EFFICIENT SEMIPARAMETRIC ESTIMATION OF DYNAMIC NONLINEAR SYSTEMS UNDER ELLIPTICAL SYMMETRY

Bryan W. Brown and Douglas J. Hodgson*
Rice University and University of Rochester

January 14, 2000

Abstract

We analyze semiparametric efficient estimation of nonlinear simultaneous equations models in a time series context and derive a feasible estimator that is efficient when the errors are iid from an unknown elliptically symmetric density. We also consider efficient estimation in the case of errors that are elliptically symmetric conditional on predetermined variables. Our procedures avert the curse of dimensionality problem that would arise in the absence of the elliptical symmetry restriction.

1 Introduction

Nonlinear simultaneous equations models, both static and dynamic, have enjoyed a prominent role in econometric theory and applications. Since our economic theories seldom suggest a complete specification of the distribution of these models it is of obvious interest to take a semiparametric approach which allows this distribution to be nonparametric. Newey (1989) has established semiparametric efficiency bounds for the static case for the special cases where the errors are either independent of, or symmetric conditional on, the exogenous variables. Although his approach develops efficient scores which might form the basis for estimators that attain the bounds, implementation involves the nonparametric estimation of a multivariate density, so the curse of dimensionality becomes a problem. In order to alleviate this problem,

*We acknowledge financial assistance from the NSF under CAREER grant #SBR-9701959, and comments from participants in workshops at Cornell, Penn, Guelph, Toronto, Queen's, Rice, Rochester, the 1999 Summer Meetings of the European and Latin American Econometric Societies, and the 1999 Cowles Foundation Econometrics Conference.

Newey (1989) proposes feasible estimators with the unknown density replaced by a known parametric density. He argues that these estimators are locally efficient.

Treating the disturbances of a linear simultaneous equation model nonparametrically introduces essentially the same problem and techniques used successfully there hold promise for the nonlinear model. In the context of a linear seemingly unrelated regression time series model, Hodgson, Linton, Vorkink, and Choo (1999) introduce a dimensionality-reducing approach by assuming that the disturbances are elliptically symmetric independent of the predetermined variables. Under this assumption they show that the model can be adaptively estimated using a feasible estimator that involves the nonparametric estimation of a univariate density. It is relatively straightforward to show that this approach applies directly to the simultaneous equation time-series context with elliptically symmetric disturbances.

In this paper, we propose to combine and extend these two strands of research to develop feasible semiparametrically efficient estimators for nonlinear dynamic systems. Specifically, we examine nonlinear dynamic systems with disturbances that may be autocorrelated with white noise residuals that are elliptically symmetric either independent of or conditional on the predetermined variables. In the next section, we will present the model and develop the semiparametric efficiency bounds and corresponding efficient scores for these two cases. In the third section, following Hodgson, Linton, Vorkink, and Choo (HLVC), we develop feasible estimators for the two cases that attain the semiparametric efficiency bound in some sense for both cases. In the fourth section, we discuss potential applications of the techniques.

Among the major contributions of this paper are the following. Using the approach of Ibragimov and Khas'minskii (1991), we develop semiparametric efficiency bounds for nonlinear dynamic systems under both independent and conditional elliptical symmetry. This work shows how to extend Newey's (1989) findings for the static nonlinear systems to the dynamic case. We develop feasible estimators for both the independent and conditional elliptical symmetry cases that are immune to the curse of dimensionality, thereby extending the previous work by HCL on linear dynamic models to nonlinear dynamic systems. The estimators are shown to attain the semiparametric efficiency bound globally for the independent elliptical symmetry case and locally, as defined by Newey (1989), in the case of conditional elliptical symmetry. The estimator for the conditional elliptical symmetry case is shown to generalize the multivariate GARCH error process with constant correlation to the semiparametric context.

2 The Model and Efficiency Bound

Suppose that we observe the stationary and ergodic $(s+k)$-vector $z_t = \left(y_t^T, x_t^T\right)^T$ for $t = 1, ..., n$, where y_t is an s-vector of endogenous variables and x_t is a k-vector of explanatory variables that are exogenous in a sense to be defined below. Suppose the following nonlinear simultaneous equations time series model holds:

$$\widetilde{\rho}\left(z_t, z_{t-1}, ..., z_{t-\ell}; \beta\right) = u_t, \tag{1}$$

where $\widetilde{\rho}$ is a possibly non-linear relationship of known functional form, u_t is an s-vector of possibly autocorrelated disturbances, and β is an unknown q-dimensional vector of parameters of interest with true value β_0. We assume that we can invert the function $\widetilde{\rho}$ and solve for y_t in terms of the following reduced form model:

$$y_t = \widetilde{\pi}\left(u_t, y_{t-1}, ..., y_{t-\ell}, x_t, x_{t-1}, ..., x_{t-\ell}; \beta\right). \tag{2}$$

In some instances, this inverse function may not be available in closed form and numerical techniques will be required to find the solution.

We incorporate autocorrelation by allowing $\{u_t\}$ to follow a VAR process of known order p :

$$u_t = \sum_{j=1}^{p} A_j u_{t-j} + e_t \tag{3}$$

where $\{e_t\}$ is a serially uncorrelated s-dimensional time series and the matrix polynomials are such that $\{u_t\}$ is covariance stationary. Define the vector of VAR parameters by $\alpha = \text{vec}\left(A^T\right)$, where $A = (A_1 : ... : A_p)$, then repeated substitution of (1) into (3) yields:

$$\rho^*\left(y_t, w_t, \alpha, \beta\right) = e_t, \tag{4}$$

where $w_t = \left(y_{t-1}^T, ..., y_{t-\nu}^T, x_t^T, x_{t-1}^T, ..., x_{t-\nu}^T\right)^T$ and $\nu = \max[p, \ell]$. Our derivation of an efficient estimator of β will depend on the assumptions we choose to make about the innovation process $\{e_t\}$. We shall consider two possibilities in the remainder of this section.

2.1 Independent Elliptical Symmetry

In the first case, we assume that $\{e_t\}$ is iid, independent of w_t, and drawn from an unknown elliptically symmetric density. Formally, then, the conditional (and marginal) distribution can be written

$$f(e|w) = |\Sigma|^{-1/2}\, g\left(e^T \Sigma^{-1} e\right)$$

where Σ is the characteristic matrix and is proportional to $\text{Cov}(e)$.[1] The functional form of $g(\cdot)$ is assumed to be unknown. Now define the vector $\sigma = \text{vech}(\Sigma)$. Multiplying both sides of (4) by $\Sigma^{-1/2}$, we then have

$$\rho(y_t, w_t, \theta) = \varepsilon_t, \tag{5}$$

where $\theta = \left(\beta^T, \alpha^T, \sigma^T\right)^T \in \text{int}\,\Theta \subset R^k$ and $\varepsilon_t = \Sigma^{-1/2} e_t$. Note that $\{\varepsilon_t\}$, for this case, is iid and spherically symmetric. The existence of (2) as the inverse relationship of (1) likewise assures the existence of, say, $y_t = \pi(\varepsilon_t, w_t, \theta)$, as an inverse relationship to (5).

Assuming that we have some initial observations $\{z_0, z_{-1}, ..., z_{-\nu}\}$, and that the density of these initial conditions is asymptotically negligible in the analysis of the likelihood function, we can base our derivation of semiparametric efficient estimators on the analysis of the following likelihood for $\{z_t\}_{t=1}^n$, conditioning on the initial conditions:

$$L\left(\{z_t\}_{t=1}^n, \theta, \eta\right) = \Pi_{t=1}^n f_{y|w}\left(y_t \left| w_t\right.; \theta, \eta\right) \cdot f_{x|z^{-\nu}}\left(x_t \left| z^{t-\nu}\right.; \eta\right), \tag{6}$$

where $z^{t-\nu} = (z_{t-1}, ..., z_{t-\nu})$. Note that we are assuming that the densities $f_{y|w}$ and $f_{x|z^{-\nu}}$ are unknown to the investigator, so we have written the likelihood for a parametric submodel (cf. Newey (1989, 1990)) in which η represents some parameterization that contains the true densities. Rewriting the likelihood in terms of the density of ε, using the change of variables formula, and noting that ε_t is assumed to be independent of w_t, we obtain:

$$L\left(\{z_t\}_{t=1}^n, \theta, \eta\right) = \Pi_{t=1}^n \widetilde{J}(y_t, w_t, \theta) f_\varepsilon(\varepsilon_t; \eta) \cdot f_{x|z^{-\nu}}\left(x_t \left| z^{t-\nu}\right.; \eta\right),$$

where $\widetilde{J}(y, w, \theta) = |\det(\partial\rho(y, w, \theta)/\partial y)|$. Using our spherical symmetry assumption on ε, we can write $f_\varepsilon(\varepsilon; \eta) = g\left(\varepsilon^T\varepsilon, \eta\right)$.

We can then write the score of the loglikelihood for a single observation (y, w) with respect to η as

$$s_\eta(y, w) = \frac{\partial \ln g\left(\varepsilon^T\varepsilon, \eta\right)}{\partial\eta} + \frac{\partial \ln f_{x|z^{-\nu}}\left(x \left| z^{-\nu}\right.; \eta\right)}{\partial\eta},$$

and that with respect to θ as

$$s_\theta(y, w, \theta_0) = J_\theta(y, w, \theta_0) + \rho_\theta^T(y, w, \theta_0)\,\varepsilon \cdot d\left(\varepsilon^T\varepsilon\right),$$

where $J(y, w, \theta) = \ln|\det(\partial\rho(y, w, \theta)/\partial y)|$, $d\left(\varepsilon^T\varepsilon\right) = 2g'\left(\varepsilon^T\varepsilon\right)/g\left(\varepsilon^T\varepsilon\right)$, and the θ subscripts denote partial derivatives. Note that the two terms in $s_\eta(y, w)$ are unrestricted (except for the zero mean property of scores) functions of their arguments.

[1] Note that Σ is identified only up to a scalar transformation. Its parameters can thus be estimated only up to such a transformation. In our computation of the estimator, we use the normalization that $\det(\Sigma) = 1$.

The efficient scores for estimation of θ are given by the part of $s_\theta(y, w, \theta_0)$ that is orthogonal to the nuisance scores $s_\eta(y, w)$. Since the nuisance scores could come from any parametric submodel that includes the truth, this requires orthogonalization with respect to the space spanned by suitable linear transformations of all such nuisance parameter scores. This space is known as the tangent set and, for the present problem, is the linear Hilbert space given by

$$\mathcal{T} = \{t_1(\varepsilon^T \varepsilon) + t_2(x, z^{-\nu}) : E[t_1] = E[t_2] = 0\}.$$

The efficient score is given by the residual of $s_\theta(y, w)$ less its projection on this space, which, for an arbitrary function $R(y, w)$, is given by

$$\text{Proj}(R(y, w)|\mathcal{T}) = E[R(y, w)|\varepsilon^T \varepsilon] + E[R(y, w)|x] - 2E[R(y, w)].$$

Due to ancillarity of $(x, z^{-\nu})$ and the mean zero property of scores we can show that $\text{Proj}(s_\theta(y, w)|\mathcal{T}) = E[s_\theta(y, w)|\varepsilon^T \varepsilon]$, whereupon we have

$$\begin{aligned} S(y, w) &= J_\theta\left(y, w, \theta_0\right) - E[J_\theta\left(y, w, \theta_0\right)|\varepsilon^T \varepsilon] \\ &\quad + \{\rho_\theta^T\left(y, w, \theta_0\right)\varepsilon - E[\rho_\theta^T\left(y, w, \theta_0\right)\varepsilon|\varepsilon^T \varepsilon]\} d\left(\varepsilon^T \varepsilon\right) \end{aligned} \tag{7}$$

as the efficient score for the independent elliptically symmetric case.

The efficiency (variance) bound for semiparametric estimation of θ for ε distributed elliptically symmetric independent of x is then

$$V_\theta^* = (E[S(y, w)S(y, w)^T])^{-1}. \tag{8}$$

This bound can be written more explicitly in terms of expectations involving the components of $S(y, w)$ given in (7), but little insight is gained through such an exercise. We should note, however, that

$$E[S(y, w)S(y, w)^T] = E[s_\theta(y, w)s_\theta(y, w)^T] - E\{E[s_\theta(y, w)|\varepsilon^T \varepsilon] \cdot E[s_\theta(y, w)|\varepsilon^T \varepsilon]^T\}$$

so V_θ^* exceeds the parametric efficiency bound $(E[s_\theta(y, w)s_\theta(y, w)^T])^{-1}$ by a positive definite matrix unless

$$E[s_\theta(y, w)|\varepsilon^T \varepsilon] = E[\{J_\theta\left(y, w, \theta_0\right) + \rho_\theta^T\left(y, w, \theta_0\right)\varepsilon \cdot d\left(\varepsilon^T \varepsilon\right)\}|\varepsilon^T \varepsilon] = 0.$$

Thus adaptive estimation is not possible except for special cases, such as the linear simultaneous model, that meet this condition.

We now state the following theorem, proved in the Appendix, which formalizes the result obtained above by couching our model within the general time series framework considered by Ibragimov and Khas'minskii (1991):

THEOREM 1: *Assume that $J\left(y, w, \theta\right)$ and $\rho\left(y, w, \theta\right)$ are twice differentiable with respect to θ, and that $g\left(\cdot\right)$ is twice differentiable. Then the model (5) under independent elliptical symmetry falls into the locally asymptotically normal (LAN) family and the semiparametric efficiency bound is given by V_θ^* as defined in equations (7) and (8).*

2.2 Conditional Elliptical Symmetry

The second case we consider involves the assumption that e is elliptically symmetrically distributed, conditional on w, with the functional form of the density again unknown. We continue to assume that the characteristic matrix is constant, so the conditional density can now be written

$$f\left(e_t|w_t\right) = |\Sigma|^{-1/2}\, g_c\left(e_t^T\Sigma^{-1}e_t, w_t\right),$$

where $g_c(\cdot)$ is of unknown form.[2] Analogous to the independence case, we can then define the spherically symmetric sequence $\varepsilon_t = \Sigma^{-1/2}e_t$, the model $\rho\left(y_t, w_t, \theta\right) = \varepsilon_t$, and its inverse $y_t = \pi(\varepsilon_t, w_t, \theta)$, where, as before, $\theta = \left(\beta^T, \alpha^T, \sigma^T\right)^T$.

This model allows the height of the elliptical level sets of the distribution to not only be non-Gaussian, as in the independence case, but also dependent on the conditioning variables. This specification implies $\mathrm{Cov}(e_t) = h(w_t)\Sigma$, where $h(w_t)$ is a scalar nonparametric function consistent with $f(e_t|w_t)$, so the untransformed disturbances have constant conditional correlations but a common scalar conditional heteroskedasticity factor of unknown form. Thus the disturbance specification presented here generalizes the multivariate GARCH approaches where constant conditional correlation has been found convenient in reducing the parameter dimensionality problem to manageable proportions (see Bollerslev (1987) for an application and Jeantheau (1998) for some relevant theory).

The conditional elliptical symmetry model has additional advantages. First, under our constant conditional correlation specification, the unconditional distribution of the disturbances is elliptically symmetric if the conditional distribution is elliptical symmetric. This enables a simplified estimator for the conditional elliptical symmetric case in the next section based on an estimator of the density for the unconditional distribution, which turns out to be the same estimator as used for estimating the elliptically symmetric density under independence. Second, given the dependence of the innovations on the predetermined variables, it is possible to avoid the VAR specification of the disturbance process, which introduces a large number of parameters. Thus the conditional elliptically symmetric case will be much more parsimonious with respect to number of parameters, when we use the simplified estimator just mentioned. Of course, there may be efficiency gains from approximating the dependence with a VAR and use the independence estimator.

Following our development for the previous case, we can write the likelihood function for a parametric submodel as in (6). However, the fact that ε need not be independent of w leads us to rewrite the likelihood as follows:

$$L\left(\{z_t\}_{t=1}^n, \theta, \eta\right) = \Pi_{t=1}^n J\left(y_t, w_t, \theta\right) f_{\varepsilon|w}\left(\varepsilon_t\left|w_t\right.; \eta\right)\cdot f_{x|z^{-\nu}}\left(x_t\left|z^{t-\nu}\right.; \eta\right).$$

[2]As in the independence model, Σ is identified only up to a scalar transformation, so we use the normalization $\det(\Sigma) = 1$ in our computations.

Now, using our conditional spherical symmetry assumption on ε, we have

$$f_{\varepsilon|w}\left(\varepsilon\,|w\,;\eta\right)=g_c\left(\varepsilon^T\varepsilon,w,\eta\right).$$

Accordingly, the score with respect to η is

$$s_\eta(y,w)=\frac{\partial\ln g_c\left(\varepsilon^T\varepsilon,w,\eta\right)}{\partial\eta}+\frac{\partial\ln f_{x|z^{-\nu}}\left(x_t\,|z^{t-\nu}\,;\eta\right)}{\partial\eta},$$

and that with respect to θ is

$$s_\theta\left(y,w\right)=J_\theta\left(y,w,\theta_0\right)+\rho_\theta^T\left(y,w,\theta_0\right)\varepsilon\cdot d_c\left(\varepsilon^T\varepsilon,w\right),$$

where $d_c\left(\varepsilon^T\varepsilon,w\right)=2g_c^{'}\left(\varepsilon^T\varepsilon,w\right)/g_c\left(\varepsilon^T\varepsilon,w\right)$ and $g_c^{'}\left(\varepsilon^T\varepsilon,w\right)$ denotes the partial derivative with respect to the first argument. As before, the two terms in $s_\eta(y,w)$ are unrestricted functions of their arguments.

Accordingly, the tangent set for the conditional elliptical symmetry case is given by the linear Hilbert space

$$\mathcal{T}=\{t_1(\epsilon^T\epsilon,w)+t_2(x,z^{-\nu}):E[t_1]=E[t_2]=0\}.$$

Since $t_1(\epsilon^T\epsilon,w)$ is an unrestricted function of its arguments, then any function orthogonal to the set of t_1 must also be orthogonal to the set of $t_2(x,z^{-\nu})$. Formally, then, the projection for an arbitrary function $R(y,w)$, is given by

$$\text{Proj}(R(y,w)|\mathcal{T})=E[R(y,w)|\varepsilon^T\varepsilon,w]-E[R(y,w)].$$

Due to the mean zero property of scores we have $\text{Proj}(s_\theta(y,w)|\mathcal{T})=E[s_\theta(y,w)|\varepsilon^T\varepsilon,w]$, whereupon we now have

$$\begin{aligned}S(y,w)\;=\;&J_\theta\left(y,w,\theta_0\right)-E[J_\theta\left(y,w,\theta_0\right)|\varepsilon^T\varepsilon,w]\\&+\{\rho_\theta^T\left(y,w,\theta_0\right)\varepsilon-E[\rho_\theta^T\left(y,w,\theta_0\right)\varepsilon|\varepsilon^T\varepsilon,w]\}d_c\left(\varepsilon^T\varepsilon,w\right)\end{aligned}\qquad(9)$$

as the efficient score. The semiparametric efficiency bound for the current (conditional elliptical symmetry) problem is, therefore,

$$V_\theta^*=(E[S(y,w)S(y,w)^T])^{-1},\qquad(10)$$

where the current definition of $S(y,w)$ is used. An analogue of Theorem 1 can be established for this model as well. The proof is similar to that of Theorem 1 and is not given.

Theorem 2: *Assume that $J\left(y,w,\theta\right)$ and $\rho\left(y,w,\theta\right)$ are twice differentiable with respect to θ, and that $g_c\left(\cdot\right)$ is twice differentiable with respect to its first argument. Then the model (5) under conditional elliptical symmetry falls into the locally asymptotically normal (LAN) family and the semiparametric efficiency bound is given by V_θ^* as defined in equations (9) and (10).*

3 Efficient Estimation

In this section, we develop feasible estimators that attain the lower bounds set out in the previous section. In general, an asymptotically efficient estimator is an asymptotically linear estimator whose influence function is the efficient influence function, which is given by

$$\psi_\theta^*(y,w) = V_\theta^* S(y,w).$$

An approach that generally attains the bound is an m-estimator based on the efficient score. Suppose we can write the efficient score as a closed form function of the observable (y_t, w_t) and unknown θ, then the m-estimator $\widehat{\theta}$ is given as the solution to

$$0 = n^{-1}\sum_{t=1}^{n} S(y_t, w_t, \widehat{\theta}) = \overline{S}(\widehat{\theta}).$$

Under general conditions, we find that $n^{1/2}(\widehat{\theta}-\theta_0) \xrightarrow{d} N(0, V_\theta^*)$. For purposes of inference the inverse of the covariance estimator $\overline{C}(\widehat{\theta}) = n^{-1}\sum_{t=1}^{n} S(y_t, w_t, \widehat{\theta})S(y_t, w_t, \widehat{\theta})'$ will provide a consistent estimator of V_θ^*.

In practice, we will use a discretized, preliminary $\sqrt{n}$-consistent estimator, say, $\widetilde{\theta}$ to obtain the following Newton-type approximation to the solution

$$\widehat{\theta} = \widetilde{\theta} - [\overline{H}(\widetilde{\theta})]^{-1}\overline{S}(\widetilde{\theta})$$

where $\overline{H}(\theta) = n^{-1}\sum_{t=1}^{n} \partial S(y_t, w_t, \theta)/\partial\theta^T$. If the score is correctly specified then this two-step estimator will also attain the semiparametric efficiency bound under general conditions. Moreover, we can further simplify by using the approximation $\overline{C}(\widehat{\theta}) = \overline{H}(\widehat{\theta}) + o_p(1)$. If, however, the score $S(y_t, w_t, \theta)$ is misspecified but still has expectation zero, which will be relevant for a case considered below, then the estimator will remain consistent but not attain the bound. The covariance matrix for this case will have the usual sandwich form and is consistently estimated by $[\overline{H}(\widetilde{\theta})]^{-1}\overline{C}(\widehat{\theta})[\overline{H}(\widetilde{\theta})']^{-1}$. The latter covariance estimator is, of course, also consistent under correct specification.

3.1 Independent Elliptical Symmetry

The problem, of course, is that we do not have closed form expressions for the efficient score for either of the cases considered above. In the independent elliptically symmetric case, whose efficient score is given by (6), we do not have closed form expressions for $E[J_\theta(z,\theta)\,|\varepsilon^T\varepsilon]$, $E[\rho_\theta^T(z,\theta)\,\varepsilon|\varepsilon^T\varepsilon]$, and $d\left(\varepsilon^T\varepsilon\right)$. The obvious solution is to substitute estimates of these components and proceed using an estimated form of the efficient score. Estimators for nonparametric functions such as $d(\varepsilon^T\varepsilon)$ are developed in Hodgson, Choo, and Linton (1998) (HCL) using kernel estimators of $g(\varepsilon^T\varepsilon)$ and $g'\left(\varepsilon^T\varepsilon\right)$ based on estimated residuals and can be applied directly to the current case using the residuals $\widetilde{\varepsilon}_t = \rho(z_t, \widetilde{\theta})$ to obtain the estimator $\widehat{d}(\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t)$ for all t.

Here we describe an algorithm for computing a consistent estimator $\widehat{d}(\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t)$ that takes advantage of our elliptical symmetry assumption to reduce the dimensionality of our density estimation problem to one. This estimator and its properties are analyzed in detail in HCL. Here, we only describe its computation. Define the random variable $\widetilde{v}_t = \widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t$ and the transformed random variable $\widetilde{\omega}_t = \tau(\widetilde{v}_t)$. The transformation $\tau(\cdot)$ belongs to the Box-Cox family $\tau(v;\zeta) = \left(v^\zeta - 1\right)/\zeta$; HCL analyze three tansformations in depth - the identity $(\tau(v) = v)$, the logarithmic $(\tau(v) = \ln(v))$, and the "$\frac{m}{2}$" transformation $\left(\tau(v) = v^{m/2}\right)$. Define the Jacobian $J_\tau(\omega) = \left|\frac{\partial\tau^{-1}(\omega)}{\partial\omega}\right|$. Define the Gaussian kernel as

$$\phi(\omega,\sigma) = \frac{1}{2\sigma\sqrt{\pi}}\exp\left\{-\frac{\omega^2}{2\sigma^2}\right\}$$

and compute the following estimate of the density of the transformed random variable ω :

$$\widehat{\gamma}_t(\omega) = (n-1)^{-1}\sum_{\substack{s=1\\ s\neq t}}^{n}\phi(\omega - \widetilde{\omega}_s, \sigma_n)$$

and of its derivative

$$\widehat{\gamma}_t'(\omega) = (n-1)^{-1}\sum_{\substack{s=1\\ s\neq t}}^{n}\phi'(\omega - \widetilde{\omega}_s, \sigma_n),$$

where $\sigma_n \to 0$ is a bandwidth sequence. We will make use of the following trimming conditions:

(i) $\widehat{\gamma}_t(\widetilde{\omega}_t) \geq d_n$;
(ii) $|\widetilde{\omega}_t| \leq e_n$;
(iii) $|\lambda(\widetilde{\omega}_t)| \leq b_n$;
(iv) $\left|\vartheta^{1/2}(\widetilde{\omega}_t)\,\widehat{\gamma}_t'(\widetilde{\omega}_t)\right| \leq c_n\widehat{\gamma}_t(\widetilde{\omega}_t)$,

where $\vartheta(\omega) = v\tau'(v)\,J_\tau^{-1}(\omega)$ [recall that $v = \tau^{-1}(z)$] and $\lambda(\omega) = (d/d\omega)^{-1}\vartheta^{1/2}(\omega)$. The constants in (i)-(iv) satisfy $c_n \to \infty$, $e_n \to \infty$, $b_n \to \infty$, $\sigma_n \to 0$, $d_n \to 0$, $\sigma_n c_n \to 0$, $e_n\sigma_n^{-3} = o(n)$, and $b_n\sigma_n^{-3} = o(n)$.

Now compute the following trimmed score estimator:

$$\widehat{d}_t(\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t) = \begin{cases} \left[\mu(\widetilde{v}_t) + \tau'(\widetilde{v}_t)\frac{\widehat{\gamma}_t'}{\widehat{\gamma}_t}(\widetilde{\omega}_t)\right] & \text{if (i)-(iv) all hold} \\ 0 & \text{otherwise,} \end{cases}$$

where $\mu(v) = (1 - m/2)v^{-1} - \frac{J_\tau'}{J_\tau}\{\tau(v)\}\,\tau'(v)$.

We now consider estimation of the unknown conditional expectation function. Let $R(y,w,\theta)$ be a general function of the data, then we seek to estimate

$$E[R(y,w,\theta)\,|\varepsilon^T\varepsilon] = \int\int R(\pi(\varepsilon,w,\theta),w,\theta)\cdot f_2(w)\cdot f(\varepsilon|\varepsilon^T\varepsilon)\cdot d\varepsilon\cdot dw.$$

Note that the level set for the conditional density $f(\varepsilon|\varepsilon^T\varepsilon)$ lie on a hypersphere with radius $(\varepsilon^T\varepsilon)^{1/2}$, so all the points on the hypersphere are equiprobable. By analogy of averages to expectations, therefore, we propose the estimator

$$\widehat{E}[R(z,\widetilde{\theta})|\varepsilon^T\varepsilon = \widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t] = m^{-1}\sum_{i=1}^{m} n^{-1}\sum_{\tau=1}^{n} R(\pi(\varepsilon_i^*, w_\tau, \widetilde{\theta}), w_\tau, \widetilde{\theta}),$$

where ε_s^* are drawn uniformly from the hypersphere with radius $(\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t)^{1/2}$. Strictly speaking, as will be shown below, under fairly general conditions, we only need $m \to \infty$ as $n \to \infty$ to attain the SEB. However, by choosing m sufficiently large, we can approximate the integration with respect to ε arbitrarily closely.

Substituting these estimators into (7) to obtain estimates of the efficient score:

$$\begin{aligned}\widehat{S}(\widetilde{\theta}) \quad &= \quad J_\theta(y_t, w_t, \widetilde{\theta}) - \widehat{E}[J_\theta(y, w, \widetilde{\theta})|\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t] \\ &\quad + \{\rho_\theta^T(y_t, w_t, \widetilde{\theta}) - \widehat{E}[\rho_\theta^T(y, w, \widetilde{\theta})\widetilde{\varepsilon}_t|\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t]\}\widehat{d}_t\left(\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t\right)\end{aligned}$$

and average efficient score:

$$\begin{aligned}\overline{S}(\widetilde{\theta}) \quad &= \quad n^{-1}\sum_{t=1}^{n}\Big\{J_\theta(y_t, w_t, \widetilde{\theta}) - \widehat{E}[J_\theta(y, w, \widetilde{\theta})|\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t] \\ &\quad + \{\rho_\theta^T(y_t, w_t, \widetilde{\theta}) - \widehat{E}[\rho_\theta^T(y, w, \widetilde{\theta})\widetilde{\varepsilon}_t|\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t]\}\widehat{d}_t\left(\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t\right)\Big\}.\end{aligned} \tag{11}$$

Using the theory of V-statistics, under general conditions, we can show that

$$n^{-1}\sum_{t=1}^{n}\widehat{E}[J_\theta(y, w, \widetilde{\theta})|\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t] - E[J_\theta(y, w, \widetilde{\theta})|\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t] = O_p(n^{-1}) \tag{12}$$

and

$$n^{-1}\sum_{t=1}^{n}\{\widehat{E}[\rho_\theta^T(y, w, \widetilde{\theta})\widetilde{\varepsilon}_t|\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t] - E[\rho_\theta^T(y, w, \widetilde{\theta})\widetilde{\varepsilon}_t|\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t]\}\widehat{d}_t\left(\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t\right) = O_p(n^{-1}). \tag{13}$$

Thus using the estimated conditional expectations will not impact the asymptotic limiting distribution. HCL prove the following consistency result for the score estimator:

$$n^{-1/2}\sum_{t=1}^{n}\left\{\widehat{d}_t\left(\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t\right) - d\left(\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t\right)\right\} = o_p(1). \tag{14}$$

Equations (12), (13), and (14) together ensure that our semiparametric score estimator $\widehat{\overline{S}}(\widetilde{\theta})$ given in (11) can be used in place of the true score $\overline{S}(\widetilde{\theta})$ in our computation of the one-step iterative estimator.

3.2 Conditional Elliptical Symmetry

In the conditional elliptical symmetry case, whose efficient score is given by (9), the unknown components are $\mathrm{E}[J_{\theta}\left(y,w,\theta_0\right)|\varepsilon^T\varepsilon,w]$, $\mathrm{E}[\rho_{\theta}^T\left(y,w,\theta_0\right)\varepsilon|\varepsilon^T\varepsilon,w]$, and $d\left(\varepsilon^T\varepsilon,w\right)$. The estimation of the conditional expectations is a special case of the conditional expectation estimator introduced for the independence case. For $R(y,w,\theta)$ a general function of the data, we now seek to estimate

$$E[R\left(y,w,\theta\right)|\varepsilon^T\varepsilon,w]=\int R(\pi(\varepsilon,w,\theta),w,\theta)\cdot f(\varepsilon|\varepsilon^T\varepsilon,w)\cdot d\varepsilon.$$

Analogous to above, the level set for the conditional density $f(\varepsilon|\varepsilon^T\varepsilon,w)$ lie on a hypersphere with radius $(\varepsilon^T\varepsilon)^{1/2}$, so all the points on the hypersphere are equiprobable. Accordingly, we propose the simplified estimator

$$\widehat{E}[R(z,\widetilde{\theta})|\varepsilon^T\varepsilon=\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t,w_t]=m^{-1}\sum_{i=1}^{m}R(\pi(\varepsilon_i^*,w_t,\widetilde{\theta}),w_t,\widetilde{\theta}),$$

where, as before, ε_s^* are drawn uniformly from the hypersphere with radius $(\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t)^{1/2}$. Under general conditions, we can show that

$$n^{-1}\sum_{t=1}^{n}\widehat{E}[J_\theta(y,w,\widetilde{\theta})|\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t,w_t]-E[J_\theta(y,w,\widetilde{\theta})|\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t,w_t]=O_p(n^{-1/2}m^{-1/2})$$

and

$$n^{-1}\sum_{t=1}^{n}\{\widehat{E}[\rho_\theta^T(y,w,\widetilde{\theta})\widetilde{\varepsilon}_t|\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t,w_t]-E[\rho_\theta^T(y,w,\widetilde{\theta})\widetilde{\varepsilon}_t|\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t,w_t]\}\widehat{d}_t\left(\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t,w_t\right)=O_p(n^{-1/2}m^{-1/2}).$$

Thus using the estimated conditional expectations will not impact the asymptotic limiting distribution.

The estimation of $d\left(\varepsilon^T\varepsilon,w\right)$ is not so easy. A kernel estimation approach to estimation of $g_c\left(\varepsilon^T\varepsilon,w\right)$ and its first derivative would involve a very high dimensional problem since w is likely a long vector. Thus the objective of reducing the dimensionality of the nonparametric component to manageable proportions is not directly attainable. Fortunately, an alternative feasible estimation approach is available for the conditional elliptical symmetry case. Consider the pseudo-score

$$\begin{aligned}S^*(y,w) &= J_\theta\left(y,w,\theta_0\right)-E[J_\theta\left(y,w,\theta_0\right)|\varepsilon^T\varepsilon,w] \\ &\quad+\{\rho_\theta^T\left(y,w,\theta_0\right)\varepsilon-E[\rho_\theta^T\left(y,w,\theta_0\right)\varepsilon|\varepsilon^T\varepsilon,w]\}d\left(\varepsilon^T\varepsilon\right)\end{aligned} \tag{15}$$

which is just the score for the conditional elliptical symmetry case with $d\left(\varepsilon^T\varepsilon\right)$ based on the unconditional distribution substituted for $d_c\left(\varepsilon^T\varepsilon,w\right)$. Since the component in braces is uncorrelated with any function of $\varepsilon^T\varepsilon$, by construction, then this pseudo-score will have expectation zero.

Accordingly, we propose an estimator based on the average pseudo-score

$$\begin{aligned}\overline{S^*}(\widetilde{\theta}) &= n^{-1}\sum_{t=1}^{n}\Big\{J_\theta(y_t,w_t,\widetilde{\theta}) - \widehat{E}[J_\theta(y,w,\widetilde{\theta})|\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t,w_t] \\ &\quad +\{\rho_\theta^T(y_t,w_t,\widetilde{\theta}) - \widehat{E}[\rho_\theta^T(y,w,\widetilde{\theta})\widetilde{\varepsilon}_t|\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t,w_t]\}\widehat{d}_t\left(\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t\right)\Big\}. \end{aligned} \tag{16}$$

where $\widehat{d}_t\left(\widetilde{\varepsilon}_t{}^T\widetilde{\varepsilon}_t\right)$ is the estimator introduced for the independent case above. Setting this average to zero and finding the solution for $\widetilde{\theta}$ yields what we call the "hybrid" estimator. An asymptotically equivalent estimator can be obtained using the two-step approach. A consistent covariance matrix for this estimator would be provided by the sandwich form, in which we are required to compute consistent semiparametric estimates of both the expected outer product of the efficient score and the Hessian. The former computation is straightforward using methods outlined above. The Hessian estimation, however, seems to be a more difficult problem that has not yet been addressed in the literature (to our knowledge).

The hybrid estimator does not generally attain the semiparametric efficiency bound for the conditional elliptical symmetry model. It does however, have a certain mini-max property which Newey (1989) has termed local semiparametric efficiency. Specifically, within the class of estimators that are guaranteed to be consistent against conditional dependence given elliptical symmetry, it will be the most efficient under independence. That is, it will attain the SEB for the conditional elliptical symmetry case if the disturbances turn out to be independent. Of course, the independence estimator will generally perform better under independence but will be inconsistent against conditional dependence. Beyond these asymptotic optimality properties, the estimator is relatively simple and generalizes the popular GARCH approaches in a nonparametric fashion.

4 Application

In this section we briefly discuss a planned application of our estimator that, for reasons discussed below, is beyond the scope of the present paper. The theory of exchange rate behavior developed by Krugman (1991) for currencies that are permitted to fluctuate within a fixed target zone is based on the premise that monetary authorities in the countries concerned only intervene in the exchange market when an exchange rate hits the boundary of the target zone. In practice, it is reasonable to expect that intervention will occur before the exchange rate hits the boundary, with the strength of the reaction being stronger the closer the rate is to the boundary. The reaction function should therefore have a nonlinear form. If we treat the interest rate differential for the two countries concerned as the principal policy instrument of the respective monetary authorities, then one could posit a policy equation in which the interest rate differential is a nonlinear function of the position of the exchange rate within the target zone. Estimating such a model would be complicated by the fact that it would not be reasonable to treat the exchange rate as exogenous; we would require a second equation modeling the reaction of the exchange rate to interest rate movements in order to complete a two-equation simultaneous equations model.

Consider a European country whose currency's exchange rate with regard to the German mark belonged to a target zone for the sample period. Suppose that we have a sample of exchange rate and interest rate data for the two countries. Denote by c_t the price of a unit of the country's currency (in terms of German marks) in period t. Let c_t^U and c_t^L denote the upper and lower boundaries, respectively, of the target zone for this exchange rate in period t, and let $\overline{c}_t$ denote the target rate. Let $d_t^U = c_t^U - \overline{c}_t$ and $d_t^L = \overline{c}_t - c_t^L$, i.e. the distances in period t between the target and the upper and lower boundaries, respectively, of the target zone, and define $\widetilde{c}_t = c_t - \overline{c}_t$. Krugman's (1991) model specifies no policy intervention if $c_t^L < c_t < c_t^U$, whereas we suggest estimating a nonlinear model in which monetary policy reacts to the exchange rate whenever $\widetilde{c}_t \neq 0$, but with a strength of reaction that is a nonlinear increasing function of $|\widetilde{c}_t|$. Let r_t denote the difference in period t between the German interest rate and the comparable interest rate of the country under consideration.

We could specify a monetary policy reaction function as follows:

$$r_t = \alpha_1 + \delta\Gamma\left(\widetilde{c}_t\right) + \sum_{j=1}^{q} \phi_j r_{t-j} + e_{1t}, \tag{17}$$

for some specified parametric nonlinear reaction function $\Gamma\left(\widetilde{c}_t\right)$, while assuming the following linear equation characterizing the reaction of the exchange rate to present and lagged interest rates:

$$\widetilde{c}_t = \alpha_2 + \sum_{j=1}^{\ell} \psi_j \widetilde{c}_{t-j} + \sum_{j=0}^{p} \zeta_j r_{t-j} + e_{2t}. \tag{18}$$

Note that Krugman's (1991) hypothesis of no policy intervention taking place would correspond to $\delta = 0$ in equation (17). We have estimated this model for Italy, Denmark, and Ireland for a daily data sample from the 1989-1990 period, using the following specification of the reaction function:

$$\Gamma\left(\widetilde{c}_t\right) = sgn\left(\widetilde{c}_t\right)\left\{I\left(\widetilde{c}_t \geq 0\right)\left|\frac{\widetilde{c}_t}{d_t^U}\right|^{\gamma} + I\left(\widetilde{c}_t < 0\right)\left|\frac{\widetilde{c}_t}{d_t^L}\right|^{\gamma}\right\},$$

where $I(\cdot)$ denotes the indicator function, and the parameter γ characterizes the degree of nonlinearity of the reaction function. Estimating the model by nonlinear instrumental variables and using our efficient estimator, our basic finding is that δ is very close to zero and that the γ estimates are highly unstable depending on the starting value used in the nonlinear iteration. These results strongly suggest that $\delta = 0$, in which case γ is unidentified, hence the instability in the estimates of this parameter. These results do not constitute a rigorous test of the non-intervention hypothesis, however, since the fact that γ is unidentified under the null implies that the usual standard errors are meaningless (hence we do not report them here) and standard t or Wald tests cannot be used to test this null (see Andrews and Ploberger (1994) and Hansen (1996)).

A more rigorous analysis of the no-intervention hypothesis entails an extension of Andrews-Ploberger (1994) and Hansen (1996) to allow for simultaneous equations and will also consider various possible specifications of the policy reaction function (such as a threshold model, for example). Such a study is a topic for further investigation.

5 Appendix

PROOF OF THEOREM 1: Our proof of local asymptotic normality and our derivation of the bound will follow the reasoning of Ibragimov and Khas'minskii (1991). We begin by defining the vector of unknown parameters

$$\vartheta = \left(\theta, g\left(\varepsilon^T \varepsilon\right), f_{x|z^{-\nu}}\left(x\left|z^{-\nu}\right.\right)\right) \in \Lambda = \Theta \times \Xi_1 \times \Xi_2,$$

where Ξ_1 is the space of density functions $g\left(\varepsilon^T \varepsilon\right)$ and Ξ_2 is the space of conditional densities $f_{x|z^{-\nu}}\left(x\left|z^{-\nu}\right.\right)$. We define the sequence of probability measures $\{P_{\vartheta,n}\}$, which represent the distribution of the sample of size n when ϑ is the parameter vector.

We first show that the family of measures $\{P_{\vartheta,n}; \vartheta \in \Lambda\}$ is locally asymptotically normal (LAN) according to the definition of Ibragimov and Khas'minski (1991, p. 1682). We define the Hilbert space $\mathbf{H} = \overline{\mathbf{H}}_1 + \overline{\mathbf{H}}_2 + \overline{\mathbf{H}}_3$, where $\overline{\mathbf{H}}_1$ is the Hilbert space containing functions of the form

$$h_1(y,w,\theta) = \kappa^T s_\theta(y,w,\theta)\sqrt{g\left(\varepsilon^T \varepsilon\right) \cdot f_{x|z^{-\nu}}\left(x\left|z^{-\nu}\right.\right) \cdot f_{-\nu}\left(z^{-\nu}\right)},$$

where κ is a vector of constants with dimensionality equal to that of θ, and $f_{-\nu}\left(z^{-\nu}\right)$ denotes the marginal density of $z^{-\nu}$. We further define $\mathbf{H}_2$ as the set of all bounded, integrable functions $h_2(y,w,\theta)$ having the form

$$h_2(y,w,\theta) = t_1\left(\varepsilon(\theta)^T \varepsilon(\theta)\right),$$

such that $\int t_1\left(\varepsilon^T \varepsilon\right)\sqrt{g\left(\varepsilon^T \varepsilon\right)}d\varepsilon = 0$, and $\mathbf{H}_3$ as the set of all bounded, integrable functions $h_3(y,w)$ having the form

$$h_3(y,w) = t_2\left(x, z^{-\nu}\right),$$

such that $\int t_2\left(x, z^{-\nu}\right)\sqrt{f_{x|z^{-\nu}}\left(x\left|z^{-\nu}\right.\right)}dx = 0$. We define the norm of an element $h \in \mathbf{H}$ by

$$\|h\|_{\mathbf{H}} = \left\{\int \left(h_1(y,w,\theta) + h_2(y,w,\theta) + h_3(y,w)\right)^2 dydw\right\}^{1/2}.$$

The following sequence of linear operators $\{A_n\}$ maps $\mathbf{H}$ into $R^k \times L_2 \times L_2$:

$$A_n(h) = n^{-1/2}\left[\begin{array}{c} V_\theta^* \int S(y,w)h_1^T(y,w,\theta)\sqrt{g\left(\varepsilon^T \varepsilon\right) \cdot f_{x|z^{-\nu}}\left(x\left|z^{-\nu}\right.\right) \cdot f_{-\nu}\left(z^{-\nu}\right)}d\varepsilon dw \\ h_2(y,w,\theta)\sqrt{g\left(\varepsilon^T \varepsilon\right)} \\ h_3(y,w)\sqrt{f_{x|z^{-\nu}}\left(x\left|z^{-\nu}\right.\right) \cdot f_{-\nu}\left(z^{-\nu}\right)} \end{array}\right].$$

For every $h \in \mathbf{H}$, we have

$$\vartheta + A_n(h) = \vartheta + n^{-1/2}\left[\begin{array}{c} \kappa \\ h_2(y,w,\theta)\sqrt{g\left(\varepsilon^T \varepsilon\right)} \\ h_3(y,w)\sqrt{f_{x|z^{-\nu}}\left(x\left|z^{-\nu}\right.\right) \cdot f_{-\nu}\left(z^{-\nu}\right)} \end{array}\right].$$

To obtain our LAN theory, we must now verify that Conditions 1-3 of Ibragimov and Khas'minskii (1991, p. 1682) hold for our model. Condition 1 states that $\lim_{n\to\infty} \|A_n(h)\| = 0 \; \forall h \in \mathbf{H}$, which holds since $h_2(y,w,\theta)$ and $h_3(y,w)$ are bounded and integrable. Condition 2 will follow if we can show that, for every $h \in \mathbf{H}$, there exists n sufficiently large that $\vartheta + A_n(h) \in \Lambda$. We first note that, for n sufficiently large, we have $\theta + n^{-1/2}\kappa \in \Theta$, since $\theta \in \operatorname{int}\Theta$. It is not hard to see that for n sufficiently large, $g\left(\varepsilon^T\varepsilon\right) + n^{-1/2}h_2(y,w,\theta)\sqrt{g(\varepsilon^T\varepsilon)}$ is a spherically symmetric density, and that, conditional on $z^{-\nu}$, $f_{x|z^{-\nu}}(x\,|z^{-\nu}) + n^{-1/2}h_3(y,w)\sqrt{f_{x|z^{-\nu}}(x\,|z^{-\nu}) \cdot f_{-\nu}(z^{-\nu})}$ is a density function.

To check Condition 3, we must analyze the asymptotic behavior of the likelihood ratio $\Lambda_n(\vartheta + A_n(h), \vartheta) = \frac{dP_{\vartheta + A_n(h)}}{dP_\vartheta}$. With appropriate assumptions on the distribution of the initial conditions, and defining $\theta_n = \theta + n^{-1/2}\kappa$, we can approximate this likelihood ratio by

$$\begin{aligned}\Lambda_n(\vartheta + A_n(h), \vartheta) \cong \prod_{t=1}^{n} \Bigg\{ & \frac{\widetilde{J}(y_t, w_t, \theta_n)\, g\left(\varepsilon_t(\theta_n)^T \varepsilon_t(\theta_n)\right) \cdot f_{x|z^{-\nu}}(x_t\,|z^{t-\nu})}{\widetilde{J}(y_t, w_t, \theta)\, g\left(\varepsilon_t(\theta)^T \varepsilon_t(\theta)\right) \cdot f_{x|z^{-\nu}}(x_t\,|z^{t-\nu})} \\ & + n^{-1/2} \frac{\widetilde{J}(y_t, w_t, \theta_n)\, h_2(y_t, w_t, \theta_n) \sqrt{g\left(\varepsilon_t(\theta_n)^T \varepsilon_t(\theta_n)\right)} f_{x|z^{-\nu}}(x_t\,|z^{t-\nu})}{\widetilde{J}(y_t, w_t, \theta)\, g\left(\varepsilon_t(\theta)^T \varepsilon_t(\theta)\right) \cdot f_{x|z^{-\nu}}(x_t\,|z^{t-\nu})} \\ & + n^{-1/2} \frac{\widetilde{J}(y_t, w_t, \theta_n)\, h_3(y_t, w_t)\, g\left(\varepsilon_t(\theta_n)^T \varepsilon_t(\theta_n)\right) \sqrt{f_{x|z^{-\nu}}(x_t\,|z^{t-\nu}) \cdot f_{-\nu}(z^{t-\nu})}}{\widetilde{J}(y_t, w_t, \theta)\, g\left(\varepsilon_t(\theta)^T \varepsilon_t(\theta)\right) \cdot f_{x|z^{-\nu}}(x_t\,|z^{t-\nu})} \\ & + n^{-1} \frac{\widetilde{J}(y_t, w_t, \theta_n)\, h_2(y_t, w_t, \theta_n)\, h_3(y_t, w_t) \sqrt{g\left(\varepsilon_t(\theta_n)^T \varepsilon_t(\theta_n)\right) f_{x|z^{-\nu}}(x_t\,|z^{t-\nu}) \cdot f_{-\nu}(z^{t-\nu})}}{\widetilde{J}(y_t, w_t, \theta)\, g\left(\varepsilon_t(\theta)^T \varepsilon_t(\theta)\right) \cdot f_{x|z^{-\nu}}(x_t\,|z^{t-\nu})} \Bigg\}. \end{aligned} \tag{A.1}$$

Recall that we can write

$$\widetilde{J}(y_t, w_t, \theta)\, g\left(\varepsilon_t(\theta)^T \varepsilon_t(\theta)\right) \cdot f_{x|z^{-\nu}}\left(x_t \,\middle|\, z^{t-\nu}\right) = f_{z|z^{-\nu}}\left(z_t \,\middle|\, z^{t-\nu}; \theta\right).$$

We have the following Taylor expansion:

$$\begin{aligned} f_{z|z^{-\nu}}\left(z_t \,\middle|\, z^{t-\nu}; \theta_n\right) = {} & f_{z|z^{-\nu}}\left(z_t \,\middle|\, z^{t-\nu}; \theta\right) - n^{-1/2}\kappa^T \frac{\partial f_{z|z^{-\nu}}(z_t\,|z^{t-\nu}; \theta)}{\partial\theta} \\ & + 2n^{-1}\kappa^T \frac{\partial^2 f_{z|z^{-\nu}}(z_t\,|z^{t-\nu}; \theta)}{\partial\theta\partial\theta^T}\kappa + O_p\left(n^{-3/2}\right). \end{aligned} \tag{A.2}$$

Now define

$$r_{1t}(\theta) = \widetilde{J}(y_t, w_t, \theta)\, h_2(y_t, w_t, \theta) \sqrt{g\left(\varepsilon_t(\theta)^T \varepsilon_t(\theta)\right)} f_{x|z^{-\nu}}\left(x_t \,\middle|\, z^{t-\nu}\right),$$

$$r_{2t}(\theta) = \widetilde{J}(y_t, w_t, \theta)\, h_3(y_t, w_t)\, g\left(\varepsilon_t(\theta)^T \varepsilon_t(\theta)\right) \sqrt{f_{x|z^{-\nu}}(x_t | z^{t-\nu}) \cdot f_{-\nu}(z^{t-\nu})},$$

and

$$r_{3t}(\theta) = \widetilde{J}(y_t, w_t, \theta)\, h_2(y_t, w_t, \theta)\, h_3(y_t, w_t) \sqrt{g\left(\varepsilon_t(\theta)^T \varepsilon_t(\theta)\right) f_{x|z^{-\nu}}(x_t | z^{t-\nu}) \cdot f_{-\nu}(z^{t-\nu})}.$$

We have

$$n^{-1/2} r_{1t}(\theta_n) = n^{-1/2} r_{1t}(\theta) - n^{-1}\kappa^T r_{1\theta t}(\theta) + O_p\left(n^{-3/2}\right), \tag{A.3}$$

$$n^{-1/2} r_{2t}(\theta_n) = n^{-1/2} r_{2t}(\theta) - n^{-1}\kappa^T r_{2\theta t}(\theta) + O_p\left(n^{-3/2}\right), \tag{A.4}$$

and

$$n^{-1} r_{3t}(\theta_n) = n^{-1} r_{3t}(\theta) + O_p\left(n^{-3/2}\right). \tag{A.5}$$

Substituting (A.2)-(A.5) into (A.1), we obtain

$$\Lambda_n(\vartheta + A_n(h), \vartheta) \cong \prod_{t=1}^{n} \Big\{ 1 - n^{-1/2}\kappa^T \left[J_\theta(y_t, w_t, \theta) + \rho_\theta^T(y_t, w_t, \theta)\, \varepsilon_t(\theta)\, d\left(\varepsilon_t(\theta)^T \varepsilon_t(\theta)\right) \right]$$

$$+ n^{-1/2} \left[\frac{h_2(y_t, w_t, \theta)}{\sqrt{g\left(\varepsilon_t(\theta)^T \varepsilon_t(\theta)\right)}} + \frac{h_3(y_t, w_t)}{\sqrt{f_{x|z^{-\nu}}(x_t | z^{t-\nu}) / f_{-\nu}(z^{t-\nu})}} \right] \tag{A.6}$$

$$+ (2n)^{-1} \frac{\kappa^T \left[\partial^2 f_{z|z^{-\nu}}(z_t | z^{t-\nu}; \theta) / \partial\theta\partial\theta^T\right] \kappa}{f_{z|z^{-\nu}}(z_t | z^{t-\nu}; \theta)}$$

$$- \frac{n^{-1}\kappa^T (r_{1\theta t}(\theta) + r_{2\theta t}(\theta)) + n^{-1} r_{3t}(\theta)}{\widetilde{J}(y_t, w_t, \theta)\, g\left(\varepsilon_t(\theta)^T \varepsilon_t(\theta)\right) \cdot f_{x|z^{-\nu}}(x_t | z^{t-\nu})} \Big\}.$$

We next note that

$$\exp(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \cdots, \tag{A.7}$$

that

$$n^{-1} \sum_{t=1}^{n} \frac{\partial^2 f_{z|z^{-\nu}}(z_t | z^{t-\nu}; \theta) / \partial\theta\partial\theta^T}{f_{z|z^{-\nu}}(z_t | z^{t-\nu}; \theta)} = o_p(1), \tag{A.8}$$

that

$$n^{-1} \sum_{t=1}^{n} \frac{r_{1\theta t}(\theta) + r_{2\theta t}(\theta)}{\widetilde{J}(y_t, w_t, \theta)\, g\left(\varepsilon_t(\theta)^T \varepsilon_t(\theta)\right) \cdot f_{x|z^{-\nu}}(x_t | z^{t-\nu})} = o_p(1), \tag{A.9}$$

and that

$$n^{-1} \sum_{t=1}^{n} \frac{r_{3t}(\theta)}{\widetilde{J}(y_t, w_t, \theta)\, g\left(\varepsilon_t(\theta)^T \varepsilon_t(\theta)\right) \cdot f_{x|z^{-\nu}}(x_t | z^{t-\nu})} = o_p(1). \tag{A.10}$$

Combining (A.6)-(A.10), we obtain

$$\Lambda_n(\vartheta + A_n(h), \vartheta) = \exp\Bigg\{-n^{-1/2}\sum_{t=1}^{n}\kappa^T\left[J_\theta(y_t, w_t, \theta) + \rho_\theta^T(y_t, w_t, \theta)\,\varepsilon_t(\theta)\,d\left(\varepsilon_t(\theta)^T\varepsilon_t(\theta)\right)\right]$$

$$+n^{-1/2}\sum_{t=1}^{n}\left[\frac{h_2(y_t, w_t, \theta)}{\sqrt{g\left(\varepsilon_t(\theta)^T\varepsilon_t(\theta)\right)}} + \frac{h_3(y_t, w_t)}{\sqrt{f_{x|z^{-\nu}}(x_t\,|z^{t-\nu})\,/f_{-\nu}(z^{t-\nu})}}\right]$$

$$-(2n)^{-1}\sum_{t=1}^{n}\Bigg\{\kappa^T\left[J_\theta(y_t, w_t, \theta) + \rho_\theta^T(y_t, w_t, \theta)\,\varepsilon_t(\theta)\,d\left(\varepsilon_t(\theta)^T\varepsilon_t(\theta)\right)\right]$$

$$+\frac{h_2(y_t, w_t, \theta)}{\sqrt{g\left(\varepsilon_t(\theta)^T\varepsilon_t(\theta)\right)}} + \frac{h_3(y_t, w_t)}{\sqrt{f_{x|z^{-\nu}}(x_t\,|z^{t-\nu})\,/f_{-\nu}(z^{t-\nu})}}\Bigg\}^2 + o_p(1)\Bigg\}.$$

Defining the quantity

$$\Delta_n(h) = -n^{-1/2}\sum_{t=1}^{n}\Bigg\{\kappa^T\left[J_\theta(y_t, w_t, \theta) + \rho_\theta^T(y_t, w_t, \theta)\,\varepsilon_t(\theta)\,d\left(\varepsilon_t(\theta)^T\varepsilon_t(\theta)\right)\right]$$

$$+\frac{h_2(y_t, w_t, \theta)}{\sqrt{g\left(\varepsilon_t(\theta)^T\varepsilon_t(\theta)\right)}} + \frac{h_3(y_t, w_t)}{\sqrt{f_{x|z^{-\nu}}(x_t\,|z^{t-\nu})\,/f_{-\nu}(z^{t-\nu})}}\Bigg\},$$

we can show that

$$\Delta_n(h) \xrightarrow{d} N\left(0, \|h\|_H^2\right),$$

and that

$$n^{-1}\sum_{t=1}^{n}\Bigg\{\kappa^T\left[J_\theta(y_t, w_t, \theta) + \rho_\theta^T(y_t, w_t, \theta)\,s\left(\varepsilon_t(\theta)^T\varepsilon_t(\theta)\right)\right]$$

$$+\frac{h_2(y_t, w_t, \theta)}{\sqrt{g\left(\varepsilon_t(\theta)^T\varepsilon_t(\theta)\right)}} + \frac{h_3(y_t, w_t)}{\sqrt{f_{x|z^{-\nu}}(x_t\,|z^{t-\nu})\,/f_{-\nu}(z^{t-\nu})}}\Bigg\}^2 \xrightarrow{p} \|h\|_H^2.$$

It follows that

$$\Lambda_n(\vartheta + A_n(h), \vartheta) = \exp\left\{\Delta_n(h) - \frac{1}{2}\|h\|_H^2 + o_p(1)\right\},$$

so that the LAN conditions of Ibragimov and Khas'minskii (1991) are satisfied.

Ibragimov and Khas'minskii (1991) derive the semiparametric efficiency bound for estimation of some parameter of interest $\varsigma(\vartheta)$ when a model belongs to the LAN family. Since our parameter of interest is θ, $\varsigma(\vartheta)$ is the function that identifies the first k elements of ϑ. The semiparametric efficiency bound is the correlation operator KK^*, where the operator K is defined as

$$K = \lim_{n\to\infty} n^{1/2} \frac{\partial \varsigma(\vartheta)}{\partial \vartheta'} A_n P_H,$$

where P_H is the projection operator onto the space $\mathbf{H}$. For our model, $\frac{\partial \varsigma(\vartheta)}{\partial \vartheta'} = [I_k, 0, 0]$, so K will be the leading k elements of the operator $\widetilde{K} = \lim_{n\to\infty} n^{1/2} A_n P_H$. From Theorem A.4.5 of Bickel, Klaassen, Ritov, and Wellner (1993, p. 444), we can write $P_H = P_1 + P_2 + P_3$, where P_1 is the projection operator onto the space spanned by $S(y,w)$ and P_2 and P_3 are the projection operators onto the spaces $\mathbf{H}_2$ and $\mathbf{H}_3$, respectively. We can then write $K = \widetilde{A}_1 (P_1 + P_2 + P_3)$, where

$$\widetilde{A}_1(\lambda) = V_\theta^* \int S(y,w)\,\lambda(y,w) \sqrt{g(\varepsilon^T\varepsilon)\, f_{x|z^{-\nu}}(x\,|z^{-\nu})\,/f_{-\nu}(z^{-\nu})} d\varepsilon dw,$$

so that we have

$$K(\lambda) = V_\theta^* \int S(y,w)\, P_1\lambda(y,w),$$

since

$$\int S(y,w)\, h_2(y,w)\, dydw = \int S(y,w)\, h_3(y,w)\, dydw = 0$$

for every $h_2 \in \mathbf{H}_2$ and $h_3 \in \mathbf{H}_3$. It follows that the bound is $KK^* = V_\theta^*$.

References

[1] Andrews, D.W.K. and Ploberger, W. 1994. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62:1383-1414.

[2] Bickel, P.J., Klaassen, C.A.J., Ritov, Y., and Wellner, J.A. 1993. *Efficient and Adaptive Estimation for Semiparametric Models.* Springer-Verlag.

[3] Bollerslev, T.P. 1987. Modeling the coherence in short run nominal exchange rates: A multivariate generalized ARCH approach. *Review of Economics and Statistics* 72:498-505.

[4] Hansen, B.E. 1996. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64:413-430.

[5] Hodgson, D. J., Linton, O., Vorkink, K. and Choo, E.. 1999. Estimation in multivariate time series regression models with elliptically symmetric errors. Manuscript, University of Rochester.

[6] Ibragimov, I.A. and Khas'minskii. 1991. Asymptotically normal families of distributions and efficient estimation. *Annals of Statistics* 19:1681-1724.

[7] Jeantheau, T. 1998. Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* 14:70-86.

[8] Krugman, P. R. 1991. Target zones and exchange rate dynamics. *Quarterly Journal of Economics* 106:669-682.

[9] Newey, W. K. 1989. Locally efficient, residual-based estimation of nonlinear simultaneous equations. Bellcore Economics Discussion Paper #59.

[10] Newey, W. K. 1990. Semiparametric efficiency bounds. *Journal of Applied Econometrics* 5:99-135.