ON THE ASYMPTOTIC EFFECT OF SUBSTITUTING ESTIMATORS FOR NUISANCE PARAMETERS IN INFERENTIAL STATISTICS

By Z. L. Yang, Y. K. Tse and Z. D. Bai

This paper studies the general problem of making inferences for a set of parameters θ in the presence of another set of (nuisance) parameters λ, based on the statistic $T(\mathbf{y}; \hat{\lambda}, \theta)$, where $\mathbf{y} = \{y_1, y_2, \cdots, y_n\}$ represents the data, $\hat{\lambda}$ is an estimator of λ and the limiting distribution of $T(\mathbf{y}; \lambda, \theta)$ is known. We provide general methods for finding the limiting distributions of $T(\mathbf{y}; \hat{\lambda}, \theta)$ when $\hat{\lambda}$ is either a constrained estimator (given θ) or an unconstrained estimator. The methods will facilitate hypothesis testing as well as confidence-interval construction. We also extend the results to the cases where inferences may concern a general function of all parameters (θ and λ) and/or some weakly exogenous variables. Applications of the theories to testing serial correlation in regression models and confidence-interval construction in Box-Cox regressions are given.

KEYWORDS: Analytical correction, asymptotic independence, classical inference, limiting distribution, nuisance parameter.

1. INTRODUCTION

In a variety of econometric problems, the models for the data $\mathbf{y} = \{y_1, y_2, \cdots, y_n\}$ often involve two sets of parameters: the parameters of interest θ and the nuisance parameters λ. When λ is known, inferences for θ are usually simple as the limiting distribution of an inferential statistic $T(\mathbf{y}; \lambda, \theta)$ can often be derived. When λ is unknown, one temptation is to conduct inferences for θ based on $T(\mathbf{y}; \hat{\lambda}, \theta)$, which is obtained by substituting an estimate $\hat{\lambda}$ of λ into $T(\mathbf{y}; \lambda, \theta)$. This raises a major question: What is the limiting distribution of $T(\mathbf{y}; \hat{\lambda}, \theta)$? Are there simple ways to adjust the asymptotic distribution of $T(\mathbf{y}; \hat{\lambda}, \theta)$ so as to allow inferences for θ to proceed in the usual manner?

This paper studies these problems. One of the key factors in determining the limiting distribution of $T(\mathbf{y}; \hat{\lambda}, \theta)$ is whether $\hat{\lambda}$ is a constrained (given θ) or unconstrained estimator. The constrained case was considered by Pierce (1982). This case occurs often in hypothesis testing such as goodness-of-fit tests, residual-based diagnostics and Lagrange multiplier tests. It contains the case of no nuisance parameter as a special case. However, many classical inference methods, such as confidence-interval construction, Wald test, likelihood ratio test, etc., require the nuisance parameters to be estimated and substituted by their unconstrained estimators. In confidence-interval construction, it is well known that if the intervals fail to account for the estimation of the nuisance parameters they usually have lower-than-nominal coverages. Thus, it should be useful to provide results for the case where the nuisance parameters are replaced by their unconstrained estimates so as to allow for analytical adjustments for the limiting distribution of the inferential statistics. Before presenting the main results in the next section, we give a simple example to further motivate and understand the problems.

1.1. *An Example: The Weibull Duration Model*

Weibull distribution is one of the most popular models for modelling economic durations (Kiefer, 1988). For illustrative purpose, we consider a simple situation where $y_1, y_2, \cdots, y_n$ are independent and identically distributed (iid) Weibull random variables with probability density function (pdf) $\lambda\theta^{-\lambda}y^{\lambda-1}\exp(-(y/\theta)^\lambda)$, $\lambda > 0$. We are interested in inferences concerning θ, the scale parameter, with λ, the shape parameter, treated as a nuisance parameter. Define

$$T(\mathbf{y}; \lambda, \theta) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\theta}\right)^{\lambda} - 1.$$

Then, $\sqrt{n}\,T(\mathbf{y}; \lambda, \theta) \xrightarrow{D} N(0, 1)$. Also, the finite sample distribution of $2\sum_{i=1}^{n}(y_i/\theta)^\lambda$ is chi-square with $2n$ degrees of freedom. Thus, if λ is known, exact inference about θ can be conducted based on $2\sum_{i=1}^{n}(y_i/\theta)^\lambda$. Denote the constrained (for a given θ) maximum likelihood estimator (MLE) of λ by $\hat{\lambda}_c$ and the unconstrained MLE by $\hat{\lambda}_u$. Define

$$T(\mathbf{y}; \hat{\lambda}_c, \theta) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\theta}\right)^{\hat{\lambda}_c} - 1$$

and

$$T(\mathbf{y}; \hat{\lambda}_u, \theta) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\theta}\right)^{\hat{\lambda}_u} - 1.$$

Here, the standard asymptotic results of the maximum likelihood theory apply. Furthermore, some tedious but straightforward calculations show that $\sqrt{n}\,T(\mathbf{y}; \hat{\lambda}_c, \theta)$ is asymptotically independent of $\sqrt{n}\,(\hat{\lambda}_c - \lambda)$, and $\sqrt{n}\,T(\mathbf{y}; \lambda, \theta)$ is asymptotically independent of $\sqrt{n}\,(\hat{\lambda}_u - \lambda)$. Some further calculations show that

$$\sqrt{n}\,T(\mathbf{y}; \hat{\lambda}_c, \theta) \xrightarrow{D} N(0, 1 - c_1^2)$$

and

$$\sqrt{n}\,T(\mathbf{y}; \hat{\lambda}_u, \theta) \xrightarrow{D} N(0, 1 + c_2^2),$$

where $c_1^2 = (1-\gamma)^2/((1-\gamma)^2 + (\pi^2/6)) = 0.0980$, $c_2^2 = 6(1-\gamma)^2/\pi^2 = 0.1087$, and $\gamma = 0.5772$ is Euler's constant. See the Appendix for the detailed calculations of the above results. As we shall see, these results can be obtained as direct applications of Theorems 1 and 3 below.

Hence, the use of $\hat{\lambda}_c$ deflates the asymptotic variance, whereas the use of $\hat{\lambda}_u$ inflates the asymptotic variance. In both cases, it is very easy to adjust the statistics to give standard normal limiting distributions. Thus, we may have $\sqrt{n}\,T^*(\mathbf{y}; \hat{\lambda}_c, \theta) = \sqrt{n}\,T(\mathbf{y}; \hat{\lambda}_c, \theta)/\sqrt{0.902}$, and $\sqrt{n}\,T^*(\mathbf{y};\hat{\lambda}_u,\theta) = \sqrt{n}\,T(\mathbf{y};\hat{\lambda}_u,\theta)/\sqrt{1.1087}$. To test for $H_0 : \theta = \theta_0$, both statistics can be used, but to construct a confidence interval for θ, it is appropriate to use the latter.

The above example clearly demonstrates the different effects on the limiting distribution of the inferential statistic, depending on whether the constrained or unconstrained estimator is substituted. It also shows the way to correct the statistic to account for the estimation of the nuisance parameters. General results along these lines are clearly desirable. Furthermore, situations often arise in practical applications when (i) inference concerning a general function of both θ and λ, and (ii) inference concerning only a subset of θ, are desired. Thus, it will be useful to extend the methods to cover these cases.

The rest of the paper is organized as follows. Sections 2 presents general methods for finding the limiting distributions of $T(\mathbf{y};\hat{\lambda},\theta)$, and some extensions of the methods. Applications of the theorems to testing serial correlation in a regression model with lagged dependent variables and confidence interval construction in Box-Cox regressions are given in Section 3. Section 4 concludes. The proofs of the results can be found in the Appendix.

2. THE MAIN RESULTS

Denote the likelihood function of the data by $p(\mathbf{y};\lambda,\theta)$. It is desired to find the limiting distribution of $T(\mathbf{y};\hat{\lambda},\theta)$ with $\hat{\lambda}$ denoting generically an estimator of λ, which may be the constrained MLE given θ, or the unconstrained MLE. We now state the major assumptions and present some preliminary results.

2.1. *Assumptions and Preliminaries*

Assumption I: *For every λ, there is a joint convergence in law to normality:*

$$\begin{bmatrix} \sqrt{n}\,T(\mathbf{y};\lambda,\theta) \\ \sqrt{n}\,(\hat{\lambda}-\lambda) \end{bmatrix} \xrightarrow{D} N\left[0, \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}\right].$$

The dispersion matrix is assumed to be nonsingular.

Assumption II: There is a matrix $B = \lim_{n\to\infty} \mathrm{E}[\partial T(\mathbf{y};\lambda,\theta)/\partial\lambda']$, such that

$$\sqrt{n}\,T(\mathbf{y};\hat{\lambda},\theta) = \sqrt{n}\,T(\mathbf{y};\lambda,\theta) + B\sqrt{n}\,(\hat{\lambda}-\lambda) + o_p(1).$$

Assumptions I and II are similar to those of Pierce (1982). Under these assumptions, we can easily see that $\sqrt{n}\,T(\mathbf{y};\hat{\lambda},\theta)$ is asymptotically normal with mean zero and asymptotic variance

$$\mathrm{AVar}[\sqrt{n}\,T(\mathbf{y};\hat{\lambda},\theta)] = V_{11} + BV_{22}B' + BV_{21} + V_{12}B'. \tag{1}$$

In the above expression, B is the same whether $\hat{\lambda}$ is the constrained or unconstrained MLE. However, V_{22} and V_{12} will be different in the two situations. For this reason, we study the two cases separately.

ASSUMPTION III: $\mathrm{E}[T(\mathbf{y};\lambda,\theta)]=0$ *(or a function that is continuous in* λ *and* θ*) and there is an integrable function* $h(\mathbf{y})$ *such that, in a neighbourhood of* (λ,θ)*, the following conditions hold:*

$$
\begin{array}{ll}
(a) & |[\frac{\partial}{\partial\lambda}T(\mathbf{y};\lambda,\theta)]p(\mathbf{y};\lambda,\theta)|\le h(\mathbf{y});\quad |T(\mathbf{y};\lambda,\theta)[\frac{\partial}{\partial\lambda}p(\mathbf{y};\lambda,\theta)]|\le h(\mathbf{y})\\
(b) & |[\frac{\partial}{\partial\theta}T(\mathbf{y};\lambda,\theta)]p(\mathbf{y};\lambda,\theta)|\le h(\mathbf{y});\quad |T(\mathbf{y};\lambda,\theta)[\frac{\partial}{\partial\theta}p(\mathbf{y};\lambda,\theta)]|\le h(\mathbf{y}).
\end{array}
$$

We denote the score function by $U(\lambda,\theta)$ and write $U(\lambda,\theta)=(U_\lambda(\lambda,\theta)',U_\theta(\lambda,\theta)')'=(\partial L(\lambda,\theta)/\partial\lambda',\partial L(\lambda,\theta)/\partial\theta')'$, where $L(\lambda,\theta)$ is the log-likelihood function. Note that as a consequence of Assumption III we have the following Lemma.

LEMMA 1: *Under Assumption III, we have*

$$
\begin{aligned}
-\lim_{n\to\infty}\mathrm{E}[T(\mathbf{y};\lambda,\theta)U_\lambda(\lambda,\theta)'] &= \lim_{n\to\infty}\mathrm{E}[\partial T(\mathbf{y};\lambda,\theta)/\partial\lambda']\equiv B,\\
-\lim_{n\to\infty}\mathrm{E}[T(\mathbf{y};\lambda,\theta)U_\theta(\lambda,\theta)'] &= \lim_{n\to\infty}\mathrm{E}[\partial T(\mathbf{y};\lambda,\theta)/\partial\theta']\equiv C.
\end{aligned}
$$

We denote the constrained MLE of λ given θ by $\hat\lambda_c$ and the unconstrained MLE by $\hat\lambda_u$. Similarly, the constrained (given λ) and unconstrained MLE of θ are denoted, respectively, by $\hat\theta_c$ and $\hat\theta_u$. We denote the Fisher information matrix $-\mathrm{E}[\partial U(\lambda,\theta)/\partial(\lambda',\theta')]$ by $I(\lambda,\theta)$ and let

$$
A=\lim_{n\to\infty}\left(\frac{1}{n}I(\lambda,\theta)\right),
$$

which is partitioned according to (λ,θ) into sub-blocks A_{ij}, $i,j=1,2$. Assume the usual regularity conditions for the MLE hold. We have, for the constrained estimations,

$$
\sqrt{n}\,(\hat\lambda_c-\lambda)=\frac{1}{\sqrt{n}}\,A_{11}^{-1}U_\lambda(\lambda,\theta)+o_p(1), \tag{2}
$$

$$
\sqrt{n}\,(\hat\theta_c-\theta)=\frac{1}{\sqrt{n}}\,A_{22}^{-1}U_\theta(\lambda,\theta)+o_p(1), \tag{3}
$$

and for the unconstrained estimation of λ,

$$
\sqrt{n}\,(\hat\lambda_u-\lambda)=\frac{1}{\sqrt{n}}\,A_{11.2}^{-1}U_\lambda(\lambda,\theta)-\frac{1}{\sqrt{n}}\,A_{11.2}^{-1}A_{12}A_{22}^{-1}U_\theta(\lambda,\theta)+o_p(1), \tag{4}
$$

where $A_{11.2}^{-1}=(A_{11}-A_{12}A_{22}^{-1}A_{21})^{-1}$, which is the upper-left-corner block in A^{-1}.

The following result has been generally neglected in the literature.

LEMMA 2: *Under usual regularity conditions,* $\hat\theta_c$ *and* $\hat\lambda_u$ *are asymptotically independent.*

Lemma 2 is a fundamental result for classical likelihood inference. This lemma is essential in the derivation of some of the key results regarding the limiting distribution of $T(\mathbf{y};\hat\lambda,\theta)$. Section 3 presents some interesting applications. We are now ready to state the main results.

We first discuss the case of constrained estimation of λ, followed by the case of unconstrained estimation and some extensions.

2.2. *Substituting the Constrained Estimator*

The constrained estimator $\hat{\lambda}_c$ solves $U_\lambda(\hat{\lambda}_c, \theta) = 0$ for a given θ. Thus, its asymptotic expansion given in (2) involves only the λ-component of the score function.

Theorem 1: *Under Assumptions I-II and III(a), $\sqrt{n}\, T(\mathbf{y}; \hat{\lambda}_c, \theta)$ is asymptotically distributed as a normal variate with mean zero and asymptotic variance given by*

$$\mathrm{AVar}[\sqrt{n}\, T(\mathbf{y}; \hat{\lambda}_c, \theta)] = V_{11} - BA_{11}^{-1}B'. \tag{5}$$

Theorem 1 has been proved by Pierce (1982) under slightly different conditions. Pierce's result, however, has been largely neglected in the econometrics literature until recently. Bera and Kim (2002) used Pierce's theorem to derive a test for constant correlation in a bivariate conditional heteroscedasticity model. Tse (2002) applied it to examine residual-based diagnostics for univariate and multivariate conditional heteroscedasticity models.

2.3. *Substituting the Unconstrained Estimator*

Theorem 1 works mainly for hypothesis testing when θ is completely specified under the null. However, there are many practical situations where the null hypothesis does not give a complete specification of θ, or the null hypothesis involves a function of both θ and λ, etc. It is much easier in these cases to replace λ in $T(\mathbf{y}; \lambda, \theta)$ by an unconstrained estimator. Furthermore, in certain applications one may be interested in constructing a confidence region for θ. Thus, the use of $\hat{\lambda}_u$ is necessary. Unconstrained estimation involves the estimation of λ and θ simultaneously, and hence the asymptotic expansion (4) applies.

Theorem 2: *Under Assumptions I-III, $\sqrt{n}\, T(\mathbf{y}; \hat{\lambda}_u, \theta)$ is asymptotically distributed as a normal variate with mean zero and asymptotic variance given by*

$$\mathrm{AVar}[\sqrt{n}\, T(\mathbf{y}; \hat{\lambda}_u, \theta)] = V_{11} - BA_{11.2}^{-1}B' + CA_{22}^{-1}A_{21}A_{11.2}^{-1}B' + BA_{11.2}^{-1}A_{12}A_{22}^{-1}C'. \tag{6}$$

Though the result of Theorem 2 looks a bit complicated compared to the result of Theorem 1, it is still implementable as all the quantities in the asymptotic variance expression can be estimated from the data so that the asymptotic variance can be corrected easily. It is interesting to note that in many classical inference problems, $T(\mathbf{y}; \lambda, \theta)$ possesses certain special structure. For example, $T(\mathbf{y}; \lambda, \theta)$ may depend on $\mathbf{y}$ only through $\hat{\theta}_c$ (see the applications in Section 3). In this case, we have a simpler result.

Theorem 3: *Under Assumption II, if $T(\mathbf{y}; \lambda, \theta)$ depends on $\mathbf{y}$ only through $\hat{\theta}_c$ and is measurable in $\hat{\theta}_c$, then $T(\mathbf{y}; \lambda, \theta)$ is asymptotically independent of $\hat{\lambda}_u$, and $\sqrt{n}\, T(\mathbf{y}; \hat{\lambda}_u, \theta)$ is as-*

ymptotically distributed as a normal variate with mean zero and asymptotic variance given by

$$\text{AVar}[\sqrt{n}\,T(\mathbf{y};\hat{\lambda}_u,\theta)] = V_{11} + BA_{11.2}^{-1}B'. \tag{7}$$

For the Weibull duration example in Section 1.1, we can see, from Appendix A.1, that

$$\hat{\theta}_c = \left(\frac{1}{n}\sum_{i=1}^{n} y_i^{\lambda}\right)^{1/\lambda},$$

so that

$$T(\mathbf{y};\lambda,\theta) = \left(\frac{\hat{\theta}_c}{\theta}\right)^{\lambda} - 1,$$

which is a function of $\mathbf{y}$ through $\hat{\theta}_c$ only. Thus, $T(\mathbf{y};\lambda,\theta)$ and $\hat{\lambda}_u$ are asymptotically independent. Furthermore, as $B = (1-\gamma)/\lambda$ and $A_{11}^{-1} = \lambda^2/((1-\gamma)^2 + (\pi^2/6))$, from Theorem 1 we have $\sqrt{n}\,T(\mathbf{y};\hat{\lambda}_c,\theta) \xrightarrow{D} N(0, 1-c_1^2)$. Also, as $A_{11.2}^{-1} = (\pi^2/6)/\lambda^2$, from Theorem 3 we have $\sqrt{n}\,T(\mathbf{y};\hat{\lambda}_u,\theta) \xrightarrow{D} N(0, 1+c_2^2)$.

In summary, the effect of substituting the nuisance parameter λ in $T(\mathbf{y};\lambda,\theta)$ by its constrained MLE $\hat{\lambda}_c$ is to reduce the asymptotic variance (Theorem 1). When λ is substituted by its unconstrained MLE $\hat{\lambda}_u$, the effect on the asymptotic variance is (whether increase or decrease) uncertain (Theorem 2). Under the special case that $T(\mathbf{y};\lambda,\theta)$ depends on $\mathbf{y}$ only through the constrained MLE of θ, namely $\hat{\theta}_c$, the effect of substituting the nuisance parameter λ by its unconstrained MLE $\hat{\lambda}_u$ is to increase the asymptotic variance (Theorem 3).

2.4. *Some Extensions*

The above results refer to the case where λ represents the parameters to be substituted while the inference concerns all the elements in θ. We now address three other interesting cases: (i) there is no parameter of interest, i.e., no θ involved in the statistic, (ii) the inference concerns a subset of θ, and (iii) the inference concerns a function of θ or a function of both θ and λ. Clearly, in these cases, it is not meaningful to use $\hat{\lambda}_c$ defined for a given θ as the null hypothesis does not completely specify the value of θ. These issues are discussed in a unified manner as follows.

Let $g(\lambda,\theta)$ be a function of the parameters and is the focus for inference. Suppose the statistic used for inference concerning g when λ is known is $T(\mathbf{y};\lambda,g)$ with its limiting distribution completely specified. When λ is unknown it is replaced by $\hat{\lambda}_u$ to give $T(\mathbf{y};\hat{\lambda}_u,g)$.[1] In addition, g can be made even more flexible by allowing it to depend on some weakly exogenous variables. The following corollaries extend the results of Theorems 2 and 3.

[1]We note that λ in the function g is not replaced. In fact, it cannot be replaced as $g(\lambda,\theta)$ is the value that the inference is concerned.

Corollary 1: *Under the assumptions of Theorem 2, $T(\mathbf{y}; \hat{\lambda}_u, g)$ is asymptotically normally distributed with mean zero and asymptotic variance given by equation (6), with*

$$C = (\lim_{n\to\infty} \mathrm{E}[\partial T(\mathbf{y}; \lambda, g)/\partial g'])[\partial g(\lambda, \theta)/\partial \theta'].$$

Hence, when g is a constant in θ, $C = 0$ and $\mathrm{AVar}[\sqrt{n}\, T(\mathbf{y}; \hat{\lambda}_u, g)] = V_{11} - BA_{11.2}^{-1}B'$.[2]

Corollary 2: *Under the conditions of Theorem 3, if g is a function of λ and θ, and possibly some weakly exogenous variables, then*

$$\sqrt{n}\, T(\mathbf{y}; \hat{\lambda}_u, g) \xrightarrow{D} N(0, V_{11} + BA_{11.2}^{-1}B').$$

If g is constant in θ, then $C = 0$ and thus $B = 0$. Hence, $\sqrt{n}\, T(\mathbf{y}; \hat{\lambda}_u, g)$ has the same asymptotic distribution as $\sqrt{n}\, T(\mathbf{y}; \lambda, g)$.

3. APPLICATIONS

We now consider some applications to illustrate the use of the theorems.

3.1. *Regression Models with Residual Serial Correlation*

Consider a linear regression model with AR(1) errors given by

$$y_t = x_t'\lambda + \varepsilon_t,$$

where

$$\varepsilon_t = \theta\varepsilon_{t-1} + u_t$$

and $\{u_t\}$ are a sequence of white noise with variance 1 (this assumption simplifies the example without losing the essence). We first assume the regressor consists of only exogenous variables. The hypothesis of interest is $H_0 : \theta = 0$, and λ is a nuisance parameter. Using the notations above, we define the following statistic:

$$T(\mathbf{y}; \lambda, \theta) = \rho^* - \theta,$$

where

$$\rho^* = \frac{\sum(y_t - x_t'\lambda)(y_{t-1} - x_{t-1}'\lambda)}{\sum(y_t - x_t'\lambda)^2} = \frac{\sum \varepsilon_t\varepsilon_{t-1}}{\sum \varepsilon_t^2}.$$

Note that ρ^* is also the constrained (under known λ) MLE of θ, namely, $\hat{\theta}_c$. When λ is known and $\{\varepsilon_t\}$ are observed under the null, then $\sqrt{n}\, \rho^*$ is asymptotically distributed as a standard normal variate under H_0. When λ is unknown and is estimated by the full MLE $\hat{\lambda}_u$, we denote $\hat{\varepsilon}_t = y_t - x_t'\hat{\lambda}_u$. Thus, upon substituting $\hat{\lambda}_u$, the test statistic becomes

$$T(\mathbf{y}; \hat{\lambda}_u, \theta) = \hat{\rho} - \theta = \frac{\sum \hat{\varepsilon}_t\hat{\varepsilon}_{t-1}}{\sum \hat{\varepsilon}_t^2} - \theta.$$

[2] As λ is also involved in g, it should be clarified that $B = \lim_{n\to\infty} \mathrm{E}[\partial T(\mathbf{y}; \lambda, g)/\partial \lambda']$, where g is treated as a constant in the differentiation of $T(\mathbf{y}; \lambda, g)$ with respect to λ.

It can be checked that $B = 0$ when $\theta = 0$, so that from Theorem 3, $\sqrt{n}\,\hat{\rho}$ is asymptotically normal with mean 0 and variance 1 on H_0.

On the other hand, if the restricted (under $\theta = 0$) MLE (i.e., the OLS) is used for λ and we denote the OLS residual by $\tilde{\varepsilon}_t$, the test statistic becomes

$$T(\mathbf{y}; \hat{\lambda}_c, \theta) = \tilde{\rho} - \theta = \frac{\sum \tilde{\varepsilon}_t \tilde{\varepsilon}_{t-1}}{\sum \tilde{\varepsilon}_t^2} - \theta.$$

Applying Theorem 1, $\sqrt{n}\,\tilde{\rho}$ is also asymptotically normal with mean 0 and variance 1 on H_0.

We now consider the case with lagged dependent variables. For illustration, we consider the simple model with one lag:

$$y_t = x_t'\lambda_1 + y_{t-1}\lambda_2 + \varepsilon_t,$$

with $|\lambda_2| < 1$, and

$$\varepsilon_t = \theta\varepsilon_{t-1} + u_t.$$

For $H_0 : \theta = 0$, the nuisance parameters are $\lambda = (\lambda_1', \lambda_2)'$. We consider the statistic $T(\mathbf{y}; \lambda, \theta) = \theta^* - \theta$, with

$$\theta^* = \frac{\sum (y_t - x_t'\lambda_1 - y_{t-1}\lambda_2)(y_{t-1} - x_{t-1}'\lambda_1 - y_{t-2}\lambda_2)}{\sum (y_t - x_t'\lambda_1 - y_{t-1}\lambda_2)^2}. \tag{8}$$

Again, $\theta^* = \hat{\theta}_c$ and $\sqrt{n}\,\theta^*$ is asymptotically distributed as a standard normal variate under H_0. Denote $\mathbf{X}$ as the regression matrix, then $A_{11} = \lim_{n\to\infty} \mathrm{E}[\mathbf{X}'\mathbf{X}/n]$. Let

$$\lim_{n\to\infty} \mathrm{E}[\mathbf{X}'\mathbf{X}/n] = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}' & \sigma_{yy} \end{pmatrix}.$$

It can be shown that $A_{21} = (0, ..., 0, 1)$ and $A_{22} = 1$. Furthermore, $B = (0, ..., 0, 1)$ on H_0. Thus, if we substitute the OLS estimate of λ into $T(\mathbf{y}; \lambda, \theta)$ to obtain $\tilde{\theta}$, we conclude from Theorem 1 that $\sqrt{n}\,\tilde{\theta}$ is asymptotically normally distributed with mean 0 and variance $1 - v$, where v is the bottom corner element of A_{11}^{-1}, namely, $1/(\sigma_{yy} - \Sigma_{xy}'\Sigma_{xx}^{-1}\Sigma_{xy})$. This result has been proved by Durbin (1970) in a more general context.

Now, we consider

$$A_{11.2} = A_{11} - A_{12}A_{22}^{-1}A_{21} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}' & \sigma_{yy} - 1 \end{pmatrix}.$$

Suppose we substitute the unrestricted MLE of λ into equation (8) to obtain $T(\mathbf{y}; \hat{\lambda}_u, \theta)$. Then, from Theorem 3, on H_0 the asymptotic variance of $\sqrt{n}\,T(\mathbf{y}; \hat{\lambda}_u, \theta)$ is given by

$$1 + BA_{11.2}^{-1}B' = 1 + \frac{1}{(\sigma_{yy} - 1) - \Sigma_{xy}'\Sigma_{xx}^{-1}\Sigma_{xy}}. \tag{9}$$

Note that on the null, $T(\mathbf{y};\hat{\lambda}_u,\theta)$ is indeed the unrestricted MLE of θ, namely, $\hat{\theta}_u$. From standard MLE theory, the asymptotic variance of $\sqrt{n}\,\hat{\theta}_u$ is

$$(A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1}=\left(1-\frac{1}{\sigma_{yy}-\Sigma_{xy}'\Sigma_{xx}^{-1}\Sigma_{xy}}\right)^{-1},$$

which reduces to the expression in equation (9).

It is well known that the tests based on $\tilde{\theta}$ and $\hat{\theta}_u$ are asymptotically equivalent under local alternatives, due to the asymptotic equivalence of the Lagrange multiplier and likelihood ratio tests. However, the estimated asymptotic variance of $\sqrt{n}\,\tilde{\theta}$ may be negative in small samples, especially when the exogenous variables are highly trended (see, for example, Tse, 1985). In contrast, the estimated asymptotic variance of $\sqrt{n}\,\hat{\theta}_u$ is always positive.

The above results can be easily extended to cases when the residual variance is unknown and there are multiple lags in the dependent variable. While many model diagnostics are constructed based on the constrained MLE, mainly due to its simplicity in calculation, our results provide a way to obtain the asymptotic distribution of a diagnostic when unconstrained MLE is used. In some cases, such as the tests for dynamic specification suggested by Sargan (1980), unconstrained MLE may be more convenient.

3.2. *Box-Cox Regression*

The usual Box-Cox transformation model (Box and Cox, 1964) has the following form

$$h(\mathbf{y},\lambda)=\mathbf{X}\beta+\sigma\mathbf{e},$$

where $\mathbf{y}$ is an $n\times 1$ vector of original observations, $h(\mathbf{y},\lambda)$ is a vector of transformed observations, and $\mathbf{X}$ is an $n\times k$ matrix the columns of which contain the values of the explanatory variables $X_1,X_2,\ldots,X_k$, β is a $k\times 1$ vector of regression coefficients, σ is the error standard deviation, $\mathbf{e}$ is an $n\times 1$ vector of $N(0,1)$ variates, and $h(\cdot,\lambda)$ is a general monotonically increasing function, known except λ, called the *transformation parameter*.

In this application, we write $\theta=(\beta',\sigma^2)'$ and λ is the nuisance parameter. If λ is known, inferences concerning θ become simple. We now define

$$\hat{\beta}(\lambda)=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'h(\mathbf{y},\lambda)$$

and

$$\hat{\sigma}^2(\lambda)=\|\mathbf{M}h(\mathbf{y},\lambda)\|^2/n,$$

where $\|\cdot\|$ is the Euclidian norm and $\mathbf{M}=\mathbf{I}_n-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ with $\mathbf{I}_n$ being the $n\times n$ identity matrix. The unrestricted MLE (see Yang and Tse, 2002, for the details) of λ is $\hat{\lambda}_u=\arg\min_\ell \dot{J}^{-1}(\ell)\|\mathbf{M}h(\mathbf{y},\ell)\|$, where $\dot{J}(\ell)$ is the geometric mean of $\{h_y(Y_i,\ell)=\partial h(Y_i,\ell)/\partial Y_i, i=1,\cdots,n\}$. Likewise, the unrestricted MLE of β and σ^2 are, respectively, $\hat{\beta}(\hat{\lambda}_u)$ and $\hat{\sigma}^2(\hat{\lambda}_u)$.

First, we consider the inferences for $a'\beta$, a general linear function of β for a fixed vector a. Let $g = a'\beta$. When λ is assumed known, we consider the following statistic:

$$\sqrt{n}\,T(\mathbf{y};\lambda,g) = \frac{a'\hat{\beta}(\lambda) - g}{\{a'(\mathbf{X}'\mathbf{X})^{-1}a\}^{\frac{1}{2}}\hat{\sigma}(\lambda)},$$

which is asymptotically a standard normal. Tests and confidence intervals can be easily constructed for g. When λ is unknown and is substituted by its unconstrained MLE $\hat{\lambda}_u$, we have

$$\sqrt{n}\,T(\mathbf{y};\hat{\lambda}_u,g) = \frac{a'\hat{\beta}(\hat{\lambda}_u) - g}{\{a'(\mathbf{X}'\mathbf{X})^{-1}a\}^{\frac{1}{2}}\hat{\sigma}(\hat{\lambda}_u)}.$$

It is easy to verify that the conditions of Corollary 2 are satisfied. Hence, $\sqrt{n}\,T(\mathbf{y};\hat{\lambda}_u,g)$ is asymptotically normal with mean zero and $\text{AVar}[\sqrt{n}\,T(\mathbf{y};\hat{\lambda}_u,g)] = 1 + B^2 A_{11.2}^{-1}$, where

$$B = \lim_{n\to\infty} \frac{a' E[\hat{\beta}_\lambda(\lambda)]}{\sqrt{n}\{a'(\mathbf{X}'\mathbf{X})^{-1}a\}^{\frac{1}{2}}\sigma},$$

and $\hat{\beta}_\lambda(\lambda)$ is the derivative of $\hat{\beta}(\lambda)$ with respect to λ. In practice, the above *variance inflation factor*, $B^2 A_{11.2}^{-1}$, can be easily estimated and $\sqrt{n}\,T(\mathbf{y};\hat{\lambda}_u,g)$ can be corrected to have a $N(0,1)$ limiting distribution, so that inference about $a'\beta$ based on the corrected statistic is asymptotically valid. Bickel and Doksum (1981) showed that the asymptotic variance of $\hat{\beta}(\hat{\lambda}_u)$ is larger than that of $\hat{\beta}(\lambda)$, and thus it is not valid for making inference concerning β in the usual way. However, they did not provide ways to correct for the asymptotic variance of $\sqrt{n}\,T(\mathbf{y};\hat{\lambda}_u,g)$.

Suppose now we want to construct a confidence interval for the pth quantile of y_0 at a given observation x_0, denoted by g. It is easy to see that $g = h^{-1}[(x_0'\beta + \sigma z_p),\lambda]$, where z_p is the pth quantile of the standard normal variate. Note that g is a function of all the parameters. To state the problem in the framework of our theory, we need to find a statistic $T(\mathbf{y};\lambda,g)$ with a known asymptotic distribution. A natural choice is:

$$\sqrt{n}\,T(\mathbf{y};\lambda,g) = \frac{x_0'\hat{\beta}(\lambda) + \hat{\sigma}(\lambda)z_p - h(g,\lambda)}{\{x_0'(\mathbf{X}'\mathbf{X})^{-1}x_0\}^{\frac{1}{2}}\hat{\sigma}(\lambda)}.$$

With proper adjustment for the degrees of freedom $T(\mathbf{y};\lambda,g)$ is distributed as a noncentral t random variable (see Yang and Tse, 2002, for the details), so that confidence interval for $h(g,\lambda)$ can be easily constructed. Applying inverse transformations to the lower and upper confidence limits for $h(g,\lambda)$ gives the confidence limits for g. When λ is unknown, substituting $\hat{\lambda}_u$ for λ in the confidence limits results in a *plug-in* type of confidence interval. The validity of this interval clearly depends on whether the statistic

$$\sqrt{n}\,T(\mathbf{y};\hat{\lambda}_u,g) = \frac{x_0'\hat{\beta}(\hat{\lambda}_u) + \hat{\sigma}(\hat{\lambda}_u)z_p - h(g,\hat{\lambda}_u)}{\{x_0'(\mathbf{X}'\mathbf{X})^{-1}x_0\}^{\frac{1}{2}}\hat{\sigma}(\hat{\lambda}_u)},$$

has the same limiting distribution as $\sqrt{n}\,T(\mathbf{y};\hat{\lambda}_u, g)$. It can be verified that this problem fits into the framework of Corollary 2. Hence, $\sqrt{n}\,T(\mathbf{y};\hat{\lambda}_u, g)$ is asymptotic normal with mean zero and $\text{AVar}[\sqrt{n}\,T(\mathbf{y};\hat{\lambda}_u, g] = V_{11} + B^2 A_{11.2}^{-1}$, where

$$B = \lim_{n\to\infty} \frac{x_0'\text{E}[\hat{\beta}_\lambda(\lambda)] + z_p \text{E}[\hat{\sigma}_\lambda(\lambda)] - h_\lambda(g,\lambda)}{\sqrt{n}\{x_0'(\mathbf{X}'\mathbf{X})^{-1}x_0\}^{\frac{1}{2}}\sigma},$$

and $h_\lambda(g,\lambda)$ and $\hat{\sigma}_\lambda(\lambda)$ are the derivatives of $h(g,\lambda)$ and $\hat{\sigma}(\lambda)$ with respect to λ. This result and an extended version of it (for a heteroscedastic Box-Cox regression) formed the base of the general *corrected plug-in method* for constructing confidence intervals for a regression quantile proposed by Yang and Tse (2002). They showed that this method has a clear advantage over the commonly used delta method in terms of finite sample coverage probability.

It is clear from the second application that, even if one is concerned with the hypothesis testing on g, it is not feasible to use the constrained estimator $\hat{\lambda}_c$ as a replacement of λ in the statistics, as the hypothesis on g does not completely specify the values of θ. This becomes even more problematic for the case of quantile estimation as the g function also involves λ, the nuisance parameter. One could of course argue to use the an estimator under the constraint imposed on the g function instead of requiring to know all the values of θ. However, doing so induces at least two problems: (i) the estimation process becomes more complicated and (ii) Theorem 1 and its related results are no longer applicable.

4. CONCLUSIONS

We have examined the effects of substituting unknown parameters using the constrained or unconstrained MLE into an inferential statistic in a very general set-up. When the constrained MLE is used, there is variance deflation, as is well known in the literature. When the unconstrained MLE is used, the effects are uncertain, and we provide a formula for its calculation. However, if the inferential statistic depends only on the constrained MLE of the parameter of interest, substituting the unconstrained MLE of the nuisance parameter results in variance inflation. This result provides a way of conveniently adjusting for the variance of the inferential statistic when the unconstrained MLE of the nuisance parameters are used. We illustrate two examples of the applications of the results: testing for residual correlation in regression models and confidence-interval construction in Box-Cox transformed regressions.

School of Economics and Social Sciences, Singapore Management University, 469 Bukit Timah Road, Singapore 259756; zlyang@smu.edu.sg

School of Economics and Social Sciences, Singapore Management University, 469 Bukit Timah Road, Singapore 259756; yktse@smu.ed.sg

Dept. of Statistics and Applied Probability, National University of Singapore, 6 Science Drive 2, Singapore 117543; stabaizd@nus.edu.sg

APPENDIX

A.1 Derivation for Section 1 (Weibull Duration Model): First, from the Taylor's expansion and the law of large numbers we have, for any $\sqrt{n}$-consistent estimator $\hat{\lambda}$,

$$\begin{aligned}
\sqrt{n}\,T(\mathbf{y};\hat{\lambda},\theta) &= \sqrt{n}\,T(\mathbf{y};\lambda,\theta) + \frac{1}{\lambda n}\sum_{i=1}^{n}\left[\left(\frac{y_i}{\theta}\right)^{\lambda}\log\left(\frac{y_i}{\theta}\right)^{\lambda}\right]\sqrt{n}(\hat{\lambda}-\lambda)+o_p(1)\\
&= \sqrt{n}\,T(\mathbf{y};\lambda,\theta) + \frac{1}{\lambda n}\sum_{i=1}^{n}\mathrm{E}\left[\left(\frac{y_i}{\theta}\right)^{\lambda}\log\left(\frac{y_i}{\theta}\right)^{\lambda}\right]\sqrt{n}(\hat{\lambda}-\lambda)+o_p(1)\\
&= \sqrt{n}\,T(\mathbf{y};\lambda,\theta) + \frac{1-\gamma}{\lambda}\sqrt{n}(\hat{\lambda}-\lambda)+o_p(1).
\end{aligned}$$

The last equation follows from the result $\mathrm{E}[w\log w] = 1-\gamma$, where w is an exponential variable with mean 1 and γ is Euler's constant.

Now, the log-likelihood function is

$$L(\lambda,\theta) = n\log\lambda - n\lambda\log\theta + (\lambda-1)\sum_{i=1}^{n}\log y_i - \sum_{i=1}^{n}\left(\frac{y_i}{\theta}\right)^{\lambda},$$

which gives the score functions

$$U_\lambda(\lambda,\theta) = \frac{\partial L(\lambda,\theta)}{\partial\lambda} = \frac{n}{\lambda} + \sum_{i=1}^{n}\log\left(\frac{y_i}{\theta}\right) - \sum_{i=1}^{n}\left(\frac{y_i}{\theta}\right)^{\lambda}\log\left(\frac{y_i}{\theta}\right)$$

and

$$U_\theta(\lambda,\theta) = \frac{\partial L(\lambda,\theta)}{\partial\theta} = -\frac{n\lambda}{\theta} + \frac{\lambda}{\theta}\sum_{i=1}^{n}\left(\frac{y_i}{\theta}\right)^{\lambda}.$$

The Fisher information matrix $I(\lambda,\theta)$ has the following elements: $I_{\lambda\lambda} = n[(1-\gamma)^2+(\pi^2/6)]/\lambda^2$, $I_{\lambda\theta} = I_{\theta\lambda} = -n(1-\gamma)/\theta$, and $I_{\theta\theta} = n(\lambda/\theta)^2$.

The constrained estimator $\hat{\lambda}_c$ involves only $U_\lambda(\lambda,\theta)$ and hence has the following first-order approximation

$$\sqrt{n}\,(\hat{\lambda}_c-\lambda) = \sqrt{n}\,I_{\lambda\lambda}^{-1}U_\lambda(\lambda,\theta)+o_p(1).$$

This gives, by noticing $T(\mathbf{y},\lambda,\theta) = \theta U_\theta(\lambda,\theta)/(n\lambda)$,

$$\mathrm{ACov}[\sqrt{n}\,T(\mathbf{y};\hat{\lambda}_c,\theta),\sqrt{n}\,(\hat{\lambda}_c-\lambda)] = \frac{\theta}{\lambda}I_{\lambda\lambda}^{-1}I_{\theta\lambda} + \frac{1-\gamma}{\lambda}nI_{\lambda\lambda}^{-1} = 0.$$

Hence, $\sqrt{n}\,T(\mathbf{y};\hat{\lambda}_c,\theta)$ is asymptotically independent of $\sqrt{n}\,(\hat{\lambda}_c-\lambda)$, which gives

$$\mathrm{AVar}[\sqrt{n}\,T(\mathbf{y};\hat{\lambda}_c,\theta)] = 1-\left(\frac{1-\gamma}{\lambda}\right)^2\mathrm{AVar}[\sqrt{n}\,(\hat{\lambda}_c-\lambda)] = 1-\frac{(1-\gamma)^2}{(1-\gamma)^2+(\pi^2/6)}.$$

Now, the unconstrained estimator $\hat{\lambda}_u$ involves both U_λ and U_θ. It can be easily seen to have the following first-order approximation

$$\sqrt{n}\,(\hat{\lambda}_u-\lambda) = \sqrt{n}\,I^{\lambda\lambda}\left[U_\lambda(\lambda,\theta)+\frac{\theta(1-\gamma)}{\lambda^2}U_\theta(\lambda,\theta)\right]+o_p(1),$$

where $I^{\lambda\lambda}$ is the upper-left-corner block of $I^{-1}(\lambda, \theta)$. This, together with the fact that $T(\mathbf{y}, \lambda, \theta) = \theta U_\theta(\lambda, \theta)/(n\lambda)$, leads immediately to $\text{ACov}[T(\mathbf{y}; \lambda, \theta), \sqrt{n}\,(\hat{\lambda}_u - \lambda)] = 0$, and hence

$$\text{AVar}[\sqrt{n}\,T(\mathbf{y}; \hat{\lambda}_u, \theta)] = 1 + \left(\frac{1-\gamma}{\lambda}\right)^2 \text{AVar}[\sqrt{n}\,(\hat{\lambda}_u - \lambda)] = 1 + \frac{6(1-\gamma)^2}{\pi^2}.$$

A.2 Proof of Lemma 1: Under Assumption III, we apply the dominated convergence theorem (DCT) to obtain

$$\int \frac{\partial}{\partial \lambda'}(T(\mathbf{y}; \lambda, \theta)p(\lambda, \theta))\,d\mathbf{y} = \frac{\partial}{\partial \lambda'}\int T(\mathbf{y}; \lambda, \theta)p(\lambda, \theta)d\mathbf{y} = \frac{\partial}{\partial \lambda'}0 = 0.$$

Thus, we have

$$\int \left(\frac{\partial}{\partial \lambda'}T(\mathbf{y}; \lambda, \theta)\right) p(\lambda, \theta)d\mathbf{y} + \int T(\mathbf{y}; \lambda, \theta)\left(\frac{\partial}{\partial \lambda'}p(\lambda, \theta)\right) d\mathbf{y} = 0.$$

As the second term on the RHS of the above equation is $\int T(\mathbf{y}; \lambda, \theta)\,(\partial L(\lambda, \theta)/\partial \lambda')\,p(\lambda, \theta)d\mathbf{y}$, the first part of the lemma follows. Similarly, we can prove the second part of the lemma.

A.3 Proof of Lemma 2: It suffices to show that $\text{AVar}[n(\hat{\lambda}_u - \lambda)(\hat{\theta}_c - \theta)'] = 0$, which follows directly from the asymptotic expansions given in equations (3) and (4):

$$\begin{aligned}
\lim_{n\to\infty} \text{E}[n(\hat{\lambda}_u - \lambda)(\hat{\theta}_c - \theta)'] &= \lim_{n\to\infty} \text{E}[\frac{1}{n}A_{11.2}^{-1}(U_\lambda(\lambda, \theta) - A_{12}A_{22}^{-1}U_\theta(\lambda, \theta))U_\theta(\lambda, \theta)'A_{22}^{-1}] \\
&= A_{11.2}^{-1}(A_{12} - A_{12})A_{22}^{-1} = 0.
\end{aligned}$$

A.4 Proof of Theorem 1: Using the asymptotic expansion in equation (2) and Assumption II, we obtain, following the first result in Lemma 1,

$$V_{12} = \lim_{n\to\infty} n\text{E}[T(\mathbf{y}; \lambda, \theta)(\hat{\lambda}_c - \lambda)'] = \lim_{n\to\infty} \text{E}[T(\mathbf{y}; \lambda, \theta)U_\lambda(\lambda, \theta)'A_{11}^{-1}] = -BA_{11}^{-1}.$$

Substituting $V_{22} = A_{11}^{-1}$ and $V_{12} = -BA_{11}^{-1}$ into (1), we obtain the result of Theorem 1.

A.5 Proof of Theorem 2: Assumptions I and II ensure that $T(\mathbf{y}, \hat{\lambda}_u, \theta)$ is asymptotically normal with mean zero. For the asymptotic variance, we have, from the asymptotic expansion in equation (3),

$$\begin{aligned}
V_{12} &= \lim_{n\to\infty} n\text{E}[T(\mathbf{y}; \lambda, \theta)(\hat{\lambda}_u - \lambda)'] \\
&= \lim_{n\to\infty} \text{E}[T(\mathbf{y}; \lambda, \theta)U_\lambda(\lambda, \theta)']A_{11.2}^{-1} - \lim_{n\to\infty} \text{E}[T(\mathbf{y}; \lambda, \theta)U_\theta(\lambda, \theta)']A_{22}^{-1}A_{21}A_{11.2}^{-1} \\
&= -BA_{11.2}^{-1} + CA_{22}^{-1}A_{21}A_{11.2}^{-1}.
\end{aligned}$$

Substituting this back into equation (1), we obtain the result of Theorem 2.

A.6 Proof of Theorem 3: Since $T(\mathbf{y};\lambda,\theta) = T(\hat{\theta}_c;\lambda,\theta)$ and T is a measurable function of $\hat{\theta}_c$, we conclude, from Lemma 2, that $T(\mathbf{y};\lambda,\theta)$ is asymptotically independent of $\hat{\lambda}_u$, i.e., $V_{12} = 0$. The result of Theorem 3 thus follows from equation (1), noting $V_{22} = A_{11.2}^{-1}$.[3]

A.7 Proof of Corollary 1: Assumption III and the DCT lead to

$$\int \left(\frac{\partial}{\partial\theta'} T(\mathbf{y};\lambda,g(\lambda,\theta))\right) p(\lambda,\theta) d\mathbf{y} = -\int T(\mathbf{y};\lambda,\theta) \left(\frac{\partial}{\partial\theta'} p(\lambda,\theta)\right) d\mathbf{y}.$$

Thus, $C = -\lim_{n\to\infty} \mathrm{E}[T(\mathbf{y};\lambda,\theta)U_\theta(\lambda,\theta)'] = (\lim_{n\to\infty} \mathrm{E}[\partial T(\mathbf{y};\lambda,g)/\partial g'])\,[\partial g(\lambda,\theta)/\partial\theta']$.

A.8 Proof of Corollary 2: The main statement follows directly from Lemma 1. For the last statement, $B = 0$ as C and V_{12} are zero.

References

[1] Bera, A. K. and Kim, S. (2002). Testing constancy of correlation and other specifications of the BEARCH model with an application to international equity return. *Journal of Empirical Finance*, **9**, 171-195.

[2] Bickel, P. J. and Doksum, K. A. (1981). An analysis of transformation revisited. *Journal of the American Statistical Association*, **76**, 296-311.

[3] Box, G. E. P. and Cox, D. R. (1964). An analysis of transformation (with discussion). *Journal of the Royal Statistical Society B*, **26**, 211-252.

[4] Durbin, J. (1970). Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica*, **38**, 410-421.

[5] Kiefer, N. M. (1988). Economic Duration Data and Hazard Functions. *Journal of Economic Literature*, **26**, 646-679.

[6] Pierce, D. A. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. *The Annals of Statistics*, **10**, 475-478.

[7] Sargan, J. D. (1980). Some tests of dynamic specification for a single equation. *Econometrica*, **48**, 879-897.

[8] Tse, Y. K. (1985). Some modified versions of Durbin's h-statistic. *Review of Economics and Statistics*, **67**, 534-538.

[3] When T is a differentiable function of $\hat{\theta}_c$, Theorem 3 can also be proved by directly working with the result of Theorem 2. However, the use of Lemma 2 makes the proof easier, and more importantly, it allows a easy generalization of the result to other cases.

[9] Tse, Y. K. (2002). Residual-based diagnostics for conditional heteroscedasticity models. *Econometrics Journal*, **5**, 358-373.

[10] Yang, Z. L. and Tse, Y. K. (2002). A corrected plug-in method for the quantile confidence interval of a transformed regression. *SMU Economics & Statistics Working Paper Series, 22-2002.* available at: `www.sess.smu.edu.sg/Research/w2002.html`