Flexible Nonlinear Inference with Endogenous Explanatory Variables

Dong Heon Kim

Hamilton's (2001) flexible nonlinear inference is not valid with endogenous explanatory variables. Hence, this paper proposes a framework to approach endogeneity problems in the flexible non-linear inference. We develop two estimation procedures, namely, joint estimation and two-step estimation procedures. The parameters in both models can be estimated by maximum likelihood or numerical Bayesian method. Our approach can be used in handling endogeneity and nonlinearity in the oil-macro relationship or in the monetary policy rule.

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

JEL Classification: C13, C32

I. Introduction

A natural approach to estimate a typical economic model is by using the linear relationship between relevant variables. Hamilton (2001) reveals that nonlinear models may improve forecasts and provide economic insights. Generally, parametric or nonparametric approaches are employed when estimating nonlinear models. A crucial aspect in using parametric approaches is deciding which parametric model to use. Meanwhile,

* Professor, Department of Economics, Korea University, 145 Anam-ro, Seongbuk-Gu, Seoul, 136-701, South Korea, (Tel): +82 2 3290 2226, (Fax): +82 2 3290 2661, (Email): dongkim@korea.ac.kr. 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. The Korea University Research Grant is also acknowledged.

popular nonparametric approaches sacrifice many of the benefits associated with parametric methods, such as the provision of a system for adjusting a bandwidth or series expansion length, improving unclear interpretation of inferences, and convenient hypothesis testing.

To develop a parametric approach along with a nonparametric philosophy, Hamilton (2001) proposes a new framework that can determine whether a given relationship is nonlinear, what the nonlinear function is, and whether a particular model can adequately describe it. 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 is $y_t=\mu(\mathbf{x}_t)+\varepsilon_t$ and the functional form of $\mu(\cdot)$ is unknown. The paper denotes $\mu(\cdot)$ as the outcome of a random process and introduces a stationary random field $m(\cdot)$, whose realizations can represent a broad class of possible forms for $\mu(\cdot)$. The proposed approach considers the parameters that characterize the relationship 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 method.

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 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 solve the endogeneity in Markov-switching regression models; these procedures are based on the control function approach (Heckman, and Vytlacil 1998; Heckman, and Navarro 2004; Altonji, and Matzkin 2005; Florens *et al.* 2007). Kim, and Nelson (2006) illustrate that the two-step MLE procedure is expedient in the estimation of a forward-looking monetary policy rule in the U.S.A.

This paper aims to develop a flexible nonlinear inference with endogenous regressors in the framework of Hamilton's methodology (2001). In this new approach, we apply the control function approach to Hamilton's (2001) flexible nonlinear framework. An appropriate transformation of the model allows us to employ Hamilton's (2001) approach directly. To estimate a flexible nonlinear model with endogenous regressors, both joint and two-step estimation procedures are considered in this paper. The parameters in these procedures are estimated by maximum likelihood or numerical Bayesian method.

The rest of the paper is organized into the following sections. Section II discusses a nonlinear form with endogenous explanatory variables, section III describes the joint estimation procedure, section IV derives the two-step estimation procedure, and section V concludes the paper.

II. Nonlinear form with endogenous explanatory variables

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

Consider the general nonlinear regression model

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

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}, $\mathbf{x}_{t-j}$ $(j=1, 2, \ldots)$. Equation (1) allows a subset of variables $\mathbf{x}_t$, which ones tend to assume linearity, thereby gaining efficiency. The form of the function $\mu(\cdot)$ is unknown, and we seek to represent it with 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 τ is treated as a Gaussian random variable with a mean value that equal to the linear component $\alpha_0+\boldsymbol{\alpha}'\tau$ and variance λ^2, where α_0, $\boldsymbol{\alpha}$, and λ are population parameters to be estimated.[2] In the special case of $\lambda=0$, $\mu(\mathbf{x}_t)$ is fixed, and Equation (1) becomes the usual linear regression model. In general, the parameter λ measures the overall extent of nonlinearity.

[1] A random field is a generalization of a stochastic process, such that the underlying parameter does not have to be a simple real or integer valued "time," but instead, can be multidimensional vectors or points on a certain manifold. At its most basic and 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, adjacent values (*i.e.*, values with adjacent indices) do not differ as greatly as values that are further apart, which is an example of a covariance structure. Numerous types of this structure may be modeled in a random field, which is known as a "function valued" random variable. (Vanmarcke 2010, Wikipedia)

[2] We do not know the functional form of $\mu(\cdot)$; thus, the final outcome $\mu(\tau)$ evaluated at the realized value τ can be treated as a random variable.

In the random field $\mu(\cdot)$, ones need to know how the random variable $\mu(\tau_1)$ is correlated with $\mu(\tau_2)$, for τ_1 and τ_2 arbitrary k-dimensional vectors. Hamilton (2001) parameterizes this correlation based on the distance measure $h_{st}=(1/2)[\sum_{i=1}^{k} g_i^2(\mathbf{x}_{is}-\mathbf{x}_{it})^2]^{1/2}$, where $\mathbf{x}_{it}$ denotes the ith element of the vector $\mathbf{x}_t$ and $g_1, g_2, \ldots, g_k$ 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 $\mathbf{x}_t$. To be precise,

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

However, when $0\leq h_{st}\leq 1$, this correlation should increase as h_{st} decreases, with the correlation reaching unity as h_{st} becomes 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

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

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

where $m(\cdot)$ is the realization of a scalar-valued Gaussian random field with mean zero and unit variance and covariance function given by Equations (2) to (4). Nonlinearity of the functional form $\mu(\cdot)$ implies a correlation between u_t and u_s, which are 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 given by

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

where $y=(y_1, y_2, ..., y_T)'$, $\mathbf{X}$ is the $T\times(k+1)$ matrix with th row $(1, \mathbf{x}_t')'$, β is the $(1+k)$-dimensional vector $(\alpha_0, \alpha')'$, 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. The function $H_k(.)$ is specified in Equation (4) if $k=2$. The indicator function $\delta_{[.]}$ is unity when the condition [.] holds, and zero otherwise.

In addition to the linear regression parameters (α_0, α) and σ^2 the parameters to be estimated are the variances of the nonlinear regression error, λ^2, which governs the overall importance of the nonlinear component, and the parameters $(g_1, g_2, ..., g_k)$ determine the variability of the nonlinear component with respect to each explanatory variable in $\mathbf{x}_t$. As the aforementioned discussion implies, estimation and inference can be achieved by a GLS Gaussian regression or numerical Bayesian method.

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

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

$$\mathbf{x}_t = \delta'\mathbf{z}_t + \mathbf{v}_t, \mathbf{v}_t \sim i.i.d.\ N(0, \Sigma_v), \tag{9}$$

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

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 $r\geq k$, δ 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.

To employ Hamilton's (2001) methodology in the estimation of Equations (8) and (9), we must 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 $\mathbf{v}_t^*=\Sigma_v^{-1/2}\mathbf{v}_t$ to rewrite $[\mathbf{v}_t^{*\prime}\varepsilon_t]'$ as a function of independent shocks, which is given by

$$\begin{aligned}\begin{bmatrix}\mathbf{v}_t^* \\ \varepsilon_t\end{bmatrix} &= \begin{bmatrix}\mathbf{I}_k & \mathbf{0}_k \\ \rho_{v,\varepsilon}'\sigma_\varepsilon & \sqrt{(1-\rho_{v,\varepsilon}'\rho_{v,\varepsilon})}\sigma_\varepsilon\end{bmatrix}\begin{bmatrix}\mathbf{u}_t \\ w_t\end{bmatrix} \\ \begin{bmatrix}\mathbf{u}_t \\ w_t\end{bmatrix} &\sim i.i.d.\ N\left(\begin{pmatrix}\mathbf{0}_k \\ 0\end{pmatrix}, \begin{pmatrix}\mathbf{I}_k & \mathbf{0}_k \\ \mathbf{0}_k' & 1\end{pmatrix}\right),\end{aligned} \tag{11}$$

where $\mathbf{O}_k$ is $(k\times 1)$ zero vector, $\rho_{v,\varepsilon}$ 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 Equation (11), we can respectively rewrite Equations (8) and (9) as

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

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

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

where $\gamma = \rho_{v,\varepsilon}\sigma_\varepsilon$, $e_t^* = \sqrt{(1-\rho_{v,\varepsilon}{}'\rho_{v,\varepsilon})}\sigma_\varepsilon w_t$, $\sigma_{e^*}^2 = (1-\rho_{v,\varepsilon}{}'\rho_{v,\varepsilon})\sigma_\varepsilon^2$, and $\mathbf{v}_t^* = \mathbf{u}_t$. Solving Equation (13) for $\mathbf{u}_t$ and substituting the outcome with Equation (12) results in the transformation of Equation (8) given by

$$y_t = \mu(\mathbf{x}_t) + [\Sigma_v^{-1/2}(\mathbf{x}_t - \delta'\mathbf{z}_t)]'\gamma + e_t^*, \tag{14}$$

$$= \mu(\mathbf{x}_t) + (\mathbf{x}_t - \delta'\mathbf{z}_t)'\gamma^* + e_t^*, \tag{15}$$

$$= \mu(\mathbf{x}_t) + \gamma^{*\prime}\mathbf{v}_t + e_t^*, \tag{16}$$

where $\gamma^* = \Sigma_v^{-1/2}\gamma$. In Equation (16), the new disturbance term e_t^* is independent of either $\mathbf{x}_t$ or $\mathbf{v}_t$. In addition, $[\Sigma_v^{-1/2}(\mathbf{x}_t - \delta'\mathbf{z}_t)]'\gamma(=\gamma^{*\prime}\mathbf{v}_t)$ works as a bias correction term, and we can apply Hamilton's (2001) methodology to have a flexible nonlinear inference. Following Hamilton (2001), the systems of Equations (12) and (13) can be respectively rewritten as

$$y_t = \mu(\mathbf{x}_t) + \gamma^{*\prime}\mathbf{v}_t + e_t^*, \tag{17}$$

$$= \alpha_0 + \alpha'\mathbf{x}_t + \gamma^{*\prime}\mathbf{v}_t + \lambda m(\mathbf{x}_t) + e_t^*, \tag{18}$$

$$\mathbf{x}_t = \delta'\mathbf{z}_t + \mathbf{v}_t, \tag{19}$$

$$e_t^* \sim i.i.d.N(0, \sigma_{e^*}^2), \mathbf{v}_t \sim i.i.d.\ N(\mathbf{o}_k, \Sigma_v)$$

where $\mathbf{v}_t = (\mathbf{x}_t - \delta'\mathbf{z}_t)$, and $\mu(\mathbf{x}_t) = \alpha_0 + \alpha'\mathbf{x}_t + \lambda m(\mathbf{x}_t)$. In Equation (18), $\mathbf{x}_t$, $\mathbf{v}_t$ and $m(\mathbf{x}_t)$ are independent of e_t^*, and e_t^* is $i.i.d.N(0, \sigma_{e^*}^2)$, thus we can apply Hamilton's (2001) procedure to the equation conditional on $\mathbf{x}_t$ and $(\mathbf{x}_t - \delta'\mathbf{z}_t)$. The result from the estimation of Equation (11) using Hamilton's

(2001) procedure is not $\mu(\mathbf{x})$, which is the function of interest, instead it is $\mu(\mathbf{x})+(\mathbf{x}_t-\delta'\mathbf{z}_t)'\gamma^*$. To obtain the estimate of $\mu(\mathbf{x})$ we need to take out the bias correction term.

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

III. Joint estimation procedure: FIML

The disturbance terms in Equations (18) and (19) are independent, providing a basis to construct the log likelihood function for the joint estimation of the models.

We define that $\mathbf{y}=(y_1\, y_2 \ldots y_T)'$, $\beta=(\alpha_0\, \alpha'\, \gamma^{*\prime})'$, $\mathbf{X}$ is $(T\times(2k+1))$ matrix with tth row $(1\, \mathbf{x}_t'\, \mathbf{v}_t')$, $\mathbf{x}=(\mathbf{x}_1', \ldots, \mathbf{x}_T')'$, $\mathbf{Z}=(\mathbf{Z}_1\, \mathbf{Z}_2 \ldots \mathbf{Z}_T)'$, 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. Moreover, define $\mathbf{u}=(\mathbf{u}'_1\, \mathbf{u}'_2 \ldots \mathbf{u}'_T)'$, where $\mathbf{u}_i$ for $i=1, 2, \ldots, T$, is $(k\times 1)$ vector, $\delta=(\delta_1' \ldots \delta_k')'$, where δ_i, for $i=1, \ldots, k$, is $[(r+1)\times 1]$ vector, and $\mathbf{V}=\mathbf{I}_T\otimes\Sigma_v$, where $\mathbf{I}_k$ is $(k\times k)$ identity matrix. The regression error e_t^* in Equation (17) is assumed to be $i.i.d.N(0, \sigma_{e^*}^2)$, and $(\mathbf{x}_t', \mathbf{z}_t')$ are strictly exogenous; thus, the specifications of Equations (18) and (19) imply a GLS regression model of the forms

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

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

where

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

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

where $\oplus$ denotes an element-by-element multiplication.

We define the parameters associated with Equations (18) and (19) as

$$\theta = [\theta_1'\, \theta_2']', \tag{24}$$

where $\theta_1=[\alpha_0, \alpha', \lambda, \mathbf{g}', \gamma^{*\prime}, \sigma_{e^*}^2]'$ is the vector of parameters associated with Equation (18), and $\theta_2=[\delta', vech(\Sigma_v)']'$ is the vector of parameters associated with Equation (19). For consistent and efficient joint estimation of Equations (18) and (19), we maximize the following log likelihood function with respect to θ:

$$
\begin{aligned}
L(\theta) &= \ln f(\mathbf{y}, \mathbf{X} : \theta) \\
&= \ln f(\mathbf{y} \mid \mathbf{X}; \theta) + \ln f(\mathbf{x}; \theta_2),
\end{aligned}
\tag{25}
$$

where

$$
\begin{aligned}
\ln f(\mathbf{y} \mid \mathbf{X}; \theta) = &-\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln \mid \mathbf{P}_0 + \sigma_{e^*}^2 \mathbf{I}_T \mid \\
&- \left\{ (-\frac{1}{2})(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{P}_0 + \sigma_{e^*}^2 \mathbf{I}_T)^{-1}(\mathbf{y} - \mathbf{X}\beta) \right\},
\end{aligned}
\tag{26}
$$

$$
\ln f(\mathbf{x}; \theta_2) = -\frac{Tk}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{V}| - \left\{ (-\frac{1}{2})(\mathbf{x} - \mathbf{Z}\delta)'\mathbf{V}^{-1}(\mathbf{x} - \mathbf{Z}\delta) \right\}. \tag{27}
$$

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 Equations (20)−(23). Let $\theta_1 = (\theta_1{}'_1, \theta_1{}'_2)'$, where $\theta_{11} = (\alpha_0, \alpha', \gamma^{*\prime}, \sigma_{e^*}^2)'$ contains the parameters from the linear part of Equation (18) and $\theta_{12} = (\mathbf{g}', \zeta)'$ 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

$$
\mathbf{W}(\mathbf{x}; \theta_{12}) \equiv \zeta^2 \mathbf{H}(\mathbf{g}) + \mathbf{I}_T, \tag{28}
$$

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

$$
\begin{aligned}
\ln f(\mathbf{y} \mid \mathbf{X}; \theta_{12}, \theta_2) = &-\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma_{e^*}^2) - \frac{1}{2}\ln \mid \mathbf{W}(\mathbf{x}; \theta_{12}) \mid \\
&- \left\{ (-\frac{1}{2\sigma_{e^*}^2})(\mathbf{y} - \mathbf{X}\beta)'\mathbf{W}(\mathbf{x}; \theta_{12})^{-1}(\mathbf{y} - \mathbf{X}\beta) \right\}.
\end{aligned}
\tag{29}
$$

For given θ_{12}, θ_2, the value of θ_{11} that maximizes Equation (29) can be calculated analytically as

$$\tilde{\beta}(\theta_{12}, \theta_2) = \left[\mathbf{X}'\mathbf{W}(\mathbf{x}; \theta_{12})^{-1}\mathbf{X}\right]^{-1}\left[\mathbf{X}'\mathbf{W}(\mathbf{x}; \theta_{12})^{-1}\mathbf{y}\right], \tag{30}$$

$$\tilde{\sigma}^2_{e^*}(\theta_{12}, \theta_2) = \left[\mathbf{y} - \mathbf{X}\tilde{\beta}(\theta_{12}, \theta_2)\right]' \mathbf{W}(\mathbf{x}; \theta_{12})^{-1}\left[\mathbf{y} - \mathbf{X}\tilde{\beta}(\theta_{12}, \theta_2)\right]/T. \tag{31}$$

Equations (30) and (31) allow us to concentrate the log likelihood in Equation (29) as

$$\begin{aligned} L(\theta_{12}; \mathbf{y}, \mathbf{X}) &= \sum_{t=1}^{T} \ln f(y_t \mid \mathbf{X}_t, \mathbf{Y}_{t-1}; \tilde{\theta}_{11}(\theta_{12}), \theta_{12}, \theta_2) \\ &= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\tilde{\sigma}^2_{e^*}) - \frac{1}{2}\ln|\mathbf{W}(\mathbf{x}; \theta_{12})| - (T/2), \end{aligned} \tag{32}$$

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

IV. Two-step estimation procedure

Although the joint estimation procedure can deliver the most asymptotically efficient estimator, this procedure may be subject to the curse of dimensionality as highlighted by Kim (2009). Hence, a reasonable alternative is a two-step estimation procedure. Here, θ_1 is the vector of parameters associated with Equation (18), and θ_2 is the vector of parameters associated with Equation (19). To obtain 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 θ_2 and θ_1 by maximizing $\ln f(\mathbf{x}; \theta_2)$ and $\ln f(\mathbf{y} \mid \mathbf{x}; \theta) = \ln f(\mathbf{y} \mid \mathbf{X}; \theta_1, \hat{\theta}_2)$, respectively, depending on the estimates for θ_2. The associated cost of a two-step procedure, however, is the potential loss of efficiency. We summarize the two-step estimation procedure below.

Step 1:

In the first step, the equation to be estimated as

$$\mathbf{x}_t = \delta'\mathbf{z}_t + \mathbf{v}_t \qquad \mathbf{v}_t \sim i.i.d.\, N(\mathbf{0}_k, \Sigma_v). \tag{33}$$

Equation (27) is the log likelihood function associated with Equation (33); thus, the log likelihood function is maximized with respect to θ_2 and then, we obtain the consistent estimates for $\hat{\theta}_2 = [\hat{\delta}', vech(\hat{\Sigma}_v)']'$.

Step 2:

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

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

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

The log likelihood function to be maximized is given by

$$\begin{aligned} \ln f(\mathbf{y} \mid \tilde{\mathbf{X}}; \theta_1, \hat{\theta}_2) = &-\frac{T}{2}(2\pi) - \frac{1}{2} \mid \mathbf{P}_0 + \sigma_e^2 \mathbf{I}_T \mid \\ &- \left\{ (-\frac{1}{2})(\mathbf{y} - \tilde{\mathrm{X}}\beta)'(\mathbf{P}_0 + \sigma_e^2\mathbf{I}_T)^{-1}(\mathbf{y} - \tilde{\mathbf{X}}\beta) \right\}, \end{aligned} \tag{35}$$

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

As the case of the convenient reparameterization in the joint estimation procedure, we estimate Equation (34) with a convenient reparameterization. Let $\theta_1^* = (\theta_{11}^{*\prime}, \theta_{12}')'$, where $\theta_{11}^* = (\alpha_0, \alpha', \gamma^{*\prime}, \sigma_e^2)'$ contains the parameters from the linear part of Equation (34) and $\theta_{12} = (\mathbf{g}', \zeta)'$ is the nonlinear parameters. From Equation (29), the log likelihood conditional on $\hat{\theta}_2$ can be written as

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

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

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

$$\tilde{\sigma}_e^2(\theta_{12}, \hat{\theta}_2) = \left[\mathbf{y} - \tilde{\mathbf{X}}\tilde{\beta}(\theta_{12}, \hat{\theta}_2)\right]' \mathbf{W}(\mathbf{x}; \theta_{12})^{-1}\left[\mathbf{y} - \tilde{\mathbf{X}}\tilde{\beta}(\theta_{12}, \theta_2)\right]/T. \tag{38}$$

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

$$\begin{aligned} L(\theta_{12}; \mathbf{y}, \mathbf{X}) &= \sum_{t=1}^{T} \ln f(y_t \mid \mathbf{X}_t, \mathbf{Y}_{t-1}; \theta_{11}^*(\theta_{12}), \theta_{12}, \hat{\theta}_2) \\ &= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\tilde{\sigma}_e^2) - \frac{1}{2}\ln|\mathbf{W}(\mathbf{x}; \theta_{12})| - (T/2). \end{aligned} \tag{39}$$

Now, the numerically maximizing Equation (39) conditional on $\hat{\theta}_2$ gives the MLE $\hat{\theta}_{12}$, which from Equations (37) and (38), gives $\hat{\theta}_{11}^*$. Although the two-step procedure provides a consistent estimation of the model, the covariance matrix of $\hat{\beta}$, which is obtained by inverting the negative of the Hessian matrix, is biased because of the generated regressors $\mathbf{x}_t - \hat{\delta}'\mathbf{z}_t$ that replace $\mathbf{x}_t - \delta'\mathbf{z}_t$ in the second-step regression of Equation (18).

V. Concluding remarks

A linear regression is not a good choice in analyzing certain cases, such as the relationship between oil prices and business cycle (Hamilton 2003; Kim 2012). In such cases, scholars tend to consider nonlinear specifications. The crucial issue, however, is selecting a proper specification among all the possible nonlinear relationships. Hamilton (2001) proposes a flexible nonlinear inference wherein the philosophy of his methodology is nonparametric, but the estimation is parametric. Hamilton (2003) and Kim (2012) demonstrate that this methodology is expedient in addressing the nonlinear relationship 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 reveal that an appropriate transformation of the model allows us to employ Hamilton's (2001) approach directly. In this paper, 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 method.

Our new methodology is useful in macroeconometric models wherein endogenous explanatory variables exist, and a true relationship between a 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 this new methodology for future research.

(Received 15 April; Revised 27 April 2015; Accepted 27 April 2015)

References

Altonji, Joseph G., and Matzkin, Rosa L. "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors." *Econometrica* 73 (No. 4 2005): 1053-102.

Florens, Jean-Pierre., Heckman, James J., Meghir, Costas., and Vytlacil, Edward J. "Identification of Treatment Effects Using Vontrol Functions in Models with Continuous, Endogenous Treatment and Heterogeneous Effects." *Econometrica* 76 (No. 5 2008): 1191-206.

Hamilton, James D. "A Parametric Approach to Flexible Nonlinear Inference." *Econometrica* 69 (No. 3 2001): 537-73.

Hamilton, James D. "What Is an Oil Shock?" *Journal of Econometrics* 113 (No. 2 2003): 363-98.

Heckman, James J., and Vytlacil, Edward J. "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 (No. 4 1998): 974-87.

Heckman, James J., and Navarro, Salvador. "Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models." *Review of Economics and Statistics* 86 (No. 1 2004): 30-57.

Kim, Chang Jin. "Markov-Switching Models with Endogenous Explanatory Variables." *Journal of Econometrics* 122 (No. 1 2004): 127-36.

Kim, Chang Jin. "Markov-Switching Models with Endogenous Explanatory

Variables II: A Two-Step MLE Procedure." *Journal of Econometrics* 148 (No. 1 2009): 46-55.

Kim, Chang Jin, and Nelson, Charles R. "Estimation of a Forward-Looking Monetary Policy Rule: A Time-Varying Parameter Model Using ex Post Data." *Journal of Monetary Economics* 53 (No. 8 2006): 1949-66.

Kim, Dong Heon. "What Is an Oil Shock? Panel Data Evidence." *Empirical Economics* 43 (No. 1 2012): 121-43.

Kim, Dong Heon, Osborn, Denise R., and Sensier, Marianne. "Nonlinearity in the FED's Monetary Policy Rule." *Journal of Applied Econometrics* 20 (No. 5 2005): 621-39.

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

Wolman, Alexander L. "The Zero Bound and the Term Structure in a Nonlinear Macro Model." *Seoul Journal of Economics* 19 (No. 1 2006): 147-70.