Econometrics Journal (2008), volume **11**, pp. 349–376.
doi: 10.1111/j.1368-423X.2008.00242.x

Generalized LM tests for functional form and heteroscedasticity

ZHENLIN YANG† AND YIU-KUEN TSE†

†*School of Economics, Singapore Management University, Singapore 178903.*
E-mail: zlyang@smu.edu.sg; yktse@smu.edu.sg

First version received: May 2005; final version accepted: December 2007

Summary We present a generalized LM test of heteroscedasticity allowing the presence of data transformation and a generalized LM test of functional form allowing the presence of heteroscedasticity. Both generalizations are meaningful as non-normality and heteroscedasticity are common in economic data. A joint test of functional form and heteroscedasticity is also given. These tests are further 'studentized' to account for possible excess skewness and kurtosis of the errors in the model. All tests are easy to implement. They are based on the expected information and are shown to possess excellent finite sample properties. Several related tests are also discussed and their finite sample performances assessed. We found that our newly proposed tests significantly outperform the others, in particular in the cases where the errors are non-normal.

Keywords: *Box-Cox transformation*, *Double length regression*, *Functional form*, *Heteroscedasticity*, *LM tests*, *Robustness*.

1. INTRODUCTION

Non-normality and heteroscedasticity are common in economic data. A popular approach to modelling these data is to apply a non-linear transformation to the response and some of the regressors, with the anticipation that the transformed model is of independent and homoscedastic normal errors, and a simple model structure. In practice, however, it may not be the case that all of these goals can be achieved simultaneously by a single transformation. Typically, when genuine heteroscedasticity is present in the data, it may not be possible to find a transformation to bring the data to normality as well as homoscedasticity. A more proper and realistic approach is perhaps to directly model the heteroscedasticity while allowing the presence of data transformation in the model. Thus, the role of transformation is basically to induce normality and a relatively simpler model structure (or a correct functional form). This model, termed as Box-Cox heteroscedastic regression (BCHR) in the literature, has found interesting applications in economics (see, e.g. Yang and Tse, 2006).

This paper presents three LM tests for the BCHR model based on the expected information (EI). We first derive a simple but general LM test for heteroscedasticity allowing the presence of data transformation in the model. There is a large literature on tests for heteroscedasticity, and most of these tests are based on the assumption that the observations are normal.[1] Some authors have relaxed the normality condition and provided robust tests for heteroscedasticity (see, e.g. Koenker, 1981 and Ruppert and Carroll, 1981). Allowing a normalizing data transformation in the model is perhaps another way to account for the non-normality of the data. Also, most of these tests concern only a null hypothesis of homoscedastic errors (e.g. Breusch and Pagan, 1979). The need for a more general test is evident: when the null hypothesis of homoscedasticity is rejected, one would like to know which heteroscedastic variables are responsible for it. Hence, our test generalizes that of Breusch and Pagan (1979) in two dimensions: (i) from a null hypothesis of homoscedasticity to a null hypothesis of a certain form of heteroscedasticity and (ii) from a regular linear regression model to a transformed regression model. To further safeguard against non-normality, we provide a studentized LM test which generalizes that of Koenker (1981).

We then derive a generalized LM test for functional form allowing the presence of heteroscedasticity in the model. This test generalizes that of Yang and Abeysinghe (2003). Most of the functional form tests concern either a specific functional form (linear or log-linear) or a model with homoscedastic errors.[2] Our test allows for a general Box-Cox functional form that includes linear, log-linear, square-root, cubic-root, etc. as special cases, and the presence of a general heteroscedastic structure in the model. Interestingly, this test is shown, through Monte Carlo simulations, to be fairly robust against non-normality. Finally, a joint test of functional form and heteroscedasticity is given, which generalizes Lahiri and Egy (1981), and a robust version of it follows from the studentization or the robustness property of the two marginal tests.

There are other tests one could use such as the LM test based on the Hessian, LM test based on outer-product-of-gradient (OPG), LM test based on double length regression (DLR), and the likelihood ratio (LR) test.[3] They are all much easier to derive than the EI-based LM test, but not necessarily easier to implement in practical applications. More importantly, their finite sample performance remains unknown, at least in the context of the BCHR model. In this paper, we present empirical evidence on the finite sample performance of the tests discussed above, including the newly proposed ones, through extensive Monte Carlo simulations. In terms of size, some general observations are in order: (i) the three EI-based LM tests generally outperform all the others; (ii) the tests are ranked in the following order: LM-EI, LM-DLR, LR, LM-Hessian and LM-OPG; (iii) LM-DLR performs reasonably well especially considering the fact that it is based on only the first derivatives of the log-likelihood function; (iv) LM-OPG often performs very poorly and (v) the studentized LM test for heteroscedasticity, the LM-EI for functional form, and the studentized joint test are all quite robust against non-normality. In terms of size-adjusted power of the tests, it is observed that the EI-based tests always have better or similar power compared with others.

Section 2 presents the model and the estimation procedure. Section 3 presents the three tests. Section 4 contains the Monte Carlo simulation results and Section 5 concludes the paper. Appendix A contains the score and Hessian functions, Appendix B discusses some related tests, and Appendix C contains the proofs of the theorems and corollaries.

[1]See, for example, Goldfeld and Quant (1965), Glejser (1969), Harvey (1976), Amemiya (1977), Breusch and Pagan (1979), Ali and Giaccotto (1984), Griffiths and Surekha (1986), Farebrother (1987), Maekawa (1987), Evans and King (1988), Kalirajan (1989), Evans (1992), Wallentin and Agren (2002), Dufour et al. (2004) and Godfrey et al. (2006).

[2]See, for example, Box and Cox (1964), Godfrey and Wickens (1981), Tse (1984), Davidson and MacKinnon (1985), Lawrance (1987), Baltagi (1997) and Yang and Abeysinghe (2003).

[3]For a comparison of the observed and expected Fisher information, see Lindsay and Li (1997).

2. MODEL ESTIMATION

The BCHR model takes the following general form:

$$\begin{aligned} h(y_i, \lambda) &= \sum_{j=1}^{k_1} x_{ij}\beta_j + \sum_{j=k_1+1}^{k} h(x_{ij}, \lambda)\beta_j + \sigma\,\omega(v_i, \gamma)\,e_i, \\ &\equiv x_i'(\lambda)\beta + \sigma\,\omega(v_i, \gamma)\,e_i, \qquad i = 1, \ldots, n, \end{aligned} \tag{2.1}$$

where $h(\cdot, \lambda)$ is a monotonic increasing transformation dependent on a parameter vector λ with p elements, $\beta = \{\beta_1, \ldots, \beta_k\}'$ is $k \times 1$ vector of regression coefficients, and x_{ij} is the ith value of the jth regressor, $\omega\,(v_i, \gamma) \equiv \omega_i(\gamma)$ is the weight function, v_i is a set of q weighting variables, γ is a $q \times 1$ vector of weighting parameters, σ is a constant, and $\{e_i\}$ are independent and identically distributed (i.i.d.) with zero mean and unit variance. The first k_1 of the k regressors are not transformed as they correspond to the intercept, dummy variables, etc.

Let $\psi = \{\beta', \sigma^2, \gamma', \lambda'\}'$, $\Omega^{\frac{1}{2}}(\gamma) = \mathrm{diag}\{\omega_1(\gamma), \ldots, \omega_n(\gamma)\}$, $\Omega(\gamma) = \Omega^{\frac{1}{2}}(\gamma)\Omega^{\frac{1}{2}}(\gamma)$, $\mathbf{X}(\lambda)$ be the $n \times k$ regression matrix, and $\mathbf{Y}$ be the $n \times 1$ vector of (untransformed) dependent variable. The Gaussian log-likelihood function of model (2.1), ignoring the constant, is

$$\ell(\psi) = -\frac{n}{2}\log\sigma^2 - \sum_{i=1}^{n}\log\omega_i(\gamma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[\frac{h(y_i, \lambda) - x_i'(\lambda)\beta}{\omega_i(\gamma)}\right]^2 + \sum_{i=1}^{n}\log h_y(y_i, \lambda), \tag{2.2}$$

where $h_y(y, \lambda) = \partial h(y, \lambda)/\partial y$.

Define $\mathbf{M}(\gamma, \lambda) = I_n - \Omega^{-\frac{1}{2}}(\gamma)\mathbf{X}(\lambda)[\mathbf{X}'(\lambda)\Omega^{-1}(\gamma)\mathbf{X}(\lambda)]^{-1}\mathbf{X}'(\lambda)\Omega^{-\frac{1}{2}}(\gamma)$ where I_n is the $n \times n$ identity matrix. Maximizing (2.2) under given γ and λ results in constrained estimates:

$$\hat{\beta}(\gamma, \lambda) = [\mathbf{X}'(\lambda)\Omega^{-1}(\gamma)\mathbf{X}(\lambda)]^{-1}\mathbf{X}'(\lambda)\Omega^{-1}(\gamma)h(\mathbf{Y}, \lambda), \tag{2.3}$$

$$\hat{\sigma}^2(\gamma, \lambda) = \frac{1}{n}h'(\mathbf{Y}, \lambda)\Omega^{-\frac{1}{2}}(\gamma)\mathbf{M}(\gamma, \lambda)\Omega^{-\frac{1}{2}}(\gamma)h(\mathbf{Y}, \lambda), \tag{2.4}$$

which upon substitution gives the concentrated Gaussian log-likelihood,

$$\ell_p(\gamma, \lambda) = n\log[\dot{J}(\lambda)/\dot{\omega}(\gamma)] - \frac{n}{2}\log\hat{\sigma}^2(\gamma, \lambda), \tag{2.5}$$

where $\dot{\omega}(\gamma)$ and $\dot{J}(\lambda)$ are the geometric means of $\omega_i(\gamma)$ and $J_i(\lambda) = h_y(y_i, \lambda)$, respectively.

When $\{e_i\}$ are exactly normal, maximizing $\ell_p(\gamma, \lambda)$ over λ gives the constrained maximum likelihood estimate (MLE) $\hat{\lambda}_c$ of λ for a given γ, maximizing $\ell_p(\gamma, \lambda)$ over γ gives the constrained MLE $\hat{\gamma}_c$ of γ for a given λ, and maximizing $\ell_p(\gamma, \lambda)$ jointly over γ and λ gives the unconstrained MLEs $\hat{\gamma}$ and $\hat{\lambda}$ of γ and λ, respectively. Substituting these constrained or unconstrained MLEs into equations (2.3) and (2.4) gives the constrained or unconstrained MLEs of β and σ^2. When $\{e_i\}$ are not exactly normal, the above procedure leads to Gaussian quasi-MLEs (QMLEs) of the model parameters. Under mild conditions, these MLEs or QMLEs of the model parameters are consistent and asymptotic normal with the same mean but different variance-covariance matrices.[4]

[4] See Hernandez and Johnson (1980), Bickel and Doksum (1981), Carroll and Ruppert (1984) and Chen et al. (2002) for asymptotic results for some related models.

3. GENERALIZED LM TESTS

We first introduce some general notations. Define $D_\circ(\gamma) = \{\omega'_{i\gamma}(\gamma)/\omega_i(\gamma)\}_{n\times q}$ and $D(\gamma) = \{1_n, D_\circ(\gamma)\}$, where 1_n is the $n \times 1$ vector of ones, and $\omega_{i\gamma}(\gamma) = \partial\omega_i(\gamma)/\partial\gamma$. Let $\epsilon(\gamma, \lambda) = \{\epsilon_i(\gamma, \lambda)\}_{n\times 1}$, where $\epsilon_i(\gamma, \lambda) = [h(y_i, \lambda) - x_i'(\lambda)\hat{\beta}(\gamma, \lambda)]/[\omega_i(\gamma)\hat{\sigma}(\gamma, \lambda)]$, and $g(\gamma, \lambda) = \{g_i(\gamma, \lambda)\}_{n\times 1}$, where $g_i(\gamma, \lambda) = \epsilon_i^2(\gamma, \lambda) - 1$. Let $h_\lambda(y_i, \lambda)$ and $g_\lambda(\gamma, \lambda)$ be, respectively, the partial derivatives of the h and g functions with respect to λ.

Some basic assumptions are as follows. We assume $\omega(v_i, 0)$= constant (as commonly assumed in the literature) so that $\gamma = 0$ represents a model with homoscedastic errors. Without loss of generality, we take $\omega(v_i, 0) = 1$. We assume that $\omega_i(\gamma)$ is twice differentiable, and that $h(y_i, \lambda)$ is differentiable once with respect to y_i and twice with respect to λ. Some general technical assumptions are as follows. Proofs of all results are given in Appendix C.

ASSUMPTION 3.1. *The disturbances $\{e_i\}$ are independent and identically distributed with mean zero, variance one, skewness α, and finite kurtosis κ.*

ASSUMPTION 3.2. *The limit $\lim_{n\to\infty} \frac{1}{n}\mathbf{X}'(\lambda)\Omega^{-1}(\gamma)\mathbf{X}(\lambda)$ exists, and is positive definite.*

ASSUMPTION 3.3. *The limit $\lim_{n\to\infty} \frac{1}{n}D'(\gamma)D(\gamma)$ exists, and is positive definite. Further, the elements of $D(\gamma)$ are uniformly bounded.*

3.1. A generalized LM test for heteroscedasticity

THEOREM 3.1. Under Assumptions 3.1–3.3, assume further that (i) $\alpha = 0$ and $\kappa = 3$, (ii) $\frac{1}{\sqrt{n}}D'(\gamma)g_\lambda(\gamma, \ell) = O_p(1)$ uniformly in ℓ in a neighborhood of λ, and (iii) $\tilde{\lambda}$ is a consistent estimator of λ.[5] The LM statistic for testing H_0: $\gamma = \gamma_0$ versus H_a: $\gamma \neq \gamma_0$ takes the form

$$\mathrm{LM_E}(\gamma_0) = \frac{1}{2}g'(\gamma_0, \tilde{\lambda})D(\gamma_0)[D'(\gamma_0)D(\gamma_0)]^{-1}D'(\gamma_0)g(\gamma_0, \tilde{\lambda}), \tag{3.1}$$

which has an asymptotic χ^2_q distribution under H_0.

It turns out that this new test statistic is very simple. It is just one half of the explained sum of squares of the regression of $g_i(\gamma_0, \tilde{\lambda}) + 1$ on $D_i(\gamma_0)$, the ith column of $D'(\gamma_0)$. On the other hand, the test is very general as it works with any smooth transformation function h and weighting function ω. Robustness of (3.1) against non-normality of the original data $\mathbf{Y}$ is enhanced as the test allows the normalizing transformation to be chosen according to the data. Furthermore, If $\omega(v_i, \gamma) = \omega(v_i'\gamma)$, the special test for homoscedasticity takes a simpler form, and the test (like that of Breusch and Pagan 1979) does not depend on the exact form of the ω function. We have the following corollary.

[5] $\tilde{\lambda}$ could be $\hat{\lambda}_c$, or $\hat{\lambda}$, or any other estimator which converges in probability to λ as $n \to \infty$. For example, such an estimator could be constructed by adapting the method proposed by Powell (1996).

COROLLARY 3.1. Under the conditions of Theorem 3.1, assume further that $\omega(v_i, \gamma) = \omega(v_i'\gamma)$. Then, the LM statistic for testing H_0: $\gamma = 0$ becomes

$$\mathrm{LM_E}(0) = \frac{1}{2} g'(0, \tilde{\lambda}) V (V'V)^{-1} V' g(0, \tilde{\lambda}), \tag{3.2}$$

where $V = \{1, v_i'\}_{n\times(q+1)}$.

The test statistic for homoscedasticity in Corollary 3.1 is simply one half of the explained sum of squares of the regression of $g_i(0, \tilde{\lambda}) + 1$ on $V_i = (1, v_i')'$. It gives a one-step generalization to that of Breusch and Pagan (1979) by allowing a normalizing transformation to be present in the model, and hence it is more robust against the non-normality of the data. The test in Theorem 3.1 gives a two-step generalization by allowing for both a normalization transformation and a non-zero null vector γ_0. Hence, the test is not only more robust against the non-normality of the data, it also allows for easy identifications of truly heteroscedastic variables. It turns out that the asymptotic distribution of the test statistic does not depend on whether the λ parameter is pre-specified or estimated from the data.

3.2. Studentizing the LM test for heteroscedasticity

The LM tests given in Theorem 3.1 and Corollary 3.1 require that $\alpha = 0$ and $\kappa = 3$, which means that the disturbances $\{e_i\}$ are essentially Gaussian. This is in line with the aims of a data transformation: to induce normality, homoscedasticity as well as a simple model structure (or correct functional form). However, in many practical applications, it may not be possible to achieve these three goals simultaneously with a single transformation, in particular the exact normality in the errors. In this case, it might be more reasonable to assume that after the transformation, one has a correct functional form for the model while the errors obey Assumption 3.1 with arbitrary α and κ.

In this subsection we explore generalizations of the results given in Theorem 3.1 and Corollary 3.1 by dropping the assumptions that $\alpha = 0$ and $\kappa = 3$. Koenker (1981) generalized the result of Breusch and Pagan (1979) by providing a studentized version of the LM test for homoscedasticity, which is robust against non-normality of the errors in terms of excess kurtosis. Very recently, Dufour et al. (2004) and Godfrey et al. (2006) presented simulation-based tests for heteroscedasticity in linear regression models. While allowing the presence of data transformations and general heteroscedastic structure in the model complicates the matter, we are able to provide a result that very much parallels that of Koenker (1981).[6]

COROLLARY 3.2. Under Assumptions of Theorem 3.1 with arbitrary α and κ, the LM statistic for testing H_0: $\gamma = \gamma_0$ versus H_a: $\gamma \neq \gamma_0$ takes the form

$$\mathrm{LM_E^*}(\gamma_0) = \frac{1}{\tilde{\kappa} - 1} g'(\gamma_0, \tilde{\lambda}) D(\gamma_0)[D'(\gamma_0)D(\gamma_0)]^{-1} D'(\gamma_0) g(\gamma_0, \tilde{\lambda}), \tag{3.3}$$

[6] We are very grateful to a referee for directing our attention to the robustness issue of the LM tests for heteroscedasticity, which directly results in a new and more useful result as stated in Corollary 3.2. This idea is further explored in Sections 3.3 and 3.4 to provide robust tests for the other two cases.

where $\tilde{\kappa} - 1 = \frac{1}{n}\sum_{i=1}^{n} g_i^2(\gamma_0, \tilde{\lambda})$. The statistic has an asymptotic χ_q^2 distribution under H_0. Furthermore, if $\gamma_0 = 0$ and $\omega_i(v_i, \gamma) = \omega(v_i'\gamma)$, then $\mathrm{LM}_{\mathrm{E}}^*(0) = \frac{1}{\tilde{\kappa}-1} g'(0, \tilde{\lambda}) V(V'V)^{-1}V'g(0, \tilde{\lambda})$.

Note that $\mathrm{LM}_{\mathrm{E}}^*(\gamma_0)$ can be written as nR^2, where R^2 is the uncentered coefficient of determination from the regression of $g(\gamma_0, \tilde{\lambda})$ on $D(\gamma_0)$. Also note that $\mathrm{LM}_{\mathrm{E}}^*(\gamma_0)$ is as simple as $\mathrm{LM}_{\mathrm{E}}(\gamma_0)$, but should be much more useful when there exist excess skewness and kurtosis even if a normalizing transformation is applied to the data. This point is later confirmed by the Monte Carlo simulation.

3.3. A generalized LM test for functional form

Unlike the LM test for heteroscedasticity which requires only the submatrix of the expected information for a given λ, the main difficulty in deriving the expected information-based LM test for functional form is that it requires the explicit expression of the full expected information matrix. This is impossible for a general transformation function. However, when h is the Box-Cox power transformation: $h(y, \lambda) = (y^\lambda - 1)/\lambda$ if $\lambda \neq 0$; $\log y$ if $\lambda = 0$ (Box and Cox 1964), we are able to derive a very accurate approximation to the full expected information matrix, based on which a simple LM test for functional form emerges. The approximation is based on the expansion:

$$\lambda \log y_i = \log(1 + \lambda\eta_i) + \theta_i e_i - \frac{1}{2}\theta_i^2 e_i^2 + \cdots + \frac{(-1)^{k+1}}{k}\theta_i^k e_i^k + \cdots, \tag{3.4}$$

where $\theta_i = \lambda\sigma\omega_i(\gamma)/(1 + \lambda\eta_i)$ and $\eta_i = x_i'(\lambda)\beta$. Typically, the θ_i's are small, and in this case, one may just need a few terms to obtain the desired degree of approximation accuracy.[7]

We need further notations. Let $u(\gamma, \lambda) = \{[h_\lambda(y_i, \lambda) - x_{i\lambda}'(\lambda)\hat{\beta}(\gamma, \lambda)]/[\omega_i(\gamma)\hat{\sigma}(\gamma, \lambda)]\}_{n\times 1}$, where $x_{i\lambda}(\lambda)$ is the first derivative of $x_i(\lambda)$. Let $h_{\lambda\lambda}(y_i, \lambda) = \frac{\partial^2}{\partial\lambda^2} h(y_i, \lambda)$. Define $\theta_0 = \max\{|\theta_i|, i = 1, \ldots, n\}$, $\theta = \{\theta_i\}_{n\times 1}$, $\phi = \{\log(1 + \lambda\eta_i)\}_{n\times 1}$, $A = I_n - \frac{1}{n}1_n 1_n'$, and $R(\gamma) = AD_\circ(\gamma)[D_\circ'(\gamma)AD_\circ(\gamma)]^{-1}\ D_\circ'(\gamma)A$. Common functions applied to a vector are operated elementwise, e.g. $\theta^2 = \{\theta_i^2\}$ and $\log\theta = \{\log\theta_i\}$. Element-by-element multiplication (or Hadamard product) of two vectors, e.g. θ and ϕ, is denoted as $\theta \odot \phi$.

THEOREM 3.2. Under Assumptions 3.1–3.3, assume further that (i) h is the Box-Cox power transformation with $\theta_0 \ll 1$; (ii) $\{e_i\}$ are Gaussian, and (iii) $E[h_\lambda^2(y_i, \lambda)]$, $E[h(y_i, \lambda)h_\lambda(y_i, \lambda)]$ and $E[h(y_i, \lambda)h_{\lambda\lambda}(y_i, \lambda)]$ exist for all i. The EI-based LM test for testing H_0: $\lambda = \lambda_0$ is

$$\mathrm{LM}_{\mathrm{E}}(\lambda_0) = \frac{1_n' \log \mathbf{Y} - \epsilon'(\hat{\gamma}_c, \lambda_0)u(\hat{\gamma}_c, \lambda_0)}{\{\xi' M(\hat{\gamma}_c, \lambda_0)\xi + \delta - 2\zeta' R(\hat{\gamma}_c)\zeta\}^{1/2}}, \tag{3.5}$$

where when $\lambda \neq 0$, $\delta = \frac{1}{\lambda^2}(\frac{3}{2}\theta'\theta - 2\phi' A\theta^2 + 2\phi' A\phi) + O(\theta_0^4)$, $\xi = \frac{1}{\lambda}(\frac{1}{2}\theta + \phi \odot \theta^{-1} + \theta^3) - \frac{1}{\sigma}\Omega^{-\frac{1}{2}}(\gamma)\mathbf{X}_\lambda(\lambda)\beta + O(\theta_0^4)$, and $\zeta = \frac{1}{\lambda}(\phi - \frac{1}{2}\theta^2) + O(\theta_0^4)$; when $\lambda = 0$, $\delta = \frac{3}{2}\sigma^2 \mathrm{tr}(\Omega(\gamma)) +$

[7]There is a well known truncation problem for the Box-Cox power transformation. Model assumption requires this truncation effect to be negligible, which in turn requires θ_i's to be small. This is seen as follows. Since $(y_i^\lambda - 1)/\lambda = x_i'(\lambda)\beta + \sigma\omega_i(\gamma)e_i$, we have $y_i^\lambda = 1 + \lambda\, x_i'(\lambda)\beta + \lambda\sigma\omega_i(\gamma)e_i$. As $y_i > 0$ implies $y_i^\lambda > 0$, this in turn implies $|\lambda\sigma\omega_i(\gamma)| \ll 1 + \lambda\, x_i'(\lambda)\beta$ for the truncation on e_i to be negligible.

$2\eta' A\eta$, $\xi = \frac{1}{2\sigma}\Omega^{-\frac{1}{2}}(\gamma)[\eta^2 + \sigma^2\Omega(\gamma)1_n - 2\log(\mathbf{X})\beta]$, and $\zeta = \eta$. All the quantities θ, ϕ, δ, ζ and ξ are evaluated at the constrained MLEs at λ_0. Under H_0, $LM_E(\lambda_0)$ is asymptotic $N(0, 1)$.

Note that the order of the remainder term in the approximation to δ, ξ and ζ is $O(\theta_0^4)$, indicating that the third-order approximation, i.e. $k = 3$ in (3.4), is used. Our simulation results show that this approximation is very accurate. Although the test statistic given in Theorem 3.2 is derived under the assumption that the errors are Gaussian, it turns out that it is fairly robust against the non-normality of the errors as long as Assumption 3.1 is satisfied. This is seen from (i) the Monte Carlo results presented in Section 4 and (ii) tedious but straightforward approximations to the numerator of (3.5) using (3.4), which show that the effects of higher-order moments of errors are involved in terms of smaller magnitude.

3.4. Joint LM test for functional form and heteroscedasticity

It is sometimes desirable to conduct a joint test first for both functional form and heteroscedasticity simply because if the null hypothesis H_0: $\gamma = 0$, $\lambda = \lambda_0$ (where λ_0 can be any of the convenient values such as 0, 1, 1/2, 1/3, etc.) is not rejected, one may just need to fit an ordinary linear regression model with response and explanatory variables appropriately transformed according to the fixed λ_0 value. Of course, it is arguable that the two one-dimensional tests given earlier are more interesting as one would typically ask: given that we have fitted a transformation model, do we still need heteroscedasticity, or given that we have fitted a heteroscedastic regression model, do we still need to transform the data? Nevertheless, a joint test should be useful in certain applications, and a strong rejection of the null would simply lead to the consideration of the full transformed heteroscedastic regression model. Following the set up in Theorem 3.2, we have our third result.

THEOREM 3.3. Under the same set of assumptions as in Theorem 3.2, the EI-based LM statistic for testing H_0: $\gamma = \gamma_0$ and $\lambda = \lambda_0$ is given by

$$\mathrm{LM_E}(\gamma_0, \lambda_0) = S_c'(\gamma_0, \lambda_0)\begin{pmatrix} 2D_\circ'(\gamma_0)AD_\circ(\gamma_0), & -2D_\circ'(\gamma_0)A\zeta \\ -2\zeta' AD_\circ(\gamma_0), & \xi' M(\gamma_0, \lambda_0)\xi + \delta \end{pmatrix}^{-1} S_c(\gamma_0, \lambda_0), \quad (3.6)$$

where the concentrated score $S_c(\gamma_0, \lambda_0) = \{D_\circ'(\gamma_0)g(\gamma_0, \lambda_0), 1_n'\log\mathbf{Y} - \epsilon'(\gamma_0, \lambda_0)\, u(\gamma_0, \lambda_0)\}'$. All the quantities ξ, ζ and δ are give in Theorem 3.2, but evaluated at the constrained MLEs at γ_0 and λ_0. Under H_0, $\mathrm{LM_E}(\gamma_0, \lambda_0)$ is asymptotic χ^2_{q+1}.

Although the derivations for the $\mathrm{LM_E}(\lambda_0)$ and $\mathrm{LM_E}(\gamma_0, \lambda_0)$ statistics are more tedious than the other forms of LM tests, their implementations are not, and may even be simpler than the other versions of the LM tests. Besides, their excellent finite sample performance as shown in Section 4 indicates that for the cases where one has only a small data set, the $\mathrm{LM_E}(\lambda_0)$ or $\mathrm{LM_E}(\gamma_0, \lambda_0)$ should be used. The point of having a test with good finite sample behaviour is further emphasized in Dufour et al. (2004) and Godfrey et al. (2006).

Following the result of Corollary 3.2 and the robustness property of the test given in (3.5), one easily generalizes the result of Theorem 3.3 to provide a studentized (robustified) version of the joint LM test, allowing the errors to be non-Gaussian satisfying Assumption 3.1.

$$\mathrm{LM}^*_{\mathrm{E}}(\gamma_0, \lambda_0) = S'_c(\gamma_0, \lambda_0) \begin{pmatrix} \bar{\tau} D'_\circ(\gamma_0) A D_\circ(\gamma_0), & -\bar{\tau} D'_\circ(\gamma_0) A \zeta \\ -\bar{\tau} \zeta' A D_\circ(\gamma_0), & \xi' M(\gamma_0, \lambda_0) \xi + \delta \end{pmatrix}^{-1} S_c(\gamma_0, \lambda_0), \quad (3.7)$$

where $\bar{\tau} = \frac{1}{n} \sum_{i=1}^{n} g_i^2(\gamma_0, \lambda_0)$.

4. MONTE CARLO RESULTS

Section 3 introduces three EI-based LM tests for three different testing situations, and Appendix B discusses some related tests. While all the tests for a given situation are asymptotically equivalent when the errors are normally distributed and hence any of them can be used when a large data set is available, their small sample performance remains an important question. The purpose of the Monte Carlo experiment is: (i) to assess the small sample performance of the three new tests, (ii) to assess the small sample performance of the related (and readily available) tests and (iii) to compare and contrast all the tests to give practical guidance on which to use when only small data set is available. We consider the following data generation process (DGP):

$$h(y_i, \lambda) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}(\lambda) + \sigma \exp(\gamma_1 x_{1i} + \gamma_2 x_{2i}) e_i, \quad i = 1, \ldots, n, \quad (4.1)$$

where the values for x_{1i} are generated from $U(0, 10)$ and the values for x_{2i} are generated from either $U(0, 10)$ or $U(0, 5)$, and then fixed throughout the whole Monte Carlo experiment. Throughout, the regression coefficient are set to $\beta_0 = 25$, $\beta_1 = 10$, and $\beta_2 = 10$.

The sample size n, transformation parameter λ, heteroscedasticity parameters γ_1 and γ_2, and the error standard deviation σ are the quantities that could potentially affect the finite sample behaviour of the LM tests. Thus, for a thorough investigation, we have considered various combinations of the values of these quantities for which $n \in \{30, 80, 200\}$, $\lambda \in \{0.0, 0.2, 0.5, 0.8, 1.0\}$, $\gamma_1 \in \{0.0, 0.1, 0.2\}$, $\gamma_2 \in \{0.0, 0.1, 0.2, 0.3\}$, and $\sigma \in \{0.1, 0.5, 1.0\}$. All parameter configurations are chosen so that the probability of truncation, i.e. the probability that $1 + \lambda[\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}(\lambda) + \sigma \exp(\gamma_1 x_{1i} + \gamma_2 x_{2i}) e_i] \leq 0$, is negligible.

The simulation process is as follows. For a given parameter configuration, i.e. each set of values of n, σ, γ_1, γ_2, and λ, a random sample of e_i's are generated from $N(0, 1)$ or a non-normal population with zero mean and unit variance, which is then converted to the values for y_i's through the DGP in (4.1). Then, we proceed with model estimation and calculation of test statistics assuming the parameters are not known. Record 1 for each test if it rejects the null hypothesis. Repeat this process 10,000 times and the proportion of rejections gives a Monte Carlo estimate of the size (empirical size) of the test. The comparison of the small-sample performance of the tests will be based on their empirical sizes. As the tests are asymptotically equivalent under the null and local alternatives, the small-sample size is the most basic criterion for performance comparison.

To examine the effects of non-normal errors on the tests, two non-normal populations are considered: a normal mixture and a normal-gamma mixture, both standardized to have zero mean and unit variance. In the case of the normal mixture, 80% of the e_i's are from $N(0, 1)$, and the remaining 20% from $N(0, 4)$; whereas in the case of the normal-gamma mixture, 80% of the e_i's are from $N(0, 1)$, and the remaining 20% from GA(1, 1), a gamma distribution with both scale and shape parameters being one.

For brevity, we report only a representative part of the Monte Carlo results. Full results are available from the authors upon request. For clarity and conciseness, we use plots to summarize the simulation results. In each plot, the vertical scale is the empirical size, and the horizontal scale is the index for the 60 possible combinations of parameter values of $\gamma_1 \in \{0.0, 0.1, 0.2\}$, $\gamma_2 \in \{0.0, 0.1, 0.2, 0.3\}$ and $\lambda \in \{0.0, 0.2, 0.5, 0.8, 1.0\}$ with λ being the fastest changing index, followed by γ_2 and then γ_1.

4.1. *Tests for heteroscedasticity*

Seven tests are investigated in this case, namely, (i) $\texttt{LME}^0$ which is (3.1) with $\tilde{\lambda}$ replaced by the true value λ, (ii) LME which is (3.1), (iii) LME* which is the studentized statistic in (3.3), (iv) LMD (LM test based on double length regression), (v) LR (likelihood ratio test), (vi) LMH (LM test based on Hessian) and (vii) LMG (LM test based on gradient). The last four tests are described in Appendix B. As these seven tests all allow for any smooth monotonic h function, we consider two transformations in this case: the Box-Cox power transformation (Box and Cox 1964) and the dual power transformation of Yang (2006), where $h(y, \lambda) = (y^{\lambda} - y^{-\lambda})/2\lambda$ if $\lambda \neq 0$; $\log y$ if $\lambda = 0$. Figure 1 summarizes the results.

From Figure 1 the following regularities are observed: (i) LME* has an excellent finite sample performance even when the sample size is as small as 30, irrespective of whether the errors are normal or non-normal, and of what transformation is used; (ii) LME and $\texttt{LME}^{o}$ have excellent finite sample performance only when the errors are normal, showing the necessity of studentizing LME to safeguard against possible departures from normality of the error distribution; (iii) LMD performs very well under normal errors when the Box-Cox transformation is used, but not well enough when the dual power transformation is used; (iv) In the case of non-normal errors, all the tests except LME* suffer from size distortions, and furthermore, their empirical sizes apparently do not converge to the nominal level 5% as n increases; (v) when errors are normal, the empirical sizes of all the seven tests converge fairly quickly to 5% as n increases, except for LMG with its empirical coverages still nearly double the nominal size when $n = 200$ and (vi) changing the error standard deviation and the ranges of the covariates' values changes the empirical sizes of the tests slightly, but not the general regularities summarized above.

4.2. *Tests for functional form*

In this case, we report the empirical sizes for five tests: LME, LMD, LMH, LMG, and LR. Selected results are summarized in Figure 2. Some general observations are in order: (i) LME generally possesses excellent finite sample properties and outperforms all the others; (ii) the tests are ranked in the following order: LME, LMD, LR, LMH and LMG, with LMG often performing very poorly; (iii) it is worthnoting that LMD performs reasonably well, especially considering the fact that it is based on only the first derivatives of the loglikelihood function; (iv) all tests are fairly robust against departures from normality of the error distribution; (v) as n increases, empirical sizes converge to 5% and (vi) changing the parameter values does not affect much the empirical sizes.

4.3. *Tests for functional form and heteroscedasticity*

Six tests, namely, LME, LMD, LMH, LMG, LR and LME* (defined in (3.7)), are compared, where when the errors are normal, LME* is excluded. Selected results are summarized in Figure 3. For

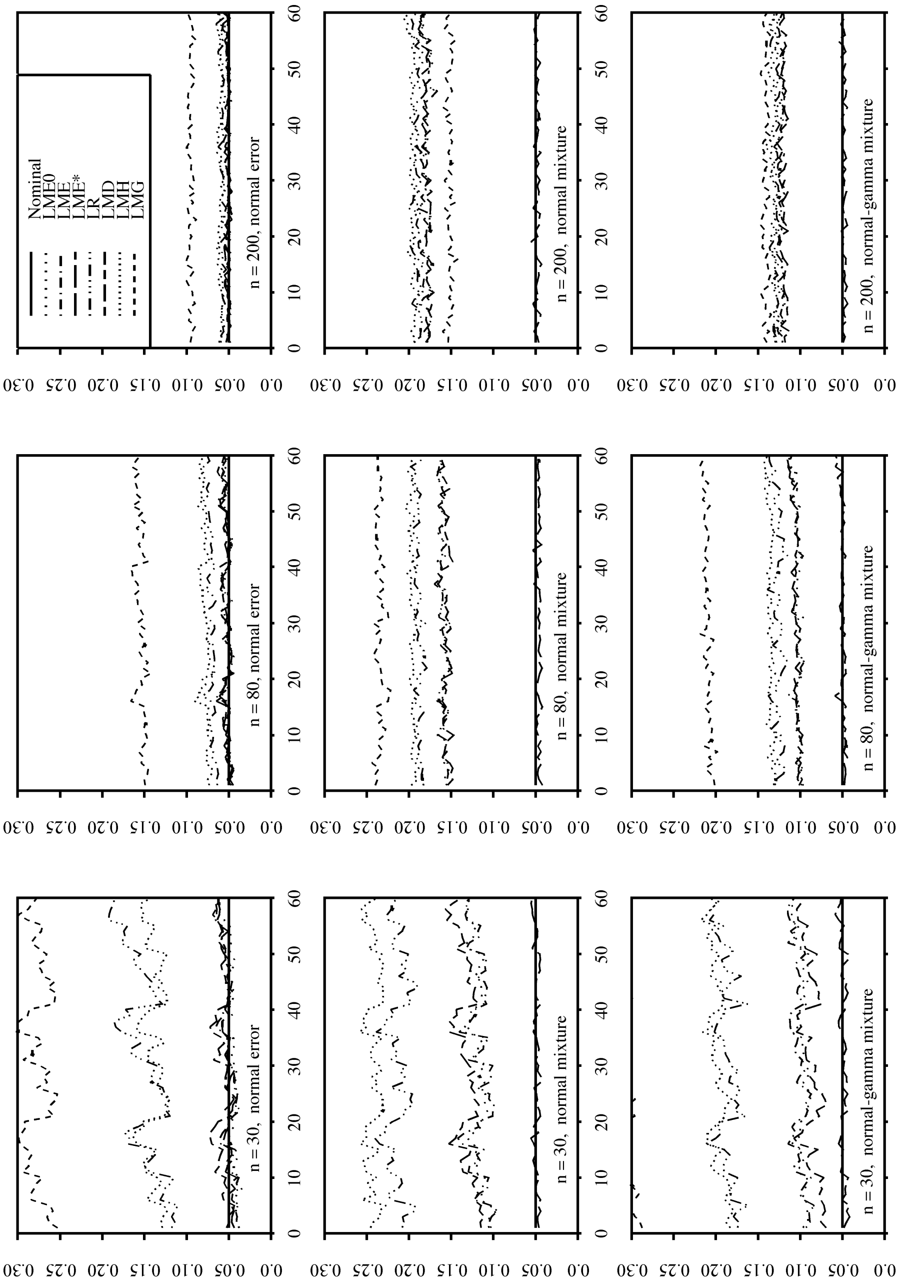

Figure 1a. Empirical sizes of LM tests for heteroscedasticity, BC transformation, $\sigma = 0.1$, $X_1 \sim U(0, 10)$ and $X_2 \sim U(0, 10)$.

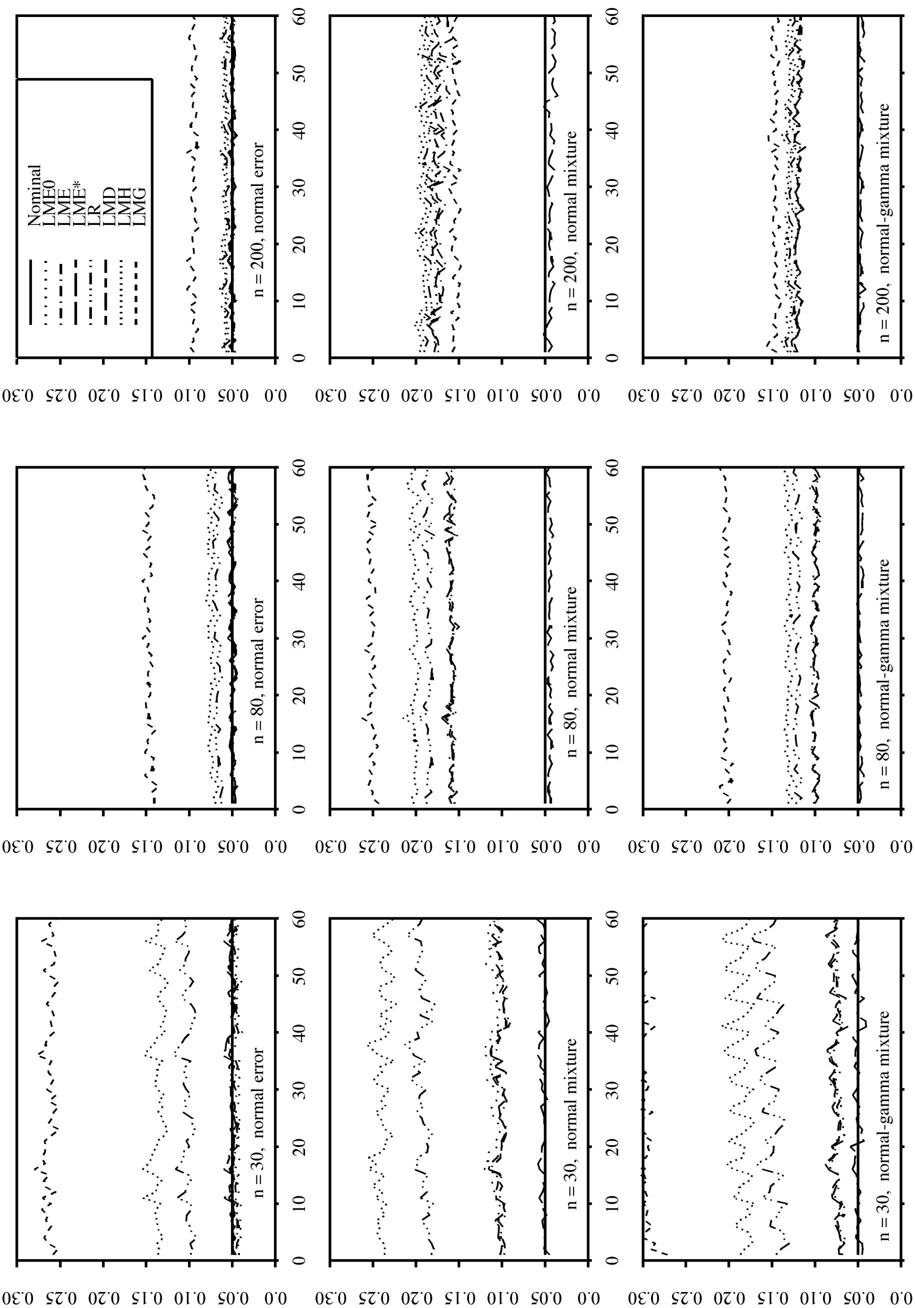

Figure 1b. Empirical sizes of LM tests for heteroscedasticity, BC transformation, $\sigma = 0.5$, $X_1 \sim U(0, 5)$ and $X_2 \sim U(0, 5)$.

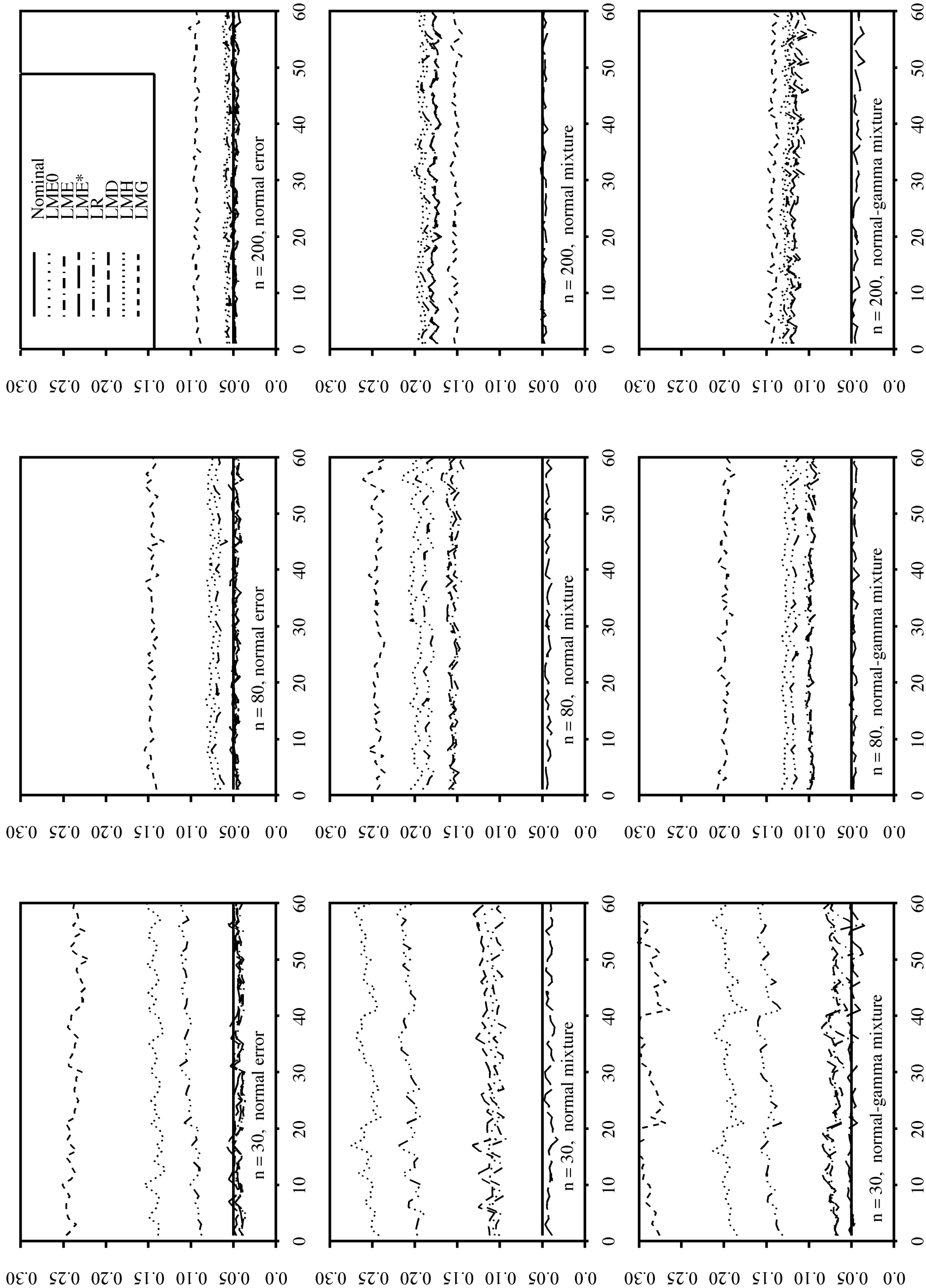

Figure 1c. Empirical sizes of LM tests for heteroscedasticity, BC transformation, $\sigma = 1.0$, $X_1 \sim U(0, 5)$ and $X_2 \sim U(0, 5)$.

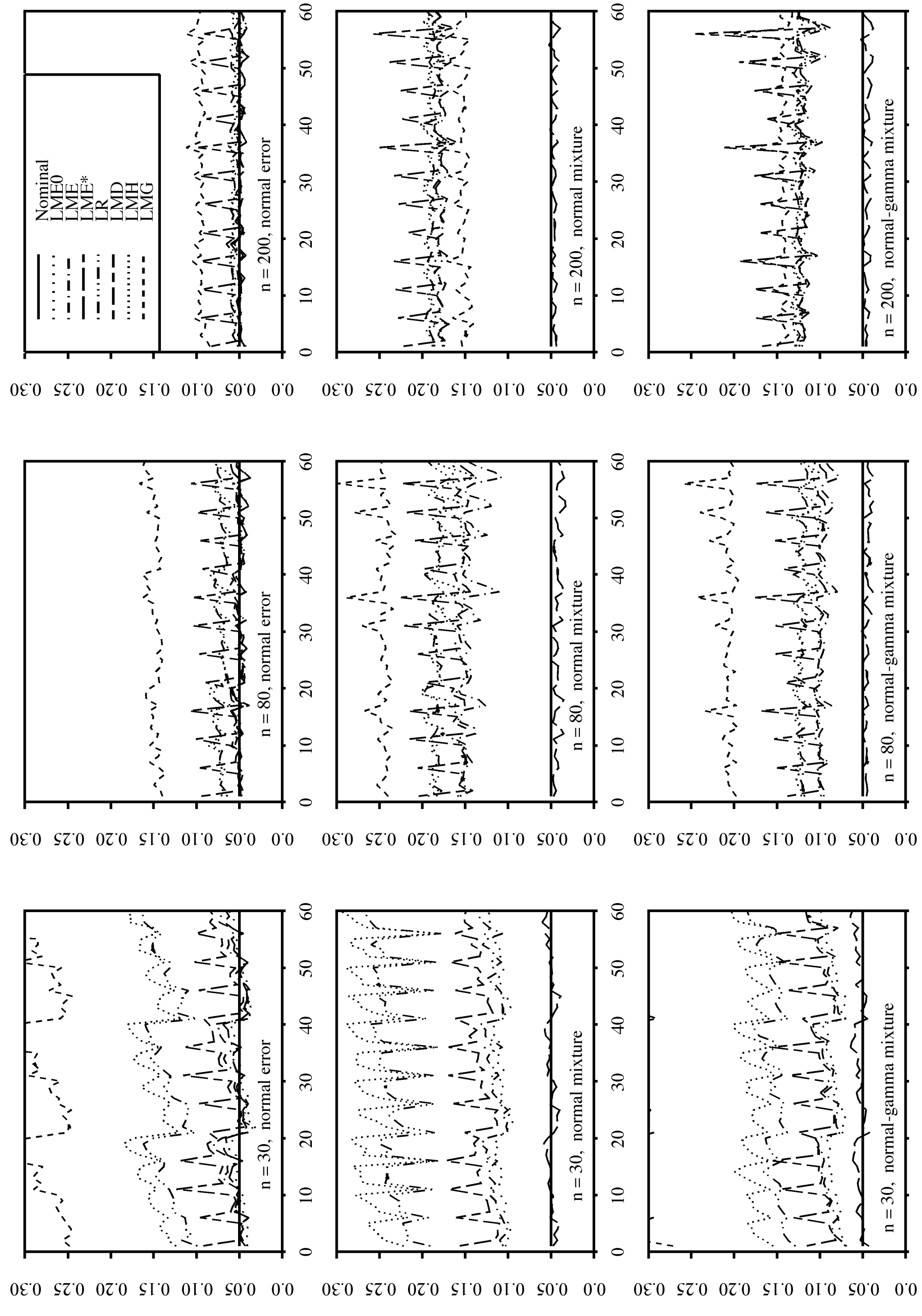

Figure 1d. Empirical sizes of LM tests for heteroscedasticity, DP transformation, $\sigma = 0.1$, $X_1 \sim U(0, 5)$ and $X_2 \sim U(0, 10)$.

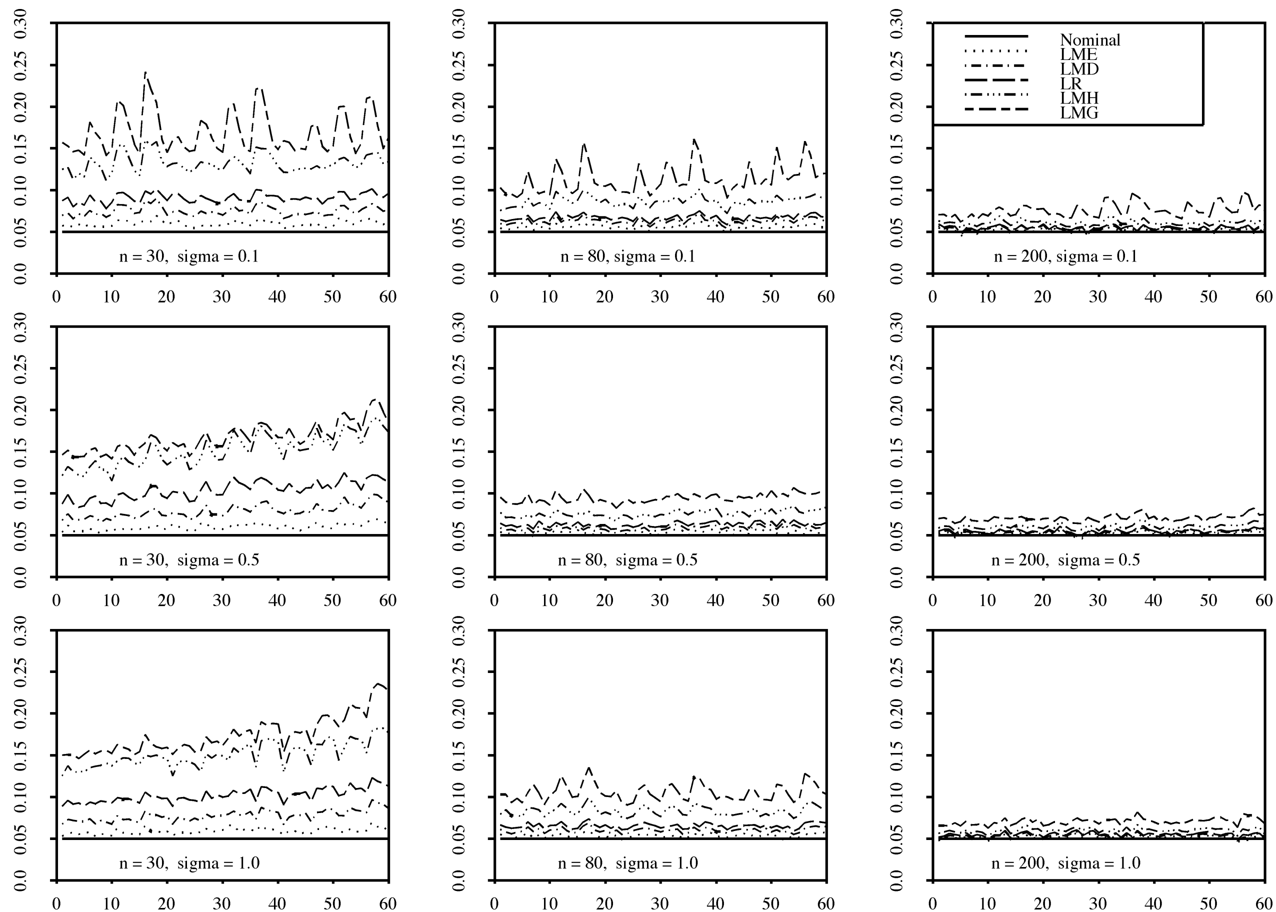

Figure 2a. Empirical sizes of LM tests for functional form, BC transformation with normal errors: $X_1 \sim U(0, 10)$, $X_2 \sim U(0, 10)$ for first row, and $X_2 \sim U(0, 5)$ for last two rows.

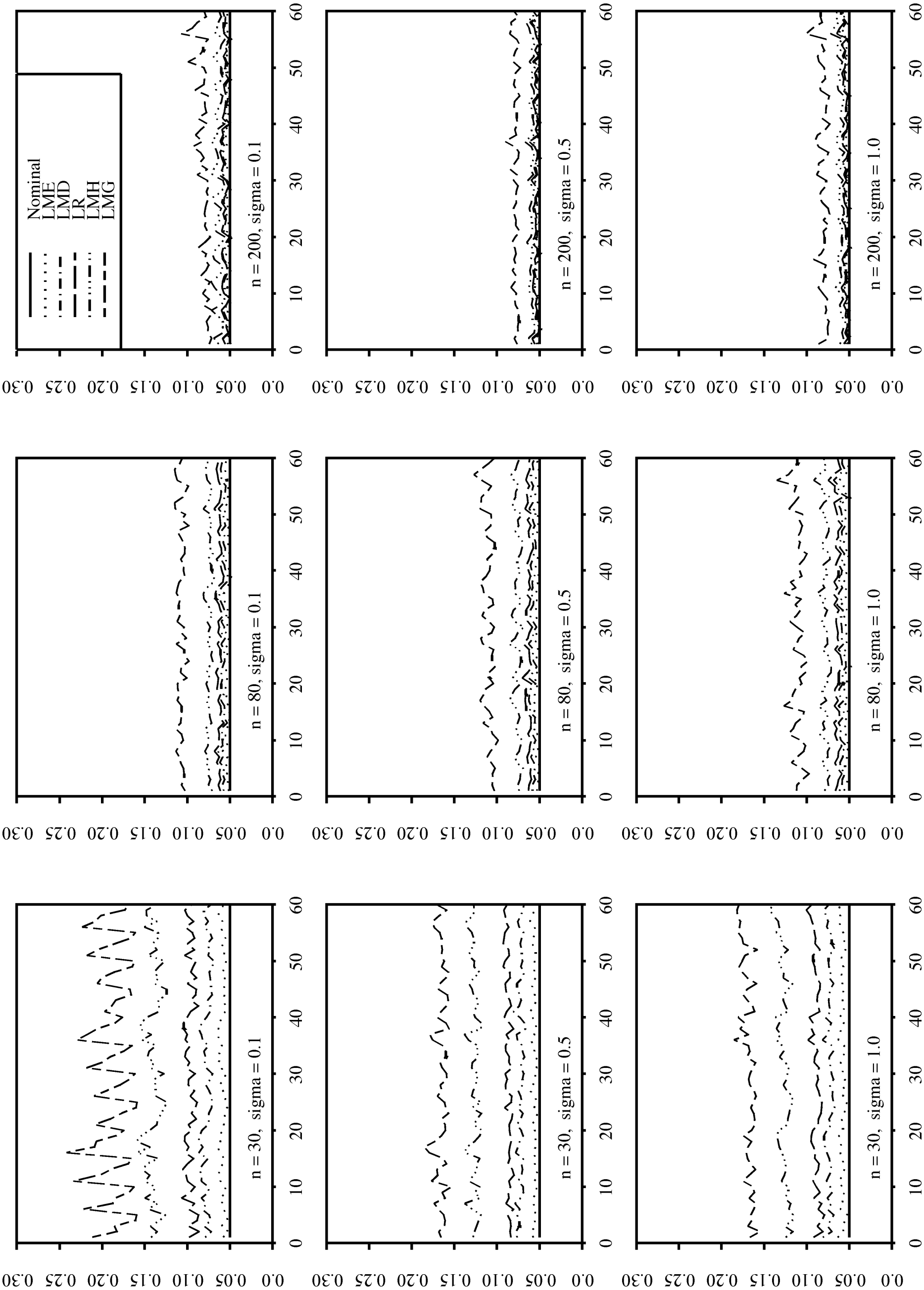

Figure 2b. Empirical sizes of LM tests for functional form, BC transformation with normal mixture: $X_1 \sim U(0, 5)$, $X_2 \sim U(0, 5)$.

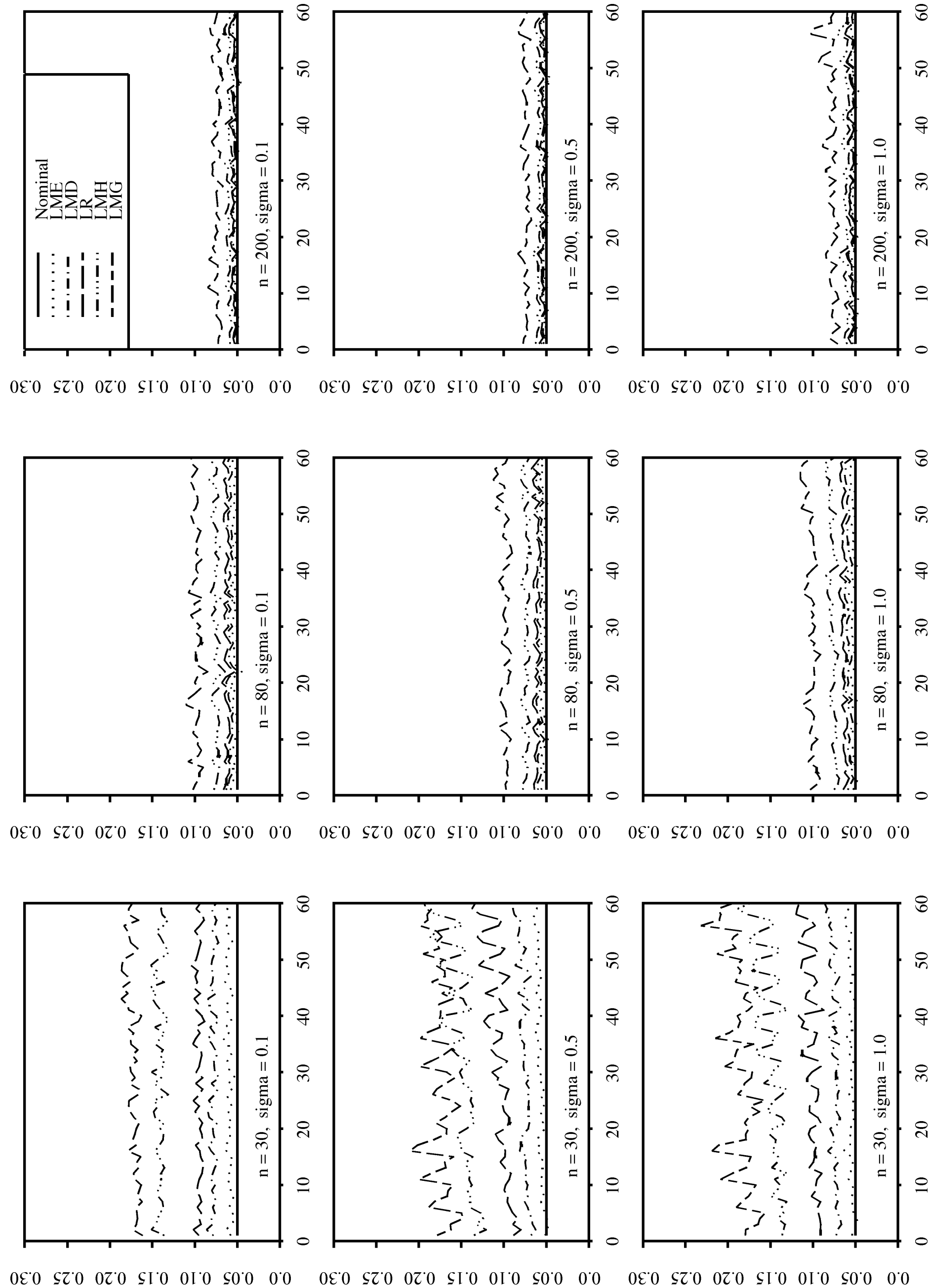

Figure 2c. Empirical sizes of LM tests for functional form, BC transformation with normal-gamma mixture, $X_1 \sim U(0, 5)$, $X_2 \sim U(0, 5)$.

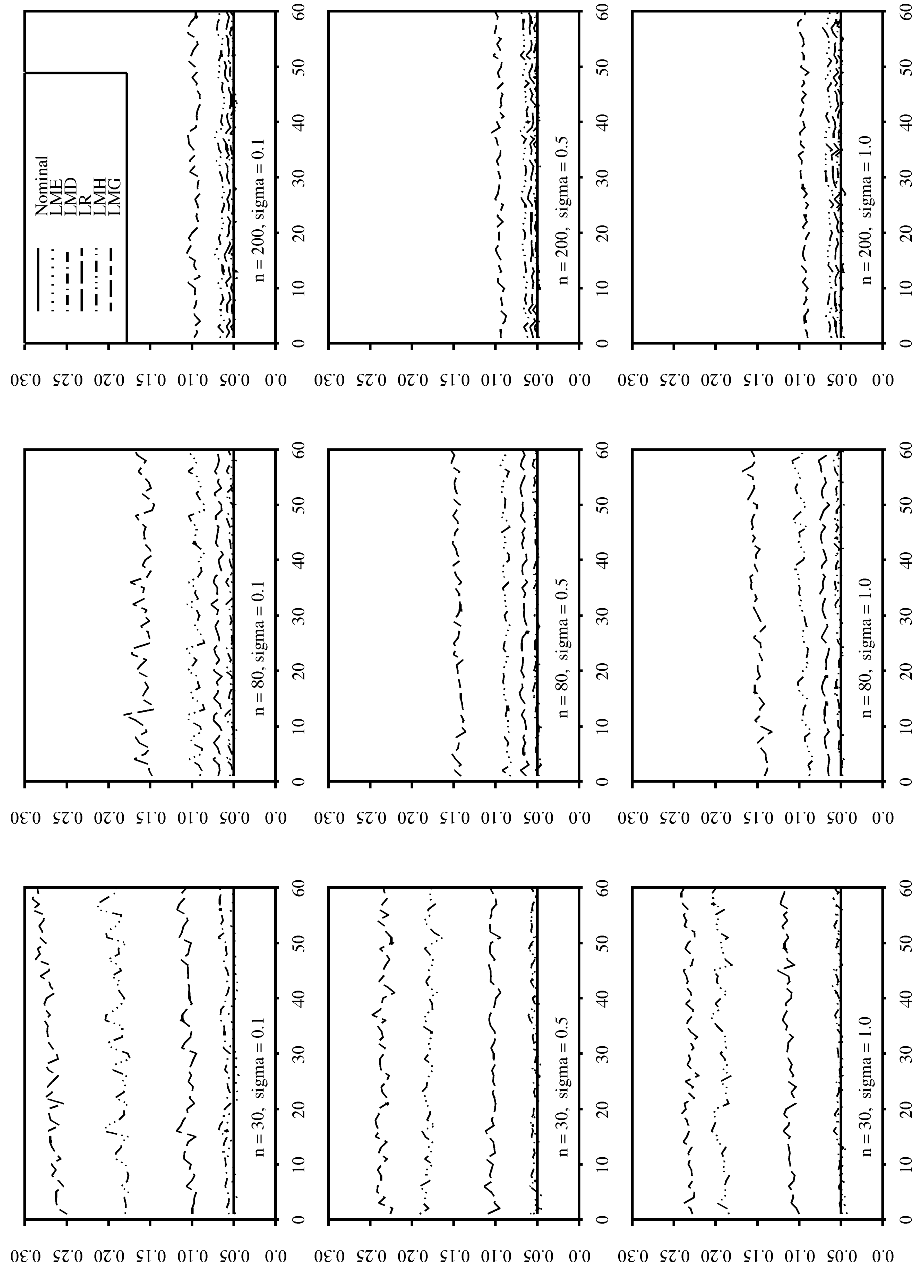

Figure 3a. Empirical sizes of the joint LM tests, BC transformation with normal errors: $X_1 \sim U(0, 10)$, $X_2 \sim U(0, 10)$ for first row, and $X_2 \sim U(0, 5)$ for last two rows.

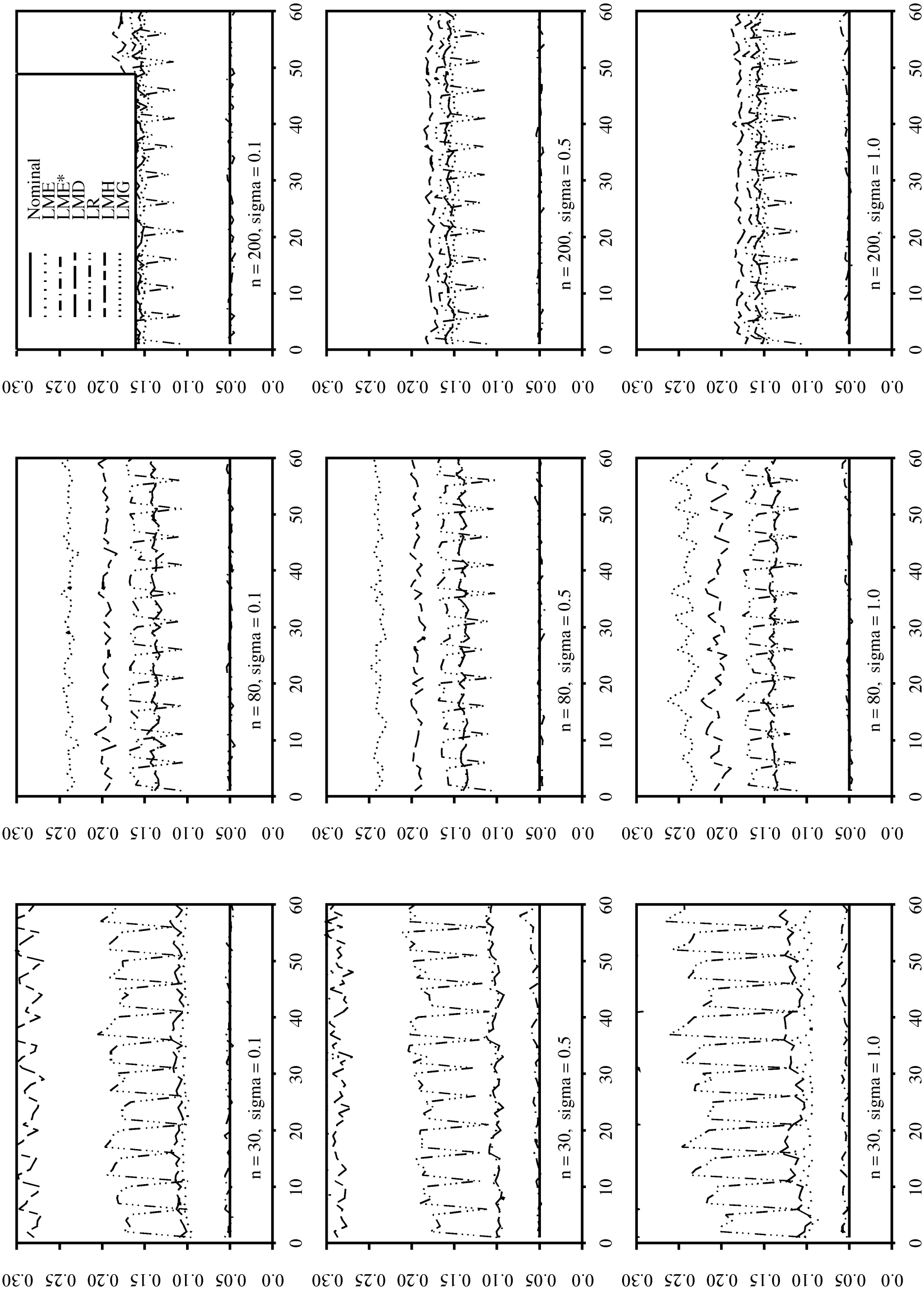

Figure 3b. Empirical sizes of the joint LM tests, BC transformation with normal mixture: $X_1 \sim U(0, 5)$, $X_2 \sim U(0, 5)$.

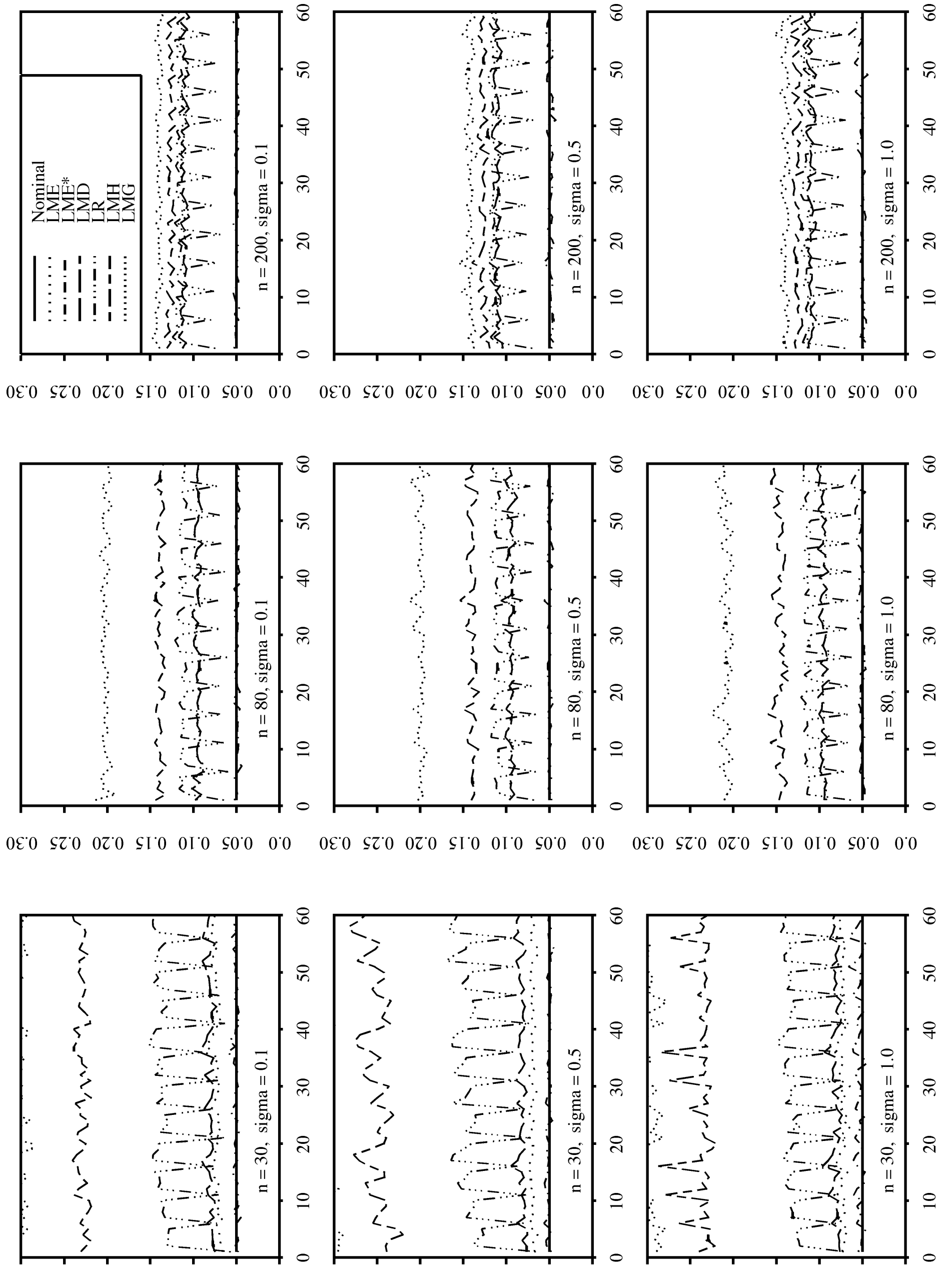

Figure 3c. Empirical sizes of the joint LM tests, BC transformation with normal-gamma mixture, $X_1 \sim U(0, 5)$, $X_2 \sim U(0, 5)$.

the case of normal errors, general observations remain the same as for testing functional form. One difference is that LMH and LMG perform notably poorer. This reinforces the necessity of using the EI-based LM test when sample size is small. Again, LMD performs reasonably well. However, unlike the EI-based LM test, LMD does not perform well uniformly for all situations. For the case of non-normal errors, LME* performs exceptionally well even when sample size is as small as 30, whereas all others perform poorly. Furthermore, the empirical sizes of the other tests apparently do not converge to the nominal level as n increases.

4.4. Power of the tests

The power of the tests is another important consideration for practitioners in choosing among the alternative tests. As the sizes of the tests can differ substantially, we use the simulated critical values to ensure fairness in making power comparison.[8] Selected results are summarized in Figure 4 with $\beta_0 = 25$, $\beta_1 = \beta_2 = 10$, and $\sigma = 1.0$. For the tests of heteroscedasticity, the null hypothesis is H_0: $\gamma_1 = \gamma_2 = 0.1$, and the alternative values are $\gamma_1 = \gamma_2 = (-0.16, -0.12, -0.08, -0.04, 0.0, 0.04, 0.07, 0.1, 0.13, 0.16, 0.2, 0.24, 0.28, 0.32, 0.36)$; for the tests of functional form, the null hypothesis is H_0: $\lambda = 0.1$ with the alternative values $\lambda = (0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17)$; and for the joint tests, the null hypothesis is H_0: $\gamma_1 = \gamma_2 = \lambda = 0.1$, and alternative values are elementwise combinations of $\gamma_1 = \gamma_2 = (-0.04, -0.02, 0.0, 0\ .02, 0.04, 0.06, 0.08, 0.10, 0.12, 0.14, 0.16, 0.18, 0.20, 0.22, 0.24)$ and $\lambda = (0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17)$, which are then indexed by the integers 1 to 15 for plotting.

From Figure 4, we see that (i) LME tests always have better or similar power compared with others, (ii) LME* for testing heteroscedasticity may have a notably lower power than the others when the sample size is small due to its robustness nature, but when the sample size increases it quickly catches up in power, (iii) The LME* for joint test performs as well as LME in terms of power and (iv) LMH and LMG may have significantly lower power than the others in the cases of functional form tests and joint tests.[9]

5. CONCLUSIONS

We provide an LM test for heteroscedasticity with the allowance of a transformation being present in the model to take care of potential non-normality of the data. With this test, one can test any specifications on the heteroscedasticity parameters so that variables attributable to heteroscedasticity can be identified. In the case of normal errors, the test compares favourably against the commonly used likelihood ratio test in both the ease of application and in the finite sample performance. The test compares also favourably against other versions of LM tests. In the case of non-normal errors, the robustified version of the EI-based LM test clearly outperforms all others.

[8]For each test, 10,000 test statistic values are generated at a given parameter configuration. The 95th percentile is calculated, which is then used in the subsequent power comparisons.

[9]Note (i) for brevity the results based on other sample sizes are not plotted and (ii) the size-adjusted tests are not feasible in practice as one does not know the true values of the model parameters.

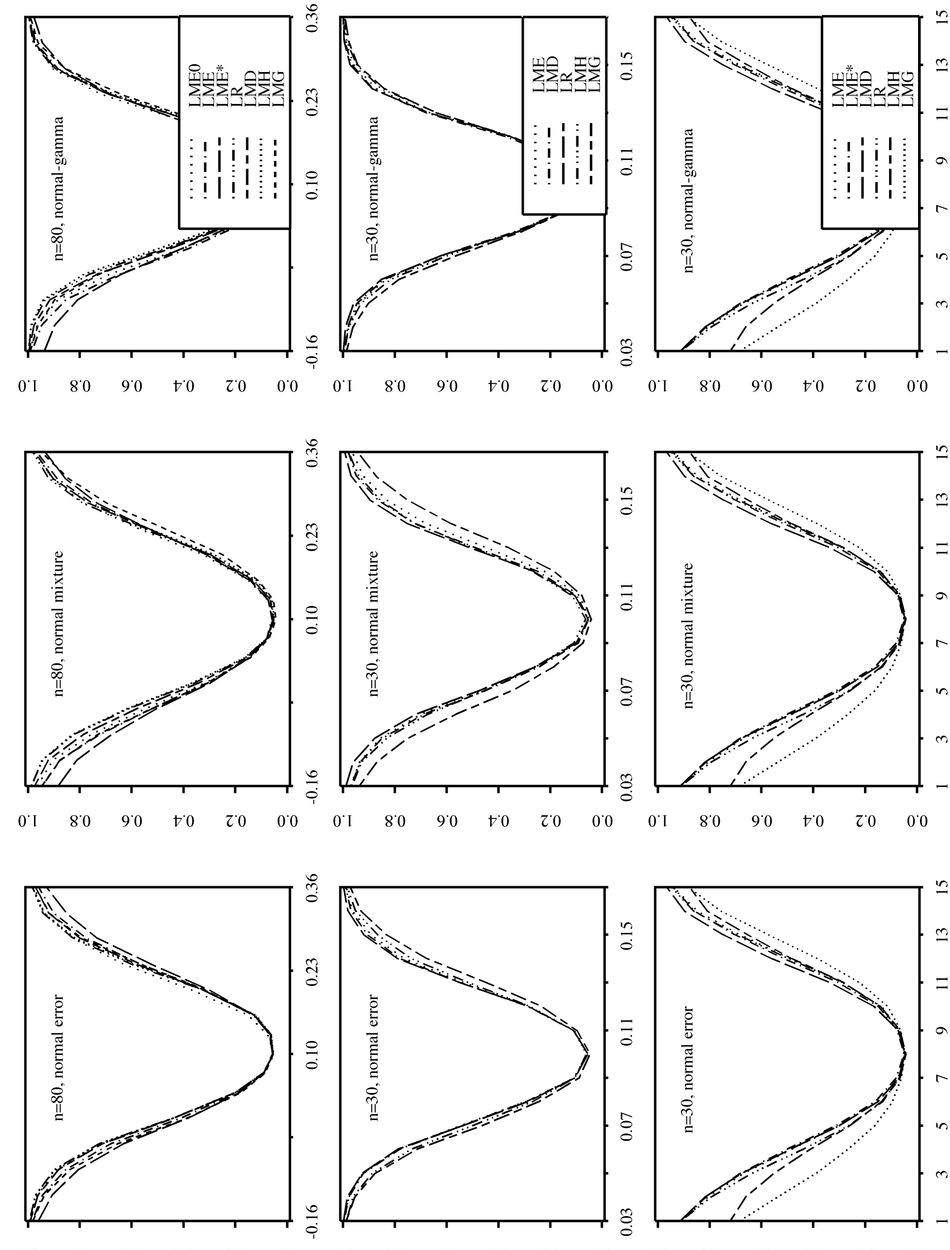

Figure 4. Power of the tests of heterscedasticity (row 1), functional form (row 2), and joint. BC transformation: $X_1, X_2 \sim U(0, 5)$.

We also provide an LM test for functional form allowing for heteroscedasticity to be present in the model. This flexibility is important as genuine heteroscedasticity often exists in the data and transformation cannot get rid of it. Monte Carlo simulations show that this test outperforms other tests. All the tests of functional form considered are quite robust against non-normality of the error distribution.

Based on the test of heteroscedasticity and the test of functional form, we provide a joint test of functional form and heteroscedasticity, and a robust version of it. Monte Carlo simulation shows excellent finite sample performance of the proposed tests, as compared with other tests. Considering the simplicity in their practical implementation and excellent small sample performance, the three proposed tests, in particular the second and the studentized versions of the first and third, should be recommended for practical applications.

ACKNOWLEDGEMENTS

We are grateful to the Editor, Pravin K. Trivedi, the Coordinating Editor, Karim Abadir, and the two anonymous referees for their very helpful comments that have led to significant improvements on an early version of this paper. We are also grateful to the comments from the seminar participants of the Far Eastern Meeting of the Econometric Society 2004 (FEMES 2004) and the Econometric Society Australasian Meeting 2004 (ESAM 2004). We gratefully acknowledge the research support from the Wharton-SMU Research Center, and the research assistance by Chenwei Li.

REFERENCES

Ali, M. M. and C. Giaccotto (1984). A study of several new and existing tests for heteroscedasticity in the general linear model. *Journal of Econometrics 26*, 355–73.

Amemiya, T. (1977). A note on a heteroscedastic model. *Journal of Econometrics 6*, 365–70.

Baltagi, B. H. (1997). Testing linear and loglinear error components regressions against Box-Cox alternatives. *Statistics and Probability Letters 33*, 63–8.

Baltagi, B. H. and D. Li (2000). Double-length regressions for the Box-Cox difference model with heteroscedasticity or autocorrelation. *Economics Letters 69*, 9–14.

Bera, A. K. and C. MacKenzie (1986). Alternative forms and properties of the score test. *Journal of Applied Statistics 13*, 13–25.

Bickel, P. J. and K. A. Doksum (1981). An analysis of transformations revisited. *Journal of the American Statistical Association 76*, 296–311.

Box, G. E. P. and D. R. Cox (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society*, Series B *26*, 211–52.

Breusch, T. S. and A. R. Pagan (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica 47*, 1287–93.

Carroll, R. J. and D. Ruppert (1984). Power transformations when fitting theoretical models to data. *Journal of the American Statistical Association 79*, 321–8.

Chen, G., R. A. Lockhart and M. A. Stephens (2002). Box-Cox transformations in linear models: Large sample theory and tests of normality (with discussion). *Canadian Journal of Statistics 30*, 177–234.

Davidson, R. and J. G. MacKinnon (1983). Small sample properties of alternative forms of the Lagrange multiplier test. *Economics Letters 12*, 269–75.

Davidson, R. and J. G. MacKinnon (1984). Model specification tests based on artificial linear regressions. *International Economic Review 25*, 485–502.

Davidson, R. and J. G. MacKinnon (1985). Testing linear and loglinear regression against Box-Cox alternatives. *Canadian Journal of Economics 18*, 499–517.

Davidson, R. and J. G. MacKinnon (1993). *Estimation and Inference in Econometrics.* Oxford: Oxford University Press.

Dufour, J.-M., L. Khalaf, J.-T. Bernard and I. Genest (2004). Simulation-based finite-sample tests for heteroscedasticity and ARCH effects. *Journal of Econometrics 122*, 317–47.

Evans, M. A. (1992). Robustness of size of tests of autocorrelation and heteroskedasticity to nonnormality. *Journal of Econometrics 51*, 7–24.

Evans, M. A. and M. L. King (1988). A further class of tests for heteroscedasticity. *Journal of Econometrics 37*, 265–76.

Farebrother, R. W. (1987). The statistical foundations of a class of parametric tests for heteroscedasticity. *Journal of Econometrics 36*, 359–68.

Glejser, H. (1969). A new test for heteroscedasticity. *Journal of the American Statistical Association 64*, 316–23.

Godfrey, L. G. (1988). *Misspecification Tests in Econometrics.* Cambridge: Cambridge University Press.

Godfrey, L. G. and M. R. Wickens (1981). Testing linear and log-linear regressions for functional form. *Review of Economic Studies 48*, 487–96.

Godfrey, L. G., C. D. Orme and J. M. C. Santos Silva (2006). Simulation-based tests for heteroskedasticity in linear regression models: Some further results. *Econometrics Journal 9*, 76–97.

Goldfeld, S. M. and R. E. Quandt (1965). Some tests for homoscedasticity. *Journal of the American Statistical Association 60*, 539–47.

Griffiths, W. E. and K. Surekha (1986). A Monte Carlo evaluation of the power of some tests for heteroscedasticity. *Journal of Econometrics 31*, 219–31.

Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica 44*, 461–65.

Hernandez, F. and R. A. Johnson (1980). The large-sample behavior of transformations to normality. *Journal of the American Statistical Association 75*, 855–61.

Kalirajan, K. P. (1989). A test for heteroscedasticity and non-normality of regression residuals. *Economics Letters 30*, 133–6.

Koenker, R. (1981). A note on studentizing a test for heteroscedasticity. *Journal of Econometrics 17*. 107–12.

Lawrance, A. J. (1987). The score statistic for regression transformation. *Biometrika 74*, 275–9.

Lahiri, K. and D. Egy (1981). Joint estimation and testing for functional form and heteroscedasticity. *Journal of Econometrics*, *15*, 299–307.

Lindsay, B. G. and B. Li (1997). On second-order optimality of the observed Fisher information. *Annals of Statistics 25*, 2172–99.

MacKinnon, J.G. and L. Magee (1990). Transforming the dependent variable in regression models. *International Economic Review 31*, 315–39.

Maekawa, K. (1988). Comparing the Wald, LR and LM tests for heteroscedasticity in a linear regression model. *Economics Letters 26*, 37–41.

Powell, J. L. (1996). Rescaled methods-of-moments estimation for the Box-Cox regression model. *Economics Letters 51*, 259–65.

Ruppert, D. and R. J. Carroll (1981). On robust tests for heteroscedasticity. *Annals of Statistics 9*, 206–10.

Tse, Y. K. (1984). Testing for linear and log-linear regression with heteroscedasticity. *Economics Letters 16*, 63–69.

Wallentin, B. and A. Agren (2002). Test of heteroscedasticity in a regression model in the presence of measurement errors. *Economics Letters 76*, 205–11.

Yang, Z. L. (2006). A modified family of power transformations. *Economics Letters 92*, 14–9.

Yang, Z. L. and T. Abeysinghe (2003). A score test for Box-Cox functional form. *Economics Letters 79*, 107–15.

Yang, Z. L. and Y. K. Tse (2006). Modelling firm-size distribution using Box-Cox heteroscedastic regression. *Journal of Applied Econometrics 21*, 641–53.

Yeo, I. K. and R. A. Johnson (2000). A new family of power transformation to improve normality or symmetry. *Biometrika 87*, 954–9.

APPENDIX A: SCORES AND OBSERVED INFORMATION

For the model with a general transformation and a general weighting function, the score function $S(\psi)$, where $\psi = \{\beta', \sigma^2, \gamma', \lambda\}'$, has the following elements:

$$
\begin{aligned}
S_\beta &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{[h(y_i,\lambda) - x_i'(\lambda)\beta]x_i(\lambda)}{\omega_i^2(\gamma)},\\
S_{\sigma^2} &= \frac{1}{2\sigma^4}\sum_{i=1}^{n}\frac{[h(y_i,\lambda) - x_i'(\lambda)\beta]^2}{\omega_i^2(\gamma)} - \frac{n}{2\sigma^2},\\
S_\gamma &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{\omega_{i\gamma}(\gamma)}{\omega_i^3(\gamma)}[h(y_i,\lambda) - x_i'(\lambda)\beta]^2 - \sum_{i=1}^{n}\frac{\omega_{i\gamma}(\gamma)}{\omega_i(\gamma)},\\
S_\lambda &= \sum_{i=1}^{n}\frac{h_{y\lambda}(y_i,\lambda)}{h_y(y_i,\lambda)} - \frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{[h(y_i,\lambda) - x_i'(\lambda)\beta][h_\lambda(y_i,\lambda) - x_{i\lambda}'(\lambda)\beta]}{\omega_i^2(\gamma)},
\end{aligned}
$$

from which the gradient matrix for use in the OPG LM test can be easily formulated. Let $e_i(\psi) = [h(y_i,\lambda) - x_i'(\lambda)\beta]/[\sigma\omega_i(\gamma)]$, and $e_{i\lambda}(\psi)$ and $e_{i\lambda\lambda}(\psi)$ be its first and second partial derivatives with respect to λ. The elements of the Hessian matrix $H(\psi) = \partial S(\psi)/\partial\psi'$ are:

$$
\begin{aligned}
H_{\beta\beta'} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{x_i(\lambda)x_i'(\lambda)}{\omega_i^2(\gamma)},\\
H_{\sigma^2\sigma^2} &= -\frac{1}{\sigma^4}\sum_{i=1}^{n}e_i^2(\psi) + \frac{n}{2\sigma^4},\\
H_{\gamma\gamma'} &= -\sum_{i=1}^{n}\left(\frac{\omega_{i\gamma\gamma'}(\gamma)}{\omega_i(\gamma)} - \frac{\omega_{i\gamma}(\gamma)\omega_{i\gamma}'(\gamma)}{\omega_i^2(\gamma)}\right) + \sum_{i=1}^{n}e_i^2(\psi)\left(\frac{\omega_{i\gamma\gamma'}(\gamma)}{\omega_i(\gamma)} - \frac{3\omega_{i\gamma}(\gamma)\omega_{i\gamma}'(\gamma)}{\omega_i^2(\gamma)}\right),\\
H_{\lambda\lambda} &= -\sum_{i=1}^{n}\left[e_{i\lambda}^2(\psi) + e_i(\psi)e_{i\lambda\lambda}(\psi)\right] + \sum_{i=1}^{n}\left(\frac{\partial^2\log h_y(y_i,\lambda)}{\partial\lambda^2}\right),\\
H_{\beta\sigma^2} &= -\frac{1}{\sigma^3}\sum_{i=1}^{n}\frac{e_i(\psi)x_i(\lambda)}{\omega_i(\gamma)},\\
H_{\beta\gamma'} &= -\frac{2}{\sigma}\sum_{i=1}^{n}\frac{e_i(\psi)x_i(\lambda)\omega_{i\gamma}'(\gamma)}{\omega_i^2(\gamma)},\\
H_{\beta\lambda} &= \frac{1}{\sigma}\sum_{i=1}^{n}\frac{e_{i\lambda}(\psi)x_i(\lambda) + e_i(\psi)x_{i\lambda}(\lambda)}{\omega_i(\gamma)},
\end{aligned}
$$

$$
\begin{aligned}
H_{\sigma^2\gamma} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{e_i^2(\psi)\omega_{i\gamma}(\gamma)}{\omega_i(\gamma)},\\
H_{\sigma^2\lambda} &= \frac{1}{\sigma^2}\sum_{i=1}^{n} e_i(\psi)e_{i\lambda}(\psi),\\
H_{\gamma\lambda} &= 2\sum_{i=1}^{n}\frac{e_i(\psi)e_{i\lambda}(\psi)\omega_{i\gamma}(\gamma)}{\omega_i(\gamma)}.
\end{aligned}
$$

Now, for the Box-Cox transformation, we have $h_y(y,\lambda) = y^{\lambda-1}$, $h_{y\lambda}(y,\lambda) = y^{\lambda-1}\log y$, $h_{y\lambda\lambda}(y,\lambda) = y^{\lambda-1}(\log y)^2$, and

$$
\begin{aligned}
h_\lambda(y,\lambda) &= \begin{cases} \frac{1}{\lambda}[1+\lambda h(y,\lambda)]\log y - \frac{1}{\lambda}h(y,\lambda), & \lambda \neq 0,\\ \frac{1}{2}(\log y)^2, & \lambda = 0,\end{cases}\\
h_{\lambda\lambda}(y,\lambda) &= \begin{cases} h_\lambda(y,\lambda)\Big(\log y - \frac{1}{\lambda}\Big) + \frac{1}{\lambda^2}[h(y,\lambda) - \log y], & \lambda \neq 0,\\ \frac{1}{3}(\log y)^3, & \lambda = 0.\end{cases}
\end{aligned}
$$

For the dual-power transformation of Yang (2006), we have $h_y(y,\lambda) = \frac{1}{2}[y^{\lambda-1} + y^{-\lambda-1}]$, $h_{y\lambda}(y,\lambda) = \frac{1}{2}(y^{\lambda-1} - y^{-\lambda-1})\log y$, $h_{y\lambda\lambda}(y,\lambda) = \frac{1}{2}(y^{\lambda-1} + y^{-\lambda-1})(\log y)^2$, and

$$
\begin{aligned}
h_\lambda(y,\lambda) &= \begin{cases} \frac{1}{2\lambda}(y^\lambda + y^{-\lambda})\log y - \frac{1}{\lambda}h(y,\lambda), & \lambda \neq 0,\\ 0, & \lambda = 0,\end{cases}\\
h_{\lambda\lambda}(y,\lambda) &= \begin{cases} h(y,\lambda)(\log y)^2 - \frac{2}{\lambda}h_\lambda(y,\lambda) & \lambda \neq 0,\\ \frac{1}{3}(\log y)^3, & \lambda = 0.\end{cases}
\end{aligned}
$$

The inverse of the dual power transformation is $y = (\lambda h + \sqrt{1+\lambda^2 h^2})^{1/\lambda}$ when $\lambda \neq 0$, and $\exp(h)$ when $\lambda = 0$, where $h = (y^\lambda - y^{-\lambda})/2\lambda$ when $\lambda \neq 0$, and $\log y$ when $\lambda = 0$.

These partial derivatives are also available for other transformations such as MacKinnon and Magee (1990), and Yeo and Johnson (2000).

APPENDIX B: SOME RELATED TEST STATISTICS

The same notations as in Appendix A are followed. Let $I(\psi)$ be the expected information matrix. If $\hat{\psi}_0$ is the constrained MLE of ψ under the constraints imposed by the null hypothesis, the LM statistic is defined as follows

$$\mathrm{LM_E} = S'(\hat{\psi}_0)I^{-1}(\hat{\psi}_0)S(\hat{\psi}_0).$$

See, for example, Godfrey (1988). In situations where the test concerns only a subvector ψ_2 of $\psi = \{\psi_1', \psi_2'\}'$, the test reduces to the following form

$$\mathrm{LM_E} = S_2'(\hat{\psi}_0)I^{22}(\hat{\psi}_0)S_2(\hat{\psi}_0),$$

where $S_2(\psi)$ denotes the relevant subvector of $S(\psi)$, and I^{22} (ψ) denotes the submatrix of $I^{-1}(\psi)$ corresponding to ψ_2.

As $I(\psi)$ may not be easily obtainable, alternative ways of estimating the information matrix have been proposed. In particular, $I(\psi)$ may be replaced by $-H(\psi)$ or the outer product of the gradient (OPG) $G(\psi)'G(\psi)$, with $G(\psi) = \{\partial\ell_i(\psi)/\partial\psi'\}$, where ℓ_i is the element of the log likelihood ℓ corresponding to the ith observation. Hence, the Hessian form and the OPG form of the LM statistic, denoted by $\mathrm{LM_H}$ and

$\mathrm{LM_G}$, respectively, can be calculated as follows:

$$\mathrm{LM_H} = -S_2'(\hat{\psi}_0)H^{22}(\hat{\psi}_0)S_2(\hat{\psi}_0),$$
$$\mathrm{LM_G} = S_2'(\hat{\psi}_0)D^{22}(\hat{\psi}_0)S_2(\hat{\psi}_0),$$

where $H^{22}(\psi)$ and $D^{22}(\psi)$ are, respectively, the submatrices of $H^{-1}(\psi)$ and $[G'(\psi)G(\psi)]^{-1}$ corresponding to ψ_2. In addition, the LM statistic can also be calculated from the double-length artificial regression proposed by Davidson and MacKinnon (1984). We denote this version of the LM statistic by $\mathrm{LM_D}$. Then, $\mathrm{LM_D}$ is the explained sum of squares of the regression of $\{e'(\hat{\psi}_0), 1_n'\}'$ on $\{-\partial e(\hat{\psi}_0)/\partial\psi', \partial(\log|\partial e(\hat{\psi}_0)/\partial y|)/\partial\psi'\}$, which has $2n$ observations and $k + p + q + 1$ regressors.

The $\mathrm{LM_D}$ statistic has been found to outperform the $\mathrm{LM_H}$ and $\mathrm{LM_G}$ statistics in finite-sample performance (Davidson and MacKinnon, 1993), and has been applied by many authors in different situations (see Tse, 1984, Baltagi and Li, 2000, among others).

Although the four forms of LM statistic are asymptotically equivalent with the same limiting chi-squared distribution under the null, $\mathrm{LM_E}$ is expected to give the best finite-sample performance.[10] This is verified empirically in our present context using Monte Carlo experiment.

The likelihood ratio (LR) test for testing, for example, heteroscedasticity is simply defined as

$$\mathrm{LR}(\gamma_0) = 2(\ell_p(\hat{\gamma}, \hat{\lambda}) - \ell_p(\gamma_0, \hat{\lambda}_c)) \tag{B.1}$$

where $\hat{\lambda}_c$ is the constrained MLE of λ at γ_0.

APPENDIX C: PROOFS OF THE THEOREMS AND COROLLARIES

Proof of Theorem 3.1: We start our derivation by first assuming that λ is known. Since λ is known, $\psi = \{\beta', \sigma^2, \gamma'\}'$, $\hat{\psi}_0 = \{\hat{\beta}'(\gamma_0, \lambda), \hat{\sigma}(\gamma_0, \lambda), \gamma_0'\}'$ and the score

$$S_\gamma(\hat{\psi}_0) = \sum_{i=1}^n \frac{\omega_{i\gamma}(\gamma_0)}{\omega_i(\gamma_0)} \frac{[h(y_i, \lambda) - x_i'(\lambda)\hat{\beta}(\gamma_0, \lambda)]^2}{\omega_i^2(\gamma_0)\hat{\sigma}^2(\gamma_0, \lambda)} - \sum_{i=1}^n \frac{\omega_{i\gamma}(\gamma_0)}{\omega_i(\gamma_0)} = D_\circ'(\gamma_0)g(\gamma_0, \lambda).$$

The elements of the expected information matrix $I(\psi)$ are: $I_{\beta\beta} = \frac{1}{\sigma^2}\mathbf{X}'(\lambda)\Omega^{-1}(\gamma)\mathbf{X}(\lambda)$, $I_{\beta\sigma^2} = 0$, $I_{\beta\gamma} = 0$, $I_{\sigma^2\sigma^2} = \frac{n}{2\sigma^4}$, $I_{\sigma^2\gamma} = \frac{1}{\sigma^2}1_n'D_\circ(\gamma)$, and $I_{\gamma\gamma} = 2D_\circ'(\gamma)D_\circ(\gamma)$. Thus, the $\gamma\gamma$-block of $I^{-1}(\psi)$ is

$$I^{\gamma\gamma} = \left(I_{\gamma\gamma} - I_{\gamma\sigma^2}I^{-1}_{\sigma^2\sigma^2}I_{\sigma^2\gamma'}\right)^{-1} = \frac{1}{2}[(D_\circ(\gamma) - 1_n\bar{D}_\circ(\gamma))'(D_\circ(\gamma) - 1_n\bar{D}_\circ(\gamma))]^{-1},$$

where $\bar{D}_\circ(\gamma) = \frac{1}{n}1_n'D_\circ(\gamma)$. These give the LM test statistic of a known λ as

$$\begin{aligned}\mathrm{LM}(\gamma_0|\lambda) &= \frac{1}{2}g'(\gamma_0, \lambda)D_\circ(\gamma_0)[(D_\circ(\gamma_0) - 1_n\bar{D}_\circ(\gamma_0))'(D_\circ(\gamma_0) - 1_n\bar{D}_\circ(\gamma_0))]^{-1}D_\circ'(\gamma_0)g(\gamma_0, \lambda)\\ &= \frac{1}{2}g'(\gamma_0, \lambda)D(\gamma_0)[D'(\gamma_0)D(\gamma_0)]^{-1}D'(\gamma_0)g(\gamma_0, \lambda).\end{aligned}$$

The proof for the asymptotic distribution of $\mathrm{LM}(\gamma_0\,|\,\lambda)$ parallels that of Koenker (1981), except that we consider only the null distribution of $\mathrm{LM}(\gamma_0\,|\,\lambda)$. It is easy to see that $g(r_0, \lambda)$ can be decomposed into $g(r_0, \lambda) = \frac{\sigma^2}{\hat{\sigma}^2(\gamma_0,\lambda)}(v_1 - 2v_2 + v_3 + v_4)$, where $v_1 = e^2 - 1$, $v_2 = e \odot (K(\gamma_0, \lambda)e)$, $v_3 = (K(\gamma_0, \lambda)e)^2$,

[10]Bera and MacKenzie (1986) has argued for the superior small-sample performance of $\mathrm{LM_E}$ over $\mathrm{LM_H}$ and $\mathrm{LM_G}$, which has been found to be empirically supported. Also, the superior performance of $\mathrm{LM_D}$ over $\mathrm{LM_H}$ and $\mathrm{LM_G}$ in small samples has been shown in many empirical studies (see Davidson and MacKinnon, 1983, 1984). We shall show below, however, that $\mathrm{LM_D}$ is dominated by $\mathrm{LM_E}$ in tests of functional form and heteroscedasticity.

and $v_4 = (1 - \frac{\hat{\sigma}^2(\gamma_0,\lambda)}{\sigma^2})1_n$ with $K(\gamma, \lambda) = I_n - M(\gamma, \lambda)$. Under Assumptions 3.1–3.3, it is easy to prove that (i) $\sqrt{n}[D'(\gamma_0)D(\gamma_0)]^{-1}D'(\gamma_0)v_k = o_p(1)$, for $k = 2, 3, 4$; (ii) $\hat{\sigma}^2(\gamma_0, \lambda) \xrightarrow{p} \sigma^2$; and (iii) $\sqrt{n}[D'(\gamma_0)D(\gamma_0)]^{-1}D'(\gamma_0)v_1 \xrightarrow{d} N(0, 2\Sigma^{-1})$, where $\Sigma = \lim_{n\to\infty} \frac{1}{n}D'(\gamma_0)D(\gamma_0)$. It follows that $\mathrm{LM}(\gamma_0 \mid \lambda) \xrightarrow{d} \chi_q^2$ under H_0.

What is being left now is to prove that $\mathrm{LM_E}(\gamma_0 \mid \tilde{\lambda})$, the LM statistic when λ is replaced by $\tilde{\lambda}$, is asymptotically equivalent to $\mathrm{LM_E}(\gamma_0 \mid \lambda)$. Under Assumption 3.2, it is sufficient to show that $\frac{1}{\sqrt{n}}D'(\gamma_0)[g(\gamma_0, \tilde{\lambda}) - g(\gamma_0, \lambda)] \xrightarrow{p} 0$. By the mean value theorem, we have

$$\frac{1}{\sqrt{n}}D'(\gamma_0)[g(\gamma_0, \tilde{\lambda}) - g(\gamma_0, \lambda)] = \frac{1}{\sqrt{n}}D'(\gamma_0)g_\lambda(\gamma_0, \lambda^*)(\tilde{\lambda} - \lambda),$$

where λ^* lies between $\tilde{\lambda}$ and λ. As $\tilde{\lambda} \xrightarrow{p} \lambda$, $\lambda^* \xrightarrow{p} \lambda$. Now, as $\frac{1}{\sqrt{n}}D'(\gamma_0)g_\lambda(\gamma_0, \ell)$ is bounded in probability uniformly in ℓ in a neighborhood of λ, we have $\frac{1}{\sqrt{n}}D'(\gamma_0)g_\lambda(\gamma_0, \lambda^*) = O_p(1)$. The result of the theorem thus follows. □

Proof of Corollary 3.1: As $\omega_i(\gamma) = \omega(v_i'\gamma)$, it must be that $\omega_{i\gamma}(0) = cv_i$ for a constant c, which directly leads to equation (3.2). □

Proof of Corollary 3.2: The proof of Corollary 3.2 is identical to the proof of Theorem 3.1, except that under the relaxed distributional assumption, $\sqrt{n}[D'(\gamma_0)D(\gamma_0)]^{-1}D'(\gamma_0)v_1 \xrightarrow{d} N(0, (\kappa - 1)\Sigma^{-1})$, where κ is consistently estimated by $\tilde{\kappa} = 1 + \frac{1}{n}\sum_{i=1}^n g_i^2(\gamma_0, \tilde{\lambda})$. □

Proof of Theorems 3.2 and 3.3: Now, $\psi = \{\beta', \sigma^2, \gamma', \lambda'\}'$. The elements of $I(\psi)$ corresponding to β, σ^2, and γ are given in the proof of Theorem 3.1. With the addition of the λ parameter and with h being the Box-Cox power transformation, the other elements of $I(\psi)$ are: $I_{\lambda\lambda} = \mathrm{E}\,[e_\lambda'(\psi)e_\lambda(\psi)] + \mathrm{E}[e_\lambda'(\psi)e_{\lambda\lambda}(\psi)]$; $I_{\beta\lambda} = -\frac{1}{\sigma}\mathbf{X}'(\lambda)\Omega^{-\frac{1}{2}}(\gamma)\mathrm{E}[e_\lambda(\psi)]$; $I_{\sigma^2\lambda} = -\frac{1}{\sigma^2}\mathrm{E}[e'(\psi)e_\lambda(\psi)]$; and $I_{\gamma\lambda} = -2D_\circ'(\gamma)\,\mathrm{E}\,[e(\psi) \odot e_\lambda(\psi)]$. These give the (γ, λ)-block and the λ-element of $I^{-1}(\psi)$ respectively as

$$\begin{pmatrix} 2D_\circ'(\gamma)AD_\circ(\gamma), & -2D_\circ'(\gamma)A\zeta \\ -2\zeta'AD_\circ(\gamma), & \xi'\mathbf{M}(\gamma, \lambda)\xi + \delta \end{pmatrix}^{-1}, \quad \text{and}$$

$$\left\{\xi'\mathbf{M}(\gamma, \lambda)\xi + \delta - 2\zeta'AD_\circ(\gamma)[D_\circ'(\gamma)AD_\circ(\gamma)]^{-1}D_\circ'(\gamma)A\zeta\right\}^{-1},$$

where $\xi = \mathrm{E}\,[e_\lambda(\psi)]$, $\zeta = \mathrm{E}\,[e(\psi) \odot e_\lambda(\psi)]$, and $\delta = \sum_{i=1}^n\{\mathrm{Var}[e_{i\lambda}(\psi)] + \mathrm{E}[e_i(\psi)e_{i\lambda\lambda}(\psi)]\} - \frac{2}{n}(1_n'\zeta)^2$. The former corresponds to the middle term of (3.6), and the latter corresponds to the denominator of (3.5). However, the three quantities ξ, ζ and δ do not possess explicit expressions in general. Thus, some approximations are desirable.

From the basic properties of the Box-Cox power transformation given at the end of Appendix A, we see that in order to obtain approximations to ξ, ζ and δ, one only needs to approximate $\log y_i$ when $\lambda \neq 0$. Using the expansion (3.4) with $k = 3$, we obtain

$$\begin{aligned}
\mathrm{E}[e_{i\lambda}(\psi)] &= \left(\frac{\theta_i}{2\lambda} + \frac{\phi_i}{\lambda\theta_i} + \frac{\theta_i^3}{\lambda} - \frac{\eta_i}{\lambda\sigma\omega_i(\gamma)} - \frac{x_{i\lambda}'(\lambda)\beta}{\sigma\omega_i(\gamma_0)}\right) + O\left(\theta_i^4\right),\\
\mathrm{E}[e_i(\psi)e_{i\lambda}(\psi)] &= \frac{1}{\lambda}\left(\phi_i - \frac{1}{2}\theta_i^2\right) + O\left(\theta_i^4\right),\\
\mathrm{Var}[e_{i\lambda}(\psi)] &= \frac{1}{\lambda^2}\left(\frac{1}{2}\theta_i^2 - \phi_i\theta_i^2 + \phi_i^2\right) + O\left(\theta_i^4\right),\\
\mathrm{E}[e_i(\psi)e_{i\lambda\lambda}(\psi)] &= \frac{1}{\lambda^2}\left(\theta_i^2 - \phi_i\theta_i^2 + \phi_i^2\right) + O\left(\theta_i^4\right),
\end{aligned}$$

for $i = 1, \ldots, n$. The first expression gives an approximation to ξ after removing the fourth term as it is absorbed by the $M(\gamma, \lambda)$ matrix, the second expression gives an approximation to ζ, and the last three expressions together give an approximation to δ. When $\lambda = 0$, exact expressions for δ, ξ and ζ follow directly from the calculations using $\log y_i = \eta_i + \sigma\omega_i(\gamma)e_i$, or from finding the limits of the above quantities when λ approaches zero. Finally, Assumptions 3.2 and 3.3 ensure that the denominator of (3.5) and the middle term of (3.6) exist for all n. This, together with Assumption 3.1 and the normality of the errors, leads to the asymptotic normal or chi-square distribution for Theorems 3.2 and 3.3, respectively. □