirical findings when the data has a hierarchical or 'stacked' representation. For example, pupils are assigned to classrooms, which are contained within schools, which in turn are contained within districts. In agriculture, we may be interested in analyzing how crop yields are related to several independent factors (e.g. soil conditions, rainfall, fertilization techniques) and our data may have a hierarchical structure: different crops contained within a larger plot of land. In each of these cases, hierarchical modeling can be justified based on the organization of the sample data. As Robert notes (Robert, 2007, p. 458):

> "Additional justifications of hierarchical Bayes modeling stem from real life, since there are settings in medicine, biology, animal breeding, economics, and so on, where the population of interest can be perceived as a sub-population of a meta-population, or even as a subpopulation of a subpopulation of a global population."

From a statistical standpoint, the hierarchical approach allows one to draw these relationships together in a probabilistically consistent, but random, fashion. It allows one to make statements about the performance of subunits or units or the sample as a whole, recognizing the inter-linkages and the intra-linkages across sample units.

There have been substantial advances in Bayesian analysis since the original publication of the Lindley and Smith paper, in particular, the advent of Markov Chain Monte Carlo techniques. These techniques allow more complicated models that are not analytically tractable to be estimated. Hierarchical models have also been helped by this computational advance, with contributions from Chib and Greenberg (2008) and Draper (2008) providing tutorials on the use of these computational techniques to the analysis of hierarchical data.

Agricultural and Related Papers

Bayesian hierarchical modeling techniques have a natural home in agricultural economics. Holloway and Paris (2002) utilize a Bayesian procedure to estimate traditional von Liebig production function models and frontier specifications across cropping sites. For example, their Gibbs sampling algorithm generates estimates of the model parameters of a von Liebig crop response model. Additionally, the Bayesian paradigm allows for extensions to the traditional model that would be computationally intractable using alternative methods. A host of additional applications reside in the literature on biological modeling of phenomena relevant to agriculture, the natural resources and to ecology, and many of these are reviewed in a survey paper by Holloway, Lacombe and LeSage (2007). Collectively, these works evidence the importance of accommodating hierarchical substructures in the data generating environment and attest to some of the imposing difficulties in estimation of complex hierarchical models.

Spatial Papers

Extensions to the basic Bayesian hierarchical linear model have been developed to address the specialized needs of various data generating environments. In particular, Bayesian hierarchical models that address the spatial dimensions of datasets have received more attention due to the fact that the aforementioned *MCMC* techniques allow the estimation of previously intractable spatial models. Certain datasets, such as agricultural yields or educational attainment by students, can have hierarchical aspects (e.g. differing crops within a larger plot of farmland or students within classrooms that are also embedded in schools) as well as spatial characteristics. For example, there can be spillover effects from one plot of land to another (e.g. fertilizer applied to one plot can spill over into adjoining plots) or there can be spillover effects in the form of 'peer effects' when discussing student performance in schools. The main point is that not only can hierarchical aspects of the problem be specified and estimated, but the additional complication of spatially correlated data can also be accounted for in a relatively straightforward fashion using Bayesian techniques. A significant overview of the literature on spatial econometrics and statistics and an introduction to some of the bio-economic hierarchical models surveyed by Holloway, Lacombe and LeSage (2007) is Banerjee, Carlin, and Gelfand (2004).

Organic farming in Italy

Since the inception of an EU common set of standards for the organic production of crops in 1991 (EU Regulation 2092/91), organic farming has experienced a rapid growth in response to a favourable market and a predominantly favourable policy environment. The financial support received first under the measures accompanying the reform of the Common Agricultural Policy in 1992 (EU Regulation 2078/92) and, since 1999, under the agro-environmental programs that are granted under the Rural Development regulation of Agenda 2000, made the number of farms turning from conventional to organic increase rapidly and promoted Italy into a leading position in terms of organic production in the world and in the EU. The adoption trend reached a maximum in 2001, both in terms of holdings and hectares. More recently the organic sector has entered a period of structural adjustment driven by the reduction of the financial support provided under the rural development programs and by the increase in the standards of quality required by the entry of organic produce into the so-called ‘Big Distribution Channels’ (for example, supermarkets and value-added chains).

The diffusion of organic farming has not followed a uniform pattern, however. The Central and Southern Regions present a higher rate of adoption. This is perhaps due to technical and political factors. First, the adoption of organic production practices is relatively simpler with Mediterranean crops, namely olives and citrus, because there is proximity of traditional production practices to organic ones and lower yield loss in comparison to other crops. This explains the rapid uptake of organic practices in central and southern regions in which the Mediterranean crops are largely diffused. In addition, the diffusion of organic is influenced by the financial support offered by the regions under the Rural Development programs. Organic schemes are defined at the Regional level and they differ either in terms of eligibility rules (e. g. in some regions only farms of a minimum size or totally converted to organic are eligible for the payment) and are also defined in terms of the level of the payments per crop. The lower organic adoption rates observed in the northern regions are also due to the fact that in the northern regions turning to organic is not the only option to get a payment for the adoption of a more environmentally sustainable production system. Northern regions have in fact activated two measures, namely granting payments to organic farming and granting payments for the adoption of low impact technologies (e.g. integrated pest management) while the southern regions have not activated this latter measure (Salvioni, 2009).

The question arising in this complex geopolitical labyrinth is precisely how adoption by farmers within one municipality affect and are affected by the adoption decisions of those in remaining municipalities when we allow for their involvement by way of treatment of municipalities via hierarchical data substructures. To our knowledge, this important question remains hitherto unanswered in the Italian organic adoption setting. Thus, considerable scope exists for nuanced empirical enquiry and the development of robust methodology for making inferences about the neighbourhood effects in organic adoption in Italian agriculture. This motivates development of a statistical methodology for which some notation will prove useful.

Notation

By way of notation, let $\boldsymbol{\theta}$ denote a vector of parameters of interest, $\pi(\boldsymbol{\theta})$ the prior probability density function (pdf) for $\boldsymbol{\theta}$ and $\pi(\boldsymbol{\theta}|\mathbf{y})$ the posterior pdf for $\boldsymbol{\theta}$, where $\mathbf{y} \equiv (y_1, y_2, .., y_N)'$ denotes data. Frequently we reference the data generating model $f(\mathbf{y}|\boldsymbol{\theta})$, which is the likelihood function when viewed as a function of $\boldsymbol{\theta}$ and, sometimes, make use of variants of the $f(\cdot|\cdot)$ notation to reference particular probability density functions. The <u>four</u> pdfs that we reference (see the appendix) are, respectively, the univariate-normal pdf, the truncated-normal pdf, the multivariate-normal pdf, and the inverted-gamma pdf. Often, we reference just the variable part of the density by noting its proportionality (integrating constant excluded) with reference to a true pdf (integrating constant included) by using the symbol '$\propto$.' Throughout, we use i = 1, 2, .., N to denote the sample. Finally, we use lower-case symbols to denote scalars and lower-case emboldened symbols to denote vectors and matrices. Parameters are denoted by Greek symbols and observed and missing data are denoted by Roman symbols.

Statistical Model

The ideas involved in linking covariate information to organic adoption in Italian agriculture and a spatial externality can be formalized within a conventional setting by considering adoption to be the observed outcome of a random process in which regional constituency, 'spatial contiguity' and other exogenous factors combine to affect adoption responses. Formalizing a little, a conventional approach is concerned with the relationship

(1) $$z_i = \rho\, \mathbf{w}_{-i}\, \mathbf{z}_{-i} + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i,$$

where i = 1, 2, .., N denotes a regional municipality; z_i denotes a latent propensity to adopt; $\mathbf{x}_i \equiv (x_{i1}, x_{i2}, .., x_{iK})'$ denotes a K-vector of covariates conditioning the latent response; $\boldsymbol{\beta} \equiv (\beta_1, \beta_2, .., \beta_K)'$ denotes the corresponding K-vector of response coefficients; $\mathbf{w}_i$ denotes an N-vector of binary values depicting contiguity between districts; ρ reflects the magnitude of correlation between districts; $\mathbf{z}_{-i}$ denotes the vector of latent responses with the i^{th} response excluded; and ε_i denotes a standard-normal random variable. In the remainder of the paper we maintain the assumption that u_i is normally distributed with zero mean and unit variance, and some additional notation will prove useful. Throughout, we use the convention that $f^a(b|c,d,..,e)$ denotes a type-'a' *probability distribution function* (pdf) for random variable 'b' conditioned by the values of parameters 'c, d, .., and e.' Hence, u_i has distribution $f^N(u_i|0,1)$. The unit-variance restriction is the standard assumption required for identification in the probit model (see, for example, Greene, 2003, p. 669). The normality assumption is a useful approximation which, in the absence of other, motivating evidence, seems reasonable to apply. We observe data $\{\mathbf{x}_i, \mathbf{w}_i, y_i\}_{i=1}^{N}$ where $y_i = 1$ denotes the adoption of organic practices of a farm unit within district i; we observe $y_i = 0$, otherwise and make inferences about $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}', \rho)'$. Stacking observations in (1) leads to

(2) $$\mathbf{z} = \rho\mathbf{W}\mathbf{z} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

where $\mathbf{z} \equiv (z_1, z_2, .., z_N)'$ denotes an N-vector of the latent responses; $\mathbf{X} \equiv (\mathbf{x}_1, \mathbf{x}_2, .., \mathbf{x}_N)'$, denotes an N×K matrix of observations on the covariates; $\mathbf{W}$ denotes the N-dimensional, square, symmetric matrix of binary contiguity indicators; and $\boldsymbol{\varepsilon} \equiv (\varepsilon_1, \varepsilon_2, .., \varepsilon_N)'$ denotes an N-vector of disturbances with distribution $f^N(\boldsymbol{\varepsilon}|\mathbf{0}_N,\mathbf{I}_N)$, where $\mathbf{0}_N$ is the length-N null vector and $\mathbf{I}_N$ is the N-dimensional identity matrix.

Bayesian estimation of this formulation is complicated by the presence of correlation across observations, which is jointly manifested by the correlation parameter 'ρ' and the design of the spatial contiguity matrix '**W**.' The conventional (non-spatial) probit model is nested as a special case of (2) under the condition $\rho = 0$. Albert and Chib (1993) present an algorithm for posterior inference for the conventional probit model and Le Sage (2000) extends their work to incorporate the spatial externality. We emphasize the two-part nature of the spatial externality namely the magnitude of the correlation, manifested by ρ, and the design of the spatial contiguity, **W**. A heritage in applied adoption studies in agricultural and development economics, many of which are relevant in the present context, constructs **W** by setting elements $w_{ij} = 1$ if observations 'i' and 'j' are 'neighbours' and $w_{ij} = 0$ otherwise and proceeds conditionally to estimate ρ. Case (1992) provides an influential technological adoption example and many others exist. The point that needs emphasis here is that usually, though not always, the definition of the 'neighbourhood' and thus the 'span' of the contiguity regions selected by the investigator are *arbitrary*.

Algorithms for comparing the evidence in favour of the competing formulations are presented in Chib (1995) and Chib and Jeliazkov (2001) and an introduction to the ideas underlying the '*Markov Chain Monte Carlo* (MCMC)' theory used there is presented in Casella and George (1992) and in Chib and Greenberg (1995). These ideas lead to an efficient algorithm for estimation and models comparisons, which has as its fulcrum, the fully conditional posterior distributions

(3) $\boldsymbol{\beta}|\mathbf{z},\rho \sim f^{N}(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}},\mathbf{C}_{\boldsymbol{\beta}})$,

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{A}\mathbf{z}+\mathbf{C}_{\boldsymbol{\beta}o}^{-1}\hat{\boldsymbol{\beta}}_o)$ and $\mathbf{A} = \mathbf{I}_N-\rho\mathbf{W}$ and $\mathbf{C}_{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X} + \mathbf{C}_{\boldsymbol{\beta}o}^{-1})^{-1}$;

(4) $z_i|\boldsymbol{\beta},\rho \sim f^{TN}(z_i|\hat{\mathbf{z}}_i,C_{zi},\mathbf{y})$, i = 1, 2, .., N,

where $\hat{\mathbf{z}}_i = \mathbf{A}^{-1}\mathbf{x}_i'\boldsymbol{\beta} - V_{ii}^{-1}\mathbf{V}_{i\text{-}i}(\mathbf{z}_{\text{-}i}-\mathbf{X}_{\text{-}i}\boldsymbol{\beta})$; $\mathbf{V} = \mathbf{A}'\mathbf{A}$; V_{ii} denotes the scalar appearing in the i^{th} row and column of **V**; $\mathbf{V}_{i\text{-}i}$ denotes the (N-1)-dimensional row vector obtained by deleting the i^{th} column from the i^{th} row of **V**; and the variance of the i^{th} latent response is $C_{zi} = V_{ii}^{-1}$; and, third,

(5) $\rho|\boldsymbol{\beta},\mathbf{z} \sim |\mathbf{A}| \exp\{-.5(\mathbf{A}\mathbf{z}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{A}\mathbf{z}-\mathbf{X}\boldsymbol{\beta})\}\times\exp\{-.5(\rho-\hat{\rho}_o)'C_{\rho o}^{-1}(\rho-\hat{\rho}_o)'\}\equiv \kappa(\rho|\boldsymbol{\beta},\mathbf{z})$,

which has an unknown integrating constant. The corresponding Metropolis-Hastings step involves drawing a proposal value, $\tau \sim f^{N}(\tau|\rho,\zeta)$, and accepting the draw with probability

$$(6) \qquad \alpha(\rho,\tau) \equiv \min\left\{\frac{\kappa(\tau|\boldsymbol{\beta},\mathbf{z})}{\kappa(\rho|\boldsymbol{\beta},\mathbf{z})},1\right\},$$

and adjusting endogenously the value of the variance parameter, ζ, in order to target an acceptance rate of 50% of the total draws. Experiments with simulated data suggest that an acceptance rate of about 50% is highly satisfactory. Hence, in summary, given arbitrary starting values, $\mathbf{z} = \mathbf{z}^{(0)}$, efficient estimates of the *Spatial Probit Model* are obtained by iterating the algorithm

> $\mathbf{A_1}$: Draw $\boldsymbol{\beta}^{(g)}$ from (3). Draw $\mathbf{z}^{(g)}$ from (4). Draw $\rho^{(g)}$ from $\tau \sim f^{N}(\tau|\rho,\zeta)$ and accept the draw with probability (6).

Although a conventional analysis proceeding along these lines is meritorious, it is computationally foreboding. There are two reasons. Both reasons relate to the need to draw conditionally dependent latent random variables contained in the vector $\mathbf{z}$. First, one must evaluate the N×N determinant $|\mathbf{A}|$ resulting from the Jacobian of the transformation from the sampling error vector $\mathbf{u}$ to the latent response vector $\mathbf{z}$. Second, in order to obtain the conditioning moments of the latter one must invert the N×N matrix

$\mathbf{V} \equiv \mathbf{A}'\mathbf{A}$. For matrices in the dimensional order of N in excess of, say, a few hundred, these manipulations often preclude efficient estimation in view of the need to run the Markov chain for a sufficiently high number of iterates before convergence is attained. This problem is even more troublesome in the context of the Chib (1995) and Chib and Jeliazkov (2001) approaches to marginal likelihood estimation which, while conceptually appealing in view of their conceptual simplicity, require the investigator to repeat the basic algorithm a number of times before precise estimates of log marginal likelihoods avail themselves.

These computational demands motivate the need for alternative statistical structures that can be suitably applied to measure the relationship between the available covariate information and the observed organic responses.

The view that farms within given municipalities are likely closely correlated in actions and are correlated indirectly with those in other municipalities endows one a basic logic for motivating the hierarchical approach to estimating the neighbourhood effects. In this context, one may view the spatial substructures as consisting of contiguous units, such as those depicted in figure 1. (Insert figure 1 about here.) In this representation, each municipality is permitted to possess a different 'neighbourhood effect' which is related to the neighbourhood effects of the remaining municipalities and the farms residing within them by an over-arching set of parametric relationships consisting of a 'sample-mean', μ, and a 'municipality based variance parameter,' ω. Obviously, with such a relationship evident within the data, estimates of μ and ω are important for policy purposes. Presently, however, interest resides in their formal estimation. This motivates a hierarchically-dependent moderation of the previous model and resulting algorithm presented in (3)-(6), above, and in algorithm $\mathbf{A_1}$. In particular, given municipalities j = 1, 2, .., N, for each subcomponent, j, we have

(7) $$z_{ij} = \rho_j \mathbf{w}_{-ij} \mathbf{z}_{-ij} + \mathbf{x}_{ij}'\boldsymbol{\beta} + \varepsilon_{ij},$$

in place of (1);

(8) $$\mathbf{z}_j = \rho_j \mathbf{W}_j \mathbf{z}_j + \mathbf{X}_j \boldsymbol{\beta} + \boldsymbol{\varepsilon}_j,$$

in place of (2); and

(9) $\mathbf{z} = \mathbf{A}^{-1}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$;

depicting the redefined statistical model. It follows that with the new definintions

(10) $\boldsymbol{\beta}|\mathbf{z},\rho \sim f^{N}(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}},\mathbf{C}_{\boldsymbol{\beta}})$,

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{A}\mathbf{z}+\mathbf{C}_{\boldsymbol{\beta}o}^{-1}\hat{\boldsymbol{\beta}}_o)$ and $\mathbf{A} = \mathbf{I}_N-\rho\mathbf{W}$ and $\mathbf{C}_{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X} + \mathbf{C}_{\boldsymbol{\beta}o}^{-1})^{-1}$ is virtually unchanged; the draws for each of the latent subcomponents of the random vector **z** are

(11) $z_{ij}|\boldsymbol{\beta},\rho_j \sim f^{TN}(z_{ij}|\hat{\mathbf{z}}_{ij},C_{zij},\mathbf{y})$, i = 1, 2, .., N,

where $\hat{\mathbf{z}}_{ij} = \mathbf{A}^{-1}\mathbf{x}_{ij}'\boldsymbol{\beta} - V_{ijij}^{-1}\mathbf{V}_{i-i}(\mathbf{z}_{-i}-\mathbf{X}_{-i}\boldsymbol{\beta})$; $\mathbf{V}_j = \mathbf{A}_j'\mathbf{A}_j$; V_{ijij} denotes the scalar appearing in the ith row and column of $\mathbf{V}_j$; $\mathbf{V}_{ij\text{-}ij}$ denotes the (N-1)-dimensional row vector obtained by deleting the ith column from the ith row of $\mathbf{V}_j$; and the variance of the ith latent response is $C_{zi} = V_{ijij}^{-1}$;

(12) $\mu|\omega,\mathbf{z},\rho \sim f^{N}(\mu|\hat{\mu},C_{\mu})$;

and

(13) $\omega|\mu,\mathbf{z},\rho \sim f^{ig}(\omega|s^2_{\omega},v_{\omega})$,

Finally, the draws for the N subunit-specific neighbourhood correlation parameters follow

(14) $\rho_j|\boldsymbol{\beta},\mathbf{z}_j \sim |\mathbf{A}_j| \exp\{-.5(\mathbf{A}j\mathbf{z}j\text{-}\mathbf{X}j\boldsymbol{\beta})'(\mathbf{A}j\mathbf{z}j\text{-}\mathbf{X}j\boldsymbol{\beta})\}$

$$\times\exp\{-.5(\rho j\text{-}\hat{\rho}_{jo})'C_{\rho jo}^{-1}(\rho j\text{-}\hat{\rho}_{jo})'\}\equiv \kappa(\rho_j|\boldsymbol{\beta},\mathbf{z}_j);$$

in terms of which one draws a proposal random variate $\tau_j \sim f^{N}(\tau_j|\rho_j,\zeta)$, and accepts the draw with probability

(15) $$\alpha(\rho_j,\tau_j) \equiv \min\left\{\frac{\kappa(\tau_j\,|\,\boldsymbol{\beta},\mathbf{z}_j)}{\kappa(\rho_j\,|\,\boldsymbol{\beta},\mathbf{z}_j)},1\right\}.$$

In summary, the modified algorithm for the hierarchical neighbourhood effects model shares many similarities with the conventional algorithm and is formalized, succinctly in the terms:

> $\mathbf{A}_2$: Draw $\boldsymbol{\beta}^{(g)}$ from (10). Draw $\mathbf{z}_j^{(g)}$, j = 1, 2, .., N, from (11). Draw μ from (12). Draw ω from (13). Draw $\rho_j^{(g)}$ from $\tau_j \sim f^N(\tau_j|\rho_j,\zeta)$ and accept the draw with probability (15).

Importantly, within these developments, especially with reference to (11) and (14), there is the need to neither invert nor evaluate determinants of matrices of order N×N.

Although the prior information concerning the alternative specifications is relatively diffuse, we investigate the sample using the proper prior $\pi(\boldsymbol{\theta}) \equiv f^N(\beta|\hat{\beta}_o,C_{\beta o}) \times f^N(\mu|\hat{\mu}_o,C_{\mu o}) \times f^{ig}(\omega|s^2_{\omega o},v_{\omega o})$,which is the product of a multivariate-normal distribution for the response coefficients, a normal distribution for the spatial correlation parameter overarching the hierarchy; and an inverted-gamma for the spatial variance parameter. IN this setting the reader should note that there is no longer the need to place prior distributions on the correlation terms within each municipality boundary which may, in some circumstances, be deemed advantageous. In view of the diffuse information about $\boldsymbol{\beta}$ and μ, we implement the estimation using prior parameter values $\beta_o = \mathbf{0}_K$, $C_{\beta o} = I_K \times 100$, $\mu_o = 0$, $C_{\mu o} = 100$, $s^2_{\omega o} = 10$ and $v_{\omega o} = 1$. Given this weakly informative prior, inference is conducted with respect to the joint posterior distribution for the parameters, which is proportional to the likelihood for the data and the prior, namely $\pi(\boldsymbol{\theta}|\mathbf{y}) \propto f(\boldsymbol{\theta}|\mathbf{y}) \times \pi(\boldsymbol{\theta}|\mathbf{y})$.

In summary, robust estimates of municipality-specific neighbourhood effects and their ancillary impacts on organic adoption in spatially distant neighbourhoods evolves conveniently and efficiently from embedding the various neighbourhood effects neatly and succinctly in a logically coherent hierarchical framework.

Data

This study uses data for year 2005 gathered by the Italian Farm Accountancy Data Network (FADN). The observation units for the sample are commercial farms – farms having a sizes in excess of 4 European Economic Size Units (ESU). The sample is stratified by region, farm size and type of farming, and it is representative at the regional level. An individual weight is applied to each farm in the sample in order to reflect their relative contributions to state agriculture.

The implicit assumption is that adoption of organic techniques coincides with adoption of process certification. This hypothesis may appear restrictive, as it does not take into account that there may be producers who use these techniques without becoming certified. Actually, however, so that a farm product can be recognised as organic by the market as well by law, it must held a certification of organic process and/or product to attest to the use of production methods defined and disciplined in Europe by Regulation EEC 2092/91 and in Italy by DM 220/95. The coincidence between adoption and certification appears even more justified in a sample limited to professional farms, like the one used in this work.

In choosing representative covariates, we make the following observations. First, we note that financial/management characteristics (income, costs, credit, debt, etc.) have not been included among the possible covariates, considering that these values are the results of adoption, rather than factors affecting adoption choices. Their exclusion from the group of explicative variables is thus justified by the desire to avoid endogeneity that could motivate departures from this initial exploratory analysis. Given this restriction, we consider the following as possibly important determinants of the adoption decision within farms.

Farmer's age

According to the literature age is one of the key variables in technology adoption (Feder et al., 1985). Younger farmers are usually more likely to invest in new technologies because of their longer planning horizons over which expected benefits can accrue (Lapar and Pardey, 1999). In addition, in the case of organic farming older farmers are less likely to be aware of the economic possibilities of these new techniques given courses started only recently to be taught in schools and universities. The sign of the coefficient of the variable 'age' is expected to be negative.

Specialization of production

The profitability of adopting organic farming differs among products. In the case of olives and citrus, for example, organic farming techniques are "closer" to the conventional ones. As a consequence conversion does not imply large yield or cost penalties. This is not the case with vegetables and wine especially due to the lower efficiency of organic pest controlling methods . As a consequence a positive coefficient is expected for olives and citrus, while a negative one is expected in case of horticultural crops. A negative sign is expected (at least in Italy and in the other EU member states) even for the coefficient livestock, this is not related to considerations on the relative profitability of different production, rather the negative sign is associated to the only recent definition of the legal framework for organic certification in this animal sector. In previous work on Italian adoption of organic farming the coefficient of specialisation in production of fruit, citrus and olives are positive,; this is partly due to the widespread use of organic farming in Southern regions, where these crops are most widely cultivated, but also the greater adaptability, hence profitability, of these crops, especially olives, to organic farming. Specialisation in milk production has a negative influence on the probability of adoption.

Other factors

The adoption of organic farming is expected to be higher among those farmers which are particularly concerned/sensitive about the conservation of natural environment, hence about the production of positive environmental externalities. The participation of the farm in agri-environmental (landscape and biodiversity conservation, use of low impact technologies, extensification) schemes can be thought of as a proxy of environmental concern/sensitivity of the farmer. The expected sign of the coefficient of these indicators is positive. In previous work we found the probability of adoption increases in correspondence of participation in support programs in favour of low-impact-technologies

In addition, organic farming can be thought of as a component of the "post-modern farming system" based on re-localization of production, lower dependence on off-farm inputs and closer relationships between production and demand (Marden, 2008; Wilson, 2008). Farmers already involved in short supply chains, i.e. direct selling, and agritourism may be more exposed to the growing demand for local, fresh and environmentally friendly products, hence, they may be keener on adopting organic farming to satisfy the requirements requested by their clients. In previous work we find a positive coefficient in the case of direct sales while a negative for agriturism.

The effect of the participation to the support program in favour of organic farming is expected to increase the probability of adoption. In previous work this is the variable that most strongly increases adoption of organic farming.

While several empirical studies support the hypothesis that land ownership increases the probability of adoption, the results are not unanimous and the issue has been widely debated (e.g., Feder et al., 1985). The apparent inconsistencies in the empirical results are due to the nature of the innovation. Land ownership is likely to increase the probability of adoption if the innovation requires investments tied to the land. This is the case of organic farming, hence a positive sign is expected.

The effect of location of the farm in upland areas on organic adoption is ambiguous. This ambiguity is partly due to the smaller size of these farms. Apart from this farms located in upland areas may be more interested in organic farming as a strategy used to contrast the lower incomes connected to the adverse natural and economic environment (dispersion, distance from market, high transport costs, etc.).

In addition, location in environmentally protected areas is considered. The reader should note that the sample shows more organic farms than conventional farms in mountain areas and environmentally less-favoured areas.

Gender

Finally, although the literature on gender-based rationales for organic adoption is growing, it remains largely unclear as to expectations concerning the sign of the impact of this specific covariate. Nevertheless we include gender as a final explanatory variable in the analysis.

Empirical Results

Tables 1 through 3 present the results of the estimation of a ‘preferred’ parsimonious specification consisting of the covariates ‘tenure (= 1 if the farm was managed by an owner operator, = 0, otherwise);’ ‘environmentally protected (= 1 if the farm is located in an environmentally protected area, = 0 , otherwise);’ ‘direct selling (= 1 if the farm is involved in direct selling, = 0 , otherwise);’ ‘cereals and oil seeds (= 1 if the farm is specialized in cereals and oil seeds production, = 0 , otherwise);’ ‘horticultural (= 1 if the farm is specialized in horticultural production, = 0 , otherwise);’ ‘wine (= 1 if the farm is specialized in cereals and oil seeds production, = 0 , otherwise);’ ‘fruits and citrus (= 1 if the farm is specialized in fruits and citrus production, = 0 , otherwise);’ ‘other permanent (= 1 if the farm is specialized in other permanent crops production, = 0 , otherwise);’ ‘olives (= 1 if the farm is specialized in olives production, = 0 , otherwise);’ ‘livestock (= 1 if the farm is specialized in livestock production, = 0 , otherwise);’ ‘milk (= 1 if the farm is specialized in milk production, = 0 , otherwise);’ ‘gender (= 1 if the farm is operated by a female, = 0 , otherwise);’ ‘age (= the operators age in years);’ and, finally, three variables depicting the type of topography in the farm and surrounds, namely ‘upland (= 1 if the farm is located in an upland area, = 0 , otherwise);’ ‘hills (= 1 if the farm is located in a hill-land area, = 0 , otherwise);’ and ‘flatland (= 1 if the farm is located in a flatland area, = 0 , otherwise).’

Table 1 reports the posterior estimates of the correlation parameters across the first ten provincial sectors within the sample. These estimates are derived under the restriction that each is bounded on the interval [-1,+1]. In this context two observations are noteworthy. First, none of the province-specific inferences about correlation appears to be ‘significantly different from zero’ with significant portions of the posterior highest posterior density intervals residing on either side of the benchmark. Second, plots of posterior densities reveal some fairly diverse locations and scales of the means, medians and modes of each neighbourhood effect. For example, the plots in figure 1 are taken from, respectively, from left-most location to right-most location, the tenth, the second and the ninth, provincial spatial externality. While these three densities are only representative, their differences raise interesting questions about the types of restrictions researchers impose in complex data-structure models such as the present one. In addition, their differences raise questions about the nature and scope of the inferences derived from restricting the neighbourhood effects to be common across the entire sample.

Table 2 reports highest posterior density intervals corresponding to the posterior densities of each of the covariate coefficients. Briefly, they paint a fairly detailed picture of a ‘representative’ adopter. In particular, operators who (a) own their own farm, (b) direct sell, (c) are specialized in horticulture, or in livestock production, and are (d) located in upland or hillside properties tend to have the greatest propensity to adopt.

Table 3 reports model estimates of the likelihood and marginal likelihood statistics and reveal that inferences are somewhat largely dependent, at least for the 1300-unit sample size, on the weight of the prior evidence about the model parameters.

It remains of course, to determine how these findings alter in the face of implementation on the entire 13,000-observation sample with the full 103 provincial districts. The preliminary results reported here raise scope for this empirical enquiry and beckon robust inferences, possibly in the extended three-layer hierarchical setup, about the nature and scope of region-specific organic adoption in Italian agriculture.

Conclusions

The hierarchical methodology presented in this preliminary application to out Italian sample shows much promise. First, it facilitates exact inference within a sample of size around 1300 observations and a sample of this size is a major impediments for non-hierarchical treatments on most standard computing platforms. Second, the hierarchical setup outlined can be used to simulate policy responses emanating from one provincial location to another in a manner that seems more logical and credible than one usually encountered in the conventional setup. Third, the hierarchical setting sheds important insights on the nature and scope of organic adoption in Italian agriculture, which we intend to pursue in further rounds of this draft document. Finally, the two-layer hierarchy pursued in this paper extends readily to more complex geographical structures including extensions to a three-layer setting; and future effort will inform whether the inferences based on this current restricted version of our target model will change dramatically.

Appendices

The four pdfs that we apply are the univariate-Normal pdf,

(A.1) $f^{\mathbf{N}}(x|\mu, \sigma) \equiv (2\pi)^{-1/2}\, \sigma^{-1} \exp\{ -½\, \sigma^{-2}\, (x-\mu)'\, (x-\mu) \}$,

$-\infty < x < +\infty$, $-\infty < \mu < +\infty$, $0 < \sigma < +\infty$; the truncated-Normal pdf,

(A.2) $f^{\mathbf{tN}}(x|\mu, \sigma, y) \equiv (2\pi)^{-1/2}\, \sigma^{-1} \exp\{ -½\, \sigma^{-2}\, (x-\mu)'\, (x-\mu) \}[1-\Phi((y-\mu)/\sigma)]^{-1}$,

$y < x < +\infty$, $-\infty < \mu < +\infty$, $0 < \sigma < +\infty$, $\Phi(\cdot)$ denotes the cdf corresponding to the standard normal pdf; the multivariate-normal pdf

(A.3) $f^{\mathbf{MN}}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \equiv (2\pi)^{-m/2}\, |\Sigma|^{-1/2} \exp\{ -½\, (\mathbf{x}-\boldsymbol{\mu})'\, \Sigma^{-1}\, (\mathbf{x}-\boldsymbol{\mu}) \}$,

$\mathbf{x} \equiv (x_1, x_2, .., x_m)'$, $\boldsymbol{\mu} \equiv (\mu_1, \mu_2, .., \mu_m)'$, $\boldsymbol{\Sigma}$ is an m×m positive definite symmetric (pds) matrix $-\infty < x_i < +\infty$, $-\infty < \mu_i < +\infty$, i = 1, 2, m; the inverted-gamma pdf, (A.4) $f^{IG}(x|\nu, s^2) \equiv (2/\Gamma(\nu/2)) (\nu s^2/2)^{\nu/2} (1/\sigma^{\nu+1}) \exp\{-\nu s^2/2\sigma^2\}$,

$\Gamma(\cdot)$ denotes the gamma function (see, for example, Mood, Graybill and Boes, pp. 534-5),

$-\infty < x < +\infty$, $0 < \nu < +\infty$, $0 < s^2 < +\infty$; the inverted-Wishart pdf

ϑ is a positive integer and for all j = 1, 2, .., k, $0 \leq \wp_j \leq 1$ and $x_j = 0$ or 1.

References

Albert, J. and S. Chib. "Bayesian Analysis of Binary and Polychotomous Data." Journal of the American Statistical Association 88(3)(1993):669-79.

Casella, G. and E. George. "Explaining the Gibbs Sampler." American Statistician 46(2)(1992):167-74.

Banerjee, S., Bradley P. Carlin, and Alan E. Gelfand. "Hierarchical Modeling and Analysis for Spatial Data" Boca Raton: CRC Press. 2004.

Chib, S. "Bayes Inference in the Tobit Censored Regression Model." Journal of Econometrics 52(1) (1992):79-99.

. "Marginal Likelihood From the Gibbs Output." Journal of the American Statistical Association 90(4) (1995):1313-21.

, and E. Greenberg. "Understanding the Metropolis-Hastings Algorithm. American Statistician 49(2) (1995):327-335.

Chib, Siddhartha and Edward Greenberg, "Hierarchical Bayes models", "The New Palgrave Dictionary of Economics", Eds. Steven N. Durlauf and Lawrence E.

Blume, Palgrave Macmillan, 2008, The New Palgrave Dictionary of Economics Online, Palgrave Macmillan. 25 May 2009, DOI:10.1057/9780230226203.0732

Robert, Christian. "The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation" Springer: New York 2007.

Diebolt, J. and C. P. Robert. "Bayesian Estimation of Finite Mixture Distributions." Journal of the Royal Statistical Society Series B 56(1994):363-75.

Draper, David. "Bayesian Multilevel Analysis and MCMC". Handbook of Multilevel Analysis. Springer: New York. 2008.

Gelfand, A. and A. Smith. "Sampling-Based Approaches to Calculating Marginal Densities." Journal of the American Statistical Association 85(3) (1990):972-985.

Geweke, J. "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds.), Proceedings of the Fourth Valencia International Conference on Bayesian Statistics. New York: Oxford University Press, 1992, pp. 169-93.

Good, I.J. "Some history of the hierarchical Bayesian methodology" Trabajos de Estadística y de Investigación Operativa, vol. 31, no. 1, February 1980, pp. 489-519.

Holloway, Garth and Quirino Paris. "Production Efficiency in the von Liebig Model" American Journal of Agricultural Economics, vol. 84, no. 5, December 2002, pp. 1271-1278.

Holloway, Garth, Donald J. Lacombe , and James P. LeSage. “Spatial Econometric Issues for Bio-Economic and Land-Use Modelling” Journal of Agricultural Economics, vol. 58, no. 3, September 2007, pp. 549-588.

Koop, G. Bayesian Econometrics. Chichester, United Kingdom: John Wiley and Sons, 2003.

Lavine, M. and M. D. West. “A Bayesian Method for Classification and Discrimination.” Canadian Journal of Statistics 20(1992):451-61.

Lindley, D. V. and Smith, A. F. M. “Bayes Estimates for the Linear Model” Journal of the Royal Statistical Society. Series B (Methodological), vol. 34, no. 1, 1972, pp. 1-41.

Mood, A. M., F. A. Graybill and D. C. Boes. Introduction to the Theory of Statistics. New York: McGraw-Hill, third edition, 1974.

Percy, D. F. “Prediction for Seemingly Unrelated Regression.” Journal of the Royal Statistical Society B 54 (1992): 243-252.

Salvioni, C. L’agricoltura biologica in Aguglia et al. Multifunctional agriculture. Enterpreneurial behaviours and strategies in the search of diversification. Roma: ESI, 2009.

Tanner, M. H. 1993. Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions. New York: Springer-Verlag.

and W. H. Wong. 1987. “The Calculation of Posterior Distributions By Data Augmentation.” Journal of the American Statistical Association 82(3):528-50.

Titterington, D. M., A. F. M. Smith and U. E. Makov. Statistical Analysis of Finite Mixture Distributions. San Diego: John Wiley & Sons, 1985.

Zellner A. 1996. An Introduction to Bayesian Inference in Econometrics. Wiley Classics Library. New York: John Wiley and Sons.

Tables

Figure 1. Province-specific spatial externalities (neighbourhood effects).

Province #	Lower 95% hpd	Posterior Mean	Upper 95% hpd
‘1’	-0.61	-0.16	0.23
‘2’	-0.09	0.00	0.16
‘3’	-0.32	0.01	0.33
‘4’	-0.31	0.02	0.30
‘5’	-0.17	0.02	0.20
‘6’	-0.22	-0.02	0.21
‘7’	-0.21	-0.04	0.10
‘8’	-0.70	-0.07	0.44
‘9’	-0.41	0.12	0.61
‘10’	-0.79	-0.21	0.32

Table 2. Covariate impacts in organic adoption

Covariate	Lower 95% hpd	Posterior Mean	Upper 95% hpd
‘Tenure’	0.04	0.99	3.03
‘Protected Area’	-0.67	0.29	1.36
‘Direct Selling’	0.16	0.36	0.59
‘Cereals and Oilseeds’	-0.26	0.09	0.41
‘Horticultural’	-17.48	-5.52	-1.42
‘Wine’	-0.71	-0.20	0.21
‘Fruit and Citrus’	-0.72	-0.20	0.23
‘Other Permanent’	-0.54	-0.19	0.18
‘Olives’	-0.12	0.45	0.96
‘Livestock’	-0.89	-0.50	-0.17
‘Milk’	-0.49	-0.05	0.37
‘Gender’	-0.37	-0.15	0.07
‘Age’	-1.50	-0.76	0.06
‘Upland’	-3.90	-1.58	-0.06
‘Hills’	-4.71	-2.31	-0.73
‘Flatland’	-4.90	-2.34	-0.75

Table 3. Likelihood statistics

Logarithm of the Likelihood	6.07
Logarithm of the Marginal Likelihood	-89.40
Numerical Standard Error	0.61

Figures

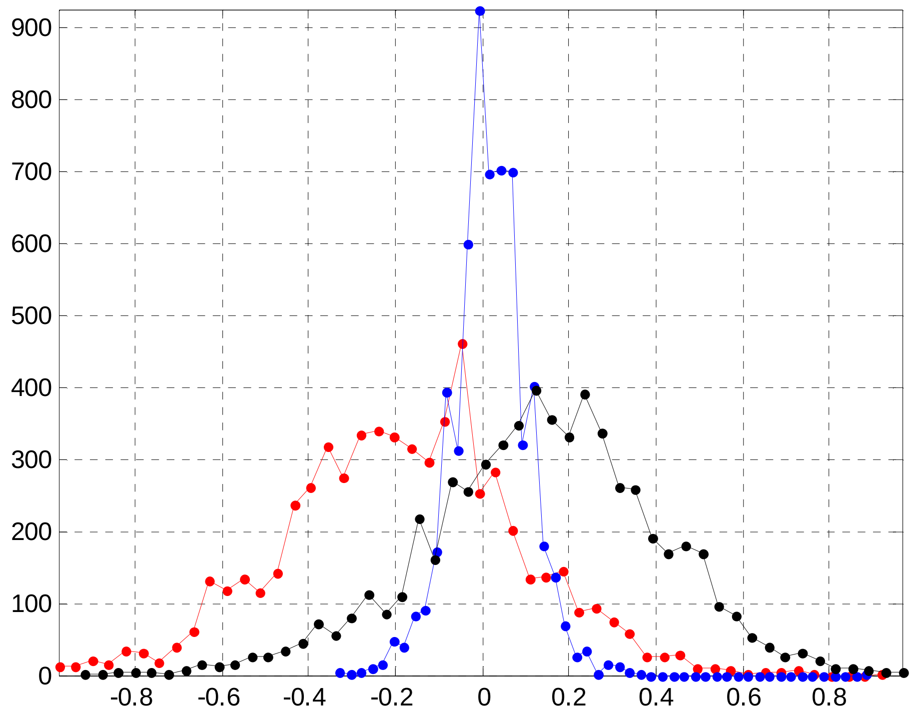

Figure 1. Representative posterior ‘neighbourhoods’ effects.