Bayesian Benefit Transfer for Choice Experiment Data: preliminary results.

Sergio Colombo, Nicholas Hanley and Garth Holloway

Abstract

Recently, several authors have proposed exploiting Bayesian methods for the purpose of inferring quantities estimated from an action at one site to those of another, proposed action, at a related policy site – an activity that has come to be known as ‘benefit transfer.’ We explore alternatives for implementing benefit transfer in the context of choice experiment data. A complete replication of the choice experiment exercise offers insights into the extent to which inferences are affected by the choice of transfer vehicle, and motivates the fundamental question guiding the spirit of the inquiry. Given assumed inter-site variability, given observed data at each of the two sites, and given the knowledge about how to proceed when the information set is ‘full,’ how would one proceed when the information set is ‘limited?’ We consider three vehicles. One vehicle is the conventional out-of-sample forecast – the ‘Bayesian predictive distribution’ – obtained from applying Jeffrey’s prior at the study site and processing the sample up until the ‘boundary’ of the policy site. The second is the ‘meta-inference’ obtained from conventional hierarchical modelling, with simulation obtained with reference to some hypothetical demography at the proposed policy site. And the third measure, with respect to the hypothetical demographic profile simulated at the policy site, involves inference obtained from a finite mixture that combines the observed data with those simulated at the policy site. A common theme throughout the exercise is the divergence between the study-site and policy-site demographic profiles, for which a variant of the Kullback-Liebler criterion proves useful. The reminder of the methodology evidences routine implementation of Markov chain Monte Carlo techniques – Gibbs sampling in particular – and demonstrates the inferential power for environmental valuation of fundamental cornerstones of Bayesian inference (290 words).

Journal of Economic Literature classifications: C11, Q51, Q53.

Keywords: Bayesian paradigm; environmental, economic intervention; choice experiments; posterior predictive distribution; natural conjugate prior; sequential Bayesian updating; hierarchical normal linear model, finite mixtures distribution; Kullback-Liebler criterion; Markov chain Monte Carlo; Gibbs sampling.

Address for correspondence: Sergio Colombo. Email: scolombo@ugr.es. We benefited from discussions with Kelvin Balcombe, Gary Koop and John Geweke about Bayesian implementation of the multinomial probit, choice experiments and convergence of the Markov chains. Full responsibility for error remains with the authors.
Recently, several authors have proposed the use of Bayesian methods for transferring the inferences obtained from a study site to those of another, related policy site, in an effort to ascertain the value of environmental economic interventions (Morrison and Bergland 2006, Moeltner Boyle and Paterson 2007). Interest in this practice, commonly referred to as ‘benefit transfer’, has intensified recently, and possibly as the result of general advances in statistical methodology (Gelfand and Smith, 1990), possibly due to the wide and broader application of choice experiments (Colombo, Calatrava and Hanley, 2007), and possibly because those advances have been proven robust under a variety of alternative data generating situations, many of particular relevance to environmental economics and to the environmental lobby in general (Morrison and Bergland 2006). In this context it is natural for investigators to explore alternative approaches to deriving the inferences that they seek and it is natural for the precision of alternative inferences to be compared to the costs one must expend in order to obtain them. Choice experiments offer a flexible, menu-driven approach to valuing a variety of environmental interventions and are attractive, not only due to the cost-efficient manner in which they can be implemented, but also due to the ease with which they offer a direct and fundamental insight into the over-arching question that economists frequently seek, namely the precise amount of money that some population is willing to forego in exchange for a particular environmental assurance. Bayesian inference in this setting has at least three attractions. One is the robustness afforded by Bayesian methodology in general (Berger, 1985); a second, due mainly to its reliance on basic laws of probability for its implementation, is the ease with which the artefacts of a full Bayesian investigation are communicated to a research audience (Jaynes, 1983); and a third attraction – the subject of the present investigation – is the great flexibility afforded methodologies that use as their basic precept the fundamental cornerstone of all Bayesian investigations, namely the assumption that the parameters directing data generation are not fixed in repeated samples but are, in contrast, random variables with associated probability distributions. And now, with most of the
prohibitive and foreboding computational problems firmly and squarely sidelined, the stage is set for a literal explosion of Bayesian methods in environmental economics; for recent evidence of similar intentions see León (2002), Moeltner Boyle and Paterson (2007); and for useful examples of conventional Bayesian analysis in environmental economics see Layton and Levine (2003) and Holloway, Tomberlin and Irz (2005).

In environmental economic investigations in which discrete-choice techniques are exploited, Bayesian inference offers at least three additional advantages. First, the multinomial probit methodology, which is known to circumvent the problem of independence of irrelevant alternatives, problematic under logit investigations (McFadden, 1984), is simple to implement. Second, the standard Bayesian implementation of multinomial probit using straight-forward Gibbs sampling algorithms is known to outperform alternative classical approaches (Geweke, Keane and Runkle, 1994). Third, whereas the logit representation rests firmly on the assumption that the errors imparted in the modelling process evolve from extreme-value distributions rendering it unable to accommodate a natural-conjugate prior (see Poirier (1998) and Koop (2003) for more detail) the multinomial probit, being based on multivariate Gaussian processes has natural conjugate prior probability density functions (pdfs) that update easily and naturally using a set of standard formulae (Imai and Van Dyk, 2005). While this may not be seen by many to be a direct advantage in implementation, this feature of the estimation framework ushers ancillary advantages that make multinomial probit estimation particularly attractive in the context of benefit transfer. There are three reasons. First, in the context of Bayesian investigations of benefit transfer, the extant literature (Morrison and Bergland (2006), León-Gonzalez and Scarpa (2006) makes clear an overarching issue, namely the extent of inter-site heterogeneity that is present and must be accounted for when ‘transferring’ inferences between sites. The natural-conjugate updating formulae, as we shall show, offer a particularly attractive feature in this regard. Second, the Gaussian assumption about the multinomial probit error structure makes tractable and natural
the implementation of data substructures into hierarchical relations that facilitate ‘meta’ or ‘supra’ policy inferences which, we argue, should be at focus in transfer investigations. Third, the Gaussian structure lends itself readily to a finite mixtures formulation, with three attractive features. First, following seminal advances at the naissance of the Markov chain Monte Carlo era (Lavine and West 1992; Diebolt and Robert 1994) the finite mixtures model is easy to implement, relying only on basic Gibbs sampling routines. Second, finite-mixtures formulations offer one of the most natural methods for combining information from composite sources. Third, when implemented through MCMC techniques, finite-mixtures formulations generate as by-products of their estimation algorithms posterior quantities that enable evaluation of divergence between the composite sources of data in the sample, which, we argue, is particularly germane in the context of transfer investigations. In short, the multinomial probit offers many advantages that motivate its use in Bayesian benefit transfer with choice experiment data.

In the analysis that follows, we indicate in cogent terms, implementation of the multinomial probit, in four distinct settings. The first setting is the natural-conjugate, sequential updating situation, which we use to benchmark our reference points and which we use throughout the paper against which we appraise the precision of the alternative transfer mechanisms. The second setting is a standard one. It is the posterior predictive inference obtained for the ‘out-of-sample’ forecast at the policy site derived from the data available at the study site – in a sense and in a manner motivated by one important graphic to follow, this is what we term ‘the study-site-policy-site boundary prediction.’ This inference, we also argue, is perhaps the most conventional of our proposed transfer mechanisms because it relies only on standard precepts in widespread use in Bayesian inference elsewhere. For examples in medicine see Paliwal and Gelfand (2006); for examples in risk assessment see Eckert et al. (2008). The third setting is the hierarchical normal-linear model setup in which the multinomial probit evidences routine calculation. The fourth and final setting is the finite
mixtures formulation where, again, computations are straightforward and ancillary statistics, collated as by-products of the estimation algorithm, offer appeal in the benefit-transfer environment.

Our principal contributions are three. First, we enrich the collection of extant approaches that have been proposed for making benefit transfers (Morrison and Bergland 2006; León, Vázquez-Polo, Guerra and Riera, 2002). Second, we contribute to the thematic developments of the literature that dates, at least to Wald (1947), focused on assessing and comparing the benefits of sequential updating to their costs. Third, we highlight the dexterity of the multinomial probit in benefit transfer with choice-experiment data and showcase basic tools of Bayesian inference, applied frequently in other realms, yet hitherto unexploited in transfer investigations.

The paper is organized as follows. Section two discusses briefly the valuation problem confronting investigators wishing to transfer inferences between a study and a related policy site. Alternative implementation vehicles, including the choice experiment setting used in our empirical analysis are discussed. Section three introduces the data availed to us and describes briefly the distinctions existing between the ‘study site’ and the hypothetical ‘policy site.’ Section four provides a brief introduction to the notation and the basic Bayesian tools that we apply and section five outlines briefly the multinomial probit methodology. Section six presents the benchmark results which we use to evaluate subsequent methodology. Section seven presents results of, respectively, predictive inference at the study-site-policy-site boundary; hierarchical normal linear modelling using a reference demography simulated at the policy site; and finite-mixtures implementation of the study-site-policy-site mixture. Conclusions and some directions for additional research are presented in section eight.
Benefit Transfer and Choice Experiments

The wider use of cost-benefit analysis is a noticeable trend in environmental policy-making in many countries. Policy makers and international agencies such as the World Bank are being challenged to assess the economic value associated with ecosystem goods and services. Additionally, legislation increasingly requires regulatory bodies to seek practical ways of estimating cost/benefit ratios for a very large number of environmental assets, as is the case for instance with the recent European Union Water Framework Directive (WATECO, 2004) where benefit assessments are required to identify ‘disproportionate cost’ cases for all EU water bodies. However, a common characteristic of environmental valuation methods such as contingent valuation is that their design and implementation for the estimation of the economic value of a particular environmental asset, or changes in this asset’s condition, is costly and time consuming. This makes the widespread use of original valuation studies impractical in many situations. As a result, benefit transfer techniques have been developed where information from past studies is used to provide a meaningful basis for directing environmental policy and management by predicting environmental values at new, ‘policy’ sites (Desvouges, Johnson and Spencer, 1998; Navrud and Ready, 2003; Rolfe and Bennett, 2006). Such benefit transfers are increasingly being used by regulatory agencies, such as the Environment Agency in the UK, by government ministries such as DEFRA and by the European Commission. However, worries have been expressed about the nature and size of the decision errors produced by relying on benefits transfer, rather than conducting original valuation studies (Barton, 2002; Rozan, 2006).

In a frequentist statistical framework one can think of three possible approaches for transferring results from a historical study to a new site. In all cases, we wish to estimate benefits at a policy site, where no original valuation-work has been undertaken, using the results from an original study at a study site. The first approach is simple value transfer, which assumes that willingness-to-pay (wtp) for an environmental quality change observed at
the study site is equal to the willingness-to-pay at the policy site. This approach has been followed in many practical applications. Whilst this approach has the merits of simplicity of use and implementation, it neither allows for differences in environmental and socio-economic conditions between the study and policy sites, nor does it allow for different preferences of people at different sites. The second approach is to predict willingness-to-pay at the policy site by using variables thought to importantly influence willingness-to-pay from the policy site via a bid curve estimated at one or many study sites. The goal is to adjust willingness-to-pay with as much data as is available from the policy site. The third approach is benefit function transfer, where the whole willingness-to-pay function is transferred from the study to the policy site. The basic problem of this approach is that it still assumes the preferences to be the same across all sites and all seemingly similar assets. Evidence on which method works best is mixed. Sometimes simpler approaches seem to lead to smaller ‘errors’ (see, for example, Ready et al., 2004; Colombo and Hanley, 2008).

Despite the widespread use of benefit transfer techniques most of the published literature fails to support convergence validity of this technique, even under ideal circumstances (Bergland, Magnussen and Navrud, 1995; Hanley, Wright and Alvarez-Farizo, 2006). Transfer errors range from a few percentage points to high degrees of inaccuracy. Rosenberg (2005) identifies three potential sources of errors that may affect the accuracy of benefit transfers namely generalization error, measurement error and publication selection bias. The first arises when the estimated value for the good in question of the study site is adjusted to a policy site, and is enhanced by the amount by which the study site and the policy site differ. The second originates from the analyst’s decisions needed to estimate values in the first place, and thus depends on the valuation method used, the model specification and so on. The third evolves from the propensity of the published literature to report only results that conform to expectations or that make a methodological contribution. Generalization errors are expected to fall when a bigger number of primary studies will be available to the analyst.
when implementing a benefit transfer. Publication selection bias can be reduced by establishing a peer-reviewed, benefits transfer ‘journal’ to increase the quality and quantity of the valuation studies available to researchers. Measurement error can be tackled by finding methodologies especially suitable for benefit transfer purposes – the Bayesian methodologies to follow being one particular set of cases in point.

Most of the empirical applications published in the benefits transfer literature have used data based on the *travel cost method* or the *contingent valuation method*. However, Morrison *et al.* (2002) have argued the choice experiment method has a good potential for benefit transfer since it has the advantage over contingent valuation that it is easier to allow for differences in improvements in environmental quality as well as differences in socio-demographics when transferring value estimates. Recent applications of choice experiments in benefit transfer have found them suitable for benefit transfer, particularly when the transfer involves implicit prices rather than compensating surplus estimates (for a review see Morrison and Bergland, 2006).

Under the Bayesian approach we can consider benefit transfers as an updating process, where knowledge from previous studies is combined with knowledge from the policy site. This is perhaps a more natural way of thinking about the problem. Benefits transfer is seldom conducted without any information on the policy site being available, whether this is information on the individuals who will be impacted by the change, or information on the environmental change or reference conditions. A Bayesian approach allows this information to be incorporated systematically with information from previous studies in an updating procedure. This approach has indeed been tested for contingent valuation data in the past (see, for examples, León *et al.*, 2002; Lehr, 2004; Bergland, 2006). In this paper, however, we extend the Bayesian benefits transfer approach to a choice experiment setting. This is an interesting extension in that first choice experiments are becoming much more widely used in the environmental valuation literature (Colombo, Calatrava-Requena and Hanley, 2007);
second choice experiments are more suitable for environmental benefit transfer purposes than alternative methods (Morrison et al., 2002) and third the Bayesian paradigm has been suggested as an interesting approach for extending the use of choice experiments in benefit transfer (Morrison and Bergland, 2006).

Data

In this section we describe the data, considering both its geographic context and the design of the survey, commencing with the former.

Geographic context

The study and policy sites are two watersheds located in the south of Spain where a huge increase in soil erosion rates has occurred in the last few decades. These watersheds were chosen for analysis given the importance of adopting soil conservation measures in olive orchards to reduce the negative impacts of soil erosion, both on-site and off-site. Off-site impacts include deterioration in water quality, loss of biodiversity and desertification of the landscape (more details provided in Colombo, Calatrava-Requena and Hanley, 2003). The economic quantification of such effects would help public administrators designing schemes to encourage farmers to adopt soil conservation measures, and in identifying ‘disproportionately costly’ measures under the Water Framework Directive. Given the large number of watersheds where soil erosion is a problem in southern Spain, benefit transfer may represent the only practical way to estimate the benefits of soil erosion reduction in the region. The ‘study’ site, the Genil basin, extends from the peak of the Sierra Nevada to the reservoir of Iznajar, and includes the city of Granada, where most of the population of the area live. The ‘policy’ site, the Guadajoz basin, extends from the Subbetic hill system to the Guadajoz estuary and the Guadalquivir river. It covers the provinces of Córdoba and Jaén but neither of these cities lies in the watershed, so that most of its residents are rural dwellers. The watersheds have similar morphological conditions and land use patterns, with olive cultivation being by far the dominant land use.
Survey design

The intent of the choice experiment survey is to provide data to learn about the preferences of residents of the study and policy sites for the reduction of soil erosion effects. Details of the survey design can be found in Colombo, Calatrava-Requena and Hanley (2007). The soil erosion effects and levels considered in the study are listed in table 1. The effects (attributes) used are impacts on landscape desertification, impacts on water quality, impacts on flora and fauna, impacts on farm jobs, the amount of land area encompassed by the scheme, and the amount of additional local tax required to fund each option, should it be implemented. Colour handbooks were used to explain each attribute to respondents, to show them photographs of current impacts, and to show how these attributes might be affected by implementing a soil-erosion-control plan. The choice exercise described to respondents three land management options (control plans) to tackle the soil erosion problem in the area. Options one and two show the expected environmental and social conditions in the watersheds in 50 years if a soil erosion reduction project was implemented in olive orchards and degraded hill slopes. These projects involve sowing cover crops in these areas to obtain a significant reduction in soil erosion rates (Pastor and Castro, 1995). Option three is the so-called ‘status-quo-option’ in which no action is undertaken. This third option therefore contains the lowest interventions of each action, but also the lowest tax rate confronting respondents, namely zero. Finally, the survey was administered to 350 people in each watershed using face-to-face interviews. In the Genil catchment 252 surveys were deemed suitable for modelling whilst in the Guadajoz basin the number is 269. In order to outline how we process these data, it is necessary to introduce some notational conventions (insert table 1 about here).

Notation

By way of notation, let $\theta$ denote a collection of parameters of interest, $\pi(\theta)$ the prior probability density function (pdf) for $\theta$ and $\pi(\theta|y)$ the posterior pdf for $\theta$, where $y \equiv (y_1, y_2, ..., y_N)'$ denotes data. Frequently we reference the data generating model $f(y|\theta)$, which is the
likelihood function when viewed as a function of $\theta$ and, sometimes, make use of variants of the $f(\cdot|\cdot)$ notation to reference particular probability density functions. There are nine pdfs that we apply, and detail in the appendix, namely, the univariate-Normal pdf, $f^N(\cdot|\cdot)$; the truncated-Normal pdf, $f^{TN}(\cdot|\cdot)$; the multivariate-normal pdf $f^{MN}(\cdot|\cdot)$; the matrix-normal pdf $f^{MN}(\cdot|\cdot)$; the inverted-Gamma pdf, $f^{iG}(\cdot|\cdot)$; the inverted-Wishart pdf $f^{iW}(\cdot|\cdot)$; the Multinomial distribution $f^M(\cdot|\cdot)$; the Dirichlet distribution $f^D(\cdot|\cdot)$; and the uniform pdf, $f^u(\cdot|\cdot)$. 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 ‘$\asymp$.’ Throughout, we use $i = 1, 2, \ldots, N$ to denote the sample; Greek letters to denote the unobserved parameters; and Roman numerals to denote the observed and missing data.

**Statistical model**

The basic situation being considered is as follows. We observe a response $y_i = j$, a choice across $j = 0, 1, 2, \ldots, M$ mutually exclusive alternatives with each of the $N$ choices in the data vector $y \equiv (y_1, y_2, \ldots, y_N)'$ assumed to satisfy a particular choice criterion. The conventional choice criterion (see McCulloch and Rossi, p. 209; Geweke, Keane and Runkle, p. 610; Koop, p. 222) is that $y_i = j$ if $u_{ij} \geq u_{ik} \forall k \neq j, k = 0, 1, 2, \ldots, M$ where $u_{ij}$ denotes the utility of individual $i$ when choice $j$ is made. Within this system, as is well-known (see especially, Geweke, Keane and Runkle, p. 610; Keane, p. 195; Dansie, p. 527; and Bunch, p. 6) each choice is identified only with respect to a numeraire. Consequently, we focus on the choices across the $M$ alternatives in the latent, normalized utilities $z_i \equiv (z_{i1}, z_{i2}, \ldots, z_{iM})'$, where $z_{ij} = u_{ij} - u_{i0}, j = 1, 2, \ldots, M$. Consequently, if $u_{i0} \geq u_{ij}, j = 1, 2, \ldots, M$, then each $z_{ij} < 0$, and if $u_{ij} \geq u_{ik} \forall k \neq j, k = 1, 2, \ldots, M$, then each $z_{ik} < z_{ij}$. Because each of the conditional utilities depends on sets of known factors, some being alternative-dependent and some not, we distinguish between these set-types in the system of equations that we use to depict choice. Our choice-experiment data arrive in a balanced panel with $t = 1, 2, \ldots, T$ replications across the
respondents $i = 1, 2, \ldots, N$. In this context, developments relegated to the appendix, lead to the remarkably simple system

\[ z = \mathbf{x} \beta + \varepsilon, \]

where $\mathbf{z} = (z_1', z_2', \ldots, z_N')'$; $\mathbf{w} = (w_1', w_2', \ldots, w_N')$ denotes the NTM×PM matrix of alternative-dependent covariates; $\mathbf{x} = (x_1', x_2', \ldots, x_N')$ denotes the NTM×K matrix of alternative-dependent covariates; and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_M)'$ denotes the NTM-dimensional random disturbance vector. Given the assumed multivariate-Gaussian structure, $\varepsilon$ is distributed $f^N(\varepsilon|0, \mathbf{I}_{NT} \otimes \Sigma)$ and it follows from the Jacobian transformation of $\varepsilon$ that $\mathbf{z}$ is distributed $f^N(\mathbf{z}|\mathbf{x}\beta, \mathbf{I}_{NT} \otimes \Sigma)$. Given this Gaussian structure on the unobserved data, $\mathbf{z}$, estimation strategies follow naturally the pathway sown by seminal Bayesian contributions of the multivariate probit and multinomial probit formulations (see, in particular, Albert and Chib, 1993; Geweke, Keane and Runkle, 1994; McCulloch and Rossi, 1994; McCulloch, Polson and Rossi, 2000). The observed-data likelihood,

\[ f(\mathbf{y}|\theta) = \prod_{i=1}^{N} \prod_{t=1}^{T} \left( \frac{N!}{\prod_{j=1}^{M} y_{ij}} \right) \prod_{j=1}^{M} \varphi(\text{rob}_{ij}^{y_i}), \]

is complicated by the integrals implicit in the probabilities $\varphi(\text{rob}_{ij}^{y_i}) = \Phi((z_{ij} - x_{i}' \beta)/\sigma_{ij})$, where $\Phi(.)$ denotes the cumulative distribution function (cdf) value corresponding to the standard normal distribution; $z_{ij}$ denotes the latent response for individual ‘i’ under replicate ‘t’ and choice alternative ‘j;’ $x_{i}$ denotes the K-vector of choice-independent covariates; and $\sigma_{ij} = \Sigma_{ij}^{-1}$ denotes the element in the $j^{\text{th}}$ column and $j^{\text{th}}$ row of the inverse matrix $\Sigma^{-1}$. An estimation strategy that proves very useful (see, for example, Wong, 1987; and Tanner, 1993) is to augment the observed-data likelihood with the latent normalized utilities, $\mathbf{z}$, and focus on the complete-data likelihood,

\[ f(\mathbf{y}|\theta, \mathbf{z}) = \frac{1}{2} \pi^{NTM} |\Sigma|^{-NT/2} \exp\{- \frac{1}{2} (\mathbf{z} - \mathbf{x}\beta)' \Sigma^{-1} (\mathbf{z} - \mathbf{x}\beta)\}. \]
The advantage of (3) as opposed to (2) is that, despite the fact that (2) contains only the observed data \( y \), whereas (3) contains both the observed \( y \) and the non-observed \( z \), a highly efficient algorithm exists for estimating the unobserved \( z \), rendering the model in (3) tractable, from an estimation viewpoint. In particular, if we let \( \pi(\theta) \) denote the prior probability density function summarizing the investigator’s prior information about the unknown quantities \( \theta = (\beta, \Sigma) \) and use \( \pi(\theta|y,z) \) to denote the resulting posterior, conditional on the known quantities, \( y \), and the unknown quantities, \( z \), the fully conditional distributions corresponding to the resulting posterior

\[
\pi(\theta|y,z) \propto f(y|\theta,z) \pi(\theta),
\]

have particularly simple forms. In particular, a simple algorithm for estimating the unknown quantities in \( \theta \) consists of only three, sequential steps (Albert and Chib, 1993; McCulloch and Rossi, 1994; Geweke, Keane and Runkle, 1994).

As is well known (see Geweke, Keane and Runkle (1994) and the references they cite) even under the normalization of latent utilities, the joint posterior pdf for \( \theta \), (4), is not likelihood-identified. Poirier (1998), among others, derives the implications of such settings. Several alternative approaches have appeared in the literature in order to remedy this situation, including fixing all of the diagonal element of \( \Sigma \) to unity (Chib and Greenberg, 1998); placing proper, non-diffuse priors on the elements of \( \Sigma \) (McCulloch and Rossi, 1994) and restricting a single diagonal element to unity while allowing the off-diagonal elements of \( \Sigma \) to vary according to their conditional distributions (Dawid, 1988; Nobile, 2000; Linardakis and Dellaportas, 2003). The origins of this restriction are attributed to Dempster (1961). Because our focus in this paper lies in the use of the multinomial probit in benefit transfer, rather than nuances stemming from implementation of alternative covariance structures in \( \Sigma \), we adopt arguably the simplest approach, which is to constrain the covariance matrix that is relevant to the normalized system, call it \( \Omega \), to equal a known quantity. This quantity is the
the (M-1)-dimensional matrix that results from assuming that the normalized system is the
one appearing in the third scenario in Geweke, Keane and Runkle’s ‘Experiment One’ in
which the covariance is set such that \( \Omega = \frac{1}{2} (I_{m-1} + u_{m-1}u_{m-1}') \), where \( I_{m-1} \) denotes the (M-1)-dimensional identity matrix and where \( u_{m-1} \) denotes the (m-1)-dimensional unit vector. With
this restriction in place, attention resides solely with the elements in the unknown parameter
vector \( \theta \equiv \beta \). Consequently, the basic algorithm for estimating the system in (6) is
considerable simplified.

\[ A_1: \text{Multinomial Probit Algorithm. Draw } \beta \text{ from a multivariate-normal distribution. Draw } z \text{ from the truncated-normal distribution.} \]

Precise details of the draws are relegated to the appendix. All of the estimations performed
henceforth in this paper are derived from these two fundamental steps of this basic
multinomial probit algorithm. For a background to the theory underlying the Markov chain
Monte Carlo approach to estimation see Gelfand and Smith (1990), Gelfand et al. (1990) and
Robert and Casella (2001); and for the intuition underlying the approach see Casella and

In this context, environmental applications of choice experiment data focus on the
‘partworths’ corresponding to the marginal willingness to pay for a particular characteristic
comprising one of the choice-experiment alternatives. For an interesting application in the
context of GM food, see James and Burton (2003) and for a general treatment see Train
(2003) and the references cited there. Consequently, in the remainder of this work, we focus
attentions on the partworth quantity corresponding to attribute ‘\( \kappa \),’

\[ \varphi_\kappa \equiv \beta_\kappa \div \beta_m, \] (5)

which is the ratio of the relevant regression coefficient \( \beta_\kappa \) to the negative of the corresponding
monetary measure, $\beta_m$ (Bennet and Blamey, 2001). We note, in passing, that the measure in (5) is, of course a function of the model parameters and write $\varphi_x(\Theta)$ to signify this. In Bayesian inference our objective is to obtain the marginal posterior probability density function of the partworth that is derived from the estimation procedures, which we denote, $\pi(\varphi_{|\Theta}|y,z)$, and, in turn, derive the posterior density for $\varphi_x(\Theta)$ once the latent data have been marginalized out, that is $\pi(\varphi_{|\Theta}|y)$. Here, once again, we denote the explicit dependence of the partworth densities on the parameters underlying the data generating process, $\Theta$; the latent data, $z$; and the observed data, $y$. A comparable measure is a focus of attention in Holloway, Teklu and Ehui (2008) where a similar normalization of parameters is necessary in order to estimate the distance that a household resides from a market. In fact, the quantity (5), obtained from normal data has received considerable attention in the early literature. In view of its importance in ensuing developments, a small discussion of its evolution seems appropriate at this point.

The distribution of the so-called ‘ratio-of-normal-random-variates’ is presented in Hinkley (1969, equations (1) and (2), p. 636), who attributes results to Fieller (1932). Fieller (1932) studies the existence of moments and offers relevant comment (p. 432, paragraph at the bottom of the page; p. 434, below equation (35); and p. 435, entire discussion) about the use of the normal approximation when the ratio of the mean of the distribution (for $\beta_m$ in our context) is ‘large’ relative to its standard deviation—a result he attributes to Geary. Hinkley (1969) studies the approximation and concludes that it is very accurate for moderate sample sizes under the restriction that the latter standard deviation is less than or equal to the former mean. The conclusion that the density of the relevant quantity is proper is stated below equation (8). Other relevant works on the topic are Merrill (1928) and Marsaglia (1965). Thus, we have the results that the marginal density, $\pi(\varphi_{|\Theta}|y)$, is proper; that, under certain conditions, its moments may not be finite; but that, when those conditions are met, the normal
distribution provides a useful approximation to the true distribution.

All of our attentions in the remainder of the paper are directed toward robust posterior inference about the location and scale of the partworth measure $\pi(\psi_\theta | y)$. In this context, choice is available, because the measure in (5) is computable for any of the characteristics involved in our choice experiments. The reader will recall that there are three qualitative characteristics namely desertification, water quality and flora and fauna density; and there are two quantitative characteristics, namely the number of additional rural jobs being created and the total land area encompassed by the project. In order to focus attentions we direct discussions on the partworth measure corresponding to the second water quality characteristic which is the fourth covariate in our empirical analysis). The motivation for this choice is simple when the objectives of the project are recalled. The posterior density corresponding to the coefficient of this particular covariate is centered the farthest from zero, making sensitive transfer estimates derived with respect to this quantity. Consequently, if the transfer methodology holds up well under this circumstance, it is likely to hold up better with reference to the remaining environmental interventions. Empirical results derived in the next section provide further motivation.

**Benchmarking a Base Line for the Partworth**

At this point it is worth emphasizing that we actually have available data from the two sites across which we wish to make transfer inference. Such full information will not usually be available in a real benefit transfer setting. The complete replication of the choice experiment exercise offers insights into the extent to which inferences are affected by the choice of transfer vehicle, and motivates the fundamental question guiding the spirit of the inquiry. Given assumed inter-site variability, given observed data at each of the two sites, and given the knowledge about how to proceed when the information set is ‘full,’ how would one proceed when the information set is ‘limited?’ We consider three vehicles. One vehicle employs the conventional out-of-sample forecast – the ‘Bayesian predictive distribution’ –
obtained from applying a weakly informative prior at the study site and processing sequentially each of the sample observations up until the ‘boundary’ of the policy site; the second is the ‘meta-inference’ obtained from conventional hierarchical modelling, with simulation obtained with reference to some hypothetical demography at the proposed policy site; and, third, with respect to the hypothetical demographic profile simulated at the policy site, inference is obtained from a finite mixture that combines the observed data with those simulated at the policy site. A common theme throughout the exercise is the divergence between the study-and-policy-site demographic profiles, for which a variant of the Kullback-Liebler criterion proves useful. The reminder of the methodology evidences routine implementation of Markov chain Monte Carlo techniques – Gibbs sampling in particular – and demonstrates the inferential power for environmental valuation of fundamental cornerstones of Bayesian inference.

The complete set of covariates consist of two desertification characteristics; two water quality characteristics; two flora-fauna characteristics; jobs; area; a tax instrument; and a constant, leading to a total of ten covariates within each regression. Pooling the observed data supposing that there may be significant inter-site variability leads naturally to two benchmarks, which, for want of a better terminology, we refer to as ‘classical.’ These are reported in the second and third columns of table 2 and we consider these settings prior to embarking on approaches that we refer to as ‘fully Bayesian’ in spirit. (Insert table 2 about here).

Column 2 of table 2 reports the results of estimation from the pooled sample and column 3 reports the results of independent estimation of the respective samples—in other words, estimation of $\beta$ distinguished by a set of dummy variables, as one would employ in the case of the familiar Chow (1960) test. The reports are posterior means computed from the Gibbs sample generated from applying the algorithm in $A_1$, respectively, with and without the dummy variables. The numbers in parentheses, below the posterior means are the 95%
posterior density intervals from the sample of 10,000 simulated draws from the posterior, after discarding the first 10,000 iterations. Analyses using five diagnostic measures (respectively, the diagnostics proposed by Gelman and Rubin (1992), Yu and Mykland (1998), Raftery and Lewis (1992), Geweke (1992), Phillipe and Robert (2001) and Giakoumatos et al. (1999)) suggest that sequences of Gibbs iterations converge very rapidly and that sample sizes in the order of 1,000 iterations (with 1,000 prior discards) are sufficient to make stable inferences. Comparisons of these alternative convergence diagnostics are presented in El-Adlouni, S., C. Favre and B. Bobée (2006), Brooks and Gelman (1998) and Brooks and Roberts (1999); and a broader discussion of convergence of Markov Chains is presented in Tierney (1994). In this context, our chosen iteration size is quite conservative.

Three features of the results are particularly noteworthy. First, each of the covariates is ‘significant’ in explaining choices across the pooled sample in the sense that each of the posterior density intervals – computed at the 95% level – do not contain zero. Second, each of the respective coefficient estimates possesses the predicted sign. Third, due to the precision of the report for the posterior density for the tax rate, the inferred partworth values should be measured with a high degree of precision. We note, in particular, that the posterior pdf for the tax rate is not centred near zero, making stable inferences about the values that the respondents place on each of the environmental interventions. In short, the combined Genil- and-Guadajoz data lend themselves well to the benefit transfer exercise.

Turning to the reports obtained from the independent regressions – the Chow regressions reported in column three – two additional insights are noteworthy. First, each of the covariate posterior densities has shifted somewhat from the centering values reported in the pooled regression. Second, although the shift is discernible, in most cases it is quite negligible. Third, the inter-site heterogeneity that is present is clearly observable but raises a question about how it affects inference between the two sites and raises scope for methodology that is explicitly designed to investigate this feature of the pooled policy and
study site samples. In short, the results of the pooled and Chow regressions raise questions about the extent to which we are able to accommodate inter-site variability in transferring inferences across the two sites and, inevitably, raise scope for additional empirical enquiry.

The heterogeneity identified under the Chow-regression model estimation is detected in this study using two standard approaches that conventional Bayesian inference would assess across the samples. The first of these is a standard hierarchical normal model investigation and the second is a conventional finite mixture methodology. Because we will focus important attentions on these in the subsequent section of the paper, we present details of the estimation algorithms. Presently, however, there does not appear to be too much dissimilarity between the ‘policy’ and the ‘study’ sites, as the schematics in figure 1 seem to indicate. Figure 1 plots the posterior density estimates from the pooled and Chow regressions for the water quality covariate, level high. The partworth pdfs corresponding to the pooled regression is depicted in blue and the partworths from the Genil and Guadajoz sites are depicted in green and red, respectively. The figure illustrates cogently that it is almost impossible to disentangle the three estimates, each having a similar location and very similar scale. Hence, at the present level of generality there would not seem too much hindrance in evaluation if the investigator were simply to apply the genil-based estimates as a transfer, directly, to the Guadajoz site. However, additional insights evolve from further methodological evaluation. (Insert figure 1 about here.)

Although these results suggest a key conclusion between the two sites, namely that data generation at the two sites is not greatly dissimilar, the multinomial probit methodology leads naturally to two alternative methodologies for encompassing inter-site unobserved heterogeneity, which we now evaluate.

Hierarchical Normal Linear Modeling

A second approach to inference lies in the notion of exchangeability originating from DeFinetti (1974) but now in wider use in a variety of situations. The notion that subsamples
are *a priori* exchangeable was first espoused in an application to gambling. Since then, exchangeability has become something of a cornerstone in Bayesian inference. Examples of its use include applications to pollution control benefits (Atkinson 1992) health (Agee and Crocker 2004) and biology (Corander et al. 2006).

The implication of exchangeability in our particular setting is the following. Given the subsamples the study site and the policy site, we seek meta-inference about the locations and scales of the over-arching densities from which the subsamples are exchangeable and are therefore able to be drawn at random. Mathematically, we depict the hierarchical structure by considering data generation of the observed sample at the two, respective sites through the respective data densities $f(y_1|\theta_1)$ and $f(y_2|\theta_2)$ where, the upper level of the hierarchy is given by the assumptions $f(\theta_1|\theta)$ and $f(\theta_2|\theta)$. Here, $y_1$ and $y_2$ represent data generated respectively at the study and policy sites and $f(\theta_1|\theta)$ and $f(\theta_2|\theta)$ denote the stochastic relations depicting the draws for the study-site-specific and policy-site-specific parameters, $\theta_1$ and $\theta_2$, respectively. In a hierarchical enquiry of this form, with actual data at the two sites, inferences for policy purposes reside in the locations and scales of the posterior pdfs for the meta-parameters $\theta$, namely $\pi(\theta|y_1,y_2)$. Here we make explicit that the inferences are drawn upon observing data at the two sites. In a non-hypothetical benefit transfer setting we will not have the data $y_2$ from the policy site. However, this should not trouble us too much. In the same spirit of the enquiry in the previous section we can generate draws for $y_2$ given assumptions about the covariate composition in the multinomial probit. The situation is then the following. Given covariate information at the policy site, say $x_2$, we can generate a set of predictions $y_2$ based on the covariate information and assumptions about $\theta_2$. Several alternatives exist for the latter sampling. One natural way to proceed is to use the posterior pdf $\pi(\theta_1|y_1)$ that one obtains from processing alone the observations available from the study site. Then using the simulated sample $y_1$, we can proceed to process $\pi(\theta|y_1,y_2)$ which, of
course is conditional upon the covariate information that one uses prescriptively. The Gaussian basis of the multinomial probit specification makes tractable the implementation of this technique through the combination of the Normal kernels in the posterior. However, availability of real data at the so-called ‘policy’ site, facilitates direct inference and posterior predictive checking. Specifically, with $z_1$ denoting the latent data relevant to the study site and $z_2$ denoting the latent data at the policy site, the conditional posterior pdf for $\theta = \beta$ in the Gibbs chain has the specific form:

$$
\pi(\theta|y_1,y_2,z_1,z_2,\theta_1,\theta_2) \propto \prod_j \exp\{-\frac{1}{2}(z_j'x_j\beta_j)'(z_j'x_j\beta_j)\} \times \exp\{-\frac{1}{2}(\beta - \beta_0)'C_{\beta_0}^{-1}(\beta - \beta_0)\},
$$

here $\beta_0$ and $C_{\beta_0}$ denote the prior mean and covariance respectively of the unknown meta-inference parameters $\beta$. This formula shows clearly the relationship between the contributions of the study and policy sites, the influence of prior information and the influence of the simulated data in the two situations. The question of how well the model performs arises naturally in this setting. For this purpose we present in the fourth column of table 2, the results of iterating the following algorithm (details are relegated to an appendix).

**A_2: Hierarchical Normal Linear Modeling Algorithm.** Draw $\beta_1$ from a multivariate-Normal distribution. Draw $\beta_2$ from a multivariate-Normal distribution. Draw $\beta$ from a multivariate-Normal distribution. Draw $z_1$ from a truncated-Normal. Draw $z_2$ from a truncated Normal.

The results in table 2 demonstrate clearly that the regression coefficients depart, but that this departure is rather modest.

**Finite Mixture Modeling**

Our final approach for benchmarking a reference point in the Genil and Guadajoz samples is the one that follows immediately from extending two seminal works on Bayesian finite
mixtures namely Lavine and West (1992) and Diebolt and Robert (1994). Subsequent extensions of the basic ideas embedded in these two seminal papers that are particularly relevant to present purposes include Green (1995) and Richardson and Green (1997). The mixture formulation of the problem is important for two reasons. First, it provides another familiar line of investigation for perceived inter-site variability. Second, it provides a highly intuitive basis for transferring benefit estimates that is hitherto unexploited. The finite mixtures algorithm follows naturally from observing the posterior in the presence of a set of unknown latent classification variables, \( \chi = (\chi_1, \chi_2, \ldots, \chi_N)' \), where, for each \( i = 1, 2, \ldots, N \), \( \chi_i = (\chi_{i1}, \chi_{i2}, \ldots, \chi_{iM}) \), denotes a collection of binary indicators such that, if \( \chi_{ij} = 1 \) then \( y_i \) emanates from component \( j \) of the mixture; and \( \chi_{ij} = 0 \), otherwise. Hence, the posterior emerging from the complete-data posterior has the form

\[
\pi(\theta|y,z,\chi,\beta) \propto \prod_i \prod_j \omega_j \exp\{-\frac{1}{2} (z_j - x_i \beta)' (z_j - x_i \beta)\}^{\chi_{ij}} \times \exp\{-\frac{1}{2} (\beta - \beta_0)' C \beta_0^{-1} (\beta - \beta_0)\},
\]

And the Gibbs sampling algorithm for deriving inferences in this context follows naturally from extending results in Lavine and West (1992) and Diebolt and Robert (1994). Relegating details to the appendix, the algorithm has the simple structure:

**A3: Finite Mixtures Modeling Algorithm.** Commence with a random classification, \( \chi \), of the sample observations into two groups. Within each group process the sample according to algorithm \( A_1 \). Subsequently, compute the probability that each respective observation lies in a designated subcomponent.

Then, reclassify, by drawing \( \chi \) from a multinomial distribution.

The reports in the fifth column of data in table 2 make clear that the level of inter-site heterogeneity is quite small. This is evidenced in two ways. First, the posterior means reports, relative to their intervals, are not too dissimilar. Second, the subcomponent sample
weights are close as the third-last and second-last entries in the column, are close to one half. Third, the percent-correct classifications, reported as the last entry in the column, is also significantly different from zero. Thus, the two-component mixture processes the sample being rather indifferent as to which of the two sites from which each actual observation is obtained.

Before turning to consider these algorithms in more detail, we consider another feature of the pooled sample data, namely the results of inferences obtained from sequentially expanding the sample size across the ‘study’ and the hypothetical ‘policy’ sites.

**Sequential Updating Through the Sample**

In order to motivate the sequential updating rule, consider the following thought experiment. Suppose, beginning from a non-informative situation, we have \( \pi(\theta) \propto \text{constant} \). A single observation avails itself and it is processed through the algorithm \( A_1 \). The posterior pdf \( \pi(\theta|y) \) is derived. Now, a second observation arrives. In this second case, the investigator now employs as the prior, the posterior pdf obtained from the first analysis, that is, she sets \( \pi(\theta) \propto \pi(\theta) \) and the analysis is executed again employing \( A_1 \). Repeating this process an unspecified number of times, two important questions emerge – each having fundamental significance for the Bayesian benefit transfer exercise. First, how many observations must we process before our inferences are the ‘same’ – in a probabilistic sense – to the ones that would evolve had the full sample been available to the investigator at the outset? Second, how many observations must be processed before inferences stabilize? This rule, the rule of natural conjugate updating in Gaussian linear models, has roots that date at least to Wald (1947). Since then, a number of interesting examples have arisen with implications for the present investigation. A particularly interesting set of examples in the context of a mixture formulation is presented in Titterington, Smith and Makov (1985). The actual estimation procedure is a simple extension of the basic multinomial probit algorithm. Relegating details to the appendix, it consists of the essential steps:
\textbf{A4: Sequential Updating Algorithm.} Iterate $A_1$, sequentially adding observations in the vectors $y$, $z$, and in the covariate matrix $x$.

A set of formulae that proves extremely useful in this regard is the formulae that defines the relationship between prior and sample information during natural-conjugate updating of the unknown parameters $\theta \equiv \beta$. Given a prior pdf for $\beta$, $f^{mN}(\beta|\beta_t, C_{\beta t})$, where $\beta_t$ and $C_{\beta t}$ denote hyper-parameters obtained from a previous hypothetical sample, one can proceed to complete the square in $\beta$ and infer that the posterior pdf has the form that is proportional to the right-hand-side of the expression relating the prior and the new sample information

\begin{equation}
\pi(\theta|y_{t+1}, z_{t+1}) \propto \exp\{-\frac{1}{2}(z_{it+1} - x_{it+1}'\beta)'(z_{it+1} - x_{it+1}'\beta) \times \exp\{-\frac{1}{2}(\beta_{t+1} - \beta_t)'C_{\beta t}(\beta_{t+1} - \beta_t)\} \right.
\end{equation}

whereupon, completing the square in $\beta_{t+1}$ (see Zellner, pp 381), one finds that $\pi(\theta|y_{t+1}, z_{t+1}) \propto \exp\{-\frac{1}{2}(\beta - \beta_{t+1})'C_{\beta t+1}(\beta - \beta_{t+1})\}$, where $\beta_{t+1}$ and $C_{\beta t+1}$ are determined from the formulae in Zellner (B13a and B14a, respectively). The natural-conjugate updating framework also allows us to answer the ancillary thought experiments. Beginning at the study site, are the inferences obtained from the model that processes the entire data set disparate from those arrived at from processing the data sequentially and stopping at the end of the study-site observations set. And finally the following thought experiments serve to complete a set of benchmarks obtained from processing the entire sample. Suppose now that the order of the data sequence is reversed in the sense that the policy-site sample is processed first. Are the inferences obtained from processing the policy-site sample the same as those that would be derived from processing the entire sample simultaneously? Additionally, suppose now that the ordering of the data is randomized in the sense that each of the observations within the panel is chosen at random from the two sites; when is it sufficient to stop in the sense that the inferences obtained from the policy and study site are the same?

Figures 2 provides the answers to each of these questions posed above. In the figure there
are three lines. The horizontal line reports the posterior mean value obtained from processing the sample sequentially and drawing inferences at the end point of the policy site. This inference combines in a formal probabilistic manner all of the relevant information – both prior and sample information – that one would employ had they actually had access to the entire sample. This baseline (0.3905 in the figure) is the posterior mean from the pooled regression for the relevant partworth. The posterior mean differs substantially from the estimate obtained form the pooled processing of the sample. This divergence is inexplicable, at present and is, perhaps, dues to the lower sample size of the Gibbs iterations required to process the sequential sample. However, what is more germane to the present task is the fact that the sequential updates of the posterior mean for the partworth converge quickly to the endpoint value. This observation motivates two additional thought experiments. Of more interest than the value itself is how we arrive at it. And of considerably more interest is the following question in this context. At what point could the investigator stop gathering information such that the disparities between the predicted partworth value and the actual partworth value are more-or-less the same. In this context the vertical line in the graphic makes clear that after relatively few observations – perhaps in the order of 500 or so – the inference about the partworth value converges quite closely to the one obtained from processing the entire sample. In this regard, the vertical line at the sample value, 1008, is very relevant. It is the endpoint of the Genil sample and thus the beginning point for the Guadajoz sample. It is precisely the study-site-policy-site boundary. Thus an immediate conclusion available from the natural conjugate updating exercise is that, at least with respect to the partworth in question, there is relatively little bias induced in transferring inferences between the two sites.

In summary, across each of the models implemented within this section, there does not appear to be very significant inter-sample site variability, raising scope for a formal assessment of the precision with which a formal benefit transfer can be made. We take this
point up in the next section.

**Posterior Predictive Inference at the Study-Site-Policy-Site Boundary**

In this section, we present the formulae that we use in order to make an actual benefit transfer between the two sites. A related approach to exploring posterior inference across the two sites is to exploit the posterior predictive distribution for the data. The posterior predictive distribution is useful as an indicator of the fit of a model and whether outliers exist within the sample. One additional reason for employing this construction is that it enables the investigator to evaluate the posterior distribution of a hypothetical sample of observations generated by the same data generating experiment. The situation under consideration is the following. Given observed data \( y \), we seek inference about a hypothetical sample, say \( y_h \). We seek to characterize the form of the posterior predictive distribution for this hypothetical sample, derived after processing the observed sample containing only \( y \). To formalize things, let \( f(y_h|y) \) denote the density obtained for the prediction \( y_h \) after observing the sample data \( y \) and let \( \pi(\theta|y) \) denote the posterior pdf for \( \theta \) after observing the sample \( y \) (equation (3) above). A set of standard results (see, for examples, Zellner (1971), pp. 28-31; Koop (2003), p. 5; Gelman et al. (1995), pp. 8-9; Robert (2001), pp. 171-173; and Ibrahim, Chen and Sinha (2001), p. 18; Tanner (1993), pp. 90, 136-137; and Berger (1985), 94-95, 156-159) shows that the posterior predictive pdf for \( y_2 \) is obtained by marginalizing over the posterior pdf for \( \theta \) in the formula

\[
(9) \quad f(y_h|y) = \int_\theta f(y_h|y, \theta) \pi(\theta|y) \, d\theta.
\]

In this formula it is shown explicitly that the posterior predictive pdf can be viewed as an average of the conditional predictive pdf \( f(y_h|y, \theta) \) with the posterior pdf for the parameters \( \pi(\theta|y) \) serving as the weighting function. Of course, in our particular exercise we are interested primarily in the predictive distribution of the partworth, which one obtains in a logical manner, using the predictive pdf for the hypothetical sample. In particular, one can
obtain such inferences using the formula in (9) as follows. Generate data $y_h$ from a hypothetical sample according to (9). Subsequently, use the sample so generated to obtain an inference about $\theta$ conditional on this simulation, say $\pi(\theta_h|y_h)$. And then use the posterior so derived to simulate draws for the partworth in (5). At this point, a computational issue deserves emphasis. This is the fact that it is very easy to sample from (9) in the process of sampling from $\pi(\theta_h|y_h)$ as a by-product of the Gibbs algorithm. The presence of the complete data $z$ in $\pi(\theta|y,z)$ does not really complicate things very much. In short a robust set of inferences about the quantity in (9) can be derived by Gibbs sampling from the joint posterior conditional on the observed data, and using the posterior predictive formulae to derive inferences about (9).

At this point is seems appropriate to detail the steps involved in evaluating the partworth quantities derived under the simulated data. The process involves two steps.

First, we select a chosen ‘demography’ – a set of covariates – to mimic the scenario in question. In accordance with the notation just applied, we refer to the reference covariates as $x_h^{(m)}$, where $m = 1, 2, 3$, denote the experiment, or model in question. In the first case, we set $x_h^{(1)}$ as the actual set of covariates observed at the Guadajoz ‘policy’ site and evaluate the posterior predictive power of the model. In the second situation we set $x_h^{(2)}$ to simulate a two-fold increase in the real values in the Gudajoz covariates. In other words, we simulate a new set of covariates such that the numbers of rural jobs, the area of land encompassed by the project, and the tax amount levied in order to support the intervention have a total variation two times that existing in the observed policy-site data. In the third scenario we set $x_h^{(3)}$ so that the real values in the Guadajoz covariate set is three times the variation in the observed data. In each of the latter two simulations we retain the same binary structure to the binary attributes.

Second, given the simulated covariates $x_h^{(1)}$, $x_h^{(2)}$, $x_h^{(3)}$, we simulate predicted responses using the predictive distribution derived in (9). This turns out to be a simple matter
and is, in fact, one of the broader attractions frequently cited about the use of Markov chain Monte Carlo in applied Bayesian inference. The steps are three. Once again, using the data from the study site – the Genil data – we run a Gibbs sampler on the multinomial probit, using

\[ A_1 \]

Conditional on the estimates that we obtain for \( \theta \), and the simulated covariates \( x_h^{(m)} \), we derive a set of conditional predicted responses \( y_h^{(m)} \), which we record in comparison with the observed data in the Guadajoz data set. Second, using these predictions, and the set of simulated covariates, we derive to make an updated inference about the parameter vector which is relevant to the simulated setting, say \( \theta_h^{(m)} \). We then use this simulated parameter vector in another Gibbs draw to derive an estimate of the part-worth at the Guadajoz (policy) site. Repeating this procedure a large number of times provides a conditional posterior estimate of the simulated partworth quantity, which is the target of the exercise. Alternatively, the simulated data can be sued, just as we have done before, applying the now familiar hierarchical linear and finite mixtures methodologies to ascertain the actual divergence of the simulated data from the study-site observations.

**Results from three simulation exercises**

The simulated data is depicted in table 3. The first two columns report the means and standard deviations obtained from using the original Guadaz covariates in the estimation. In comparison with the response variables recorded in table 1, the predictive power of the model is immediately seen to be very robust. The predicted pattern of the response variables is almost identical to the observed data in the third and fourth columns of table 1. The remaining columns in table 3 depict the simulated covariate and inferred response variables from the second and third experiments. Of note is the alteration in the response measures at the base of the table, raising scope for evaluation of the robustness of the predicted transfer quantities. (Insert table 3 about here.)

Figure 3 depicts the posterior predictive pdf for the partworth using the predictive-distribution technique. The posterior pdfs obtained from the three simulations are very
similar. The posterior pdf for the base run (the blue indications) is almost identical to the posterior pdf for the second experiment (the green indications) which, in turn, is not too distinct from the posterior pdf reported under the third experiment (the red indications). Taken together, the results of the simulations exercise motivate two conclusions which are restricted to our data and the context in which the environmental evaluations are being motivated. First, because the plots indicate clearly that there is little divergence between the study and policy site inferences, we can conclude satisfactorily that the posterior predictive pdf in (9) seems to be able to mimic adequately the true value of the partworth derived from the ‘policy-site’ data. This is, of course, is the reason that posterior predictive densities are used so commonly in Bayesian inference. Second, at least with reference to the Genil-versus-Guadajoz comparisons, Bayesian benefit transfer does not appear to be too troubled by the extant level of unobserved intersite heterogeneity. To the extent that the detected levels of this heterogeneity appear to be relatively minor, this leaves open the door for additional nuanced analysis about the validity of our methods applied to more extreme circumstance. However, at the present level of generality, there do not appear to me too many deterrents from using the study-site-policy-site predictive inference in transferring benefit quantities between the study and the policy sites. Further analyses with additional data are necessary in order to evaluate whether this conclusion is idiosyncratic, or whether it is part of a conclusion that is applicable to a wide and broader set of conclusion. (Insert figure 3 about here.)

Conclusions

This paper develops alternative approaches to obtaining Bayesian benefit transfers. Each of the methods is a standard approach in a typical Bayesian investigation in which data evolve naturally at distinct locations. Broadly defined, the techniques refer to what we normally consider to be unobserved inter-site heterogeneity. The extensions we suggest and evaluate here are mandated by the non-availability – hypothetical in our Genil-Guadajoz samples – that data will not be available at the policy site. Inevitably, therefore, the techniques involve
prediction and inevitably simulations and subjective assessments about the nature and form of specific covariate information. Naturally inferences are conditioned by the covariate information that the investigator chooses to provide in his/her subjective assessment. The techniques explored here indicate a natural way to disentangle at least the extent of heterogeneity across sites and the implausibility of various predictions based only on the sample from the study site. Future work should aim to identify additional limitations of the sampling framework in an effort to improve the robustness of inferences derived from Bayesian benefit transfer simulations.

Appendix

The nine pdfs that we apply are the univariate-Normal pdf, $f^N(x|\mu, \sigma) \equiv (2\pi)^{-1/2} \sigma^{-1} \exp\{-\frac{1}{2} \sigma^{-2} (x-\mu)^2\}$, $-\infty < x < +\infty$, $-\infty < \mu < +\infty$, $0 < \sigma < +\infty$; the truncated-Normal pdf, $f^T(x|\mu, \sigma, I_{[a,b]}) \equiv (2\pi)^{-1/2} \sigma^{-1} \exp\{-\frac{1}{2} \sigma^{-2} (x-\mu)^2\} \Phi((b-\mu)/\sigma)-\Phi((a-\mu)/\sigma)\} \times I_{[a,b]}, I_{[a,b]}$ is the indicator function for the event $a < x < b$, $-\infty < \mu < +\infty$, $0 < \sigma < +\infty$, $\Phi(\cdot)$ denotes the cdf corresponding to the standard normal pdf; the multivariate-normal pdf $f^{MN}(x|\mu, \Sigma) \equiv (2\pi)^{-m/2} |\Sigma|^{-1/2} \exp\{-\frac{1}{2} (x-\mu)\Sigma^{-1}(x-\mu)\}$, $x \equiv (x_1, x_2, \ldots, x_m)$, $\mu \equiv (\mu_1, \mu_2, \ldots, \mu_m)$, $\Sigma$ is an $m \times m$ positive definite symmetric (pds) matrix $-\infty < x_i < +\infty$, $-\infty < \mu_i < +\infty$, $i = 1, 2, \ldots, m$; the matrix-normal pdf $f^{MX}(X|M, \Omega, \Xi) \equiv (2\pi)^{-pq/2} |\Omega|^{-p/2} |\Xi|^{-q/2} \exp\{-\frac{1}{2} \text{trace } \Omega^{-1} (X-M)\Xi^{-1} (X-M)\}$, $X \equiv (x_1, x_2, \ldots, x_q)$, $X_i \equiv (x_{1i}, x_{2i}, \ldots, x_{pi})$, $\mu \equiv (\mu_{1i}, \mu_{2i}, \ldots, \mu_{pi})$, $\Omega$ is a $p \times p$ pds matrix, $\Xi$ is a $q \times q$ pds matrix, $-\infty < x_{ij} < +\infty$, $-\infty < \mu_{ij} < +\infty$, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, q$; the inverted-Gamma pdf, $f^{IG}(x|v, s^2) \equiv (2/\Gamma(v/2)) (vs^2/2)^{v/2} (1/s^{v+1}) \exp\{-vs^2/2s^2\}$, $\Gamma(\cdot)$ denotes the gamma function (see, for example, Zellner, 1996, pp. 364-5), $-\infty < x < +\infty$, $0 < v < +\infty$, $0 < s^2 < +\infty$; the inverted-Wishart pdf $f^{IW}(X|H, \nu) \equiv k^{-1} |H|^{\nu/2} |G|^{(v+n-1)/2} \exp\{-\frac{1}{2} \text{trace } X^{-1} H\}$, $k \equiv 2^{v+n/2} \pi^{m(m+1)/4} \Gamma((v+1-i)/2)$, $\nu \geq m$, $H$ is an $m \times m$ pds matrix, $X$ is an $m \times m$ pds matrix containing $m(m+1)/2$ distinct elements; the multinomial distribution $f^M(x|\phi, j = 1, 2, \ldots, k, 0) \equiv$ - 30 -
The Dirichlet distribution \( f^D(\omega|\omega, j = 1, 2, ..., k) \equiv \Gamma(\omega_0) \prod_j \Gamma(\omega_j)^{-1} \prod_j \omega_j^{(\omega_0-1)} \), \( \omega_0 = \sum \omega_j = 1 \); and the uniform pdf \( f^u(x|\alpha, \beta) \equiv 1/(\alpha - \beta), \beta \leq x \leq \alpha, -\infty < \alpha < +\infty, -\infty < \beta < +\infty \).

A general specification assumes that there exist both choice-dependent and choice-independent covariate information. Thus, let \( w_{ijt} \equiv (w_{ij1}, w_{ij2}, ..., w_{ijP})' \) denote the \( P \)-vector of alternative-dependent covariates that affect the utility derived from choice of alternative \( j \); let \( x_{it} \equiv (x_{i1}, x_{i2}, ..., x_{iK})' \) denote the \( K \)-vector of alternative-independent covariates; and let \( \alpha_j \equiv (\alpha_{j1}, \alpha_{j2}, ..., \alpha_{jP})' \) and \( \beta \equiv (\beta_1, \beta_2, ..., \beta_K)' \) denote the corresponding coefficient vectors. Then, taking account of the panel, we focus attentions on the observational system

\[
(a_1) \quad z_{ijt} = w_{ijt}' \alpha_j + x_{it}' \beta + \varepsilon_{ijt}, \quad i = 1, 2, ..., N; \quad t = 1, 2, ..., T; \quad j = 1, 2, ..., M;
\]

where \( y_{it} = j \) if \( z_{ijt} \geq z_{ikt} \forall k \neq j, k = 1, 2, ..., M \) and \( y_{it} = 0 \) if \( z_{ijt} < 0 \). Stacking over the \( M \) alternatives, yields

\[
(a_2) \quad z_{it} = w_{it}' \alpha + x_{it}' \beta + \varepsilon_{it}, \quad i = 1, 2, ..., N; \quad t = 1, 2, ..., T;
\]

where \( z_{it} \equiv (z_{it1}, z_{it2}, ..., z_{itM})' \) denotes the \( M \)-vector of latent responses; \( w_{it} \) denotes the \( M \times PM \) block-diagonal arrangement of the alternative-dependent covariate vectors \( \{ w_{it1}', w_{it2}', ..., w_{itM}' \} \); \( \textbf{1}_M \) denotes the \( M \)-dimensional unit vector; and \( \varepsilon_{it} \equiv (\varepsilon_{it1}, \varepsilon_{it2}, ..., \varepsilon_{itm})' \) denotes an \( M \)-dimensional random disturbance vector with mean the null vector \( \textbf{0}_M \) and covariance matrix the positive-definite symmetric matrix \( \Sigma \). Stacking over the replications obtained from each respondent yields

\[
(a_3) \quad z_i = w_i \alpha + x_i \beta + \varepsilon_i, \quad i = 1, 2, ..., N;
\]

where \( z_i \equiv (z_{i1}', z_{i2}', ..., z_{iT}')' \) denotes the \( MT \)-vector of latent responses; \( w_i \equiv (w_{i1}', w_{i2}', ..., w_{im}') \) denotes the \( TM \times PM \) matrix of alternative-dependent covariates; \( x_i \equiv (x_{i1}', x_{i2}', ..., x_{iT}') \) denotes the \( TM \times K \) matrix of alternative-independent covariates; and \( \varepsilon_i \equiv (\varepsilon_{i1}, \varepsilon_{i2}, ..., \varepsilon_{iT})' \) denotes the \( TM \)-dimensional random disturbance vector. Finally, stacking over respondents leads to the system
(a4) \[ z = w \alpha + x \beta + \epsilon. \]

Test equation (1) follows form assuming that there is only choice-independent covariate information.

Details of algorithm A₁ follow from combining the complete-data likelihood, \( f(y|\theta,z) \), specified in text equation (3) with a multi-variate Normal prior for the regression coefficients, \( f^{\text{mN}}(\beta|\beta_o,C_{\beta_o}) \), using \( \pi(\theta|y,z) \propto f(y|\theta,z) \pi(\theta) \), and noting that the posterior has the form
\[
\pi(\theta|y,z) \propto \exp\{- \frac{1}{2} (z - x\beta)^T \Sigma^{-1} \mathbf{I}_{NT} (z - x\beta) \} \times \exp\{- \frac{1}{2} (\beta - \beta_o)^T C_{\beta_o} (\beta - \beta_o) \}.
\]

Where we have used the fact that the \((M-1) \times (M-1)\) psd covariance matrix \( \Sigma \) is a known quantity. Upon completing the square in \( \beta \), and taking account of the sign restrictions in \( z \), algorithm A₁ can be seen to consist of the steps: \( A_1: \text{Multinomial Probit Algorithm} \). Draw \( \beta \) from \( \pi(\beta|y,z) \equiv f^{\text{mN}}(\beta|\beta^*,C_{\beta^*}), \beta^* \equiv C_{\beta^*} (x'(\Sigma^{-1} \mathbf{I}_{NT}) z + C_{\beta_o}^{-1} \beta_o), \beta^* \equiv (x'(\Sigma^{-1} \mathbf{I}_{NT}) x + C_{\beta_o}^{-1})^{-1} \). Draw \( z \) from \( \pi(z|\beta,y) \equiv f^{\text{N}}(z|z^*,C_{z^*}), z^* \equiv x\beta, C_{z^*} \equiv \Sigma^{-1} \mathbf{I}_{NT} \) and truncate the draws according to the restrictions implied by the observed choice at each observation.

Details of algorithm A₂ follow from observing the posterior in text equation (6). Given
\[
\pi(\theta|y_1,y_2,z_1,z_2,\theta_1,\theta_2) \propto \prod_j \exp\{- \frac{1}{2} (z_j - x_j\beta_j)^T (z_j - x_j\beta_j) \} \times \exp\{- \frac{1}{2} (\beta - \beta_o)^T C_{\beta_o} (\beta - \beta_o) \},
\]
algorithm A₂ is seen to consist of the steps: \( A_2: \text{Hierarchical Normal Linear Modeling Algorithm} \). Draw \( \beta_1 \) from \( \pi(\beta_1|y_1,z_1) \equiv f^{\text{mN}}(\beta_1|\beta_{1^*},C_{\beta_{1^*}}), \beta_{1^*} \equiv C_{\beta_{1^*}} (x_1'(\Sigma^{-1} \mathbf{I}_{N_{1T}}) z_1 + C_{\beta_o}^{-1} \beta_o), C_{\beta_{1^*}} \equiv (x_1'(\Sigma^{-1} \mathbf{I}_{N_{1T}}) x_1 + C_{\beta_o}^{-1})^{-1} \). Draw \( \beta_2 \) from \( \pi(\beta_2|y_2,z_2) \equiv f^{\text{mN}}(\beta_2|\beta_{2^*},C_{\beta_{2^*}}), \beta_{2^*} \equiv C_{\beta_{2^*}} (x_2'(\Sigma^{-1} \mathbf{I}_{N_{2T}}) z_2 + C_{\beta_o}^{-1} \beta_o), C_{\beta_{2^*}} \equiv (x_2'(\Sigma^{-1} \mathbf{I}_{N_{2T}}) x_2 + C_{\beta_o}^{-1})^{-1} \). Draw \( \beta \) from \( \pi(\beta|\beta_1,\beta_2) \equiv f^{\text{mN}}(\beta|\beta^*,C_{\beta^*}), \beta^* \equiv C_{\beta^*} (C_{\beta_o}^{-1} \beta_1 + C_{\beta_o}^{-1} \beta_2 + C_{\beta_o}^{-1} \beta_o), C_{\beta^*} \equiv (C_{\beta_o}^{-1} + C_{\beta_o}^{-1} + C_{\beta_o}^{-1})^{-1} \). Draw \( z_1 \) from \( \pi(z_1|\beta_1,y_1) \equiv f^{\text{N}}(z_1|z_{1^*},\beta_{1^*}), z_{1^*} \equiv x_1\beta_1, \beta_{1^*} \equiv \Sigma^{-1} \mathbf{I}_{N_{1T}}; \) draw \( z_2 \) from \( \pi(z_2|\beta_2,y_2) \equiv f^{\text{N}}(z_2|z_{2^*},\beta_{2^*}), z_{2^*} \equiv x_2\beta_2, \beta_{2^*} \equiv \Sigma^{-1} \mathbf{I}_{N_{2T}}; \) truncate the draws according to the restrictions implied by the observed choice at each observation.

Details of algorithm A₃ follow from observing the posterior in text equation (7). Given
\[
\pi(\theta|y,z,x,\beta,\omega) \propto \prod_j \beta_j \exp\{- \frac{1}{2} (z_j - x_j\beta_j)^T (z_j - x_j\beta_j) \} \times \exp\{- \frac{1}{2} (\beta_j - \beta_{jo})^T C_{\beta_{jo}^{-1}} (\beta_j - \beta_{jo}) \},
\]
\( A_3 \) consists of the steps \( A_3: \text{Finite Mixtures Modeling Algorithm} \). Given classification between the two groups, \( \chi \), let \( N_1 \) denote the number of observations in subset-1 and let \( N_2 \) denote the sample in subset-2. In addition, partition, \( y = (y_1', y_2') \), \( x = (x_1', x_2') \) and \( z = (z_1', z_2') \) and define \( \beta_1 \) as the set of coefficients conditioning responses in the first subset and define \( \beta_2 \) as the set of coefficients conditioning responses in the second subset. And use \( \omega \equiv (\omega_1', \omega_2') \) to denote the ‘mixture weights’ appearing in (a7). Then, draw \( \beta_1 \) from \( \pi(\beta_1|y_1,z_1) \equiv f^{mN}(\beta_1|\beta_{1*},C_{\beta_{1*}}) \), \( \beta_{1*} \equiv C_{\beta_{1*}}(x_1'(|\Sigma^{-1}\otimes I_{N1T})z_{1*} + C_{\beta_{10}^{-1}}\beta_{10}^0) \), \( C_{\beta_{1*}} \equiv (x_1'(|\Sigma^{-1}\otimes I_{N1T})x_1 + C_{\beta_{10}^{-1}})^{-1} \). Draw \( \beta_2 \) from \( \pi(\beta_2|y_2,z_2) \equiv f^{mN}(\beta_2|\beta_{2*},C_{\beta_{2*}}) \), \( \beta_{2*} \equiv C_{\beta_{2*}}(x_2'(|\Sigma^{-1}\otimes I_{N2T})z_{2*} + C_{\beta_{20}^{-1}}\beta_{20}^0) \), \( C_{\beta_{2*}} \equiv (x_2'(|\Sigma^{-1}\otimes I_{N2T})x_2 + C_{\beta_{20}^{-1}})^{-1} \). Draw \( z_1 \) from \( \pi(z_1|\beta_{1*},y_1) \equiv f^{N}(z_1|z_{1*},C_{z_{1*}}) \), \( z_{1*} \equiv x_1\beta_{1*} \), \( C_{z_{1*}} \equiv \sum \) \( ^{-1}\otimes I_{N1T} \); draw \( z_2 \) from \( \pi(z_2|\beta_{2*},y_2) \equiv f^{N}(z_2|z_{2*},C_{z_{2*}}) \), \( z_{2*} \equiv x_2\beta_{2*} \), \( C_{z_{2*}} \equiv \sum \) \( ^{-1}\otimes I_{N2T} \); truncate the draws according to the restrictions implied by the observed choice at each observation. Draw \( \omega_1 \) from \( \pi(\omega_1|\chi) \equiv f^{D}(\omega_1|\alpha_{1*},\alpha_{2*}) \), draw \( \omega_2 \) from \( \pi(\omega_2|\chi) \equiv f^{D}(\omega_2|\alpha_{1*},\alpha_{2*}) \), \( \alpha_{1*} \equiv (N_1+\alpha_o) \div (N_1+N_2+\alpha_o) \), \( \alpha_{2*} \equiv (N_2+\alpha_o) \div (N_1+N_2+\alpha_o) \), and \( \alpha_o \) defines a prior belief about the parameters in the natural conjugate pdf for the parameters of the Dirichlet distribution. Finally, each row and column entry in the NTxM matrix of probabilities leading to the classification variable \( \chi \) has probability proportional to the kernel of the relevant Normal distribution, namely \( \phi_{ij} \propto \omega_j^{1/2} \pi^{-M} |\Sigma|^{-1/2} \exp\left\{-1/2(z_{ij'}-x_i'\beta_j)^T\Sigma^{-1}(z_{ij'}-x_i'\beta_j)\right\} \).

Finally, the algorithm for estimating the parameters evolving from the natural-conjugate, sequential updating specification follows naturally from completing the square in text equation (8), namely

(a8) \( \pi(\theta_{t+1}|y_{t+1},z_{t+1}) \propto \exp\{-1/2(z_{dt+1}-x_{dt+1}|\beta_{t+1}^d)^T\Sigma^{-1}(z_{dt+1}-x_{dt+1}|\beta_{t+1}^d)\} \exp\{-1/2(\beta_{t+1}-\beta_0)^T C_{\beta^{-1}}(\beta_{t+1}-\beta_0)\} \).

Consequently, we have \( A_4: \text{Sequential Updating Algorithm} \). Draw \( \beta_{t+1} \) from \( \pi(\beta_{t+1}|y_{t+1},z_{t+1}) \equiv f^{mN}(\beta_{t+1}|\beta_{t+1*},C_{\beta_{t+1*}}) \), \( \beta_{t+1*} \equiv C_{\beta_{t+1*}}(x_{t+1}'(\Sigma^{-1}z_{t+1} + C_{\beta^{1_0}}^{-1}\beta_0), \, C_{\beta_{t+1*}} \equiv (x_{t+1}'(\Sigma^{-1}x_{t+1} + C_{\beta^{1_0}}^{-1})^{-1} \). Draw \( z_{t+1} \) from \( \pi(z_{t+1}|\beta_{t+1*},y_{t+1}) \equiv f^{N}(z_{t+1}|z_{t+1*},C_{z_{t+1*}}) \), \( z_{t+1*} \equiv x_{t+1}\beta_{t+1}* \), \( C_{z_{t+1*}} \equiv \sum \) and truncate the draws according to the restrictions implied by the observed choice at each observation.
Footnotes

1 As pointed out by Bergland (2006) the practice of benefit transfer has a long history of being conducted behind closed doors, but only recently has been subjected to academic reviews. The publication of a special issue of Water Research Resource in March of 1992 and the AERE workshop on “benefit transfer” held in June of 1992, clearly place the issue of benefit transfer on the agenda.

2 Defined as ((transferred estimate - original site estimate)/original site estimate) * 100.

3 The Environmental Valuation Reference Inventory database (EVRI) has currently 1300 studies and it is growing at the rate at about 300 studies per year.

4 All the surveys where respondents expressed a protest answer were excluded from the analysis. The main reasons of protest were the lack of confidence in the government’s capacity to carry out the proposed project or because they thought they should not have to pay to reduce soil erosion since it is a government duty.

5 A small, but important point here is that, by stacking the data in the said order, the error structure has a particularly appealing form that simplifies computational demands during estimation. This point is ably illustrated with reference to the Seemingly-Unrelated-Regressions situation in Koop (pp. 138 and 22).

6 A good deal of discussion (references) has centred around the ways to obtain efficient one-for-one draws from truncated-Normal densities. The approach used throughout this paper is the one advocated in Chib (1992) in the context of censored regression as outlined originally in Geweke (Valencia meetings paper) and uses the probability-integral transform technique. Specifically, for the scalar normal random variate $z$, distributed $f_{\mathcal{N}}(z|\mu,\sigma,I_{[\alpha,\beta]})$, we draw $u$ from $f_{U}(u|0,1)$ and set $z = \Phi^{-1}((u\times(\Phi((\beta-\mu)/\sigma)-\Phi((\alpha-\mu)/\sigma)+\Phi((\alpha-\mu)/\sigma)-\Phi((\alpha-\mu)/\sigma)))$, in which $\Phi(\cdot)$ denotes the cdf corresponding to the standard normal distribution function and $\Phi^{-1}(\cdot)$ denotes its inverse.

References


Dempster, A. P. Elements of Continuous Multivariate Analysis. Reading; Addison-Wesley, 1969.


Edward Elgar..


*Stochastic Environmental Research and Risk Assessment* 22:185-206


76:609-32.


Yu, B. and P. Mykland. “Looking at Markov Samplers through CUSUM Path Plots; A Simple


Journal of Econometrics 137:260-76.


### Table 1. Summary statistics

<table>
<thead>
<tr>
<th>Definition</th>
<th>Description</th>
<th>Genil</th>
<th>Guadajoz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desertification</td>
<td>Slight improvement</td>
<td>0.35</td>
<td>0.32</td>
</tr>
<tr>
<td>Desertification</td>
<td>Moderate improve.</td>
<td>0.31</td>
<td>0.34</td>
</tr>
<tr>
<td>Water quality</td>
<td>Medium</td>
<td>0.33</td>
<td>0.32</td>
</tr>
<tr>
<td>Water quality</td>
<td>High</td>
<td>0.35</td>
<td>0.33</td>
</tr>
<tr>
<td>Flora density</td>
<td>Medium</td>
<td>0.33</td>
<td>0.32</td>
</tr>
<tr>
<td>Flora density</td>
<td>High</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>Rural jobs creation</td>
<td>100, 200; 65, 130</td>
<td>99.11</td>
<td>64.72</td>
</tr>
<tr>
<td>Area km²</td>
<td>0, 330, 660, 990 km²</td>
<td>663.60</td>
<td>309.77</td>
</tr>
<tr>
<td>Tax rates</td>
<td>0, 6, 12, 18, 24, 30, 36 Euros</td>
<td>20.83</td>
<td>21.25</td>
</tr>
<tr>
<td>Responses</td>
<td>Chosen Alternative</td>
<td>2.26</td>
<td>2.26</td>
</tr>
<tr>
<td>Definition</td>
<td>Pooled</td>
<td>Chow</td>
<td>Hierarchical</td>
</tr>
<tr>
<td>-------------</td>
<td>--------</td>
<td>-------</td>
<td>--------------</td>
</tr>
<tr>
<td>Landscape2</td>
<td>0.61</td>
<td>0.49</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>0.51</td>
<td>0.72</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.87</td>
</tr>
<tr>
<td>Landscape3</td>
<td>0.76</td>
<td>0.74</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>0.66</td>
<td>0.87</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.69</td>
</tr>
<tr>
<td>Water2</td>
<td>0.58</td>
<td>0.48</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>0.47</td>
<td>0.69</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.24</td>
</tr>
<tr>
<td>Water3</td>
<td>0.86</td>
<td>0.71</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.96</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.62</td>
</tr>
<tr>
<td>Flora2</td>
<td>0.34</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>0.24</td>
<td>0.45</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-1.89</td>
</tr>
<tr>
<td>Flora3</td>
<td>0.46</td>
<td>0.46</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>0.36</td>
<td>0.57</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.01</td>
</tr>
<tr>
<td>Jobs</td>
<td>0.76</td>
<td>0.66</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>0.64</td>
<td>0.89</td>
<td>0.51</td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Area</td>
<td>0.28</td>
<td>0.34</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>0.12</td>
<td>0.44</td>
<td>0.13</td>
</tr>
<tr>
<td>Tax</td>
<td>-0.93</td>
<td>-0.88</td>
<td>-0.89</td>
</tr>
<tr>
<td></td>
<td>-1.10</td>
<td>-0.77</td>
<td>-1.12</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.57</td>
<td>-1.58</td>
<td>-1.58</td>
</tr>
<tr>
<td></td>
<td>-1.76</td>
<td>-1.38</td>
<td>-1.88</td>
</tr>
<tr>
<td>Landscape2</td>
<td>0.80</td>
<td>0.80</td>
<td>-0.17</td>
</tr>
<tr>
<td></td>
<td>0.64</td>
<td>0.34</td>
<td>0.64</td>
</tr>
<tr>
<td>Landscape3</td>
<td>0.86</td>
<td>0.86</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>0.71</td>
<td>0.59</td>
<td>0.71</td>
</tr>
<tr>
<td>Water2</td>
<td>0.74</td>
<td>0.74</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>0.57</td>
<td>0.32</td>
<td>0.57</td>
</tr>
<tr>
<td>Water3</td>
<td>1.07</td>
<td>1.07</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>0.91</td>
<td>0.56</td>
<td>0.91</td>
</tr>
<tr>
<td>Flora2</td>
<td>0.41</td>
<td>0.41</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.26</td>
<td>0.20</td>
<td>0.26</td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Flora3</td>
<td>0.57</td>
<td>0.57</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>0.41</td>
<td>0.32</td>
<td>0.41</td>
</tr>
<tr>
<td>Jobs</td>
<td>1.26</td>
<td>1.26</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>1.02</td>
<td>0.51</td>
<td>1.02</td>
</tr>
<tr>
<td>Area</td>
<td>1.63</td>
<td>1.63</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>1.16</td>
<td>0.13</td>
<td>1.16</td>
</tr>
<tr>
<td>Tax</td>
<td>-1.19</td>
<td>-1.19</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>-1.44</td>
<td>-1.12</td>
<td>-1.44</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.19</td>
<td>-2.19</td>
<td>-0.39</td>
</tr>
<tr>
<td></td>
<td>-2.50</td>
<td>-1.88</td>
<td>-2.50</td>
</tr>
<tr>
<td>Landscape2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Landscape3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 48 -
<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water3</td>
<td>-0.73</td>
<td>1.55</td>
</tr>
<tr>
<td>Flora2</td>
<td>-0.53</td>
<td>1.73</td>
</tr>
<tr>
<td>Flora3</td>
<td>-0.75</td>
<td>1.46</td>
</tr>
<tr>
<td>Jobs</td>
<td>-0.05</td>
<td>1.79</td>
</tr>
<tr>
<td>Area</td>
<td>-0.45</td>
<td>1.82</td>
</tr>
<tr>
<td>Tax</td>
<td>-1.84</td>
<td>0.43</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.4</td>
<td>-0.09</td>
</tr>
<tr>
<td>Alpha1</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Alpha2</td>
<td></td>
<td>0.50</td>
</tr>
<tr>
<td>Correct</td>
<td>0.46</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.24</td>
</tr>
<tr>
<td>Definition</td>
<td>Status Quo</td>
<td>Experiment 1</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Desertification</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>Desertification</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>Water quality</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>Water quality</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>Flora density</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>Flora density</td>
<td>0.34</td>
<td>0.34</td>
</tr>
<tr>
<td>Rural jobs creation</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>Area km²</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>Tax rates</td>
<td>0.59</td>
<td>0.59</td>
</tr>
<tr>
<td>Responses</td>
<td>0.62</td>
<td>0.62</td>
</tr>
</tbody>
</table>
Figure 1. Partworth benchmark pdfs from pooled (blue) and Chow (red and green) specifications.
Figure 2. Natural conjugate updating and partworth policy-versus-study site boundary evaluation.
Figure 3. Posterior pdfs for the partworth from experiments one (blue), two (green) and three (red).