JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **20**: 723–747 (2005)
Published online 9 June 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jae.810

SEMIPARAMETRIC BAYESIAN INFERENCE IN MULTIPLE EQUATION MODELS

GARY KOOP,[a]* DALE J. POIRIER[b] AND JUSTIN TOBIAS[b,c]

[a] *Department of Economics, University of Leicester, Leicester, UK*
[b] *Department of Economics, University of California at Irvine, Irvine, CA, USA*
[c] *Department of Economics, Iowa State University, Ames, IA, USA*

SUMMARY

This paper outlines an approach to Bayesian semiparametric regression in multiple equation models which can be used to carry out inference in seemingly unrelated regressions or simultaneous equations models with nonparametric components. The approach treats the points on each nonparametric regression line as unknown parameters and uses a prior on the degree of smoothness of each line to ensure valid posterior inference despite the fact that the number of parameters is greater than the number of observations. We develop an empirical Bayesian approach that allows us to estimate the prior smoothing hyperparameters from the data. An advantage of our semiparametric model is that it is written as a seemingly unrelated regressions model with independent normal–Wishart prior. Since this model is a common one, textbook results for posterior inference, model comparison, prediction and posterior computation are immediately available. We use this model in an application involving a two-equation structural model drawn from the labour and returns to schooling literatures. Copyright © 2005 John Wiley & Sons, Ltd.

1. INTRODUCTION

Despite the proliferation of methods for semiparametric and nonparametric regression, the use of these techniques remains relatively rare in empirical practice. Increased computational difficulty and mathematical sophistication, and perhaps most importantly, the *curse of dimensionality*—wherein the rate of convergence of the nonparametric regression estimator slows with the number of variables treated nonparametrically—all seem to provide barriers which prevent the widespread use of nonparametric techniques.

The rapid increase in computing power and growth in nonparametric routines found in statistical software packages has helped to mitigate computational concerns. To combat the curse of dimensionality problem, many researchers have adopted the use of the *partially linear* (or *semilinear*) regression model. This model, though not fully nonparametric, provides a convenient generalization of the standard linear model which is not as susceptible to the curse of dimensionality since only one, or perhaps a few, variables are treated nonparametrically. Finally, some studies (e.g. Blundell and Duncan, 1998; Yatchew, 1998; DiNardo and Tobias, 2001) have tried to bridge the gap between theory and practice and make these techniques accessible to applied researchers.

In this paper we continue in this tradition and describe and implement simple and intuitive Bayesian methods for semiparametric and nonparametric regression. Importantly, the methods we describe can be used in the context of multiple equation models, thus generalizing the class of models for which simple Bayesian semiparametric methods are available. In our discussion

* Correspondence to: Professor Gary Koop, Department of Economics, University of Leicester, University Road, Leicester LE1 7RH, UK. E-mail: gary.koop@leicester.ac.uk

Received 24 June 2003
Revised 27 April 2004

we focus primarily on the seemingly unrelated regression (SUR) model. This model is of interest in and of itself, but is also of interest as the (possibly restricted) reduced form of a semiparametric simultaneous equations model (or the structural form of a triangular simultaneous equations model).

Before describing the contributions of this paper, it is useful to briefly review a simple method used in related work (Koop and Poirier, 2004a) in the single-equation partially linear regression model. This partially linear model divides the explanatory variables into a set which is treated parametrically, z, and a set which is treated nonparametrically, x, and relates them to a dependent variable y as:

$$y_i = z_i'\beta + f(x_i) + \varepsilon_i$$

for $i = 1, \ldots, N$, where $f(\cdot)$ is an unknown function. Because of the curse of dimensionality, x_i must be of low dimension and is often a scalar (see Yatchew, 1998 for an excellent introduction to the partial linear model). For most of this paper we will assume x_i is a scalar, although this assumption is relaxed in Section 4.

In this model we assume $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$ for $i = 1, \ldots, N$, and all explanatory variables are fixed or exogenous. Observations are ordered so that $x_1 < x_2 < \cdots < x_N$. Define $y = (y_1, \ldots, y_N)'$, $Z = (z_1, \ldots, z_N)'$ and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)'$. Letting $\gamma = (f(x_1), \ldots, f(x_N))'$, $W = (Z : I_N)$ and $\delta = (\beta', \gamma')'$, Koop and Poirier (2004a) show that the previous equation could be written as:

$$y = W\delta + \varepsilon$$

Thus, the partially linear model can be written as the standard normal linear regression model where the unknown points on the nonparametric regression line are treated as unknown parameters. This regression model is characterized by insufficient observations in that the number of explanatory variables is greater than N. However, Koop and Poirier (2004a) showed that, if a natural conjugate prior is used, the posterior is still well-defined. In fact, they showed that the natural conjugate prior did not even have to be informative in all dimensions and that prior information about the smoothness of the nonparametric regression line was all that was required to ensure valid posterior inference. Thus, for the subjective Bayesian, prior information can be used to surmount the problem of insufficient observations. Furthermore, for the researcher uncomfortable with subjective prior information, the required amount of prior information was quite small, involving the selection of a single prior hyperparameter called η that governed the smoothness of the nonparametric regression line. Koop and Poirier (2004b) went even further and showed how (under weak conditions) empirical Bayesian methods could be used to estimate η from the data.

The advantages of remaining within the framework of the normal linear regression model with a natural conjugate prior are clear. This model is very well understood and standard textbook results for estimation, model comparison and prediction are immediately available. Analytical results for posterior moments, marginal likelihoods and predictives exist and, thus, there is no need for posterior simulation. This means methods which search over many values for η (e.g. empirical Bayesian methods or cross-validation) can be implemented at a low computational cost. Furthermore, as shown in Koop and Poirier (2004a), the partial linear model can serve as a component in numerous other models which do involve posterior simulation (e.g. semiparametric tobit and probit models or the partial linear model with the errors treated flexibly by using mixtures of normals). The ability to simplify the estimation of the nonparametric component in such a complicated empirical exercise may provide the researcher with great computational benefit.

In this paper we take up the case of Bayesian semiparametric estimation in multiple equation models and adopt a similar approach for smoothing the regression functions. In particular, we consider the estimation of a semiparametric SUR model of the form:

$$y_{ij} = z'_{ij}\beta_j + f_j(x_{ij}) + \varepsilon_{ij} \tag{1}$$

where y_{ij} is the ith observation ($i = 1, \ldots, N$) on the endogenous variable in the jth equation ($j = 1, \ldots, m$), z_{ij} is a $k_j \times 1$ vector of observations on the exogenous variables which enter linearly, $f_j(x_{ij})$ is an unknown function which depends on a vector of variables, x_{ij}, and ε_{ij} is the error term. For equations which have nonparametric components, z_{ij} does not contain an intercept since the first point on a nonparametric regression line plays the role of an intercept.

The approach we describe for the estimation of this model is simple and intuitive and, hopefully, will appeal to practitioners seeking to add flexibility to their multiple equation analyses. As in Koop and Poirier (2004a,b), we employ a prior which serves to smooth the nonparametric regression functions. It is important to recognize that for the (parametric) seemingly unrelated regressions model (and the reduced form of the simultaneous equations model), the natural conjugate prior suffers from well-known criticisms (see Rothenberg, 1963 or Dreze and Richard, 1983). On the basis of these, Dreze and Richard (1983, p. 541) argue against using the natural conjugate prior (except for certain noninformative limiting cases not relevant for our class of models). Their arguments carry even more force in the present semiparametric context since the natural conjugate prior places some undesirable restrictions on the way smoothing is carried out on nonparametric regression functions in different equations (i.e., the nonparametric component in each equation is smoothed in the same way). Thus, in the present paper we do not adopt a natural conjugate prior, but rather use an independent normal–Wishart prior.

The basic ideas behind our approach are straightforward extensions of standard textbook Bayesian methods for the SUR model (see, e.g., Koop, 2003, pp. 137–142). Thus, textbook results for estimation, model comparison (including comparison of parametric to nonparametric models), model diagnostics (e.g. posterior predictive p-values) and prediction are immediately available. This, we argue, is an advantage relative to the relevant non-Bayesian literature (see, e.g., Pagan and Ullah, 1999, chapter 6; Newey *et al.*, 1999; Darolles *et al.*, 2003) and to other, more complicated, Bayesian approaches to nonparametric seemingly unrelated regression such as Smith and Kohn (2000).

We illustrate the use of our methods by estimating a two-equation simultaneous equations model in parallel with the development of our theory. This application takes data from the National Longitudinal Survey of Youth and involves estimating the returns to schooling, job tenure and ability for a cross-sectional sample of white males. Our triangular simultaneous equations model has two equations, one for the (log) wage and the other for the quantity of schooling attained. After estimating standard parametric models that have appeared in the literature, we first extend them to allow for nonparametric treatment of an exogenous variable (weeks of tenure on the current job) in the wage equation (Case 1). Subsequently, we consider Case 2 where single explanatory variables enter nonparametrically in each equation. In this model we additionally allow a measure of cognitive ability to enter the schooling equation nonparametrically. We complete our empirical work with Case 3 by giving cognitive ability a nonparametric treatment in both the wage and schooling equations (with tenure on the job also given a nonparametric treatment in the wage equation).

Our results reveal the practicality and usefulness of our approach. In some cases, our semiparametric treatment yields results which are very similar to those from simple parametric nonlinear models (e.g. quadratic). However, one advantage of a semiparametric approach is that a particular functional form such as the quadratic does not have to be chosen, either in an *ad hoc* fashion or through pre-testing. Furthermore, in some cases our semiparametric approach yields empirical results that could not easily be obtained using standard parametric methods. In terms of our application, our results reveal the empirical importance of controlling for nonlinearities in ability, particularly in the schooling equation, when trying to estimate the return to education.

The outline of our paper is as follows. In the next section, we outline our basic semiparametric SUR model, describe our data, and obtain parametric results and semiparametric results for a model where job tenure is treated nonparametrically. In Section 3, we describe the process of estimating a model with nonparametric components in both equations, and estimate the model in Case 2. Finally, in Section 4, we describe how to handle the estimation of additive models and provide estimation results for our most general Case 3. The paper concludes with a summary in Section 5.

2. CASE 1: A SINGLE NONPARAMETRIC COMPONENT IN A SINGLE EQUATION

We begin by considering a simplified version of equation (1) where a nonparametric component enters a single equation (chosen to be the mth equation) and the explanatory variable which receives a nonparametric treatment, x_{im}, is a scalar. In later sections, we consider cases where several equations have nonparametric components each depending on a different explanatory variable (or variables).

We assume that the data is ordered so that $x_{1m} < \cdots < x_{Nm}$ and define $\gamma_i = f_m(x_{im})$ for $i = 1, \ldots, N$ to be the unknown points on the nonparametric regression line in the mth equation. We also let $\gamma = (\gamma_1, \ldots, \gamma_N)'$ and ζ_i be the ith row of I_N. With these definitions, we can write the model as:

$$y_i = W_i\delta + \varepsilon_i \tag{2}$$

where $y_i = (y_{i1}, \ldots, y_{im})'$, $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{im})'$,

$$W_i = \begin{bmatrix} z_{i1}' & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & z_{i2}' & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & z_{i,m-1}' & 0 & 0 \\ 0 & \cdot & \cdot & 0 & z_{im}' & \zeta_i \end{bmatrix}$$

and $\delta = (\beta_1', \ldots, \beta_m', \gamma')'$ is a $K + N$ vector where $K = \sum_{j=1}^{m} k_j$. For future reference, we define the partition $W_i = [W_i^{(1)} : W_i^{(2)}]$ where $W_i^{(1)}$ is an $m \times (K + 2)$ matrix and $W_i^{(2)}$ is $m \times (N - 2)$. The likelihood for this model is defined by assuming $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, \Sigma)$.

We define smoothness according to second differences of points on the nonparametric regression line. In light of this, it proves convenient to transform the model. Define the $(N-2)\times N$ second-differencing matrix as:

$$D=\begin{bmatrix}1 & -2 & 1 & \cdot & \cdot & \cdot & \cdot & 0\\ 0 & 1 & -2 & 1 & 0 & \cdot & \cdot & 0\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ 0 & 0 & \cdot & \cdot & \cdot & 1 & -2 & 1\end{bmatrix} \tag{3}$$

so that $D\gamma$ is the vector of second differences, $\Delta^2\gamma_i$. Prior information about smoothness of the nonparametric regression line will be expressed in terms of $R\delta$, where the $(N-2)\times(K+N)$ matrix $R=[0_{(N-2)\times K}:D]$. For future reference, we partition $R=[R_1:R_2]$ where R_1 is an $(N-2)\times(K+2)$ matrix and R_2 is $(N-2)\times(N-2)$ (i.e., the nonsingular matrix R_2 is D with the first two columns deleted). Note that other degrees of differencing can be handled by redefining equation (3) as appropriate (see, e.g., Yatchew, 1998, pp. 695–698 or Koop and Poirier, 2004a).

Using standard transformations (see, e.g., Poirier, 1995, pp. 503–504), equation (2) can be written as:

$$y_i=V_i^{(1)}\lambda_1+V_i^{(2)}\lambda_2+\varepsilon_i=V_i\lambda+\varepsilon_i \tag{4}$$

where $\lambda=(\lambda_1',\lambda_2')'$, $\lambda_1=(\beta_1',\ldots,\beta_m',\gamma_1,\gamma_2)'$, $\lambda_2=D\gamma$, $V_i^{(1)}=W_i^{(1)}-W_i^{(2)}R_2^{-1}R_1$ and $V_i^{(2)}=W_i^{(2)}R_2^{-1}$. Note that λ_2 is the vector of second differences of the points on the nonparametric regression line and it is on this parameter vector that we place our smoothness prior.

We use an independent normal–Wishart prior for λ and Σ^{-1} which is a common choice for the parametric SUR model (see, e.g., Chib and Greenberg, 1995, 1996). Thus,

$$\lambda\sim N(\underline{\lambda},\underline{V}_\lambda) \tag{5}$$

and

$$\Sigma^{-1}\sim W(\underline{V}_\Sigma^{-1},\underline{v}) \tag{6}$$

where $W(\underline{V}_\Sigma,\underline{v})$ denotes the Wishart distribution (see, e.g., Poirier, 1995, p. 136).

Our computational work is based on a Gibbs sampler involving $p(\lambda|y,\Sigma^{-1})$ and $p(\Sigma^{-1}|y,\lambda)$. Straightforward manipulations show these to be:

$$\lambda|y,\Sigma^{-1}\sim N(\overline{\lambda},\overline{V}_\lambda) \tag{7}$$

and

$$\Sigma^{-1}|y,\lambda\sim W(\overline{V}_\Sigma^{-1},\overline{v}) \tag{8}$$

where

$$\overline{v}=\underline{v}+N \tag{9}$$

$$\overline{V}_\Sigma^{-1}=\left[\underline{V}_\Sigma+\sum_{i=1}^{N}(y_i-V_i\lambda)(y_i-V_i\lambda)'\right]^{-1} \tag{10}$$

$$\overline{V}_\lambda=\left(\underline{V}_\lambda^{-1}+\sum_{i=1}^{N}V_i'\Sigma^{-1}V_i\right)^{-1} \tag{11}$$

and

$$\overline{\lambda} = \overline{V}_\lambda \left(\underline{V}_\lambda^{-1} \underline{\lambda} + \sum_{i=1}^{N} V_i' \Sigma^{-1} y \right) \tag{12}$$

Of course, many values may be selected for the prior hyperparameters, $\underline{\lambda}$, $\underline{V}_\lambda$, $\underline{V}_\Sigma^{-1}$ and $\underline{\nu}$. Here we describe a particular prior elicitation strategy that requires a minimal amount of subjective prior information. We assume

$$\underline{V}_\lambda = \begin{bmatrix} \underline{V}_1 & 0 \\ 0 & V(\eta) \end{bmatrix} \tag{13}$$

where $\underline{V}_1$ and $V(\eta)$ are the prior covariance matrices for λ_1 and λ_2, respectively. We set $\underline{V}_1^{-1} = 0$, the noninformative choice. Since $\lambda_2 = D\gamma$ is the vector of second differences of points on the nonparametric regression line, $V(\eta)$ controls its degree of smoothness. We assume $V(\eta)$ depends on a scalar parameter, η. As discussed in Koop and Poirier (2004a), several sensible forms for $V(\eta)$ can be chosen. In this paper, we set $V(\eta) = \eta I_{N-2}$. We also set $\underline{\lambda} = 0_{K+N}$. Note that these assumptions imply we are noninformative about $\lambda_1 = (\beta_1', \ldots, \beta_m', \gamma_1, \gamma_2)'$, but have an informative prior for the remaining parameters which reflect the degree of smoothness in the nonparametric regression line. Our information about this smoothness is of the form: $\Delta^2 \gamma_i \sim N(0, \eta)$ for $i = 3, \ldots, N$.[1]

In this paper we adopt an empirical Bayesian approach where η is chosen so as to maximize the marginal likelihood. However, it is worth noting that η could be treated either as a prior hyperparameter to be selected by the researcher or as a parameter in a hierarchical prior. If the latter approach were adopted, η could be integrated out of the posterior. Our empirical Bayesian approach is equivalent to this hierarchical prior approach using a noninformative flat prior for η (and plugging in the posterior mode of η instead of integrating out this parameter).[2]

The results of Fernandez *et al.* (1997) imply that an informative prior is required for Σ^{-1} in order to ensure propriety of the posterior. However, in related work with a single-equation model (see Koop and Poirier, 2004b), we found that use of a proper, but relatively noninformative prior on a similar nuisance parameter yielded sensible (and robust) results. Accordingly, we set $\underline{\nu} = 10$ in our application.[3] Using the properties of the Wishart distribution, we obtain the prior mean $E(\sigma_{ij}^{-1}) = \underline{\nu} \underline{V}_{\Sigma_{ij}}^{-1}$, where σ_{ij}^{-1} and $\underline{V}_{\Sigma_{ij}}^{-1}$ are the ijth elements of Σ^{-1} and $\underline{V}_\Sigma^{-1}$, respectively. To centre the prior correctly, we calculate the OLS estimate of Σ based on a parametric SUR model where all variables (including x_{im}) enter linearly. We set $\nu \underline{V}_\Sigma^{-1}$ equal to the inverse of this OLS estimate.

In order to compare models or estimate/select η in our empirical Bayesian approach, the marginal likelihood (for a given value of η) is required. No analytical expression for this exists. However, we can estimate the marginal likelihood using Gibbs sampler output and the Savage–Dickey density ratio (see Verdinelli and Wasserman, 1995). Define M_1 to be the semiparametric SUR model given in equation (4) with prior given by equations (5) and (6) and a particular value for η selected. Define M_2 to be M_1 with the restriction $\lambda_2 = 0_{N-2}$ imposed (with the same prior for

[1] This approach to prior elicitation does not include any information in x_{im} other than order information (i.e., data is ordered so that $x_{1m} < \cdots < x_{Nm}$). If desired, the researcher could include x_{im} by eliciting a prior, e.g., of the form $\Delta^2 \gamma_i \sim N(0, \eta \Delta^2 x_{im})$.

[2] Further motivation for our approach can be obtained by noting that our framework is similar to a state-space model in which a parameter analogous to η is the error variance in the state equation and can be estimated from the data. See Koop and Poirier (2004b).

[3] In practice, one may wish the choice of $\underline{\nu}$ to depend on the number of equations m. See Poirier (1995, pp. 136–138) for a parameterization of the Wishart density and related properties.

λ_1 and Σ^{-1}). Using the Savage–Dickey density ratio, the Bayes factor comparing M_1 to M_2 can be written as:

$$BF(\eta) = \frac{p(\lambda_2 = 0|M_1)}{p(\lambda_2 = 0|y, M_1)} \tag{14}$$

where the numerator and denominator are the prior and posterior, respectively, of λ_2 in the semiparametric SUR model evaluated at the point $\lambda_2 = 0_{N-2}$. This Bayes factor may be of interest in and of itself since it compares the semiparametric SUR model to a sensible parametric alternative.[4] *However, this procedure can also be used in an empirical Bayesian analysis to select η and thereby determine the appropriate amount of smoothing*. That is, since η does not enter the prior for M_2, $BF(\eta)$ will be proportional to the marginal likelihood for the semiparametric SUR model for a given value of η. The empirical Bayes estimate of η can be implemented by running the Gibbs sampler for a grid of values for η and choosing the value which maximizes $BF(\eta)$. Alternative methods for selecting η include cross-validation or extensions of the reference prior approach discussed in van der Linde (2000).

Note that $BF(\eta)$ can be calculated in the Gibbs sampler in a straightforward manner. The quantity $p(\lambda_2 = 0|M_1)$ can be directly evaluated using the normal prior given in equation (5), while $p(\lambda_2 = 0|y, M_1)$ can be evaluated with the Gibbs sampler as described in Gelfand and Smith (1990). That is, if we define

$$\widehat{p}(\lambda_2 = 0|y, M_1) = \frac{1}{S}\sum_{s=1}^{S} p(\lambda_2 = 0|y, \Sigma^{(s)})$$

where $\Sigma^{(s)}$ for $s = 1, \ldots, S$ denotes draws from the Gibbs sampler (after discarding initial burn-in draws), then

$$\widehat{p}(\lambda_2 = 0|y, M_1) \rightarrow p(\lambda_2 = 0|y, M_1)$$

as $S \rightarrow \infty$. Note that the conditional posterior $p(\lambda_2 = 0|y, \Sigma^{(s)})$ is simple to evaluate since it is normal [see equation (7)].

This form of the semiparametric SUR model described above is also the same as that of a triangular simultaneous equations model, since the Jacobian in such models is unity.[5] We illustrate the use of our approach in such a model in the following sections.

Before introducing our application, however, we briefly discuss related (parametric) Bayesian work on simultaneous equations models. The literature on Bayesian analysis of simultaneous equations models is voluminous. Dreze and Richard (1983) survey the literature through the early

[4] Specifically, M_2 is nearly equivalent to a SUR model with an intercept and x_{im} entering linearly. It is not exactly equivalent since we only use ordering information about x_{im}. As suggested by a referee, it would also be possible to reparameterize the model to let the γ_i denote departures from the proposed parametric form. For example, if one wishes to test for the adequacy of a linear specification in a cross-sectional model, one could specify $y_i = \alpha_0 + \alpha_1 x_i + \gamma_i + u_i$, so that γ_i denotes deviations from linearity. Such an approach would require imposing additional restrictions for identification purposes and would imply the same type of prior on the differences of γ_i. Finally, note that preference for the semiparametric SUR need not imply preference for a more complicated model. For example, if the 'true' effect were constant, the data would select η to be very small, thus essentially restricting the semiparametric model to one that only contains an intercept parameter. The semiparametric specification (which closely approximates the 'true' model) might then be favoured relative to, say, a model with a quadratic term. This preference, however, should not be interpreted as a need for something more complicated than a quadratic specification.

[5] In the case of a nontriangular SEM, the simple Gibbs algorithm described here is not applicable and other simulation methods need to be employed.

1980s, while Kleibergen (1997), Kleibergen and van Dijk (1998), Chao and Phillips (1998, 2002) and Kleibergen and Zivot (2003) are recent references. The latter work focuses on issues of identification and prior elicitation which are of little relevance for our work. For instance, some of this recent work discusses problems with the use of noninformative priors at points in the parameter space which imply nonidentification (and show how noninformative priors based on Jeffreys' principle overcome these problems). However, these problems are less empirically relevant if the posterior is located in regions of the parameter space away from the point of nonidentification or if informative priors are used. Furthermore, parameters in the reduced form model do not suffer from these problems. Hence, we feel some of the problems discussed in, e.g., Kleibergen and van Dijk (1998) are not critical for our work. In some sense, these problems all involve prior elicitation and, with moderately large data sets, data information should predominate. In practice, we argue that our approach is sensible for practitioners, and the advantage of being semiparametric outweighs any costs associated with not eliciting priors directly from structural parameters.

2.1. The Parametric SEM

In this section we show how our techniques can be applied in practice. Our empirical example, though primarily illustrative in nature, simultaneously addresses several topics of considerable interest in labour economics. Specifically, we introduce and estimate a two-equation structural simultaneous equations model and permit various nonparametric specifications within this system. The two endogenous variables in our system are the log hourly wage received by individuals in the labour force and the quantity of schooling attained by those individuals. While many studies have recognized the potential endogeneity of schooling in standard log wage equations (see Card, 1999 for a review of recent instrumental variable studies on this issue), these studies do not typically estimate the full underlying structural model and have not allowed for nonparametric specifications within these systems.

To fix ideas, the fully parametric version of our model may be written as:

$$\begin{aligned} s_i &= z_i^{C\prime}\alpha_1^S + z_i^{S\prime}\alpha_2^S + u_i^S \\ w_i &= \alpha_0 + \rho s_i + z_i^{C\prime}\alpha_1^W + z_i^{W\prime}\alpha_2^W + u_i^W \end{aligned} \tag{15}$$

with

$$\begin{bmatrix} u_i^s \\ u_i^w \end{bmatrix} \underset{\sim}{\text{i.i.d.}} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_s^2 & \sigma_{sw} \\ \sigma_{sw} & \sigma_w^2 \end{pmatrix}\right] \equiv N(0, \Sigma)$$

In the above equation, z_i^C is a k_C-vector of exogenous variables common to both equations, z_i^S is a k_S-vector of exogenous variables which enter only the schooling equation (i.e., these are the instruments), and z_i^W is a k_W-vector of exogenous variables which enter only the wage equation. The parameters in equation (15) are structural, with ρ being the returns to schooling parameter that is often of central interest. The triangular structure of equation (15) implies that the Jacobian is unity, so that we can directly estimate the structural form using methods developed in the previous section for the semiparametric SUR model.

In our empirical work we generalize this fully parametric structural model by permitting nonparametric specifications for a few variables. We divide our empirical analysis into three cases, with each case adding a new nonparametric component. In Case 1 we add a nonparametric specification to our wage equation and treat tenure on the job nonparametrically. Several studies

in labour economics (e.g. Altonji and Shakotko, 1987; Topel, 1991; Light and McGarry, 1998; Bratsberg and Terrell, 1998) have addressed the issue of separating the effects of on-the-job tenure and total labour market experience, with all of these studies specifying parametric (typically quadratic) specifications for each of these variables. In Case 1, we include a linear experience term and a nonparametric tenure term to flexibly investigate the shape of the relationship between job tenure and log hourly wages. In Case 2, we add a nonparametric component to the schooling equation and treat the 'ability' variable nonparametrically. In this analysis, 'ability' refers to measured cognitive ability, and is proxied by a (continuous) test score that is available in our data set. In Case 3 we also treat this ability variable nonparametrically in our wage equation.[6]

Finally, it is also worth mentioning that the approach described in the previous sections is quite appealing in that it can be applied to the case where the endogenous variable itself is treated in a nonparametric fashion. We do not take up the case of this particular model in our empirical work, however, given the rather discrete nature of our schooling variable. Nonetheless, the described framework can be applied directly in other applications where it is desired to maintain a flexible specification of the endogenous variable. Before discussing the specific models that we wish to tackle and results obtained from those models, we first describe the data used in this analysis.

2.2. The Data

To estimate our models we take data from the National Longitudinal Survey of Youth (NLSY), a widely-used panel survey containing a wealth of demographic and labour market information for a sample of young males and females in the USA. In keeping with the previous literature, and to abstract from selection issues into employment, we focus exclusively on the outcomes of white males in the NLSY. We restrict our attention to a cross-section of 1992 outcomes and exclude observations for those in the military, those reporting to be currently enrolled in school, those reporting hourly wages greater than \$100 or less than \$1, and those completing fewer than 9 years of schooling. Finally, we also require complete information on all the variables used in the analysis (which are described in detail below). These sample restrictions leave us with a total of $N = 303$ observations, for which a Bayesian analysis seems particularly useful.

Our dependent variables are the log of hourly wages and years of schooling. Perhaps the most important variable in this analysis, which is central to identification of the simultaneous equations model in (15), is the instrument or exclusion restriction. In the context of our application we need to find some variable in the NLSY that affects the quantity of schooling attained, yet has no direct structural effect on wages given the other controls we employ. To this end, we use the number of years of schooling attained by the respondent's oldest sibling (SIBED) as our instrument.[7] The first argument motivating this choice of instrument is that sibling's education should be strongly correlated with one's own schooling. This correlation could arise, for example, from unobserved family attitudes towards the importance of education, or credit constraints faced by the family. Secondly, and perhaps most importantly, the only channel through which sibling's education affects

[6] In related work, Cawley *et al*. (1999) argue that ability enters the wage equation nonlinearly. Blackburn and Neumark (1995), Heckman and Vytlacil (2001) and Tobias (2003) examine if returns to schooling vary with ability. The latter two of these studies obtain results by allowing for flexible specifications of the relationship between ability and log wages.

[7] In the base year of the NLSY survey, participants are asked to report the highest grade completed by their oldest sibling. To ensure that the oldest sibling had completed his/her education, we restrict our attention to those observations where the oldest sibling was at least 24 years of age. Thus, our analysis conditions on those in the NLSY with an older sibling who is at least 24 years old.

one's own wages should be an indirect one (through the quantity of schooling attained), since conditioned on the schooling of the respondent himself and added controls for family background, the education of the sibling should play no structural role in the wage equation.[8]

To estimate our models of interest we also obtain information about the actual labour market experience of the individual, as well as his tenure on the current job in 1992. The job tenure (TENURE) variable is readily available, as the NLSY directly provides information on the total tenure (in weeks) with the current employer. Total labour market experience (EXPERIENCE) is constructed from reported weeks of work between interview dates.[9]

In both the schooling and wage equations, we include the respondent's Armed Forces Qualifying Test (AFQT) score (denoted ABILITY), which is standardized by age given that respondents varied in age at the time the test was administered.[10] In both equations of (15) we additionally include the number of years of schooling completed by the respondent's mother (MOMED) and father (DADED), and a dummy variable equal to 1 if the respondent lived with both of his parents at age 14 (NON-BROKEN). In the wage equation, we also include weeks of actual labour market experience (EXPERIENCE), weeks of tenure at the current job (TENURE), a dummy for residence in an urban area (URBAN), and a continuous measure of the local unemployment rate (UNEMP). When measured in weeks, both EXPERIENCE and TENURE can be regarded as approximately continuous variables, and in our empirical analysis we choose to model TENURE and ABILITY nonparametrically. Finally, in all our specifications, SIBED is included in the schooling equation but is excluded from the wage equation.

2.3. Parametric Results

Before presenting results from semiparametric models, we briefly present results using a fully parametric approach.[11] Specifically, we estimate (15) using the prior specification discussed in Section 2. Results are reported in Table I, and specific results from this fully parametric SEM are presented in Table I(A).

The results obtained from the parametric SEM are mostly sensible. Importantly, the coefficient on our instrument SIBED is positive in the schooling equation, as expected, with a posterior mean more than twice its posterior standard deviation.[12] We find the point estimate of the return to schooling parameter (the coefficient on schooling in the wage equation) is roughly 8%, which is consistent with the majority of reported estimates in the literature. As for the empirical importance of endogeneity, the posterior mean of the correlation between the errors in the two equations is not far from zero (i.e., it is 0.054). Although the point estimate suggests that endogeneity is not a problem, the posterior for the correlation between the residuals is quite dispersed and allocates appreciable probability to regions of the parameter space where endogeneity is a problem (i.e., its

[8] Simple regression analyses that included sibling's education along with the other controls found no significant role for SIBED in the log wage equation.

[9] This definition is not without controversy, with many researchers (e.g. Wolpin, 1992 and Bratsberg and Terrell, 1998) only considering labour market experience after the completion of high school (or looking at only 'terminal' high school graduates). Light (1998) investigates this issue and finds sensitivity of results to the definition of the career starting point. In this analysis, we do not make a distinction between pre-high school and post-high school labour market experience.

[10] The AFQT score is constructed from component tests of the Armed Services Vocational Aptitude Battery (ASVAB), and these scores are available for the majority of NLSY participants.

[11] We chose to add a linear term for EXPERIENCE as we found little evidence of nonlinearity for this variable.

[12] In fact, we find a 'significant' role for this instrument in all of our model specifications.

Table I. Coefficient posterior means and standard deviations from parametric SEM and semiparametric SEMs described in Case 1 → Case 3

	(A) Parametric SEM				(B) Semiparametric SEM (Case 1)			
	Wage equation		Schooling equation		Wage equation		Schooling equation	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
INTERCEPT	0.566	0.662	9.255	0.714	—	—	9.271	0.688
ABILITY	0.022	0.098	1.253	0.136	0.009	0.093	1.252	0.131
MOMED	−0.015	0.017	0.107	0.053	−0.017	0.016	0.108	0.054
DADED	0.019	0.011	0.043	0.037	0.019	0.010	0.042	0.036
NON-BROKEN	−0.077	0.080	0.190	0.297	−0.079	0.079	0.195	0.296
SIBED	—	—	0.116	0.049	—	—	0.116	0.050
EXPERIENCE	8.7×10^{-4}	2.7×10^{-4}	—	—	8.5×10^{-4}	2.6×10^{-4}	—	—
TENURE	1.4×10^{-3}	4.3×10^{-4}	—	—	(Fig. 1)	(Fig. 1)	—	—
TENURE^2	-1.3×10^{-4}	6.4×10^{-5}	—	—	(Fig. 1)	(Fig. 1)	—	—
URBAN	0.119	0.061	—	—	0.118	0.061	—	—
UNEMP	-6.6×10^{-3}	0.010	—	—	-6.5×10^{-3}	0.010	—	—
SCHOOL	0.080	0.068	—	—	0.090	0.064	—	—
	(C) Semiparametric SEM (Case 2)				(D) Semiparametric SEM (Case 3)			
	Wage equation		Schooling equation		Wage equation		Schooling equation	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
INTERCEPT	—	—	—	—	—	—	—	—
ABILITY	0.051	0.061	(Fig. 3)	(Fig. 3)	(Fig. 5)	(Fig. 5)	(Fig. 6)	(Fig. 6)
MOMED	−0.012	0.015	0.123	0.052	−0.010	0.015	0.202	0.056
DADED	0.021	0.010	0.050	0.035	0.022	0.010	0.047	0.039
NON-BROKEN	−0.068	0.077	0.045	0.287	−0.080	0.079	0.257	0.308
SIBED	—	—	0.087	0.047	—	0.163	0.049	—
EXPERIENCE	8.0×10^{-4}	2.7×10^{-4}	—	—	6.8×10^{-4}	2.7×10^{-4}	—	—
TENURE	(Fig. 2)	(Fig. 2)	—	—	(Fig. 4)	(Fig. 4)	—	—
URBAN	0.124	0.062	—	—	0.112	0.063	—	—
UNEMP	-6.4×10^{-3}	0.010	—	—	−0.010	0.010	—	—
SCHOOL	0.058	0.038	—	—	0.042	0.039	—	—

posterior standard deviation is 0.252). Given this uncertainty regarding the empirical importance of endogeneity, the standard errors associated with the parameters in Table I(A) (notably the return to schooling parameter) are relatively large, and are large relative to those obtained from a fully parametric model that ignores the potential endogeneity problem.[13]

The finding that appreciable posterior probability is allocated to regions where the correlation between the errors in the two equations is near zero or small in magnitude is consistent with some of the other empirical work in this literature. That is, it has often been either assumed or more formally argued that after controlling for a rich set of explanatory variables, endogeneity problems are likely to be mitigated (see, e.g., Blackburn and Neumark, 1995). Our analysis of the parametric SEM lends some additional credence to this claim, as we 'test down' from a structural

[13] Specifically, we estimated an equation-by-equation version of (15) assuming the errors were uncorrelated and using noninformative priors. Point estimates from this approach and those reported in Table I(A) were virtually identical. However, posterior standard errors increased considerably. The posterior standard deviation on the return to schooling coefficient, for example, was 0.015 in the model which ignores the endogeneity problem.

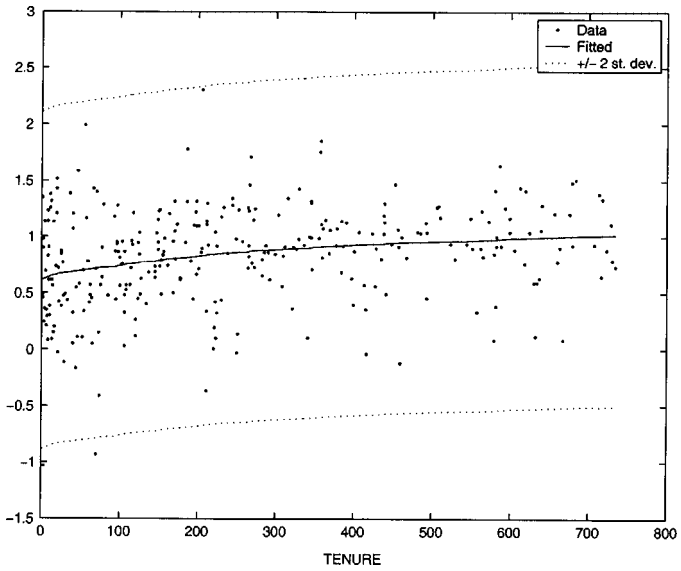

Figure 1. Fitted nonparametric regression line

model that permits endogeneity, and find little evidence that endogeneity is a serious empirical issue for this model. As we show in later sections, however, we find evidence against this basic parametric model, and in our generalized model specifications, there is some indication of a need to control for endogeneity of schooling.

2.4. Case 1: Application

In this model we elaborate (15), and allow the variable TENURE to enter the log wage equation nonparametrically. Formally, we specify:

$$\begin{aligned} s_i &= z_i^{C'}\alpha_1^S + z_i^{S'}\alpha_2^S + u_i^S \\ w_i &= \rho s_i + z_i^{C'}\alpha_1^W + z_i^{W'}\alpha 2^W + f(x_i) + u_i^W \end{aligned} \tag{16}$$

where x_i denotes the number of weeks of work on the current job (TENURE) and z_i^W no longer contains TENURE. In our analysis, we set $\eta = 5 \times 10^{-9}$, which is the empirical Bayes estimate that was found to maximize the marginal likelihood of the semiparametric SEM in equation (16). At this value of η, the log of the Bayes factor in favour of the semiparametric model over the parametric model with a linear tenure term was 0.026, indicating only slight support for the semiparametric specification. Note that as $\eta \to 0$, the nonparametric and linear models become equivalent and the log Bayes factor tends to zero. Given that our empirical Bayes estimate of η is quite small, it is not surprising that the odds reveal near indifference between these two specifications. Finally, Figure 1 plots the fitted nonparametric regression line against the data (after removing the effect of the other explanatory variables). That is, Figure 1 plots the posterior mean of the nonparametric regression line (and $+/-$ two standard deviation bands) and the 'data' points have coordinates x_i and $w_i - \rho s_i - z_i^{C'}\alpha_1^W - z_i^{W'}\alpha_2^W$ for $i = 1, \ldots, N$, where all parameters are evaluated at their posterior means.

The results in Table I(B) are very similar to the two-equation results in Table I(A) which differ in that TENURE and TENURE^2 are included parametrically. The point estimate of returns to schooling, at 9.0%, is slightly higher than with the parametric models. However, relative to its posterior standard deviation this difference is minor. The posterior mean of the correlation between the errors in the two equations is also very similar to that of the parametric model (i.e., its posterior mean is 0.031 and standard deviation is 0.231). Overall, we find our nonparametric function of TENURE to be playing a nearly identical role to the quadratic specification of this variable in the parametric model.

A comparison between a parametric SUR with TENURE entering quadratically to the semiparametric SUR can be done by first calculating the Bayes factor in favour of the quadratic SUR model of Table I(A) against the parametric SUR with TENURE entering linearly [call this Bayes factor BF^* to distinguish it from $BF(\eta)$ defined in equation (14)].[14] That is, $BF(\eta)$ compares the semiparametric SUR against a linear SUR (subject to the qualification of footnote 4), and thus the two Bayes factors BF^* and $BF(\eta)$ will each be comparing a nonlinear (either quadratic or nonparametric) specification to the linear one. However, Bayes factor calculation requires an informative prior over parameters which are not common to both models. Thus, to calculate BF^* we require an informative prior for the coefficient on TENURE^2 which we choose to be $N(0, \underline{v}_q)$. With this prior, BF^* can be calculated using the Savage–Dickey density ratio with the strategy discussed above [see the discussion around equation (14)]. The elicitation of prior hyperparameters such as $\underline{v}_q$ can be difficult (which is a further motivation for our empirical Bayesian analysis of a semiparametric model). In our application, values of $\underline{v}_q$ greater that 10^{-10} indicate support for the linear model (i.e., $BF^* < 1$). This apparently informative choice of prior variance is actually not that informative relative to the data information (note that the posterior standard deviation of this coefficient in Table I is 6.4×10^{-5}). For $\underline{v}_q < 10^{-10}$, the quadratic model is supported (i.e., $BF^* > 1$). However, there is no value for $\underline{v}_q$ for which the quadratic model receives overwhelming support. The maximum value for BF^* is 2.77, which occurs when $\underline{v}_q = 10^{-12}$.[15]

3. CASE 2: A SINGLE NONPARAMETRIC COMPONENT IN SEVERAL EQUATIONS

In this section we consider the more general semiparametric SUR model given in (1) where a nonparametric component potentially exists in every equation. That is, $\gamma_{ij} = f_j(x_{ij})$ for $j = 1, \ldots, m$ is the ith point on the nonparametric regression line in the jth equation. We maintain the assumption that x_{ij} is a scalar. Simple Bayesian methods for this model can be developed similarly to those developed for Case 1. We adopt the same strategy of treating unknown points on the nonparametric regression lines as unknown parameters and, hence, augment each equation with N new explanatory variables [as in equation (2)]. We then use a smoothness prior on each nonparametric regression line [analogous to equations (5) and (13)]. The resulting posterior can be handled using a Gibbs sampler [analogous to equations (7) and (8)]. Note, however, that we expressed our smoothness prior in terms of the second-differencing matrix D given in (3). This

[14] Of course, given the Bayes factor of the semiparametric SUR against the linear model, and the Bayes factor of the quadratic model against the linear model, one can calculate the Bayes factor of the semiparametric SUR against the quadratic model.

[15] The Bayes factor in favour of the quadratic model becomes larger if the prior mean of the quadratic coefficient is located closer to the posterior mean. However, we do not consider this case since it is common practice to centre the prior over the restriction being tested.

prior required the data to be ordered so that $x_{1m} < \cdots < x_{Nm}$. However, unless each equation has its nonparametric component depending on the same explanatory variable (i.e., $x_{ij} = x_{im}$ for $j = 1, \ldots, m-1$), the data in the jth equation (for $j = 1, \ldots, m-1$) will not be ordered in such a way that a smoothness prior can be expressed in terms of D. However, this can be corrected for by redefining the explanatory variables. This requires some new, somewhat messy, notation. Unless otherwise noted, all other assumptions and notation are as for Case 1. For future reference, define $\gamma_j = (\gamma_{1j}, \ldots, \gamma_{Nj})'$.

In Case 1, the inclusion of the nonparametric component implied that the identity matrix, I_N, was included as a matrix of explanatory variables [see equation (2) and the surrounding definitions]. Here we define I_j^* which is the identity matrix with columns rearranged to correspond to the ordering of the data in the jth equation for $j = 1, \ldots, m$. Thus, since $x_{1m} < \cdots < x_{Nm}$, I_m^* is simply I_N, but the other equations potentially involving a reordering of the columns of I_N. Also define ζ_{ij} to be the ith row of I_j^*.

A concrete example of how this works might help. Suppose we have $N = 5$ observations and the explanatory variables treated nonparametrically in the $m = 2$ equations have values in the columns of the following matrix:

$$\begin{bmatrix} 3 & 1 \\ 4 & 2 \\ 1 & 3 \\ 2 & 4 \\ 5 & 5 \end{bmatrix}$$

The data has been ordered so that the second explanatory variable is in ascending order, $x_{12} < \cdots < x_{52}$ and, hence, the Case 1 smoothness prior can be directly applied in the second equation. However, the first explanatory variable is not in ascending order. However, we can reorder the columns of the identity matrix as:

$$I_1^* = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

It can be seen that with I_1^* used to define the nonparametric explanatory variables for the first equation, γ_{11} is the first point on the nonparametric regression line, γ_{21} is the second point, etc. Thus, the smoothness prior can be expressed as restricting $D\gamma_1$.

In Case 1, we noted that the smoothness prior was only an $N-2$ dimensional distribution for the N points on each nonparametric regression line. Implicitly, this prior did not provide any information about the initial conditions (i.e., what we called γ_1 and γ_2 in Case 1), but only the second differences of points on the nonparametric regression line, $\gamma_i - 2\gamma_{i-1} + \gamma_{i-2}$. For the initial conditions, we used a noninformative prior. This need to separate out initial conditions necessitates the introduction of more notation. Define the 2×1 vector of initial conditions in every equation as γ_j^0 for $j = 1, \ldots, m$. Let γ_j^* be γ_j with these first two elements deleted. Similarly, let I_j^{**} be I_j^* with its first two columns deleted and I_j^{***} be the two deleted columns. Also define ζ_{ij}^* to be ζ_{ij} with the first two elements deleted and ζ_{ij}^0 to be the two deleted elements. Analogously, partition $D = [D^{**} : D^*]$ where D^{**} contains the first two columns of D.

With all these definitions, we can write the Case 2 semiparametric SUR as equation (2) with

$$W_i = \begin{bmatrix} z'_{i1} & \zeta^0_{i1} & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \zeta^*_{i1} & 0 & \cdot & \cdot & 0 \\ 0 & 0 & z'_{i2} & \zeta^0_{i2} & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \zeta^*_{i2} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 & z'_{i,m-1} & \zeta^0_{i,m-1} & 0 & 0 & \cdot & \cdot & 0 & \zeta^*_{i,m-1} & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & z'_{im} & \zeta^0_{im} & 0 & \cdot & \cdot & \cdot & \zeta^*_{im} \end{bmatrix} \tag{17}$$

where $\delta = (\beta'_1, \gamma_1^{0'}, \ldots, \beta'_m, \gamma_m^{0'}, \gamma_1^{*'}, \ldots, \gamma_m^{*'})'$ is a $K + mN$ vector of coefficients.

Prior information about the smoothness of the nonparametric regression lines will be expressed in terms of $R\delta$, where the $m(N-2) \times (K+mN)$ matrix R is given by

$$R = \begin{bmatrix} 0 & D^{**} & 0 & \cdot & \cdot & \cdot & \cdot & 0 & D^* & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & 0 & D^{**} & \cdot & \cdot & \cdot & \cdot & 0 & D^* & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & D^{**} & 0 & 0 & \cdot & \cdot & \cdot & D^* & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 & D^{**} & 0 & \cdot & \cdot & 0 & D^* \end{bmatrix} \tag{18}$$

The remainder of the derivations are essentially the same as for Case 1. Define the partitions $W_i = [W_i^{(1)} : W_i^{(2)}]$ where $W_i^{(1)}$ is an $m \times (K+2m)$ matrix and $W_i^{(2)}$ is $m \times m(N-2)$ and $R = [R_1 : R_2]$ where R_1 is an $m(N-2) \times (K+2m)$ matrix and R_2 is $m(N-2) \times m(N-2)$. Transform the model as:

$$y_i = V_i^{(1)}\lambda_1 + V_i^{(2)}\lambda_2 + \varepsilon_i = V_i\lambda + \varepsilon_i \tag{19}$$

where $\lambda = (\lambda'_1, \lambda'_2)'$, $\lambda_1 = (\beta'_1, \gamma_1^{0'}, \ldots, \beta'_m, \gamma_m^{0'})'$, $\lambda_2 = R\delta = [(D\gamma_1)', \ldots, (D\gamma_m)']'$, $V_i^{(1)} = W_i^{(1)} - W_i^{(2)}R_2^{-1}R_1$ and $V_i^{(2)} = W_i^{(2)}R_2^{-1}$.

This model is now in the same form as Case 1. Given an independent normal–Wishart prior as in (5) and (6), posterior analysis can be carried out using the Gibbs sampler described in (7) through (12). As in Case 1, we use a noninformative prior for λ_1. The prior for Σ^{-1} uses the same hyperparameter values as in Case 1. The smoothness prior relates to λ_2 and, for this, we extend the prior of Case 1 [see equation (13)] to be:

$$V(\eta_1, \ldots, \eta_m) = \begin{bmatrix} \eta_1 I_{N-2} & 0 & \cdot & \cdot & 0 \\ 0 & \eta_1 I_{N-2} & 0 & \cdot & \cdot \\ \cdot & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & \eta_m I_{N-2} \end{bmatrix} \tag{20}$$

Thus, the nonparametric component of each equation can be smoothed to a different degree. An empirical Bayesian analysis can be carried out as described above [see equation (14) and surrounding discussion]. The computational demands of empirical Bayes in this general case can be quite substantial since a search over m dimensions of the smoothing parameter vector must be carried out.

3.1. Case 2: Application

For Case 2, we extend the Case 1 model to allow for an exogenous variable in the schooling equation to receive a nonparametric treatment. The model we consider here is:

$$\begin{aligned} s_i &= z_i^{C'}\alpha_1^S + z_i^{S'}\alpha_2^S + f_1(x_{i1}) + u_i^S \\ w_i &= \rho s_i + z_i^{C'}\alpha_1^W + z_i^{W'}\alpha_2^W + f_2(x_{i2}) + u_i^W \end{aligned} \tag{21}$$

where all definitions are as in (16) except that x_{i1} is ABILITY and x_{i2} is TENURE and z_i^C no longer contains ABILITY in the schooling equation. Empirical Bayesian methods are used to select η_1 and η_2 which smooth the nonparametric regression lines in the two equations. This leads us to set $\eta_1 = 5 \times 10^{-6}$ and $\eta_2 = 10^{-11}$.

Table I(C) presents posterior results for the parametric coefficients in this semiparametric model. As found in our previous results, the correlation between the errors in the two equations has a point estimate near to, but now farther away from zero (i.e., its posterior mean is 0.102) and remains very imprecisely estimated (i.e., its posterior standard deviation is 0.142). Thus, we have more evidence that endogeneity is an issue in this model specification. Other entries in Table I(C) are similar to those reported in Table I(B).

Interestingly, this analysis finds rather strong evidence of nonlinearities in the relationship between ability and schooling. The log of the Bayes factor of our semiparametric model against the linear-in-schooling (and tenure) model is 4.645, which indicates substantially more support for departures from linearity than was found in Case 1. Figures 2 and 3 plot the posterior means of the two nonparametric regression lines against the data (after controlling for parametric explanatory variables). That is, the 'data' points in Figure 2 plot TENURE against $w_i - \rho s_i - z_i^{C'}\alpha_1^W - z_i^{W'}\alpha_2^W$ for $i = 1, \ldots, N$, where all parameters are evaluated at their posterior means. The comparable points in Figure 3 plot ABILITY against $s_i - z_i^{C'}\alpha_1^S - z_i^{S'}\alpha_2^S$ (evaluated at the posterior means for α_1^S and α_2^S). Figure 2 looks very similar to Figure 1 and indicates some slight nonlinearities that appear quadratic. Figure 3 indicates more interesting (and more precisely estimated) nonlinear effects that would not be captured by simple parametric methods (e.g. including ABILITY in a quadratic manner).[16] Specifically, the graph suggests that marginal increments in ability for low-ability individuals does little to increase the quantity of schooling attained (i.e., the graph is quite flat to the left of zero). However, for those individuals above the mean of the ability distribution, marginal increments in ability significantly increase the likelihood of acquiring more schooling. The fact that ability is a strong predictor of schooling has been well-documented (e.g. Heckman and Vytlacil, 2000), and here we add to this result by finding that it is relatively high-ability individuals whose schooling choices are most affected by changes in ability. Figure 3 also contains a line labelled 'Linear Fitted', which is the comparable line from the parametric model.[17]

It is also of interest to note that results for the returns for schooling parameter are slightly lower than what we have seen in either the parametric model or Case 1, with a posterior mean of 0.058 and posterior standard deviation of 0.038. We will now try to explain why this reduction

[16] If we add ABILITY^2 to the parametric SUR in (15) its coefficient has posterior mean which is roughly one posterior standard deviation from zero. Thus, a parametric analysis using a quadratic functional form for ABILITY would likely conclude that no nonlinearities existed in the ABILITY/SCHOOL relationship.

[17] This is not a perfect diagnostic for comparing semiparametric to parametric fit since the 'data' points in the plot are generated from the semiparametric model. Nevertheless, the 'data' points from the parametric model are quite similar and, hence, the patterns in Figure 3 are suggestive of the inadequacy of the linear model.

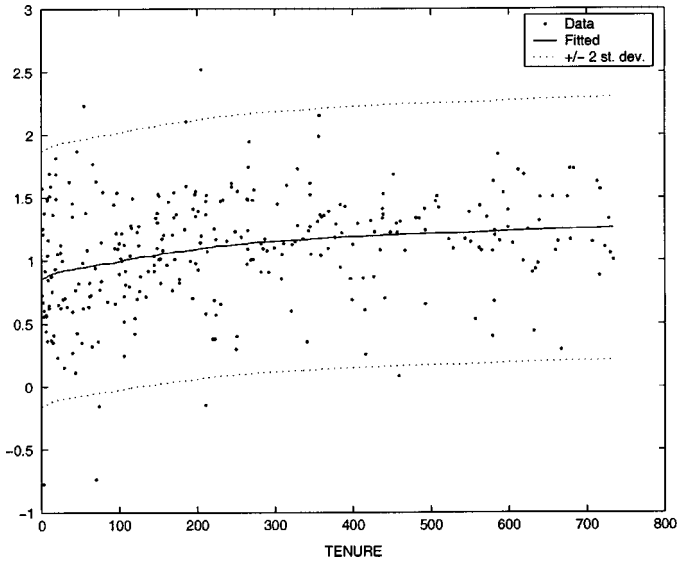

Figure 2. Fitted nonparametric regression line in wage equation

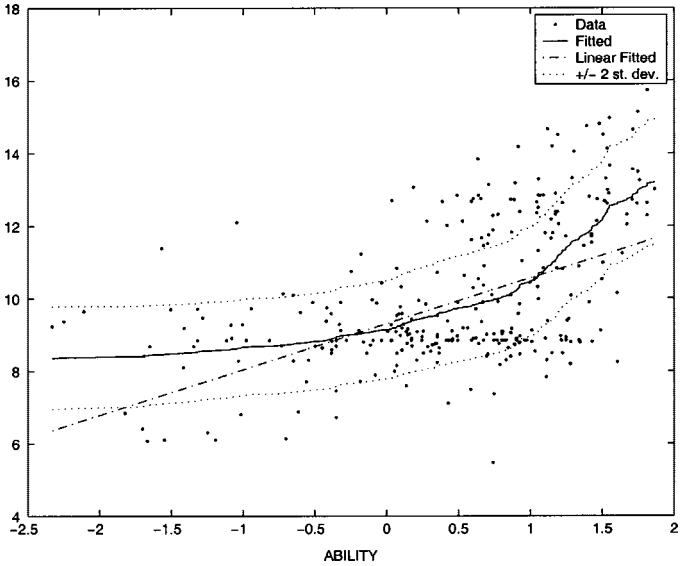

Figure 3. Fitted nonparametric regression line in schooling equation

has taken place. Our semiparametric estimation results found strong evidence of a nonlinear (and convex) relationship between ability and the quantity of schooling attained. To illustrate how this convex relationship may lead to a reduction of the schooling coefficient, let's suppose for the sake of simplicity that the actual relationship between schooling and ability is quadratic, with a positive coefficient on the squared term. Since the correlation between the errors of the structural

equations of Case 2 is nonzero (or at least most of the posterior mass is concentrated away from zero), this implies that the conditional mean of wages given schooling will now contain the nonlinear ability term that enters the education equation. This nonlinear term was, of course, not present in the conditional mean of Case 1 since that model only contained a linear ability term. So, we can regard the differences between the conditional means of Case 2 and Case 1 as essentially an omitted variable problem—in Case 2 we have an added quadratic ability term that is positively correlated with education (see Figure 3) and also positively correlated with log wages (we provide evidence of this in the next section). Using standard omitted variable bias formulas, we would thus predict a reduction in the 'reduced form' schooling coefficient upon controlling for this nonlinearity in ability. This result has potentially significant implications for this literature, as it suggests the importance of controlling for nonlinearities in ability (in both the schooling or wage equations) in order to extract accurate estimates of the return to education. Despite this result, it is also important to recognize that the shift in the posterior of this key parameter is small relative to its posterior standard deviation.

Suggestions Regarding Diagnostic Checking

Thus far, we have focused on estimation and model comparison (through Bayes factors). However, it is worth stressing that two other activities of the econometrician, prediction and diagnostic checking, can be done using standard Bayesian methods for the normal linear regression model. For instance, Koop (2003, sections 6.6 and 4.2.6) discusses prediction in the parametric SUR model and these methods can be directly applied here. Diagnostic checking can be done using, e.g., posterior predictive p-values (see, e.g., Gelman *et al*., 1995, section 6.3; Koop, 2003, section 5.6; Bayarri and Berger, 2000). To provide a simple illustration in our application, we have calculated the R^2 values (evaluated at the posterior mean of the parameters) for each equation for both the parametric and semiparametric models. For the wage equation, the two approaches yield the same value (to two decimal places) of R^2 of 0.30. This result comes as no surprise given the analysis of Case 1, which showed little distinction between the performances of the semiparametric and linear specifications. However, for the schooling equation, the semiparametric R^2 (0.48) is much higher than the parametric one (0.41). This result is consistent with Figure 3, which reveals the inadequacies of the linear model in capturing the shape of the schooling–ability relationship. Of course, such simple diagnostics should be interpreted with a degree of caution since, with a semiparametric approach, the errors can be driven to zero and thus measures of fit such as R^2 can be driven to one by letting $\eta \to \infty$.

4. CASE 3: NONPARAMETRIC COMPONENTS DEPEND ON SEVERAL EXPLANATORY VARIABLES: ADDITIVE MODELS

Up to this point we have only considered cases where the nonparametric component in a given equation depended on a single explanatory variable. That is, x_{ij} was assumed to be a scalar. In this section, we assume x_{ij} to be a vector of p explanatory variables.[18] The curse of dimensionality (see, e.g., Yatchew, 1998, pp. 675–676) implies that it is difficult to carry our nonparametric inference (whether Bayesian or non-Bayesian) when p is even moderately large. The intuition underlying our smoothness prior is that values of x_{ij} which are near one another should have

[18] The case where p varies across equations is a trivial extension of what is done in this section. We do not consider this extension to keep already messy notation from getting even messier.

points on the nonparametric regression line which are also near one another. When x_{ij} is a scalar, the definition of 'nearby' points is simple and is expressed through our ordering of the data as $x_{1m} < \cdots < x_{Nm}$. When x_{ij} is not a scalar, it is possible to order the data in an analogous way using some distance metric. If it is sensible to order the data in this way, then the approach of Case 2 can be applied directly. However, this approach is apt to be sensitive to choice of distance definition and which point to choose as the first on each nonparametric regression line. In the single equation case, Yatchew (1998, p. 697) argues that the classical differencing estimator works well provided the dimension of x_i does not exceed 3. It is likely that such an approach could work well in our Bayesian semiparametric framework when dimensionality is as low as this. Nevertheless, in this section we develop a different approach.

The curse of dimensionality is greatly reduced if it is assumed that $f_j(x_{ij})$ is additive. This, of course, is more restrictive than simply assuming $f_j(x_{ij})$ is an unknown smooth function, but it is much less restrictive than virtually any parametric model used in this literature. Furthermore, by defining x_{ij} to include interactions of explanatory variables, some of the restrictions imposed by the additive form can be surmounted. Accordingly, in this section we develop methods for Bayesian inference in the model given in (1) with:

$$f_j(x_{ij}) = f_j(x_{ij1}, \ldots, x_{jip}) = f_{j1}(x_{ij1}) + \cdots + f_{jp}(x_{ijp}) = \gamma_{ij1} + \cdots + \gamma_{ijp} \tag{22}$$

The basic idea underlying our approach to this model is straightforward: define a smoothness prior for each of the $f_{jr}(x_{ijr})$ for $j = 1, \ldots, m$ and $r = 1, \ldots, p$ and use the methods for Bayesian inference in the semiparametric SUR model with independent normal–Wishart prior described for Case 1. However, we must further complicate notation to handle this general case. In the following material, the indices run $i = 1, \ldots, N$, $j = 1, \ldots, m$ and $r = 1, \ldots, p$.

For Case 2, we defined matrices, $I_1^*, \ldots, I_m^*$ which were used as explanatory variables for the nonparametric regression lines taking into account the fact that each nonparametric explanatory variable was not necessarily in ascending order. For Case 3, we define analogously I_{jr}^* which is the reordered identity matrix needed to incorporate $f_{jr}(x_{ijr})$, taking into account that the data are not necessarily ordered so that $x_{1jr} < \cdots < x_{Njr}$. All the other Case 2 definitions can be extended in a similar fashion. Divide the vector of points on each nonparametric regression line, $\gamma_{jr} = (\gamma_{1jr}, \ldots, \gamma_{Njr})'$, into the 2×1 vector of initial conditions, γ_{jr}^0, and the remaining elements, γ_{jr}^*. Similarly, let I_{jr}^{**} be I_{jr}^* with the first two columns correspondingly deleted and I_{jr}^{***} be the two deleted columns. Furthermore, let ζ_{ijr}^* be ζ_{ijr} with the elements corresponding to the initial conditions deleted and ζ_{ijr}^0 be the two deleted elements, where ζ_{ijr} is the ith row of I_{jr}^*. Further define $\zeta_{ij}^0 = (\zeta_{ij1}^0, \ldots, \zeta_{ijp}^0)$ and $\zeta_{im}^* = (\zeta_{ij1}^*, \ldots, \zeta_{ijp}^*)$. Note that these last two definitions differ from Case 2.

With all these definitions, we can write the Case 3 model as a semiparametric SUR as in (2) with W_i as given in (17), except that the definition of some of the terms has changed slightly and

$$\delta = (\beta_1', \gamma_{11}^{0'}, \ldots, \gamma_{1p}^{0'}, \ldots, \beta_m', \gamma_{m1}^{0'}, \ldots, \gamma_{mp}^{0'}, \gamma_{11}^*, \ldots, \gamma_{1p}^{*'}, \ldots, \gamma_{m1}^{*'}, \ldots, \gamma_{mp}^{*'})'$$

is now a $K + mpN$ vector of coefficients.

As before, our smoothness prior is expressed in terms of $R\delta$, where R is now an $mp(N-2) \times (K+mpN)$ matrix:

$$R = \begin{bmatrix} 0 & D_p^{**} & 0 & \cdot & \cdot & \cdot & \cdot & 0 & D_p^{*} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & 0 & D_p^{**} & \cdot & \cdot & \cdot & \cdot & 0 & D_p^{*} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & D_p^{**} & 0 & 0 & \cdot & \cdot & \cdot & D_p^{*} & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 0 & D_p^{**} & 0 & \cdot & \cdot & 0 & D_p^{*} \end{bmatrix}$$

where

$$D_p^{**} = \begin{bmatrix} D^{**} & 0 & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & 0 & D^{**} \end{bmatrix}$$

is an $p(N-2) \times 2p$ matrix and

$$D_p^{*} = \begin{bmatrix} D^{*} & 0 & \cdot & 0 \\ 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0 \\ 0 & \cdot & 0 & D^{**} \end{bmatrix}$$

is $p(N-2) \times p(N-2)$.

The remainder of the derivations are the same as for Cases 1 and 2. That is, the model can be transformed as in (19). An independent normal–Wishart prior for the transformed parameters is used with prior hyperparameters selected as for Case 2. The Gibbs sampler described in (7) through (12) can be used for posterior inference. The only difference is that it will usually be desirable to have a different smoothing parameter for every nonparametric regression line in every equation. Thus, we choose the prior covariance matrix for λ_2 to be:

$$V(\eta_{11}, \ldots, \eta_{1p}, \ldots, \eta_{m1}, \ldots, \eta_{mp}) = \begin{bmatrix} \eta_{11} I_{N-2} & 0 & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 0 & \eta_{1p} I_{N-2} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & 0 & \eta_{mp} I_{N-2} \end{bmatrix} \tag{23}$$

There is an identification problem with this model in that constants may be added and subtracted to each nonparametric component without changing the likelihood. For instance, the equations $y_{ij} = z_{ij}'\beta_j + f_{j1}(x_{ij1}) + f_{j2}(x_{ij2}) + \varepsilon_{ij}$ and $y_{ij} = z_{ij}'\beta_j + g_{j1}(x_{ij1}) + g_{j2}(x_{ij2}) + \varepsilon_{ij}$ are equivalent if $g_{j1}(x_{ij1}) = f_{j1}(x_{ij1}) + c$ and $g_{j2}(x_{ij2}) = f_{j2}(x_{ij2}) - c$ where c is an arbitrary constant. Insofar as interest centres on the shapes of the $f_{jr}(x_{ijr})$ for $r = 1, \ldots, p$, prediction or the overall fit of the nonparametric regression line, the lack of identification is irrelevant. If desired, identification can be imposed in many ways (e.g. by setting the intercept of the rth nonparametric component in each equation to be zero for $r = 2, \ldots, p$).

4.1. Case 3: Application

In Case 3 we extend Case 2 to also allow for a nonparametric treatment of ABILITY in the wage equation. Thus, ABILITY is treated nonparametrically in both equations and TENURE is treated

nonparametrically in the wage equation. The model is:

$$s_i = z_i^{C'}\alpha_1^S + z_i^{S'}\alpha_2^S + f_{11}(x_{i11}) + u_i^S$$
$$w_i = \rho s_i + z_i^{C'}\alpha_1^{W'} + z_i^{W'}\alpha_2^W + f_{21}(x_{i21}) + f_{22}(x_{i22}) + u_i^W \quad (24)$$

where definitions are as for Case 2 except that $x_{i11} = x_{i22}$ is ABILITY and x_{i21} is TENURE and z_i^C no longer contains ABILITY in either equation. We identify the model by setting the intercept of one of the nonparametric functions in the wage equation to be zero, i.e., $f_{22}(x_{122}) = 0$.

With three nonparametric components, empirical Bayesian methods involve a three-dimensional grid search over the smoothing parameters η_1, η_2 and η_3 for terms relating to ABILITY (in the schooling equation), TENURE and ABILITY (in the wage equation), respectively. We find $\eta_1 = 10^{-6}$, $\eta_2 = 10^{-9}$ and $\eta_3 = 10^{-11}$. With these values, the log of the Bayes factor in favour of the nonparametric model is 3.837, indicating stronger support for the semiparametric model over the parametric alternative of (15) than with Case 1.

Empirical results for the regression coefficients are presented in Table I(D) and are found to be similar to those for Case 2. In addition, the posterior mean of the correlation between the errors in the two equations is 0.138 (standard deviation 0.140), values similar to Case 2. Perhaps the most interesting finding is that the posterior mean of the return to schooling parameter is, at 0.042, similar to but smaller than that found for Case 2, and approximately half the size of those reported in Case 1 and the parametric model. Again, upon controlling for nonlinearities in the relationship between ability and log wages, we find even more of a reduction in the return to schooling coefficient. However, the posterior standard deviation of this parameter is still quite large.

Figures 4, 5 and 6 plot the fitted nonparametric regression lines (after controlling for other explanatory variables in the same manner as for previous cases). Figure 6 indicates the same nonquadratic nonlinearities in the relationship between ABILITY and SCHOOL (after controlling for other explanatory variables) as Figure 3, while Figure 4 is similar to Figures 1 and 2. Figure 5 also appears to exhibit a slightly nonlinear regression relationship between log wages and ABILITY of a nonquadratic form (although the pattern is much weaker than in Figure 6). Specifically, Figure 5 suggests that marginal increments in ability do little to increase the log wages of individuals of low to moderate ability, but do begin to have a reasonable effect on the log wages of those already above the mean of the ability distribution (i.e., increasing returns to ability). It is also important to recognize that we are obtaining this result after controlling for the potentially endogenous education variable and also controlling for nonlinearities in the education–ability relationship. The fact that +/− two posterior standard deviation bands in Figure 5 are very tight for the lowest values of ABILITY is due to the identification restriction and the fact that there are very few observations in this region.

To summarize, the estimated regression functions, diagnostic checks and Bayes factors suggest the empirical importance of several potentially overlooked nonlinearities in this often-studied problem. Specifically, we see (Figures 3 and 6) a clear need to account for nonlinearities in the relationship between ability and schooling and that a simpler quadratic specification may not be fully adequate to account for these nonlinearities.

There is also evidence in support of nonlinearities in the ability–log wage relationship (Figure 5), though these are less pronounced. Though one might argue that the shapes of the curves can be well-approximated by particular parametric forms, and thus there is little need to be nonparametric for this application, it is important to recognize that this is an *ex post* determination, and it is only through the use of semiparametric techniques that we can decide on those forms.

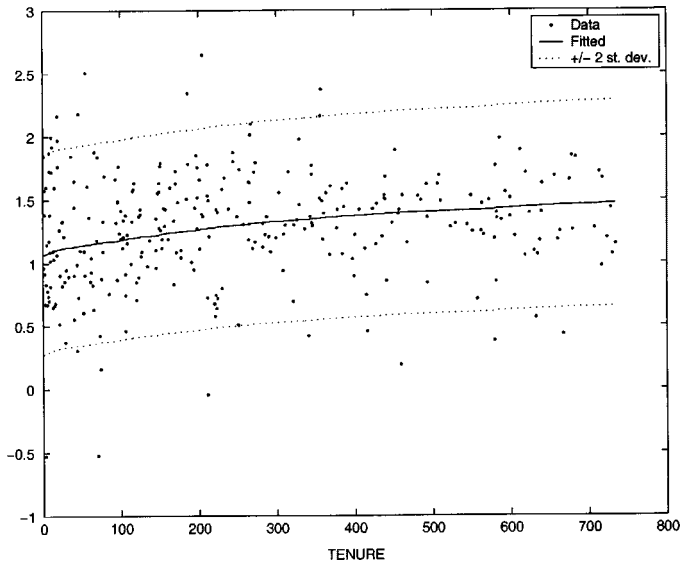

Figure 4. Fitted nonparametric regression line for TENURE in wage equation

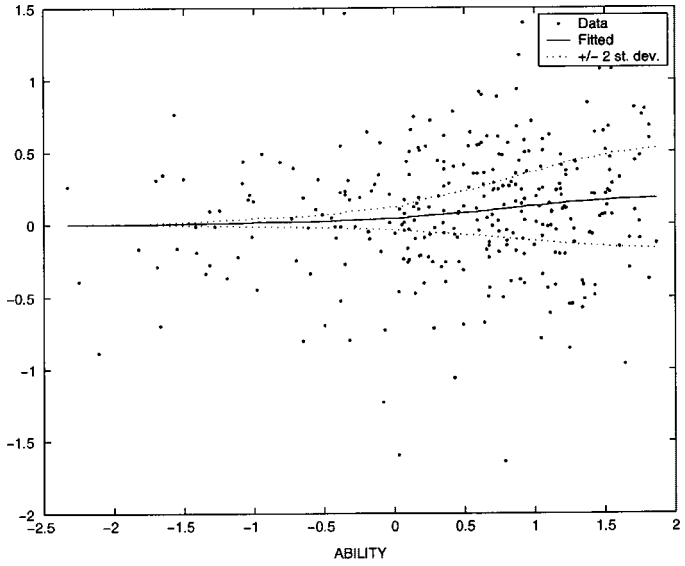

Figure 5. Fitted nonparametric regression line for ABILITY in wage equation

Further, several of the regression functions (i.e., ability–schooling and ability–log wage) appear to have curvatures that are not well-approximated by standard parametric models. Finally, the use of these semiparametric methods as an exploratory device to suggest appropriate parametric forms also invites criticism regarding pretesting, as subsequent inference from the parametric model will not be formally correct. Thus, we feel the methods described here offer a flexible

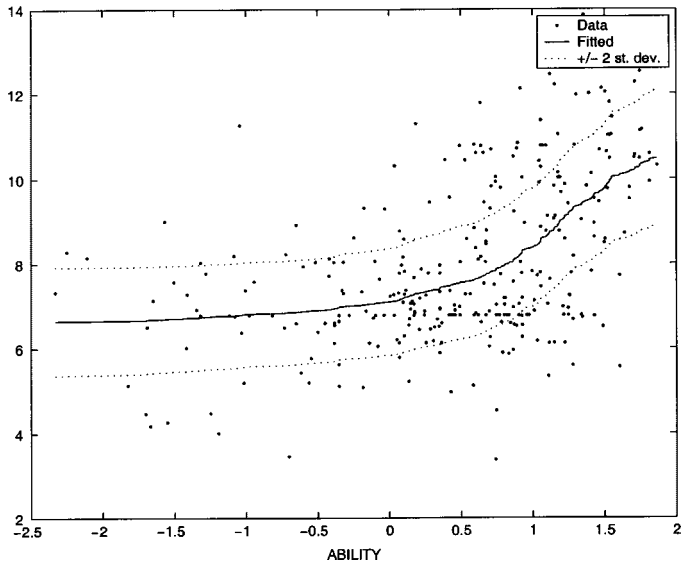

Figure 6. Fitted nonparametric regression line in schooling equation

yet computationally simple alternative, which is useful for this and other applications involving equation systems.

5. CONCLUSIONS

In this paper, we have developed methods for carrying out Bayesian inference in the semiparametric seemingly unrelated regressions model and showed how these methods can also be used for triangular semiparametric simultaneous equations models. There are, of course, other methods for carrying out Bayesian inference in semi- or nonparametric extensions of SUR models (e.g. Smith and Kohn, 2000). A distinguishing feature of our approach is that we stay within the simple and familiar framework of the SUR model with independent normal–Wishart prior. Thus, textbook results for Bayesian inference, model comparison, prediction and posterior computation are immediately available.

The focus of this paper is on prior information about the degree of smoothness in the nonparametric regression lines (although, of course, prior information about other parameters can easily be accommodated). We show how empirical Bayesian methods can be used to estimate smoothing parameters, thus minimizing the need for subjective prior elicitation.

The practicality of our approach is demonstrated in a two-equation application involving returns to schooling. In addition to parametric models, we estimate models with a single nonparametric component in one equation and a single nonparametric component in both equations. Our most general model contained an additive specification in the wage equation and a nonparametric ability component in the schooling equation. Although our semiparametric results are, in some cases, similar to those from simpler parametric nonlinear models (e.g. where explanatory variables enter in a quadratic fashion), in other cases our semiparametric approach yields empirical results which could not easily be obtained using standard parametric methods. Using our approach, we found

suggestive evidence of nonlinearities in the relationships between ability and the quantity of schooling attained, and that estimates of the return to schooling were sensitive to controlling for these nonlinear relationships. Furthermore, we stress that one clear advantage of a semiparametric approach is that a particular functional form such as the quadratic does not have to be chosen, either in an *ad hoc* fashion or through pre-testing.

Finally, it is worth noting that it is very easy to incorporate the semiparametric SUR model developed here in a more complicated multiple equation model. For instance, Bayesian inference in a multinomial semiparametric probit model can be done by adding a data augmentation step in the Gibbs sampler outlined in this paper as in, e.g., McCulloch and Rossi (1994). Bayesian inference in a semiparametric multiple equation model where one (or more) of the dependent variables is censored can be handled in a similar manner. We have assumed normal errors, but this assumption can easily be relaxed through the use of mixtures of normals. In short, Bayesian inference in semiparametric variants of a wide range of multiple equation models can be handled in a straightforward manner.

ACKNOWLEDGEMENTS

We would like to thank Luc Bauwens, two anonymous referees and the co-editor of this journal, Herman van Dijk, for helpful comments.

REFERENCES

Altonji J, Shakotko RA. 1987. Do wages rise with job seniority? *Review of Economic Studies* **54**(3): 437–459.
Bayarri M, Berger J. 2000. P-values for composite null models. *Journal of the American Statistical Association* **95**: 1127–1142.
Blackburn M, Neumark D. 1995. Are OLS estimates of the return to schooling biased downward? Another look. *Review of Economics and Statistics* **77**: 217–230.
Blundell R, Duncan A. 1998. Kernel regression in empirical microeconometrics. *Journal of Human Resources* **33**: 62–87.
Bratsberg B, Terrell D. 1998. Experience, tenure and wage growth of young black and white men. *Journal of Human Resources* **33**: 658–677.
Card D. 1999. The causal effect of education on earnings. In *Handbook of Labor Economics*, Vol. 3A, Ashenfelter O, Card D (eds). North Holland: Amsterdam; 1801–1863.
Cawley J, Vytlacil E, Heckman J. 1999. On policies to reward the value added by educators. *Review of Economics and Statistics* **81**(4): 720–728.
Chao JC, Phillips PCB. 1998. Posterior distributions in limited information analysis of the simultaneous equations model using the Jeffreys prior. *Journal of Econometrics* **87**: 49–86.
Chao JC, Phillips PCB. 2002. Jeffreys prior analysis of the simultaneous equations model in the case with $n + 1$ endogenous variables. *Journal of Econometrics* **111**: 251–283.
Chib S, Greenberg E. 1995. Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models. *Journal of Econometrics* **68**: 339–360.
Chib S, Greenberg E. 1996. Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory* **12**: 409–431.
Darolles S, Florens J-P, Renault E. 2003. Nonparametric instrumental regression. Working paper, available at www.univ-tlsel.fr/idei/Commun/Articles/Florens/renaultdarolles.pdf.
DiNardo J, Tobias JL. 2001. Nonparametric density and regression estimation. *Journal of Economic Perspectives* **15**: 11–28.
Dreze J, Richard J. 1983. Bayesian analysis of simultaneous equations systems. In *Handbook of Econometrics*, Vol. 1, Griliches Z, Intriligator M (eds). North Holland: Amsterdam.

Fernandez C, Osiewalski J, Steel M. 1997. On the use of panel data in stochastic frontier models with improper priors. *Journal of Econometrics* **79**: 169–193.
Gelfand AE, Smith AFM. 1990. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**: 398–409.
Gelman A, Carlin JB, Stern HS, Rubin DB. 1995. *Bayesian Data Analysis*. Chapman & Hall: London.
Heckman J, Vytlacil E. 2001. Identifying the role of cognitive ability in explaining the level of and change in the return to schooling. *Review of Economics and Statistics* **83**(1): 1–12.
Kleibergen F. 1997. Bayesian simultaneous equations analysis using equality restricted random variables. *American Statistical Association, 1997 Proceedings of the Section on Bayesian Statistical Science*, 141–147.
Kleibergen F, van Dijk H. 1998. Bayesian simultaneous equations analysis using reduced rank structures. *Econometric Theory* **14**: 699–744.
Kleibergen F, Zivot E. 2003. Bayesian and classical approaches to instrumental variable regression. *Journal of Econometrics* **114**: 29–72.
Koop G. 2003. *Bayesian Econometrics*. John Wiley & Sons: Chichester, UK.
Koop G, Poirier DJ. 2004a. Bayesian variants of some classical semiparametric regression techniques. *Journal of Econometrics* **123**: 259–282.
Koop G, Poirier DJ. 2004b. Empirical Bayesian inference in a nonparametric regression model. To appear in a volume from the *Conference in Honour of Professor J. Durbin on State Space Models and Unobserved Components*.
Light A. 1998. Estimating returns to schooling: when does the career begin? *Economics of Education Review* **17**(1): 31–45.
Light A, McGarry K. 1998. Job change patterns and the wages of young men. *Review of Economics and Statistics* **80**(2): 276–286.
McCulloch R, Rossi P. 1994. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics* **64**: 207–240.
Newey W, Powell J, Vella F. 1999. Nonparametric estimation of triangular simultaneous equations models. *Econometrica* **67**: 565–604.
Pagan A, Ullah A. 1999. *Nonparametric Econometrics*. Cambridge University Press: Cambridge.
Poirier DJ. 1995. *Intermediate Statistics and Econometrics*. The MIT Press: Cambridge, MA.
Rothenberg T. 1963. A Bayesian analysis of simultaneous equation systems. Econometric Institute Report 6315, Erasmus University, Rotterdam.
Smith M, Kohn R. 2000. Nonparametric seemingly unrelated regression. *Journal of Econometrics* **98**: 257–281.
Tobias JL. 2003. Are returns to schooling concentrated among the most able? A semiparametric analysis of the ability–earnings relationships. *Oxford Bulletin of Economics and Statistics* **65**(1): 1–29.
Topel R. 1991. Specific capital, mobility and wages: wages rise with job seniority. *Journal of Political Economy* **99**(1): 145–176.
van der Linde A. 2000. Reference priors for shrinkage and smoothing analysis. *Journal of Statistical Planning and Inference* **90**: 245–274.
Verdinelli I, Wasserman L. 1995. Computing Bayes factors using a generalization of the Savage–Dickey density ratio. *Journal of the American Statistical Association* **90**: 614–618.
Wolpin K. 1992. The determinants of black–white differences in early employment careers: search, layoffs, quits and endogenous wage growth. *Journal of Political Economy* **100**(3): 525–560.
Yatchew A. 1998. Nonparametric regression techniques in economics. *Journal of Economic Literature* **36**: 669–721.

