A Flexible Nonlinear Inference with Endogenous Explanatory Variables

Dong Heon Kim*

Department of Economics

Korea University

Revised: July 03, 2015

Abstract

Hamilton's (2001) flexible nonlinear inference is not valid in the case of endogenous explanatory variables. This paper proposes a framework for dealing with endogeneity problems in the flexible nonlinear inference. We develop two estimation procedures: a joint estimation procedure and a two-step estimation procedure. The parameters in both models can be estimated by maximum likelihood or numerical Bayesian method. Our approach would be useful in dealing with endogeneity and nonlinearity in the oil-macro relation or in the monetary policy rule.

Keywords Endogeneity · Nonlinear flexible inference · Control function approach · Two-step procedure

JEL Classification C13 · C32

---

[0]* This work is supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2012S1A5A2A01020228). The author would like to thank the Editor, two anonymous referees, James D. Hamilton, Chang-Jin Kim, and Sang-Soo Park for their constructive comments.

Corresponding address: Professor Dong Heon Kim, Department of Economics, Korea University, 145 Anam–ro, Seongbuk-Gu, Seoul, 136-701, South Korea, Tel: +82 2 3290 2226, Fax: +82 2 928 4948, Email: dongkim@korea.ac.kr.

1 Introduction

A natural approach to estimate a usual economic model is a linear relation between relevant variables. As pointed out in Hamilton (2001), however, nonlinear models may improve forecasts and provide economic insights. In usual estimation for nonlinear models, ones employ parametric approaches or nonparametric approaches. On the one hand, a crucial aspect for parametric approaches is deciding which parametric model to use. In other words, ones need to decide in what way the data might be nonlinear. On the other hand, popular nonparametric approaches sacrifice many of the benefits of parametric methods such as a needed system for adjusting a bandwidth or series expansion length, unclear interpretation of inferences and the difficulty of the hypothesis testing.

In order to develop a parametric approach along with nonparametric philosophy, Hamilton (2001) proposes a new framework for determining whether a given relationship is nonlinear, what the nonlinear function looks like, and whether it is adequately described by some particular model. He studies the expectation of a scalar y_t conditional on an observed vector $\mathbf{x}_t$, $E(y_t|\mathbf{x}_t) = \mu(\mathbf{x}_t)$, where a regression of the form $y_t = \mu(\mathbf{x}_t) + \varepsilon_t$ and the functional form of $\mu(\cdot)$ is unknown. The paper views $\mu(\cdot)$ itself as the outcome of a random process and introduces a stationary random field $m(\cdot)$ whose realizations could represent a broad class of possible forms for $\mu(\cdot)$. The proposed approach views the parameters that characterize the relation between a given realization of $m(\cdot)$ and the particular value of $\mu(\cdot)$ for a given sample as population parameters to be estimated by maximum likelihood or Bayesian methods.

Hamilton's (2001) parametric approach to flexible nonlinear inference, however, is not valid in the presence of endogenous explanatory variables, where $\mathbf{x}_t$ is correlated with ε_t. This endogeneity of explanatory variables is frequently observed in the macroeconometric models and results in inconsistent estimates of parameters. Recently Kim (2004, 2009) proposed a joint estimation procedure and a two-step maximum likelihood estimation (MLE) procedure to deal with the problem of endogeneity in Markov-switching regression models and these procedures are based on the control function approach (Heckman and Vytlacil 1998, Heckman and Navarro 2004, Altonji and Matzkin 2005, Florens, Heckman, Meghir, and Vytlacil 2007, among others). Kim and Nelson (2006) show that the two-step MLE procedure is very useful in the estimation of a forward-looking monetary policy rule in the U.S.A.

The purpose of this paper is to develop a flexible nonlinear inference of Hamilton (2001) in the case of endogenous regressors. In this new approach, we apply the control function approach to Hamilton's (2001) flexible nonlinear framework. We show that there exists an appropriate transformation of the model that allows us to directly employ Hamilton's (2001) approach. In estimating a flexible nonlinear model with endogenous regressors, both joint and two-step estimation procedures are considered in this paper. The parameters in these procedure can be estimated by maximum likelihood or numerical Bayesian method.

The plan of the paper is as follows. Section 2 discusses a nonlinear form with endogenous explanatory variables. Section 3 describes the joint estimation procedure and Section 4 derives the two-step estimation procedure. Section 5 concludes.

2 Nonlinear form with endogenous explanatory variables

Hamilton (2001) proposes a new framework that combines the advantages of non-parametric and parametric methods. While the procedure does not assume any specific functional form for the conditional mean function, parameters are used to characterize this function and these parameters are estimated by maximum likelihood or Bayesian methods. Inference is based on classical econometric theory.

Consider the general nonlinear regression model

$$y_t = \mu(\mathbf{x}_t) + \varepsilon_t,\ \varepsilon_t \tilde{} i.i.d.\ N(0, \sigma_\varepsilon^2), \tag{1}$$

where y_t is a scalar dependent variable, $\mathbf{x}_t$ is k-dimensional vector of explanatory variables, and ε_t is an error term with mean zero that is independent of $\mathbf{x}_t$ and of lagged values $y_{t-j},\ \mathbf{x}_{t-j},\ \mathbf{z}_{t-j} (j = 1, 2, ...)$. Equation (1) allows a subset of variables $\mathbf{x}_t$ for which the researcher is willing to assume linearity, thereby gaining efficiency by imposing this restriction. The form of the function $\mu(\cdot)$ is unknown and we seek to represent it using a flexible class. Following Hamilton (2001), we view this function as the outcome of a random field[1]. That is, if τ denotes an arbitrary, nonstochastic $k-$dimensional vector, then the value of the function $\mu(\cdot)$ evaluated at $\boldsymbol{\tau}$ is treated as being a Gaussian random variable with mean equal to the linear component $\alpha_0 + \boldsymbol{\alpha}'\boldsymbol{\tau}$ and variance λ^2, where $\alpha_0, \boldsymbol{\alpha}$, and λ are population parameters to be estimated[2]. In

[1] A random field is a generalization of a stochastic process such that the underlying parameter need no longher be a simple real or interger valued "time" but can instead take values that are multidimensional vectors or points on some manifold. At its most basic, discrete case, a random field is a list of a random numbers whose indices are mapped onto a space (of n dimensions). In its most basic form this might mean that adjacent values (i.e. values with adjacent indices) do not differe as much as values that are further apart. This is an example of a covarinace strcutre, man different types of which may be modeled in a random field. The random filed might be thought of as a "function valued" random variable. (Vanmarcke 2010, Wikipedia)

[2] Since we do not know the functional form of $\mu(\cdot)$, the final outcome $\mu(\tau)$ evaluated at the realized value τ

the special case of $\lambda = 0$, then $\mu(\mathbf{x}_t)$ is fixed and equation (1) becomes the usual linear regression model. In general, the parameter λ measures the overall extent of nonlinearity.

The basic point one needs to know about the random field $\mu(\cdot)$ is how the random variable $\mu(\tau_1)$ is correlated with $\mu(\tau_2)$, for τ_1 and τ_2 again arbitrary $k-$dimensional vectors. Hamilton (2001) parameterizes this correlation based on the distance measure $h_{st} = (1/2)\left[\sum_{i=1}^{k} g_i^2 (x_{is} - x_{it})^2\right]^{1/2}$, where x_{it} denotes the ith element of the vector $\mathbf{x}_t$ and $g_1, g_2, ..., g_k$ are k additional parameters to be estimated. Hamilton proposes that $\mu(\mathbf{x}_s)$ should be uncorrelated with $\mu(\mathbf{x}_t)$ if $\mathbf{x}_s$ is sufficiently far away from $\mathbf{x}_t$. More precisely,

$$E\{[\mu(\mathbf{x}_s) - \alpha_0 - \boldsymbol{\alpha}^{'}\mathbf{x}_s][\mu(\mathbf{x}_t) - \alpha_0 - \boldsymbol{\alpha}^{'}\mathbf{x}_t]\} = 0 \quad \text{if } h_{st} > 1 \tag{2}$$

However, when $0 \leq h_{st} \leq 1$, this correlation should increase as h_{st} decreases, with the correlation going to unity as h_{st} goes to zero. For example, in the case of two explanatory variables, $k = 2$, the correlation is assumed to be given by

$$Corr(\mu(\mathbf{x}_s), \mu(\mathbf{x}_t)) = H_2(h_{st}) \;\; \text{if } 0 \leq h_{st} \leq 1 \tag{3}$$

where

$$H_2(h_{st}) = 1 - (2/\pi)[h_{st}(1 - h_{st}^2)^{1/2} + \sin^{-1}(h_{st})]. \tag{4}$$

In the presence of nonlinearity, Hamilton writes equation (1) as

can be treated as a random variable.

$$y_t = \alpha_0 + \boldsymbol{\alpha}' \mathbf{x}_t + \lambda m(\mathbf{x}_t) + \varepsilon_t \tag{5}$$

$$= \alpha_0 + \boldsymbol{\alpha}' \mathbf{x}_t + u_t, \tag{6}$$

where $m(.)$ is the realization of a scalar-valued Gaussian random field with mean zero, unit variance and covariance function given by equation (2) to equation (4). Nonlinearity of the functional form $\mu(\cdot)$ implies a correlation between u_t and u_s, the residuals of the linear specification, whenever $\mathbf{x}_t$ and $\mathbf{x}_s$ are close together.

Assuming that the regression disturbance ε_t is i.i.d. $N(0, \sigma_\varepsilon^2)$, the composite disturbance $u_t = \lambda m(\mathbf{x}_t) + \varepsilon_t$ is also Gaussian. With independence between $\mathbf{x}_t'$ and ε_t, this specification implies a GLS regression model of the form

$$\mathbf{y}|\mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{P}_0 + \sigma^2 \mathbf{I}_T) \tag{7}$$

where $\mathbf{y} = (y_1, y_2, ..., y_T)'$, $\mathbf{X}$ is the $T \times (k+1)$ matrix with tth row $(1, \mathbf{x}_t')'$, $\boldsymbol{\beta}$ is the $(1+k)$-dimensional vector $(\alpha_0, \boldsymbol{\alpha}')'$, and $\mathbf{P}_0$ is a $(T \times T)$ matrix whose row s, column t element is given by $\lambda^2 H_k(h_{st}) \delta_{[h_{st}<1]}$ with h_{st} defined above, and the function $H_k(.)$ is specified in equation (4) for the case $k = 2$. The indicator function $\delta_{[.]}$ is unity when the condition [.] holds, and zero otherwise.

In addition to the linear regression parameters $(\alpha_0, \boldsymbol{\alpha})$ and σ^2, parameters to be estimated are the variance of the nonlinear regression error, λ^2, which governs the overall importance of the nonlinear component, and the parameters $(g_1, g_2, ..., g_k)$ determining the variability of the

nonlinear component with respect to each explanatory variable in $\mathbf{x}_t$. As the above discussion implies, estimation and inference can be achieved by a GLS Gaussian regression or numerical Bayesian methods.

Hamilton's (2001) methodology for the estimation of equation (7), however, is not valid when the regressors $\mathbf{x}_t$ are endogenous. In order to resolve this issue, in this paper, we consider the following nonlinear regression model in which the explanatory variables are correlated with the disturbance term:

$$y_t = \mu(\mathbf{x}_t) + \varepsilon_t, \ \varepsilon_t \tilde{} i.i.d. \ N(0, \sigma_\varepsilon^2), \tag{8}$$

$$\mathbf{x}_t = \boldsymbol{\delta}' \mathbf{z}_t + \mathbf{v}_t, \ \mathbf{v}_t \tilde{} i.i.d. \ N(\mathbf{0}, \boldsymbol{\Sigma}_v), \tag{9}$$

$$Cov(\mathbf{v}_t, \varepsilon_t) = \mathbf{C}_{v,\varepsilon}, \tag{10}$$

where y_t is a scalar dependent variable, $\mathbf{x}_t$ is a $(k \times 1)$ vector of explanatory variables correlated with ε_t, $\mathbf{z}_t$ is a $(r \times 1)$ vector of instrumental variables with $r \geq k$, $\boldsymbol{\delta}$ is a $(r \times k)$ coefficient matrix, $\mathbf{C}_{v,\varepsilon}$ is a constant correlation vector, and $\mathbf{v}_t$ is a $(k \times 1)$ vector.

In order to employ Hamilton's (2001) methodology in the estimation of equations (8) and (9), we need to transform the model so that the explanatory variables and the disturbance terms are uncorrelated. As in Kim (2004, 2009), the key to the approach is the Cholesky decomposition of the variance-covariance matrix of $[\mathbf{v}_t^{*\prime} \ \varepsilon_t]'$ where $\mathbf{v}_t^* = \boldsymbol{\Sigma}_v^{-1/2} \mathbf{v}_t$ in order to rewrite $[\mathbf{v}_t^{*\prime} \ \varepsilon_t]'$ as a function of independent shocks:

$$\begin{bmatrix} \mathbf{v}_t^* \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} \mathbf{I}_k & \mathbf{0}_k \\ \boldsymbol{\rho}'_{v,\varepsilon}\sigma_\varepsilon & \sqrt{(1-\boldsymbol{\rho}'_{v,\varepsilon}\boldsymbol{\rho}_{v,\varepsilon})}\sigma_\varepsilon \end{bmatrix} \begin{bmatrix} \mathbf{u}_t \\ w_t \end{bmatrix}, \tag{11}$$

$$\begin{bmatrix} \mathbf{u}_t \\ w_t \end{bmatrix} \sim i.i.d.\ N\left(\begin{pmatrix} \mathbf{0}_k \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{I}_k & \mathbf{0}_k \\ \mathbf{0}'_k & 1 \end{pmatrix}\right),$$

where $\mathbf{0}_k$ is $(k \times 1)$ zeros vector, $\boldsymbol{\rho}_{v,\varepsilon}$ is a $(k \times 1)$ vector of correlation coefficients, $\mathbf{I}_k$ is a $(k \times k)$ identity matrix and $\mathbf{u}_t$, and w_t are independent standard normal random variables. From the equation (11), we can rewrite equation (8) and (9) as follows:

$$y_t = \mu(\mathbf{x}_t) + \boldsymbol{\gamma}'\mathbf{v}_t^* + e_t^*, \tag{12}$$

$$\mathbf{x}_t = \boldsymbol{\delta}'\mathbf{z}_t + \boldsymbol{\Sigma}_v^{1/2}\mathbf{u}_t, \tag{13}$$

$$e_t^* \sim i.i.d.N(0, \sigma_{e^*}^2),$$

where $\boldsymbol{\gamma} = \boldsymbol{\rho}_{v,\varepsilon}\sigma_\varepsilon$, $e_t^* = \sqrt{(1-\boldsymbol{\rho}'_{v,\varepsilon}\boldsymbol{\rho}_{v,\varepsilon})}\sigma_\varepsilon w_t$, $\sigma_{e^*}^2 = (1-\boldsymbol{\rho}'_{v,\varepsilon}\boldsymbol{\rho}_{v,\varepsilon})\sigma_\varepsilon^2$, and $\mathbf{v}_t^* = \mathbf{u}_t$. Solving equation (13) for $\mathbf{u}_t$ and substituting this into equation (12) results in the following transformation of equation (8):

$$
\begin{aligned}
y_t &= \mu(\mathbf{x}_t) + [\boldsymbol{\Sigma}_v^{-1/2}(\mathbf{x}_t - \boldsymbol{\delta}'\mathbf{z}_t)]'\boldsymbol{\gamma} + e_t^* && (14)\\
&= \mu(\mathbf{x}_t) + (\mathbf{x}_t - \boldsymbol{\delta}'\mathbf{z}_t)'\boldsymbol{\gamma}^* + e_t^* && (15)\\
&= \mu(\mathbf{x}_t) + \boldsymbol{\gamma}^{*\prime}\mathbf{v}_t + e_t^*, && (16)
\end{aligned}
$$

where $\boldsymbol{\gamma}^* = \boldsymbol{\Sigma}_v^{-1/2}\boldsymbol{\gamma}$. Note that in equation (16), the new disturbance term e_t^* is independent of either $\mathbf{x}_t$ or $\mathbf{v}_t$. The term, $[\boldsymbol{\Sigma}_v^{-1/2}(\mathbf{x}_t - \boldsymbol{\delta}'\mathbf{z}_t)]'\boldsymbol{\gamma}$ $(= \boldsymbol{\gamma}^{*\prime}\mathbf{v}_t)$, works as a bias correction term and we can apply Hamilton's (2001) methodology to have flexible nonlinear inference. Following Hamilton (2001), the system of two equations (12) and (13) can be rewritten as follows:

$$
\begin{aligned}
y_t &= \mu(\mathbf{x}_t) + \boldsymbol{\gamma}^{*\prime}\mathbf{v}_t + e_t^* && (17)\\
&= \alpha_0 + \boldsymbol{\alpha}'\mathbf{x}_t + \boldsymbol{\gamma}^{*\prime}\mathbf{v}_t + \lambda m(\mathbf{x}_t) + e_t^* && (18)\\
\mathbf{x}_t &= \boldsymbol{\delta}'\mathbf{z}_t + \mathbf{v}_t, && (19)\\
e_t^* &\sim i.i.d.N(0, \sigma_{e^*}^2),\ \mathbf{v}_t \sim i.i.d.N(\mathbf{0}_k, \boldsymbol{\Sigma}_v)
\end{aligned}
$$

where $\mathbf{v}_t = (\mathbf{x}_t - \boldsymbol{\delta}'\mathbf{z}_t)$, and $\mu(\mathbf{x}_t) = \alpha_0 + \boldsymbol{\alpha}'\mathbf{x}_t + \lambda m(\mathbf{x}_t)$. Since, in equation (18), $\mathbf{x}_t$, $\mathbf{v}_t$ and $m(\mathbf{x}_t)$ are independent of e_t^* and e_t^* is $i.i.d.N(0, \sigma_{e^*}^2)$, we can apply Hamilton's (2001) procedure to the equation (18) conditional on $\mathbf{x}_t$ and $(\mathbf{x}_t - \boldsymbol{\delta}'\mathbf{z}_t)$. Note that the result from the estimate of equation (11) by Hamilton's (2001) procedure would not be $\mu(\mathbf{x})$, which is the function of interest, but instead of $\mu(\mathbf{x}) + (\mathbf{x}_t - \boldsymbol{\delta}'\mathbf{z}_t)'\boldsymbol{\gamma}^*$. To obtain the estimate of $\mu(\mathbf{x})$, we need to take off the bias correction term.

In estimating the model described in (18) and (19), both joint and two-step estimation are considered in this paper.

3 Joint estimation procedure:FIML

Note that the disturbance terms in (18) and (19) are independent. This provides a basis for constructing the log likelihood function for the joint estimation of the models.

Define $\mathbf{y} = (y_1 \ y_2 \ \cdots y_T)'$, $\boldsymbol{\beta} = (\alpha_0 \ \boldsymbol{\alpha}' \ \boldsymbol{\gamma}^{*\prime})'$, $\mathbf{X}$ is $(T \times (2k+1))$ matrix with tth row $(1 \ \mathbf{x}_t' \ \mathbf{v}_t')$, $\mathbf{x} = (\mathbf{x}_1', ..., \mathbf{x}_T')'$, $\mathbf{Z} = (\mathbf{Z}_1 \ \mathbf{Z}_2 \ \ldots \mathbf{Z}_T)'$, where $\mathbf{Z}_t = \mathbf{I}_k \otimes \tilde{\mathbf{z}}_t$, $\tilde{\mathbf{z}}_t = (1 \ \mathbf{z}_t')'$ and $\mathbf{I}_k$ is $(k \times k)$ identity matrix, $\mathbf{u} = (\mathbf{u}_1' \ \mathbf{u}_2' \ \ldots \mathbf{u}_T')'$, where $\mathbf{u}_i$ for $i = 1, 2, ..., T$, is $(k \times 1)$ vector, $\boldsymbol{\delta} = (\boldsymbol{\delta}_1' \ \ldots \boldsymbol{\delta}_k')'$, where $\boldsymbol{\delta}_i$, $for \ i = 1, ..., k$, is $[(r+1) \times 1]$ vector, and $\mathbf{V} = \mathbf{I}_T \otimes \boldsymbol{\Sigma}_v$, where $\mathbf{I}_k$ is $(k \times k)$ identity matrix. Since the regression errors e_t^* is assumed to be $i.i.d.N(0, \sigma_{e^*}^2)$ and $(\mathbf{x}_t', \mathbf{z}_t')$ are strictly exogenous, the specification of the equations (18) and (19) implies a GLS regression model of the form

$$\mathbf{y}|\mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{P}_0 + \sigma_{e^*}^2 \mathbf{I}_T) \tag{20}$$

$$\mathbf{x}|\mathbf{Z} \sim N(\mathbf{Z}\boldsymbol{\delta}, \mathbf{V}), \tag{21}$$

where

$$\mathbf{P}_0 = [\lambda^2 H_k(h_{ij})]_{i,j=1,2,\ldots,T}, \tag{22}$$

$$h_{ij} = (1/2)\{[\mathbf{g} \odot (\mathbf{x}_i - \mathbf{x}_j)]'[\mathbf{g} \odot (\mathbf{x}_i - \mathbf{x}_j)]\}^{1/2}. \tag{23}$$

We define the parameters associated with the models, (18) and (19) as:

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_1' \ \boldsymbol{\theta}_2']', \tag{24}$$

where $\boldsymbol{\theta}_1 = [\alpha_0, \boldsymbol{\alpha}', \lambda, \mathbf{g}', \boldsymbol{\gamma}^{*\prime}, \sigma^2_{e^*}]'$ is the vector of parameters that are associated with equation (18) and $\boldsymbol{\theta}_2 = [\boldsymbol{\delta}', vech(\boldsymbol{\Sigma}_v)']'$ is the vector of parameters that are associated with equation (19). For consistent and efficient joint estimation of the model (18) and (19), we maximize the following log likelihood function with respect to $\boldsymbol{\theta}$:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \ln f(\mathbf{y}, \mathbf{X}; \boldsymbol{\theta}) \\ &= \ln f(\mathbf{y}|\mathbf{X}; \boldsymbol{\theta}) + \ln f(\mathbf{x}; \boldsymbol{\theta}_2), \end{aligned} \tag{25}$$

where

$$
\begin{aligned}
\ln f(\mathbf{y}|\mathbf{X};\boldsymbol{\theta}) &= -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{P}_0 + \sigma_{e^*}^2\mathbf{I}_T| \\
&\quad -\left\{(-\frac{1}{2})(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{P}_0+\sigma_{e^*}^2\mathbf{I}_T)^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\} \qquad (26) \\
\ln f(\mathbf{x};\boldsymbol{\theta}_2) &= -\frac{Tk}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{V}| - \left\{(-\frac{1}{2})(\mathbf{x}-\mathbf{Z}\boldsymbol{\delta})'\mathbf{V}^{-1}(\mathbf{x}-\mathbf{Z}\boldsymbol{\delta})\right\}. \qquad (27)
\end{aligned}
$$

Define $\zeta \equiv \lambda/\sigma_{e^*}$ to be the ratio of the standard deviation of the nonlinear component $\lambda m(\mathbf{x}_t)$ to that of the regression residual e^*. As in Hamilton (2001), a convenient reparameterization can allow easier estimation structure of the equations (20) - (23). Let $\boldsymbol{\theta}_1 = (\boldsymbol{\theta}_{11}', \ \boldsymbol{\theta}_{12}')'$, where $\boldsymbol{\theta}_{11} = (\alpha_0, \boldsymbol{\alpha}', \boldsymbol{\gamma}^{*\prime}, \sigma_{e^*}^2)'$ contain the parameters from the linear part of the model (18) and $\boldsymbol{\theta}_{12} = (\mathbf{g}', \ \zeta)'$ the nonlinear parameters. Let $\mathbf{H}(\mathbf{g})$ denote the $(T \times T)$ matrix whose (t,s) element is $H_k(h_{ts}(\mathbf{g}))$ and

$$
\mathbf{W}(\mathbf{x};\boldsymbol{\theta}_{12}) \equiv \zeta^2\mathbf{H}(\mathbf{g}) + \mathbf{I}_T, \qquad (28)
$$

where for each pair of observations t and s, $\tilde{\mathbf{x}}_t = \mathbf{g} \odot \mathbf{x}_t$ and $h_{ts}(\mathbf{g}) = (1/2)[(\tilde{\mathbf{x}}_t - \tilde{\mathbf{x}}_s)'(\tilde{\mathbf{x}}_t - \tilde{\mathbf{x}}_s)]^{1/2}$. Note from (26) that the log likelihood can be written

$$
\begin{aligned}
\ln f(\mathbf{y}|\mathbf{X};\boldsymbol{\theta}_{12},\boldsymbol{\theta}_2) &= -\frac{T}{2}\ln(2\pi) - -\frac{T}{2}\ln(\sigma_{e^*}^2) - \frac{1}{2}\ln|\mathbf{W}(\mathbf{x};\boldsymbol{\theta}_{12})| \\
&\quad -\left\{(-\frac{1}{2\sigma_{e^*}^2})(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{W}(\mathbf{x};\boldsymbol{\theta}_{12})^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\}. \qquad (29)
\end{aligned}
$$

For given $\boldsymbol{\theta}_{12}, \boldsymbol{\theta}_2$, the value of $\boldsymbol{\theta}_{11}$ that maximizes (29) can be calculated analytically as

$$\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}_{12}, \boldsymbol{\theta}_2) = \left[\mathbf{X}'\mathbf{W}(\mathbf{x}; \boldsymbol{\theta}_{12})^{-1}\mathbf{X}\right]^{-1} \left[\mathbf{X}'\mathbf{W}(\mathbf{x}; \boldsymbol{\theta}_{12})^{-1}\mathbf{y}\right], \tag{30}$$

$$\tilde{\sigma}^2_{e^*}(\boldsymbol{\theta}_{12}, \boldsymbol{\theta}_2) = \left[\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}_{12}, \boldsymbol{\theta}_2)\right]' \mathbf{W}(\mathbf{x}; \boldsymbol{\theta}_{12})^{-1} \left[\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}_{12}, \boldsymbol{\theta}_2)\right] / T \tag{31}$$

The equations (30) and (31) allow us to concentrate the log likelihood (29) as

$$\begin{aligned} L(\boldsymbol{\theta}_{12}; \mathbf{y}, \mathbf{X}) &= \sum_{t=1}^{T} \ln f(y_t | \mathbf{X}_t, \mathbf{Y}_{t-1}; \tilde{\boldsymbol{\theta}}_{11}(\boldsymbol{\theta}_{12}), \boldsymbol{\theta}_{12}, \boldsymbol{\theta}_2) \\ &= -\frac{T}{2} \ln(2\pi) - -\frac{T}{2} \ln(\tilde{\sigma}^2_{e^*}) - \frac{1}{2} \ln |\mathbf{W}(\mathbf{x}; \boldsymbol{\theta}_{12})| - (T/2), \end{aligned} \tag{32}$$

where $\mathbf{Y}_t = (y_t, \mathbf{X}'_t, y_{t-1}, ..., y_1, \mathbf{X}'_1)'$ denote information observed through date t. Now, the numerically maximizing (27) and (32) gives the MLE $\hat{\boldsymbol{\theta}}_2$, and $\hat{\boldsymbol{\theta}}_{12}$, which from (30) and (31) gives $\hat{\boldsymbol{\theta}}_{11}$. Therefore, the joint estimation procedure is based on the equations (27), (30), (31) and (32) and would deliver an asymptotically most efficient estimator.

4 Two-step estimation procedure

Although the joint estimation procedure would deliver an asymptotically most efficient estimator, the joint estimation procedure may be subject to the curse of dimensionality as pointed out in Kim (2009) and a reasonable alternative to the joint estimation procedure is a two-step estimation procedure. Recall that $\boldsymbol{\theta}_1$ is parameter vectors associated with equation (18) and $\boldsymbol{\theta}_2$ is the vector of parameters associated with equation (19). To get an insight into the two-step

estimation of the model given by equations (18) and (19), consider the log likelihood function in equation (25). The basic idea for a two-step procedure is to estimate $\boldsymbol{\theta}_2$ by maximizing $\ln f(\mathbf{x};\boldsymbol{\theta}_2)$, and then to estimate $\boldsymbol{\theta}_1$ by maximizing $\ln f(\mathbf{y}|\mathbf{x};\boldsymbol{\theta}) = \ln f(\mathbf{y}|\mathbf{X};\boldsymbol{\theta}_1,\hat{\boldsymbol{\theta}}_2)$ conditional on the estimates for $\boldsymbol{\theta}_2$. The associated cost of a two-step procedure, however, is the potential loss of efficiency. We summarize the two-step estimation procedure as follows:

Step 1:

In the first step, the equation to be estimated is:

$$\begin{aligned} \mathbf{x}_t &= \boldsymbol{\delta}'\mathbf{z}_t + \mathbf{v}_t \\ \mathbf{v}_t &\sim i.i.d.N(\mathbf{0}_k, \boldsymbol{\Sigma}_v). \end{aligned} \tag{33}$$

The equation (27) is the log likelihood function associated with the equation (33), and so, the log likelihood function is maximized with respect to $\boldsymbol{\theta}_2$ and then, we obtain the consistent estimates for $\hat{\boldsymbol{\theta}}_2 = [\hat{\boldsymbol{\delta}}', vech(\hat{\boldsymbol{\Sigma}}_v)']'$.

Step 2:

In the second step, we estimate equation (18) conditional on $\hat{\boldsymbol{\theta}}_2$ which is obtained from the Step 1. The equation to be estimated is:

$$\begin{aligned} y_t &= \alpha_0 + \boldsymbol{\alpha}'\mathbf{x}_t + \boldsymbol{\gamma}^{*\prime}\hat{\mathbf{v}}_t + \lambda m(\mathbf{x}_t) + e_t \\ &= \alpha_0 + \boldsymbol{\alpha}'\mathbf{x}_t + \boldsymbol{\gamma}^{*\prime}[(\mathbf{x}_t - \hat{\boldsymbol{\delta}}'\mathbf{z}_t)] + \lambda m(\mathbf{x}_t) + e_t \\ e_t &\sim i.i.d.N(0, \sigma_e^2), \end{aligned} \tag{34}$$

where $e_t = e_t^* + (\mathbf{v}_t - \hat{\mathbf{v}}_t)'\boldsymbol{\gamma}^*,\ \hat{\mathbf{v}}_t = \mathbf{x}_t - \hat{\boldsymbol{\delta}}'\mathbf{z}_t$.

The log likelihood function to be maximized is given by:

$$\begin{aligned}\ln f(\mathbf{y}|\tilde{\mathbf{X}};\boldsymbol{\theta}_1,\hat{\boldsymbol{\theta}}_2) &= -\frac{T}{2}(2\pi) - \frac{1}{2}|\mathbf{P}_0 + \sigma_e^2\mathbf{I}_T| \\ &\quad - \left\{(-\frac{1}{2})(\mathbf{y} - \tilde{\mathbf{X}}\boldsymbol{\beta})'(\mathbf{P}_0 + \sigma_e^2\mathbf{I}_T)^{-1}(\mathbf{y} - \tilde{\mathbf{X}}\boldsymbol{\beta})\right\}, \qquad (35)\end{aligned}$$

where $\tilde{\mathbf{X}}$ =$(T \times (2k+1))$ matrix with tth row $(1\ \mathbf{x}_t'\ \hat{\mathbf{v}}_t')$.

As the case of a convenient reparameterization in the joint estimation procedure, we estimate equation (34) with a convenient reparameterization. Let $\boldsymbol{\theta}_1^* = (\boldsymbol{\theta}_{11}^{*\prime},\ \boldsymbol{\theta}_{12}')'$, where $\boldsymbol{\theta}_{11}^* = (\alpha_0, \boldsymbol{\alpha}', \boldsymbol{\gamma}^{*\prime}, \sigma_e^2)'$ contain the parameters from the linear part of the model (34) and $\boldsymbol{\theta}_{12} = (\mathbf{g}',\ \zeta)'$ the nonlinear parameters. Note from (29) that the log likelihood conditional on $\hat{\boldsymbol{\theta}}_2$ can be written

$$\begin{aligned}\ln f(\mathbf{y}|\tilde{\mathbf{X}};\boldsymbol{\theta}_{12},\hat{\boldsymbol{\theta}}_2) &= -\frac{T}{2}\ln(2\pi) - -\frac{T}{2}\ln(\sigma_e^2) - \frac{1}{2}\ln|\mathbf{W}(\mathbf{x};\boldsymbol{\theta}_{12})| \\ &\quad - \left\{(-\frac{1}{2\sigma_e^2})(\mathbf{y} - \tilde{\mathbf{X}}\boldsymbol{\beta})'\mathbf{W}(\mathbf{x};\boldsymbol{\theta}_{12})^{-1}(\mathbf{y} - \tilde{\mathbf{X}}\boldsymbol{\beta})\right\}. \qquad (36)\end{aligned}$$

For given $\boldsymbol{\theta}_{12}, \hat{\boldsymbol{\theta}}_2$, the value of $\boldsymbol{\theta}_{11}$ that maximizes (36) can be calculated analytically as:

$$\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}_{12},\hat{\boldsymbol{\theta}}_2) = \left[\tilde{\mathbf{X}}'\mathbf{W}(\mathbf{x};\boldsymbol{\theta}_{12})^{-1}\tilde{\mathbf{X}}\right]^{-1}\left[\tilde{\mathbf{X}}'\mathbf{W}(\mathbf{x};\boldsymbol{\theta}_{12})^{-1}\mathbf{y}\right], \qquad (37)$$

$$\tilde{\sigma}_e^2(\boldsymbol{\theta}_{12},\hat{\boldsymbol{\theta}}_2) = \left[\mathbf{y} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}_{12},\hat{\boldsymbol{\theta}}_2)\right]'\mathbf{W}(\mathbf{x};\boldsymbol{\theta}_{12})^{-1}\left[\mathbf{y} - \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}}(\boldsymbol{\theta}_{12},\boldsymbol{\theta}_2)\right]/T. \qquad (38)$$

The equations (37) and (38) allow us to concentrate the log likelihood (36) conditional on $\hat{\boldsymbol{\theta}}_2$ as

$$\begin{aligned} L(\boldsymbol{\theta}_{12}; \mathbf{y}, \mathbf{X}) &= \sum_{t=1}^{T} \ln f(y_t | \mathbf{X}_t, \mathbf{Y}_{t-1}; \boldsymbol{\theta}_{11}^*(\boldsymbol{\theta}_{12}), \boldsymbol{\theta}_{12}, \hat{\boldsymbol{\theta}}_2) \\ &= -\frac{T}{2} \ln(2\pi) - -\frac{T}{2} \ln(\tilde{\sigma}_e^2) - \frac{1}{2} \ln |\mathbf{W}(\mathbf{x}; \boldsymbol{\theta}_{12})| - (T/2). \end{aligned} \tag{32}$$

Now, the numerically maximizing (32) conditional on $\hat{\boldsymbol{\theta}}_2$ gives the MLE $\hat{\boldsymbol{\theta}}_{12}$, which from (37) and (38) gives $\hat{\boldsymbol{\theta}}_{11}^*$. Even though the two-step procedure provides a consistent estimation of the model, the covariance matrix of $\hat{\boldsymbol{\beta}}$ obtained by inverting the negative of the Hessian matrix would be biased, due to the generated regressors $\mathbf{x}_t - \hat{\boldsymbol{\delta}}'\mathbf{z}_t$ that replace $\mathbf{x}_t - \boldsymbol{\delta}'\mathbf{z}_t$ in the second-step regression of equation (18).

5 Concluding remarks

A linear regression is not a good choice in the certain cases such as the relationship between oil prices and business cycle (Hamilton 2003, Kim 2012) and scholars tend to consider nonlinear specifications. The crucial issue, however, is how ones choose a proper specification over all possible nonlinear relations. Hamilton (2001) proposes a flexible nonlinear inference where the philosophy of his methodology is non parametric idea but the estimation is a parametric approach and Hamilton (2003) and Kim (2012) show that this methodology is very useful in addressing nonlinear relation between oil prices and the business cycle in the time-series and in the panel

framework respectively.

Hamilton's (2001) methodology, however, is not valid in the presence of endogenous explanatory variables. This paper develops a flexible nonlinear inference with endogenous regressors. We apply the control function approach and show that there exists an appropriate transformation of the model that allows us to directly employ Hamilton's (2001) approach. We propose two estimation procedures: a joint estimation procedure and a two-step estimation procedure. The parameters in these procedures can be estimated by maximum likelihood or Bayesian methods.

Our new methodology can be useful in the macroeconometric models where there are endogenous explanatory variables and true relation between dependent variable and explanatory variables is nonlinear such as the cases of nonlinear Taylor rule (Kim et al. 2005) and nonlinear macro model (Wolman 2006). We leave the application of new methodology for future research.

References

Altonji, Joseph G., and Matzkin, Rosa L„ 2005, "Cross section and panel data estimators for nonseparable models with endogenous regressors" *Econometrica* 73(4), 1053-1102.

Florens, Jean-Pierre, Heckman, James J., Meghir, Costas, and Vytlacil, Edward J., 2007, "Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects," Columbia University working paper.

Hamilton, James D., 2001, "A parametric approach to flexible nonlinear inference" *Econometrica* 69 (No.3), 537-573.

Hamilton, James D., 2003, "What is an oil shock?" *Journal of Econometrics* 113, 363-398.

Heckman, James J., and Vytlacil, Edward J., 1998, "Instrumental variables methods for the correlated random coefficient model: Estimating the average rate of return to schooling when the return is correlated with schooling" *Journal of Human Resources* 33(4), 974-987.

Heckman, James J., and Navarro, Salvador, 2004, "Using matching, instrumental variables, and control functions to estimate economic choice models" *Review of Economics and Statistics* 86(1), 30-57.

Kim, Chang-Jin, 2004, "Markov-switching models with endogenous explanatory variables" *Journal of Econometrics* 122, 127-136.

Kim, Chang-Jin, 2009, "Markov-switching models with endogenous explanatory variables II: A two-step MLE procedure" *Journal of Econometrics* 148, 46-55.

Kim, Chang-Jin, and Nelson, Charles R., 2006, "Estimation of a forward-looking monetary policy rule: A time-varying parameter model using ex post data" *Journal of Monetary Economics* 53, 1949-1966.

Kim, Dong Heon, 2012, "What is an oil shock? Panel data evidence" *Empirical Economics* 43, 121 - 143.

Kim, Dong Heon, Osborn, Denise R., and Sensier, Marianne, 2005, "Nonlinearity in the FED's monetary policy rule" *Journal of Applied Econometrics* 20, 621-639.

Vanmarcke, Erik, 2010, *Random fields: Analysis and Synthesis*, World Scientific Publishing Company.

Wolman, Alexander L., 2006, "The Zero Bound and the Term Structure in a Nonlinear Macro Model" *Seoul Journal of Economics* 19 (1), 147 - 170.