ON WEAK EXOGENEITY OF THE STUDENT'S t AND ELLIPTICAL LINEAR REGRESSION MODELS

Jiro Hodoshima*
Graduate School of Economics, Nagoya City University
Nagoya 467-8501, Japan

December, 2003

Abstract

This paper studies weak exogeneity of conditioning variables for the inference of a subset of parameters of the conditional student's t and elliptical linear regression models considered by Spanos (1994). Weak exogeneity of the conditioning variables is shown to hold for the inference of regression parameters of the conditional student's t and elliptical linear regression models. A new definition of weak exogeneity is given which utilizes block-diagonality of the conditional information matrix. A simulation experiment is made to compare the full-likelihood and conditional maximum likelihood estimators in finite samples for the conditional student's t linear regression model. The conditional maximum likelihood estimator of the regression parameters is found to work well in finite samples.

*The author gratefully acknowledges financial support from Grant-in-Aid for Scientic Research (KAKENHI(14530036)) for this research. Address correspondence to: Jiro Hodoshima, Graduate School of Economics, Nagoya City University, 1 Yamanohata, Mizuho-cho, Mizuho-ku, Nagoya 467-8501, Japan. Tel.: +81-52-872-5715; fax: +81-52-871-9429; E-mail address: hodoshi@econ.nagoya-cu.ac.jp

1 Introduction

Spanos (1994) presented an interesting class of non-normal linear heteroskedastic models with potentially wide applicability. It was shown in Spanos (1994) that linear heteroskedastic models naturally arise as conditional models when the underlying joint distribution of all the random variables in question follows the class of elliptically symmetric non-normal distributions. We shall call them the conditional non-normal linear heteroskedastic models in this paper. Spanos rightly claimed necessity to use the information of the marginal distribution of the conditioning random variables in order to estimate efficiently the parameters of the conditional non-normal linear heteroskedastic models. This is because the conditioning random variables are not weak exogenous for the inference of the parameters of the conditional non-normal linear heteroskedastic models. In this paper, we are concerned with a situation where we are interested in regression parameters, a subset of the parameters, of the conditional non-normal linear heteroskedastic models, instead of all of the parameters as in Spanos (1994). It is often the case that we are interested in the regression parameters but not variance of the error term. We show it is not necessary to use the information of the marginal distribution of the conditioning random variables in order to estimate efficiently the regression parameters, i.e., parameters of interest, of the conditional non-normal linear heteroskedastic models. This is equivalent to weak exogeneity of the conditioning random variables for the inference of the regression parameters of the conditional non-normal linear heteroskedastic models.

Engle et al. (1983) defined weak exogeneity of the conditioning random variables in estimation of the parameters associated with the conditional models. Spanos (1994) followed Engle et al. (1983) when he considered estimation of the conditional non-normal linear heteroskedastic models. Engle et al. (1983) used the idea of sequential cut or cut in defining weak exogeneity. Although Engle et al. (1983) is a seminal paper which established the concept of weak exogeneity in econometrics, we consider the condition imposed by sequential cut or cut, on which weak exogeneity in Engle et al. (1983) is based, is rather too restrictive to be applied to various models. In other words, the concept of sequential cut or cut requires parameters of the conditional distribution and those of the marginal distribution to be variation free (cf., Definition 2.4 of Engle et al. (1983)). We consider this is so restrictive as to make the conditioning random variables hard to satisfy weak exogeneity. In this paper, we first give a new definition of weak exogeneity which utilizes block-diagonality of the conditional information matrix, i.e., information matrix of the conditional distribution, in order to show weak exogeneity of the conditioning random variables for the inference of a subset of the parameters associated with the conditional models. By introducing the new definition of weak exogeneity, we can add flexibility to the restrictive concept of weak exogeneity established by Engle et al. (1983). Based on the new definition of weak exogeneity, weak exogeneity of the conditioning random variables is proved for the inference of the regression parameters of the conditional non-normal linear heteroskedasic models, by showing the block-daigonality of the conditional information matrix with respect to the regression parameters and variance of the error term of the conditional non-normal linear heteroskedastic models. We also present a simulation experiment to compare in finite samples the full-likelihood and conditional maximum likelihood estimators (MLEs) of the regression parameters of the conditional student's t linear heteroskedastic model, following the simulation study by Spanos (1994). The conditional MLE of the regression parameters is shown to work all right in finite samples while the conditional MLE of the remaining parameters does not perform well compared to the full-likelihood MLE.

The paper is organized as follows. In Section 2, a new definition of weak exogeneity is given. In Section 3, the block-diagonality of conditional information matrix with respect to the regression parameters and variance of the error term is proved to show weak exogeneity of the conditioning random variables for the inference of the regression parameters of the conditional non-normal linear heteroskedastic models. In Section 4, a simulation study is presented which compares the full-likelihood and conditional MLEs of the regression parameters of the conditional student's t linear regression model in finite samples. Concluding comments are given in Section 5.

2 A New Definition of Weak Exogeneity

In this section, we introduce a new definition of weak exogeneity which can be used to establish weak exogeneity of the conditioning random variables for the inference of the regression parameters of the conditional non-normal linear heteroskedastic models. Spanos (1994) considered a situation where the underlying data is independently and identically distributed (i.i.d.) while Engle et al. (1983) considered dynamic models where the underlying data is generally dependent over time. We basically follow the notation of Spanos (1994) to consider the joint distribution of $\mathbf{Z}_t \equiv (y_t, \mathbf{X}_t')'$, $t \in \mathbf{N}$, which is i.i.d. with elliptically symmetric non-normal distribution with mean $\boldsymbol{\mu}$ and scale matrix $\boldsymbol{\Sigma}$ (assuming they exist), denoted by

$$\mathbf{Z}_t \sim \mathcal{C}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where dimensions of y_t and $\mathbf{X}_t$ are 1×1 and $k \times 1$ respectively and $\mathbf{N}$ denotes the set of natural numbers. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are decomposed, conformably to the decomposition of $\mathbf{Z}_t$, as follows;

$$\begin{aligned} \boldsymbol{\mu} &= \begin{pmatrix} \mu_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \\ \boldsymbol{\Sigma} &= \begin{pmatrix} \sigma_{11} & \boldsymbol{\sigma}_{12} \\ \boldsymbol{\sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}. \end{aligned}$$

We remark that $\boldsymbol{\Sigma}$ does not necessarily correspond to the covariance matrix of $\mathbf{Z}_t$[1]. Since $\mathbf{Z}_t$ is assumed to be i.i.d., it is sufficient to consider the joint distribution of $\mathbf{Z}_t = (y_t, \mathbf{X}_t')'$ to define weak exogeneity of $\mathbf{X}_t$ for the inference of parameters of the conditional non-normal linear heteroskedastic models. Hence, the original parameters associated with the full-likelihood function or the joint probability density function (pdf) of $\mathbf{Z}_t$ are $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We now consider a one-to-one transformation or reparametrization of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, i.e., $\Psi \equiv ([\mu_1 - \boldsymbol{\sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\mu}_2], [\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}], [\sigma_{11} - \boldsymbol{\sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}], \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$. When $\mathbf{Z}_t$ is normal, the first three of Ψ correspond to the parameters associated with the conditional pdf of y_t given $\mathbf{X}_t$ and the last two of Ψ correspond to the parameters associated with the marginal pdf

[1]Spanos (1994) gave at subsection 3.1 a definition of the density function of multivariate Student t distribution with ν degrees of freedom, where its covariance matrix is given by $\frac{\nu}{\nu-2}\boldsymbol{\Sigma}$ not $\boldsymbol{\Sigma}$ (cf., Appendix B.2 of Zellner (1971)). Therefore, $\boldsymbol{\Sigma}$ is scale matrix but not necessarily covariance matrix in Spanos (1994). We follow Spanos in this respect.

of $\mathbf{X}_t$. On the other hand, when $\mathbf{Z}_t$ is non-normal, all of Ψ correspond to the parameters associated with the conditional pdf of y_t given $\mathbf{X}_t$ while the last two of Ψ continue to correspond to the parameters associated with the marginal pdf of $\mathbf{X}_t$ (cf., Lemma 1 and 2 of Spanos (1994) (or Theorem 7 of Kelker (1970) and Theorem 5 of Chu (1973))). We decompose Ψ as follows; $\Psi = (\Psi_1, \Psi_2)$ where $\Psi_1 = (\Psi_{11}, \Psi_{12})$ with $\Psi_{11} = (\beta_0, \boldsymbol{\beta})$ and $\Psi_{12} = (\sigma^2, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ and $\Psi_2 = (\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$ where $\beta_0 = \mu_1 - \boldsymbol{\beta}'\boldsymbol{\mu}_2$, $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$, and $\sigma^2 = \sigma_{11} - \boldsymbol{\sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\sigma}_{21}$. Then the joint pdf of $\mathbf{Z}_t$ is decomposed as the conditional pdf of y_t given $\mathbf{X}_t$ times the marginal pdf of $\mathbf{X}_t$ as follows;

$$D(\mathbf{Z}_t; \Psi) = D(y_t \mid \mathbf{X}_t; \Psi_1) D(\mathbf{X}_t; \Psi_2)$$

where $D(\mathbf{Z}_t; \Psi)$, $D(y_t \mid \mathbf{X}_t; \Psi_1)$, and $D(\mathbf{X}_t; \Psi_2)$ denote the joint pdf of $\mathbf{Z}_t$, conditional pdf of y_t given $\mathbf{X}_t$, and marginal pdf of $\mathbf{X}_t$ respectively. The definition of a sequential cut, or actually a cut since our setup is i.i.d., on $D(\mathbf{Z}_t; \Psi)$ (cf., Definition 2.4 of Engle et al. (1983)) cannot be applied to the decomposition $D(y_t \mid \mathbf{X}_t; \Psi_1) D(\mathbf{X}_t; \Psi_2)$ when $\mathbf{Z}_t$ is non-normally elliptical becasue of the overlap between Ψ_1 and Ψ_2.

Spanos (1994) investigated the weak exogeneity of conditioning random variables $\mathbf{X}_t$ for the inference of the parameters Ψ_1 of the class of the conditional non-normal linear heteroskedastic models according to the definition of Engle et al. (1983). However, the underlying model on which the definition of weak exogeneity in Engle et al. (1983) is based has rather a simplified structure and not suitable to analyze weak exogeneity of the conditioning random variables for the inference of the regression parameters of the conditional non-normal linear heteroskedastic models in Spanos (1994). We reproduce the definition of sequential cut, equivalent to cut in our setup, in Engle et al. (1983).

Definition 1. $[(y_t|\mathbf{X}_t; \lambda_1), (\mathbf{X}_t; \lambda_2)]$ operates a sequential cut (cut) on $D(\mathbf{Z}_t; \lambda)$ if and only if

$$D(\mathbf{Z}_t; \lambda) = D(y_t \mid \mathbf{X}_t; \lambda_1) D(\mathbf{X}_t; \lambda_2)$$

where $\lambda = (\lambda_1', \lambda_2')'$ is transformed parameters of the original parameters and λ_1 and λ_2 are variation free.

In the above definition, λ_1 and λ_2 being variation free implies that the range of admissible values for λ_i should not vary with $\lambda_j (j \neq i)$ and hence no cross-restrictions between λ_1 and λ_2 should not exist (cf., p. 282 of Engle et al. (1983)). When λ_1 is parameters of interest, $\mathbf{X}_t$ is defined in Engle et al. (1983) to be weakly exogenous for the inference of λ_1 if $[(y_t|\mathbf{X}_t;\lambda_1),(\mathbf{X}_t;\lambda_2)]$ operates a sequential cut (cut) on $D(\mathbf{Z}_t;\lambda)$.

We now introduce a new definition of weak exogeneity as follows.

Definition 2. When the conditional information matrix associated with the conditional pdf $D(y_t \mid \mathbf{X}_t;\Psi_1)$ is block-diagonal with respect to Ψ_{11} and Ψ_{12} , $\mathbf{X}_t$ is weakly exogenous for the inference of Ψ_{11}, i.e., inference on Ψ_{11} based on the conditional pdf of $D(y_t \mid \mathbf{X}_t;\Psi_1)$ involves no loss of information, if Ψ_{11} and Ψ_2 are variation free.

In the above definition 2, we have introduced block-diagonality of the conditional information matrix, i.e., the information matrix associated with the conditional pdf $D(y_t \mid \mathbf{X}_t;\Psi_1)$, to handle the overlap between Ψ_1 and Ψ_2. When Ψ_{12} and Ψ_2 are variation free, $\mathbf{X}_t$ is obviously weakly exogenous in the above definition, since Ψ_{11} and Ψ_2 are variation free. Even when Ψ_{12} and Ψ_2 are not variation free, the block-diagonality of the conditional information matrix with respect to Ψ_{11} and Ψ_{12} makes $\mathbf{X}_t$ weakly exogenous, if Ψ_{11} and Ψ_2 are variation free. In Engle et al. (1983, p.286), it was stated that the block-diagonality of the information matrix between two sets of parameters is often equivalent to the condition that the parametrization should operate a sequential cut (cut). However, the block-diagonality of the information matrix was not used there to discuss the exogeneity. Here we have made use of the block-diagonality of the conditional information matrix to define weak exogeneity of $\mathbf{X}_t$ for the inference of parameters Ψ_{11} . Since sequential cut (cut), the concept on which weak exogeneity of the conditioning random variables is based in Engle et al. (1983), has a simple structure, we need to adopt a different concept to establish weak exogeneity of the conditioning random variables for a subset of the conditional model. We find the block-diagonality of the conditional information matrix convenient for this purpose.

3 Block-diagonality of the conditional information matrix with respect to Ψ_{11} and Ψ_{12}

When $\mathbf{Z}_t$ is multivariate student's t distribution with ν degrees of freedom, the conditional log-likelihood function of y_t given $\mathbf{X}_t, t = 1, \ldots, T$, is given by

$$\begin{aligned} l_{Tc}(\Psi_1) &\equiv lnLc(\Psi_{\mathbf{1}}; y_1|\mathbf{X}_1, \ldots, y_T|\mathbf{X}_T) \\ &\propto -\frac{T}{2} ln(\sigma^2) + \frac{1}{2}(\nu + k)\sum_{t=1}^{T} ln(c_t) - \frac{1}{2}(\nu + m)\sum_{t=1}^{T} ln(\gamma_t) \end{aligned} \tag{1}$$

where

$$\begin{aligned} c_t &= \{1 + [\mathbf{X}_t - \boldsymbol{\mu}_2]'\boldsymbol{\Sigma}_{22}^{-1}[\mathbf{X}_t - \boldsymbol{\mu}_2]/\nu\} \\ \gamma_t &= (c_t + u_t^2/(\nu\sigma^2)) \end{aligned}$$

where $u_t = y_t - \beta_0 - \boldsymbol{\beta}'\mathbf{X}_t$ (cf.,p.296 of Spanos (1994)). As in Spanos (1994), ν is assumed to be known in this paper. Then we have

$$\frac{\partial l_{Tc}}{\partial \boldsymbol{\beta}} = \frac{\nu + m}{\nu\sigma^2}\sum_{t=1}^{T}[1/\gamma_t]\{\mathbf{X}_t(y_t - \beta_0 - \boldsymbol{\beta}'\mathbf{X}_t)\} \tag{2}$$

$$\frac{\partial l_{Tc}}{\partial \beta_0} = \frac{\nu + m}{\nu\sigma^2}\sum_{t=1}^{T}[1/\gamma_t](y_t - \beta_0 - \boldsymbol{\beta}'\mathbf{X}_t). \tag{3}$$

From the first derivative above, it can be easily derived that

$$\frac{\partial^2 l_{Tc}}{\partial \boldsymbol{\beta}\partial\sigma^2} = -\frac{\nu + m}{\nu\sigma^4}\sum_{t=1}^{T}\frac{\mathbf{X}_t u_t}{\gamma_t} + \frac{\nu + m}{\nu^2\sigma^6}\sum_{t=1}^{T}\frac{\mathbf{X}_t u_t^3}{\gamma_t^2} \tag{4}$$

$$\frac{\partial^2 l_{Tc}}{\partial \boldsymbol{\beta}\partial\mu_2'} = 2\frac{\nu + m}{\nu^2\sigma^2}\sum_{t=1}^{T}\frac{1}{\gamma_t^2}\mathbf{X}_t u_t(\mathbf{X}_t - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_{22}^{-1} \tag{5}$$

$$\frac{\partial^2 L_{Tc}}{\partial \boldsymbol{\beta}\partial\boldsymbol{\Sigma}_{22}'} = \frac{\nu + m}{\nu^2\sigma^2}\sum_{t=1}^{T}\frac{u_t}{\gamma_t^2}[\mathbf{X}_t \otimes \boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_t - \boldsymbol{\mu}_2)(\mathbf{X}_t - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_{22}^{-1}\}]. \tag{6}$$

(4) coincides with the corresponding second derivative of the log-likelihood function for the joint density, as given in Spanos (1994). The second derivative of the log-likelihood funciton for the joint density corresponding to (5) was not given in Spanos. The second derivative corresponding to (6) was given in Spanos in a different parametrization where $\boldsymbol{\Sigma}_{22}$ was replaced by $Q = (\frac{1}{\nu})\boldsymbol{\Sigma}_{22}^{-1}$.

The following property of the class of elliptically symmetric distributions holds; if $\mathbf{Z}_t \sim \mathcal{C}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{a} + \mathbf{B}\mathbf{Z}_t \sim \mathcal{C}_m(\mathbf{a} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\Sigma\mathbf{B}')$ where $\mathbf{a}$ is an mxl vector of constants and $\mathbf{B}$ is an mxm nonsingular matrix of constants (cf., for example, Theorem 2.6.3 of Fang and Zhang(1990)). Since

$$\begin{pmatrix} -\beta_0 \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} 1 & -\boldsymbol{\beta}' \\ \mathbf{0} & \boldsymbol{I}_k \end{pmatrix} \begin{pmatrix} y_t \\ \mathbf{X}_t \end{pmatrix} = \begin{pmatrix} u_t \\ \mathbf{X}_t \end{pmatrix},$$

$\mathbf{Z}_t = (y_t, \mathbf{X}_t')'$ being in the class of elliptically symmetric distributions implies $(u_t, \mathbf{X_t'})'$ being in the class of elliptically symmetric distributions. Hence, the conditional distribution of u_t given $\mathbf{X}_t$ is also elliptically symmetric distributed by the above property of the class of elliptically symmetric distributions. When $\mathbf{Z}_t$ is multivariate student's t distribution with ν degrees of freedom, u_t given $\mathbf{X}_t$ is student's t distribution with $k+\nu$ degrees of freedom with mean 0 and variance

$$\frac{\nu}{k+\nu-2}\sigma^2(1+(\mathbf{X}_t-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_t-\boldsymbol{\mu}_2)/\nu).$$

We take conditional expectation of the second derivatives (4),(5), and (6). $\mathbf{X}_t$ is given in (4),(5), and (6). Therefore, we evaluate conditional expectation of $u_t/\gamma_t, u_t^3/\gamma_t^2$, and u_t/γ_t^2 given $\mathbf{X}_t$. Since $\gamma_t = c_t + u_t^2/(\nu\sigma^2)$ where $c_t = (1+(\mathbf{X}_t-\boldsymbol{\mu}_2)'\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_t-\boldsymbol{\mu}_2)/\nu)$, the conditional expectation of $u_t/\gamma_t, u_t^3/\gamma_t^2$ and u_t/γ_t^2 given $\mathbf{X}_t$ are all zero because of symmetry if the conditional expectations exist. This is the same property utilized in Spanos (1994) to prove the unbiasedness of the full-likelihood MLE of $(\beta_0, \boldsymbol{\beta}', \boldsymbol{\mu}_2')'$. When $\mathbf{Z}_t$ is multivariate student's t distribution with ν degrees of freedom, all of the three conditional expectations exist if $k+\nu > -3$. Hence, if $k+\nu > -3$, we have

$$plim\frac{1}{T}\frac{\partial^2 l_{Tc}}{\partial\boldsymbol{\beta}\partial\sigma^2} = \mathbf{0} \tag{7}$$

$$plim\frac{1}{T}\frac{\partial^2 l_{Tc}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\mu}_2'} = \mathbf{0} \tag{8}$$

$$plim\frac{1}{T}\frac{\partial^2 l_{Tc}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\Sigma}_{22}'} = \mathbf{0}. \tag{9}$$

Similarly, since u_t given $\mathbf{X}_t$ is i.i.d., if $k+\nu > -3$, we can show

$$plim\frac{1}{T}\frac{\partial^2 l_{Tc}}{\partial\beta_0\partial\sigma^2} = 0,\ plim\frac{1}{T}\frac{\partial^2 l_{Tc}}{\partial\beta_0\partial\boldsymbol{\mu}_2'} = \mathbf{0},\ plim\frac{1}{T}\frac{\partial^2 l_{Tc}}{\partial\beta_0\partial\boldsymbol{\Sigma}_{22}'} = \mathbf{0}.$$

Therefore, we have block-diagonality of the conditional information matrix with respect to Ψ_{11} and Ψ_{12}.

The result developed above is the case of multivariate student's t distribution. Similarly, we can prove block-diagonality of the conditional information matrix with respect to Ψ_{11} and Ψ_{12} when $\mathbf{Z}_t = (y_t, \mathbf{X}_t')'$ follows other elliptically symmetric non-normal distributions since u_t given $\mathbf{X}_t$ is elliptically symmetric.

4 A simulation study

In this section, we present a simulation study to compare the finite-sample properties of the full-likelihood and conditional MLEs for the conditional student's t linear heteroskedastic model. We use bivariate student's t distribution with 9 degrees of freedom with mean $(2.5, 1.5)'$ and scale matrix $\begin{pmatrix} 2 & 0.86 \\ 0.86 & 1.2 \end{pmatrix}$ as the underlying joint distribution of $\mathbf{Z}_t$. We set the sample size T 200, 400, 800 and the number of repetitions N 1,000. This is the same setup as Spanos (1994) except the sample size T. In our study, we use the three different sample sizes to see the finite-sample comparison of the full-likelihood and conditional MLEs.

We compute the full-likelihood and conditional MLEs using Gauss[2] with the starting value of Ψ obtained from the sample mean and sample covariance matrix estimate for $\boldsymbol{\mu}$ and $cov(\mathbf{Z}_t) = \dfrac{\nu}{\nu - 2}\boldsymbol{\Sigma}$, using the transformation from $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to Ψ. The simulation results are given in Tables 1-6, which are presented similarly as in Spanos (1994). The population parameter values are given as follows from the transformation from $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to Ψ; $\beta_0 = 1.425$, $\beta = 0.717$, $\sigma^2 = 1.384$, $\mu_2 = 1.5$, and $\delta = \frac{1}{\nu}\boldsymbol{\Sigma}_{22}^{-1} = 0.0926$.[3]

Table 1 and 2 give simulation summary statistics of the full-likelihood and conditional MLEs respectively, when T = 200. The means of the full-likelihood MLE are all close to the population values. The standard deviations are larger, the minimums are smaller, and the maximums are larger in the full-likelihood MLE of β_0, β, and μ_2 than the corresponding ones reported in Spanos. This indicates our full-likelihood MLE varies more than that in Spanos. We attribute this to the difference of the student's t random numbers in Spanos and ours. The performance of the full-likelihood MLE of σ^2 and δ is unfortunately not comparable to the result of Spanos. However, it appears not to be

[2] Gauss programs are available on request from the author.

[3] The population parameter values of σ^2 and δ were not given in Spanos (1994). The full-likelihood MLEs of σ^2 and δ reported at Table 1 of Spanos (1994) deviate significantly from the population parameter values $\sigma^2 = 1.384$ and $\delta = 0.0926$. We do not know the reason for the above fact. On the other hand, our full-likelihood and conditional MLEs of σ^2 and δ given in Table 1-6 center around the population parameter values.

qualitatively different from that of β_0, β, and μ_2. The full-likelihood MLE works well as a whole when T = 200. On the other hand, the performance of the conditional MLE is characteristically different from that of the unconditional MLE. The conditional MLE of β_0 and β performs similarly to the full-likelihood MLE while the conditional MLE of σ^2, μ_2, and δ does not work well compared to the full-likelihood MLE, when T = 200. All of the summary statistics for the conditional MLE of σ^2, μ_2, and δ indicate the poor performance of the conditional MLE of these parameters. The poor performance of the conditional MLE of μ_2 and δ seems to be due to the fact that the information from the marginal distribution of $\mathbf{X}_t$ contains most of the information of μ_2 and δ. The performance of the conditional MLE of σ^2 could be explained by that of the conditional MLE of μ_2 and δ because of the correlation. Therefore, the conditional MLE works well, as good as the full-likelihood MLE, with respect to the regression parameters β_0 and β but not with respect to other parameters. As the number of observations increases, the performance of the full-likelihood and conditional MLEs generally improves as shown in Table 3-6. However, the same characteristics of the conditional MLE continue to hold when T = 400 and 800. 800 observations are not enough to make the performance of the conditional MLE of σ^2, μ_2, and δ reliable in this bivariate student's t linear heteroskedastic model.

Overall, the conditional MLE of the regression parameters works well in finite samples compared to the full-likelihood MLE. Therefore, our simulation study, although quite limited, verifies our theoretical finding that $\mathbf{X}_t$ is weakly exogenous for estimating $(\beta_0, \boldsymbol{\beta}')'$ when the joint distribution of $\mathbf{Z}_t = (y_t, \mathbf{X}_t')'$ is elliptically symmetric, and also shows the inference of $(\beta_0, \boldsymbol{\beta}')'$ based on the conditional MLE works well in finite samples.

5 Concluding comments

We have reexamined weak exogeneity of the conditioning random variables for the inference of the conditional non-normal linear heteroskedastic models considered by Spanos (1994). When we are interested in only a subset of the parameters of the conditional models, the conditioning random variables may be weakly exogenous for the inference of the subset of the parameters of the conditional models even though they are not weakly exogenous for the inference of all of the parameters of the conditional models. Weak exogeneity defined in Engle et al. (1983) is not applicable to this situation. Thus, we have introduced a new definition of weak exogeneity of the conditioning random variables for the inference of the subset of the parameters of the conditional models based on the block-diagonality of the conditional information matrix with respect to two disjoint subsets of the parameters associated with the conditional models. Based on the new definition of weak exogeneity, we have shown weak exogeneity of the conditioning random variables for the inference of the regression parameters, a subset of all of the parameters, of the conditional non-normal linear heteroskedastic models considered by Spanos (1994), where the conditioning random variables were not weakly exogenous for the inference of all of the parameters of the conditional models. In a limited simulation study, we have shown the conditional MLE of the regression parameters works as good as the full likelihood MLE in finite samples in the conditional student's t linear heteroskedastic model although the conditional MLE of the remaining parameters varies a lot and is not reliable.

Table 1. Simulation summary statistics (T=200) (full likelihood)

	Mean	SD	Min	Max	Skewness	Kurtosis
$\hat{\beta}_0$	1.424	0.158	0.970	2.075	0.145	3.030
$\hat{\beta}$	0.718	0.082	0.363	0.957	-0.084	2.974
$\hat{\sigma}^2$	1.368	0.165	0.974	2.033	0.366	3.058
$\hat{\mu}_2$	1.495	0.085	1.243	1.776	-0.042	3.086
$\hat{\delta}$	0.094	0.011	0.062	0.140	0.405	3.322
Correlation Matrix						
$\hat{\beta}_0$	1.000					
$\hat{\beta}$	-0.807	1.000				
$\hat{\gamma}^2$	-0.046	0.014	1.000			
$\hat{\mu}_2$	0.045	-0.024	-0.020	1.000		
$\hat{\delta}$	0.044	-0.013	-0.131	0.054	1.000	

Table 2. Simulation summary statistics (T=200) (conditional likelihood)

	Mean	SD	Min	Max	Skewness	Kurtosis
$\tilde{\beta}_0$	1.424	0.156	0.889	1.937	-0.029	3.069
$\tilde{\beta}$	0.718	0.084	0.409	1.070	0.026	3.359
$\tilde{\sigma}^2$	1.236	0.332	0.000	1.931	-1.824	7.873
$\tilde{\mu}_2$	1.453	4.110	-34.598	30.130	-0.513	29.485
$\tilde{\delta}$	0.929	17.376	0.000	540.557	30.080	931.431
Correlation Matrix						
$\tilde{\beta}_0$	1.000					
$\tilde{\beta}$	-0.820	1.000				
$\tilde{\sigma}^2$	0.043	-0.026	1.000			
$\tilde{\mu}_2$	-0.007	0.003	0.011	1.000		
$\tilde{\delta}$	0.020	-0.002	-0.176	-0.061	1.000	

Table 3. Simulation summary statistics (T=400)
(full likelihood)

	Mean	SD	Min	Max	Skewness	Kurtosis
$\hat{\beta}_0$	1.423	0.109	1.076	1.720	-0.080	2.895
$\hat{\beta}$	0.719	0.059	0.543	0.913	0.089	2.878
$\hat{\sigma}^2$	1.379	0.112	1.067	1.759	0.228	3.191
$\hat{\mu}_2$	1.500	0.058	1.334	1.694	0.097	2.867
$\hat{\delta}$	0.093	0.008	0.071	0.119	0.229	3.014
Correlation Matrix						
$\hat{\beta}_0$	1.000					
$\hat{\beta}$	-0.816	1.000				
$\hat{\sigma}^2$	-0.010	0.039	1.000			
$\hat{\mu}_2$	0.030	-0.008	0.035	1.000		
$\hat{\delta}$	0.061	-0.029	-0.118	-0.001	1.000	

Table 4. Simulation summary statistics (T=400)
(conditional likelihood)

	Mean	SD	Min	Max	Skewness	Kurtosis
$\tilde{\beta}_0$	1.423	0.109	1.077	1.823	-0.037	2.952
$\tilde{\beta}_1$	0.717	0.060	0.523	0.932	0.026	2.947
$\tilde{\sigma}^2$	1.325	0.194	0.003	1.783	-2.633	19.027
$\tilde{\mu}_2$	1.431	2.293	-23.051	30.538	-1.811	76.845
$\tilde{\delta}$	0.108	0.122	0.000	2.592	12.824	239.448
Correlation Matrix						
$\tilde{\beta}_0$	1.000					
$\tilde{\beta}$	-0.832	1.000				
$\tilde{\sigma}^2$	0.001	0.001	1.000			
$\tilde{\mu}_2$	0.034	-0.008	0.240	1.000		
$\tilde{\delta}$	0.000	0.021	-0.508	-0.212	1.000	

Table 5. Simulation summary statistics (T=800) (full likelihood)

	Mean	SD	Min	Max	Skewness	Kurtosis
$\hat{\beta}_0$	1.424	0.076	1.188	1.657	0.070	2.800
$\hat{\beta}$	0.717	0.042	0.581	0.842	-0.032	2.908
$\hat{\sigma}^2$	1.385	0.080	1.125	1.674	0.111	3.090
$\hat{\mu}_2$	1.501	0.041	1.367	1.653	0.120	3.218
$\hat{\delta}$	0.093	0.005	0.075	0.110	0.118	2.862
Correlation Matrix						
$\hat{\beta}_0$	1.000					
$\hat{\beta}$	-0.796	1.000				
$\hat{\sigma}^2$	0.030	-0.054	1.000			
$\hat{\mu}_2$	0.017	-0.014	0.055	1.000		
$\hat{\delta}$	0.006	0.018	-0.113	0.024	1.000	

Table 6. Simulation summary statistics (T=800) (conditional likelihood)

	Mean	SD	Min	Max	Skewness	Kurtosis
$\tilde{\beta}_0$	1.422	0.077	1.162	1.670	-0.007	3.096
$\tilde{\beta}$	0.719	0.042	0.585	0.868	-0.083	2.912
$\tilde{\sigma}^2$	1.362	0.106	0.305	1.775	-0.839	12.694
$\tilde{\mu}_2$	1.564	1.820	-5.028	55.162	25.913	756.337
$\tilde{\delta}$	0.096	0.044	0.000	0.287	0.480	3.591
Correlation Matrix						
$\tilde{\beta}_0$	1.000					
$\tilde{\beta}$	-0.808	1.000				
$\tilde{\sigma}^2$	-0.025	0.050	1.000			
$\tilde{\mu}_2$	0.009	0.015	-0.301	1.000		
$\tilde{\delta}$	-0.012	-0.025	-0.484	-0.078	1.000	

References

Chu, K.C. (1973) Estimation and decision for linear systems with elliptical random processes. *IEEE Transactions on Automatic Control* AC18, 499-505.

Engle, R.F., D.F. Hendry & J.F. Richard. (1983) Exogeneity. *Econometrica* 51, 277-304.

Fang, K.T. & Y.T. Zhang. (1990) *Generalized Multivariate Analysis*. Springer-Verlag.

Kelker, D. (1970) Distribution theory of spherical distribution and a location scale parameter generalization. *Sankhya A* 43, 419-430.

Spanos, A. (1994) On modeling heteroskedasticity: The student's t and elliptical linear regression models. *Econometric Theory* 10, 286-315.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics.* Wiley.