Volume 178 Part 3 January 2014 ISSN 0304-4076

JOURNAL OF Econometrics

CONTENTS

CONTENTS CONTINUED ON BACK COVER

Journal of Econometrics 178 (2014) 484–494

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Testing for heteroskedasticity in fixed effects models

Ted Juhl[a,*], Walter Sosa-Escudero[b,c]

[a] University of Kansas, Department of Economics, 415 Snow Hall, Lawrence, KS 66045, United States
[b] Department of Economics, Universidad de San Andres, Argentina
[c] CONICET, Argentina

ARTICLE INFO

Article history:
Received 1 March 2011
Received in revised form
25 April 2013
Accepted 18 July 2013
Available online 6 August 2013

JEL classification:
C12
C23

Keywords:
Heteroskedasticity
Panel data

ABSTRACT

We derive tests for heteroskedasticity after fixed effects estimation of linear panel models. The asymptotic results are based on a 'large N–fixed T' framework, where the incidental parameters problem is bypassed by utilizing a (pseudo) likelihood function conditional on the sufficient statistic for these parameters. A simple 'studentization' produces distribution free tests that can easily be implemented using an artificial regression based on residuals after fixed effects estimation. A Monte Carlo exploration suggests that the tests perform well in small samples such as those encountered in practice.

1. Introduction

Simple linear panel models share many characteristics with standard linear regressions when, as is typical in Economics, the cross-sectional variability is considerably larger than the temporal variation. Consequently, cross-sectional heteroskedasticity is the rule rather than the exception in most panels, and its presence invalidates standard inference, or calls for more efficient estimation strategies.

A recent line of research has successfully produced tests for the random effects case. See, for example, Montes-Rojas and Sosa-Escudero (2011), Baltagi et al. (2010), Baltagi et al. (2006), and Lejeune (2006). Nevertheless, the vast majority of the applied research focuses on fixed effects strategies, which, unlike simple random effects treatments, help bypass the biases due to the presence of unobserved heterogeneities possibly correlated with observed covariates.

Even though some results translate easily from the cross-sectional domain to panels, fixed effects estimation usually requires some care to handle the well known incidental parameters problem, especially when the relevant asymptotic approximations should be based on the cross-sectional dimension being much larger than the temporal dimension.

In this paper, we propose simple and readily implementable tests for heteroskedasticity after fixed effects estimation, whose asymptotic properties require only the cross-sectional dimension to grow to infinity while holding the temporal dimension fixed. The tests are robust to alternative distributional assumptions. As in Inoue and Solon (2006), we derive our tests using a conditional (pseudo) likelihood function based on a sufficient statistic for the incidental parameters. The final tests are analogous to Koenker's (1981) robustified version of the classic Breusch and Pagan (1979) procedure, and can be implemented by simple artificial regressions using residuals after OLS fixed effects estimation. In addition, we develop tests that relax an assumption (known as homokurtosis) and propose another variant of these tests that is directly based on testing a theoretical moment condition. This test is implemented using an artificial regression approach proposed by Wooldridge (1990).

Panels pose an additional identification challenge with respect to the simple cross-sectional or time series case, since, in the context of the one-way model, conditional variances may vary between the cross-sectional units, within cross-sectional units, or both. This is a relevant question, since corrections aimed at guaranteeing valid inference or at improving efficiency are markedly different depending on each case. For this purpose, we propose two tests that can help distinguish among these cases.

Detecting heteroskedasticity is relevant from several perspectives. First, the presence of heteroskedastic residuals suggests potential efficiency improvements upon standard estimators and/or power gains for tests based on them. As is well known (Davidson

* Corresponding author. Tel.: +1 785 864 2849; fax: +1 785 864 5270.
E-mail addresses: juhl@ku.edu (T. Juhl), wsosa@udesa.edu.ar (W. Sosa-Escudero).

0304-4076/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jeconom.2013.07.005

and MacKinnon, 1993, Ch. 12), non-rejections of standard significance tests are compatible with their nulls being true, or with lack of power. Consequently, better procedures may lead to more efficient estimates and more powerful tests that help practitioners interpret test results. In the panel context, the extent of these gains depends on the particular nature of heteroskedasticity. The semi-parametric adaptive strategy of Li and Stengos (1994) produces significant efficiency gains, if heteroskedasticity is known to be due to the observation specific error term only. When heteroskedasticity is in the individual specific error, a different strategy should be used, as proposed by Roy (2002). Second, the common practice of preserving standard estimators (like the within estimator) and fixing standard errors is not without costs when homoskedasticity holds, as stressed in Angrist and Pischke (2008, Ch. 8). Moreover, such a procedure may ignore potentially helpful efficiency improvements and power gains. Finally, and perhaps more importantly, the hypothesis of heteroskedasticity, and the identification of its source, may be of relevant economic concern *per-se*, when couched from the perspective of suggesting heterogeneous responses. As highlighted by Zietz (2001), heteroskedasticity is akin to particular forms of parameter heterogeneity, and tests for slope heterogeneity have power in the direction of heteroskedasticity. In particular, tests for slope heterogeneity in panel data models are conducted under the assumption of no heteroskedasticity, as in Pesaran and Yamagata (2008). The related literature on quantile regressions provides a semi-parametric strategy specifically aimed at detecting heterogeneous effects and, as shown by Koenker and Bassett (1982), relevant heterogeneous patterns are compatible with heteroskedasticity.

The paper is organized as follows. Section 2 presents the analytical framework used to derive the new tests and discusses the identification issues mentioned above. In Section 3, we suggest a modified test based on the moment condition from our theorem. Finally, Section 4 explores the small sample performance of the new tests through an extensive Monte Carlo exercise.

2. The tests

Consider a simple fixed effects model

$$y_{it} = \alpha_i + x_{it}^\top \beta + v_{it},$$

with $i = 1, \ldots, N$ and $t = 1, \ldots, T$ representing individuals and periods, respectively. The vector x_{it} contains K strictly exogenous regressors. The random variables α_i are individual specific effects, possibly correlated with x_{it}, and v_{it} is an observation specific error term. In what follows, we define quantities' group means and total means as

$$\bar{x}_{i\cdot} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$$

$$\bar{x}_{\cdot\cdot} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}.$$

We list the basic assumptions below.

Assumption 1.

$$E(v_{it}|X_i, Z_i, \alpha_i) = 0$$

where X_i and Z_i are, respectively, $T \times K$ and $T \times p$ matrices containing the T observations of x_{it} and z_{it}.

Assumption 2.

$$E\left(X_i^\top M_0 X_i\right)$$

is a full rank matrix where

$$M_0 = I_T - \frac{\iota_T \iota_T^\top}{T}$$

$$\iota_T = (1, 1, \ldots, 1)^\top$$

with ι_T a T vector so that M_0 subtracts group means.

Assumption 3. $E(v_{it} v_{is}|X_i, Z_i, \alpha_i) = 0$, for all $s \neq t$.

Assumption 4. The regressors x_{it} and z_{it} have finite sixth moments.

Assumption 5. Let Z_0 be the $NT \times p$ matrix of Z variables with the mean of all Z subtracted and let $D_0 = \lim_{N\to\infty} \frac{1}{N} E(Z_0^\top Z_0)$ be full rank, let $D_G = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E(Z_i^\top M_0 Z_i)$ be full rank, and let $D_B = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} E(\bar{z}_{i\cdot} - \bar{z}_{\cdot\cdot})(\bar{z}_{i\cdot} - \bar{z}_{\cdot\cdot})^\top$.

Assumption 6. Heteroskedasticity:

$$E(v_{it}^2|X_i, Z_i, \alpha_i) = \sigma_{it}^2 = \sigma_v^2\, h(z_{it}^\top \gamma),$$

where $h(.)$ is any strictly positive, twice differentiable function such that $h(0) = 1$, $h'(0) \neq 0$, and σ_v^2 is a positive constant.[1]

We will be interested in evaluating the null hypothesis $H_0 : \gamma = 0$. In this setup, z_{it} is a vector of p strictly exogenous variables that may account for heteroskedasticity, which can be taken as a subset or all of x_{it}, and z_{it} may also include variables that are not contained in x_{it} so long as the assumptions are satisfied.

Let $v_i \equiv (v_{i1}, \ldots, v_{iT})'$, and

$$\Sigma_i = E(v_i v_i^\top|X_i, Z_i, \alpha_i) = \sigma_v^2 \begin{pmatrix} h(z_{i1}^\top\gamma) & & \ldots & 0 \\ 0 & h(z_{i2}^\top\gamma) & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & \ldots & 0 & h(z_{iT}^\top\gamma) \end{pmatrix}.$$

To avoid the incidental parameters problem associated with the fixed effects estimation of α_i, we follow Inoue and Solon (2006) (see also Chamberlain, 1980) and condition on the unbiased sufficient statistic for α_i, given by $y_i^\top \Sigma_i^{-1}\iota_T / \iota_T^\top \Sigma_i^{-1}\iota_T$ (ι_T is a $T \times 1$ vector of ones), so that the conditional likelihood is

$$\ell_C(\beta, \gamma) = \sum_{i=1}^{N}\left[-\frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}\ln(\iota_T^\top \Sigma_i^{-1}\iota_T)\right] - \frac{1}{2}\sum_{i=1}^{N}\left[v_i^\top\left(\Sigma_i^{-1} - \frac{\Sigma_i^{-1}\iota_T\iota_T^\top\Sigma_i^{-1}}{\iota_T^\top\Sigma_i^{-1}\iota_T}\right)v_i\right].$$

Our test will be based on the conditional score with respect to the parameter γ obtained in the following theorem.

Theorem 2.1. *The conditional score with respect to γ evaluated at $\gamma = 0$ is given by*

$$\frac{h'(0)}{2\sigma_v^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(v_{it} - \bar{v}_{i\cdot})^2 - \left(1 - \frac{1}{T}\right)\sigma_v^2\right] z_{it}$$

where $\sigma_v^2 = E(v_{it}^2)$.

Proof. See the Appendix.

[1] The notation σ_v^2 is used to denote the conditional variance of v_{it} under the null of no heteroskedasticity.

Notice that if there is no heteroskedasticity with respect to z_{it}, the expected value of the score is zero. However, if Assumption 3 is violated, so that there is serial correlation, the expectation may no longer be zero.

There are a variety of ways to implement an LM type test based on the above score. One simple way would be to follow a strict likelihood approach, by inserting the fixed effects residuals for each of the terms involving v_{it}. In addition, a simple estimator for σ_v^2 is available based on these same fixed effects residuals. Nevertheless, heteroskedasticity tests are well known to be severely affected by violations to the Gaussian assumption (Evans, 1992 and Montes-Rojas and Sosa-Escudero, 2011). Consequently, in order to derive a distribution free test, we will follow a conditional moment-based approach (Newey, 1985; Tauchen, 1985 and White, 1987), where the conditional score obtained above is used as a valid moment condition, and a robust test is derived by properly normalizing the score without exploiting restrictions that depend on the Gaussian assumption, as in Koenker's (1981) 'studentization' procedure for the classic Breusch and Pagan (1979) LM based test.

Consider the following artificial regression:

$$\hat{w}_{it}^2 = c_1 + z_{it}^\top c_2 + \epsilon_{it}$$

where $\hat{w}_{it}$ is an estimate of the variable $w_{it} = v_{it} - \bar{v}_{i\cdot}$. That is,

$$\begin{aligned}\hat{w}_{it} &= (\widehat{v_{it} - \bar{v}_{i\cdot}}) \\ &= (y_{it} - \bar{y}_{i\cdot}) - (x_{it} - \bar{x}_{i\cdot})^\top \hat{\beta}_{FE},\end{aligned}$$

where $\hat{\beta}_{FE}$ is the usual fixed effects estimator given by

$$\begin{aligned}\hat{\beta}_{FE} &= \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_{i\cdot})(x_{it} - \bar{x}_{i\cdot})^\top\right]^{-1} \\ &\quad \times \sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_{i\cdot})(y_{it} - \bar{y}_{i\cdot}).\end{aligned}$$

The above artificial regression can be seen as the Gauss–Newton regression corresponding to $E(v_{it}^2|X_i, Z_i, \alpha_i) = \sigma_v^2 \ h(z_{it}^\top \gamma)$; see Davidson and MacKinnon (2003, pp. 264–268) for further details.

Let $LM = NTR^2$, where R^2 is the coefficient of determination of the artificial regression. Based on this artificial regression, we obtain the following result, that guarantees consistency and correct asymptotic size for our proposed test LM, for the 'large N–fixed T' framework.

Theorem 2.2. *Under $H_0 : \gamma = 0$ and given Assumptions 1–5, $E(v_{it}^2) = \sigma_v^2$, and $E(v_{it}^4|X_i, Z_i, \alpha_i) = \kappa < \infty$, LM is asymptotically distributed as central $\chi^2(p)$, as $N \to \infty$. Under the sequence of local alternatives $H_A : \gamma = N^{-1/2}\delta$ and Assumption 6, LM is asymptotically distributed as non-central $\chi^2(p)$ with non-centrality parameter given by*

$$\frac{[h'(0)]^2}{\sigma_{w^2}^2}\left(1 - \frac{2}{T}\right)^2 \delta^\top D_0 \delta + 2\frac{[h'(0)]^2}{\sigma_{w^2}^2}\left(1 - \frac{2}{T}\right)\delta^\top D_B \delta + \frac{[h'(0)]^2}{\sigma_{w^2}^2}\delta^\top D_B D_0^{-1} D_B \delta$$

where $\sigma_{w^2}^2$ is the variance of $w_{it}^2 = (v_{it} - \bar{v}_{i\cdot})^2$.

Proof. See the Appendix.

Theorem 2.2 relies on an assumption that the conditional fourth moments are constant. That is, $E(v_{it}^4|X_i, Z_i, \alpha_i) = \kappa < \infty$, a constant. Such an assumption is referred to as homokurtosis and simplifies the analysis. We will relax this assumption in Section 3. From the theorem, we see that the non-centrality parameter depends on several quantities. First, $h'(0)$ enters in each term. Moreover, the local power depends on the matrices D_0 and D_B. Notice that if $D_B = 0$, there is no between group variation in the limit, and there is no local power if $T = 2$. That is, if there is no between group variation in z_{it}, the only source of heteroskedasticity that we could hope to observe would come from within variation of z_{it}. With only $T = 2$ observations, this is impossible since we do not know the individual specific intercept, and we use group demeaned residuals. For example, for $T = 2$, we would want to employ regressands given by

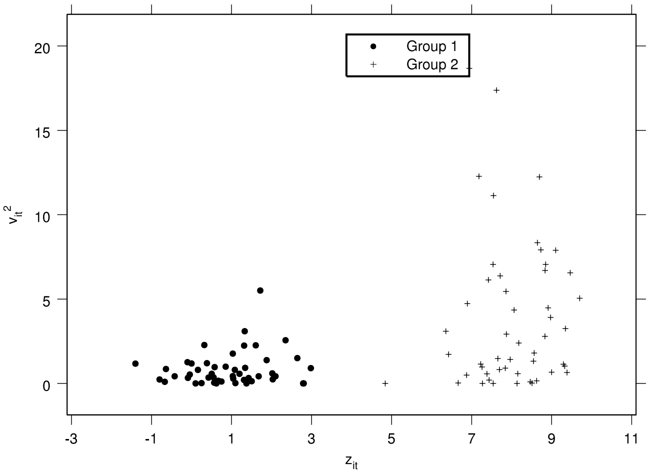

Fig. 1. Groupwise heteroskedasticity.

$$\left[v_{i1} - \left(\frac{v_{i1} + v_{i2}}{2}\right)\right]^2$$

$$\left[v_{i2} - \left(\frac{v_{i1} + v_{i2}}{2}\right)\right]^2$$

which are identical. Hence, when $T = 2$, we can only observe heteroskedasticity at the between group level.

In the context of the one-way model, heteroskedasticity may occur at the between level only, at the within level, or at both the between and within level. It is relevant to distinguish among these cases, in order to adopt appropriate strategies aimed at guaranteeing valid asymptotic inference, or at gaining efficiency by explicitly accommodating the heteroskedastic structure of the problem. For example, the recent literature dealing with the estimation of heterogeneous panels (Bresson et al., 2007; Baltagi et al., 2005; Li and Stengos, 1994 or Roy, 2002) suggests that the choice of an appropriate strategy is sensitive to specifying the correct source of heteroskedasticity (see Bresson et al., 2007).

Our LM test is based on a null hypothesis of homoskedasticity against a rather general form of heteroskedasticity induced by the vector z_{it}. Consequently, LM rejects the null if heteroskedasticity is related to any variation in z_{it}, either at the between level solely, or at both the between and within levels. Consequently, in its actual form our test can detect heteroskedasticity related to z_{it} but, if it rejects, it is silent about the specific variation driving it.

Consider the case illustrated in Fig. 1 where squared errors are plotted against the realizations of a variable z_{it}. In this case there are two individuals, with bold points corresponding to observations in one group. There appears to be a relationship between z_{it} and the squared errors, but mostly driven by the fact that the mean of z_{it} is higher in the second of the two groups. In such a case, LM would tend to reject the null of no heteroskedasticity, but is not able to let the researcher learn that this occurs only at the individual (between) level.

In order to help identify the source of heteroskedasticity, we propose a second test, based on a modified score properly centered by $E(v_{it}^2) = \sigma_i^2$ where i is the cross-sectional index. Hence, the score is proportional to

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(v_{it}-\bar{v}_{i\cdot})^2-\left(1-\frac{1}{T}\right)\sigma_i^2\right]z_{it}.$$

It is easy to see that the artificial regression is now given by

$$\hat{w}_{it}^2-\hat{\sigma}_i^2=c_1+(z_{it}-\bar{z}_{i\cdot})^\top c_2+\epsilon_{it}$$

where $\hat{\sigma}_i^2$ is different for each cross-sectional unit, with

$$\hat{\sigma}_i^2=\frac{1}{T}\sum_{t=1}^{T}\hat{w}_{it}^2,$$

and $\hat{w}_{it}$ is defined as before. The new test is denoted by $LM_g = NTR_g^2$ where R_g^2 is the coefficient of determination from the artificial regression above which now uses the regressand. The following theorem establishes its asymptotic validity.

Theorem 2.3. *Under $H_0:\gamma=0$ and given Assumptions 1–5, $E(Z_i^\top M_0 Z_i)$ full rank, $E(v_{it}^4|X_i,Z_i,\alpha_i)=\kappa<\infty$, and $E(v_{it}^2|X_i,Z_i,\alpha_i)=\sigma_i^2$, LM_g is asymptotically distributed as central $\chi^2(p)$, as $N\to\infty$. Under the sequence of local alternatives $H_A:\gamma=N^{-1/2}\delta$ and Assumption 6, LM_g is asymptotically distributed as non-central $\chi^2(p)$ with non-centrality parameter given by $\frac{h'(0)^2}{\sigma_{w^2}^2}\left(1-\frac{2}{T}\right)^2\delta' D_G\delta$ where $D_G=\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}E(Z_i^\top M_0 Z_i)$, $\sigma_{w^2}^2$ is the variance of $w_{it}^2=(v_{it}-\bar{v}_{i\cdot})^2$, and M_0 is the matrix that subtracts group means from each variable.*

Proof. See the Appendix.

This new test will also have a $\chi^2(p)$ distribution if there is no heteroskedasticity. The main difference with the LM test is that LM_g will reject only if heteroskedasticity occurs beyond the between level. The local power depends on the matrix D_G, which is the group demeaned variance of z_{it}. In practice, we propose to use both LM and LM_g to distinguish among the three relevant cases: no heteroskedasticity, heteroskedasticity at the between level only, and heteroskedasticity at one or both of the between and within level. That is, if both tests reject, this is an indication of heteroskedasticity at potentially both levels, if LM rejects and LM_g does not, this is compatible with heteroskedasticity at the between level only, and if both fail to reject, this suggests homoskedasticity related to z_{it}. Notice since D_0-D_G is positive semi-definite, the local power of LM is larger than that of LM_g when both sources of variability are present. In addition, the local power of the LM test also has two other positive terms that depend on D_B. Intuitively, when heteroskedasticity occurs beyond the between level, there is a 'power cost' in the use of LM_g, which eliminates one source of heteroskedasticity. We also develop a test that allows for heterokurtosis in Section 3.

As with the case of the tests under random effects of Montes-Rojas and Sosa-Escudero (2011), it is relevant to remark that the multiple tests can be combined in a Bonferroni approach, to produce a joint test that is compatible with both separate tests (see Savin, 1984, for further details). That is, compute both marginal tests, and reject the joint null if at least one of them lies in its rejection region, where the significance level for the marginal tests is halved, in order to guarantee that the resulting joint test has the desired asymptotic size. This is the essence of the 'multiple comparison procedure' in Bera and Jarque (1982).

3. Testing under heterokurtosis

Part of the attractiveness of the tests proposed in the last section is that they are simple to calculate using the NTR^2 representation. However, the limiting chi-square distribution depends on homokurtosis. This assumption is used to simplify the variance of the score which we described in Theorem 2.1, which in turn allows for the simple calculation of the test as NTR^2. An alternative strategy to the computation of the tests using the NTR^2 structure is to use a moment-based test interpretation and to calculate a robust version of the variance. In this way, we will not require the fourth moments to be uniform over the possible values of x_{it} that may cause heteroskedasticity.

The score from Theorem 2.1 is

$$\frac{h'(0)}{2\sigma_v^4}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(v_{it}-\bar{v}_{i\cdot})^2-\left(1-\frac{1}{T}\right)\sigma_v^2\right]z_{it}.$$

The null hypothesis is that $E(v_{it}^2|X_i,Z_i,\alpha_i)=\sigma_v^2$, a constant. If this is true, we see that the expected value of the score is zero and the variance of the score is given by

$$\frac{h'(0)^2}{4\sigma_v^8}\sum_{i=1}^{N}E\left(\sum_{t=1}^{T}\left[(v_{it}-\bar{v}_{i\cdot})^2-\left(1-\frac{1}{T}\right)\sigma_v^2\right]z_{it}\right)$$
$$\times\left(\sum_{t=1}^{T}\left[(v_{it}-\bar{v}_{i\cdot})^2-\left(1-\frac{1}{T}\right)\sigma_v^2\right]z_{it}\right)^\top.$$

An artificial regression test discussed in Wooldridge (2010) and proposed in Wooldridge (1990) that is robust to heterokurtosis is calculated by a different type of artificial regression. For our purposes, we consider the regression

$$1_{it}=c_1^\top\left[\hat{w}_{it}^2-\left(1-\frac{1}{T}\right)\hat{\sigma}_v^2\right](z_{it}-\bar{z}_{\cdot\cdot})+\epsilon_{it},$$

where 1_{it} is 1 for every observation, $\bar{z}_{\cdot\cdot}=1/NT\sum_{i=1}^{N}\sum_{t=1}^{T}z_{it}$, and $\hat{\sigma}_v^2=1/NT\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{w}_{it}^2$. The final statistic is represented as $LMS=NTR_s^2$ from the above regression, where R_s^2 is the uncentered R^2 from the artificial regression.[2]

The analogue of the LM_g test as applied to the score, say LMS_g, can be computed in a similar way, but with $\hat{\sigma}_v^2$ replaced with $\hat{\sigma}_i^2$ and $z_{it}-\bar{z}_{\cdot\cdot}$ replaced with $z_{it}-\bar{z}_{i\cdot}$ in the above equations. The distribution under the null hypothesis for both LMS and LMS_g is chi-square. The proof of this result is similar to that of Theorem 2.2, and we omit this to conserve space. The score tests will be valid even if we violate homokurtosis, so long as $E(v_{it}^4|Z_i,Z_i,\alpha_i)$ is bounded by a function with finite moments.

4. Monte Carlo experiment

In order to explore the small sample behavior of the proposed tests, we have implemented a simple Monte Carlo analysis. The setup is a linear panel model with only one explanatory variable, x_{it}:

$$y_{it}=\alpha_i+\beta x_{it}+v_{it},\quad i=1,\ldots,N,\ t=1,\ldots,T.$$

The variables α_i are generated randomly from the normal distribution with mean one and variance one, and β is set to 1. In order to allow for correlation between α_i and x_{it}, we set $x_{it}=\alpha_i+u_{it}$ where u_{it} is standard normal.

We start by exploring the size properties of our tests. We have generated 10 000 replications of the model under the null hypothesis of homoskedasticity. The error v_{it} was generated using a standard normal, student's 't'-distributions with 2 and 3 degrees of freedom, and a centered χ^2 distribution with two degrees of freedom. We note that the t-distributions do not satisfy the requirements for finite fourth moments, but these examples are included to ascertain the effects of such extremes. Sample sizes included are combinations of $N=30, 50, 100, 200, 300, 400$ and $T=5, 7, 10, 20, 30, 50$. Table 1 presents empirical rejection frequencies of our

[2] We thank a referee for suggesting the connection between the score test and the artificial regression with the uncentered R^2.

Table 1
Empirical rejection frequencies for LM NTR^2 based tests.

N	T	LM				LM_g			
		Normal	t(2)	t(3)	Chi	Normal	t(2)	t(3)	Chi
30	5	0.0623	0.0804	0.0759	0.0802	0.0745	0.0775	0.0757	0.0868
30	7	0.0570	0.0698	0.0640	0.0682	0.0675	0.0676	0.0691	0.0742
30	10	0.0570	0.0654	0.0631	0.0625	0.0628	0.0676	0.0668	0.0643
30	20	0.0546	0.0557	0.0573	0.0594	0.0560	0.0538	0.0560	0.0580
30	30	0.0494	0.0562	0.0522	0.0532	0.0535	0.0572	0.0559	0.0546
30	50	0.0508	0.0495	0.0536	0.0532	0.0531	0.0508	0.0515	0.0529
50	5	0.0642	0.0824	0.0723	0.0776	0.0819	0.0809	0.0806	0.0787
50	7	0.0615	0.0729	0.0712	0.0677	0.0694	0.0725	0.0680	0.0744
50	10	0.0594	0.0626	0.0583	0.0623	0.0636	0.0597	0.0625	0.0614
50	20	0.0529	0.0552	0.0554	0.0503	0.0527	0.0554	0.0518	0.0526
50	30	0.0538	0.0571	0.0530	0.0534	0.0536	0.0551	0.0521	0.0518
50	50	0.0459	0.0491	0.0519	0.0513	0.0516	0.0504	0.0530	0.0513
100	5	0.0617	0.0761	0.0800	0.0753	0.0773	0.0800	0.0812	0.0806
100	7	0.0608	0.0725	0.0655	0.0708	0.0717	0.0685	0.0696	0.0694
100	10	0.0573	0.0592	0.0575	0.0598	0.0608	0.0649	0.0607	0.0603
100	20	0.0521	0.0572	0.0525	0.0574	0.0530	0.0548	0.0541	0.0540
100	30	0.0504	0.0574	0.0561	0.0517	0.0527	0.0567	0.0557	0.0552
100	50	0.0518	0.0521	0.0490	0.0554	0.0517	0.0529	0.0515	0.0516
200	5	0.0638	0.0793	0.0796	0.0763	0.0811	0.0866	0.0793	0.0817
200	7	0.0601	0.0716	0.0667	0.0716	0.0677	0.0680	0.0708	0.0690
200	10	0.0576	0.0652	0.0617	0.0631	0.0630	0.0670	0.0661	0.0664
200	20	0.0541	0.0576	0.0561	0.0561	0.0574	0.0563	0.0542	0.0527
200	30	0.0524	0.0550	0.0521	0.0533	0.0564	0.0586	0.0557	0.0529
200	50	0.0528	0.0537	0.0501	0.0534	0.0509	0.0525	0.0551	0.0500
300	5	0.0646	0.0810	0.0796	0.0804	0.0807	0.0812	0.0869	0.0845
300	7	0.0557	0.0707	0.0724	0.0666	0.0671	0.0705	0.0702	0.0692
300	10	0.0589	0.0673	0.0622	0.0631	0.0626	0.0645	0.0603	0.0630
300	20	0.0534	0.0521	0.0568	0.0601	0.0571	0.0565	0.0579	0.0601
300	30	0.0522	0.0516	0.0522	0.0524	0.0537	0.0518	0.0522	0.0506
300	50	0.0497	0.0545	0.0540	0.0516	0.0522	0.0525	0.0523	0.0508
400	5	0.0674	0.0809	0.0787	0.0745	0.0774	0.0842	0.0786	0.0798
400	7	0.0630	0.0704	0.0693	0.0677	0.0712	0.0709	0.0718	0.0700
400	10	0.0558	0.0677	0.0649	0.0614	0.0552	0.0624	0.0625	0.0611
400	20	0.0566	0.0565	0.0593	0.0552	0.0574	0.0560	0.0532	0.0538
400	30	0.0486	0.0513	0.0553	0.0539	0.0524	0.0540	0.0570	0.0560
400	50	0.0480	0.0516	0.0531	0.0532	0.0501	0.0515	0.0533	0.0498

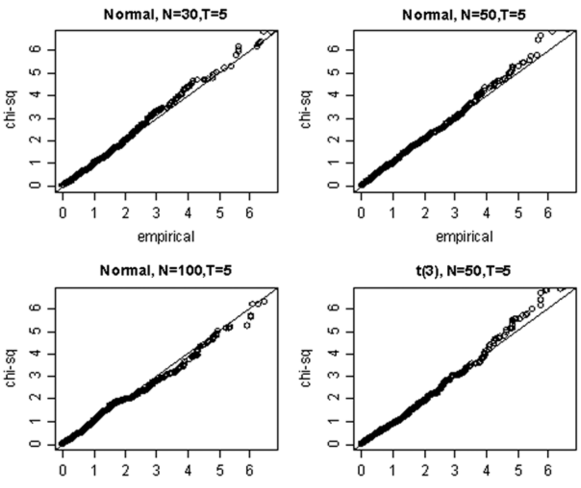

Fig. 2. Size. QQPlots.

tests, where the variable candidate to account for heteroskedasticity is x_{it}, so throughout this experiment $z_{it} = x_{it}$. Critical values correspond to the 0.95 percentile of the chi-squared distribution with 1 degree of freedom.

We separate the analysis by listing the results for the LM and LM_g tests (based on NTR^2 from the artificial regressions) in Table 1, and the score based tests using the artificial regression from Wooldridge (1990) denoted LMS and LMS_g in Table 2. In all cases, empirical sizes are very similar to the theoretical ones, even for small values like $N = 30$. Even though we require finite fourth moments, the performance of the test is robust to non-Gaussian heavy tailed distributions like the t with 2 or 3 degrees of freedom, and asymmetric distributions like the centered χ^2. In order to explore the relevance of the asymptotic chi-square approximation, Fig. 2 presents *qqplots* for some selected sample sizes and distributions for the LM test. The first three figures illustrate rejection frequencies based on normal errors, and the last one on the student's 't' case with 3 degrees of freedom, in all cases with $T = 5$, in order to highlight the performance for the 'small T' case, common in practice. All figures suggest that the asymptotic approximations perform remarkably well across most of the relevant support of the chi-squared distribution with one degree of freedom, even for a small number of cross-sectional observations ($N = 30$).

To explore power, we use two different data generating processes, each with 10 000 replications. First, we have alternatives H_1 with $v_{it} = \theta_{it}(1 + \delta u_{it})$ and $\theta_{it} \sim N(0, 1)$, and independently distributed from x_{it}, u_{it} and α_i, so this case corresponds to heteroskedasticity at the observation level. Since $x_{it} = \alpha_i + u_{it}$, we see that only the within variation of x_{it} influences the conditional variance. Alternatively, we could have specified the conditional variance as a function of x_{it} for H_1 and then made x_{it} not depend on the parameter α_i. However, since the focus of the tests is the validity under fixed effects estimation, we choose to specify H_1 as a function of u_{it} and use fixed effects estimation. The parameter δ controls the strength of heteroskedasticity. Results are presented graphically in Fig. 3, that shows the empirical power functions for LM, LM_g, LMS, and LMS_g. The graphs all increase monotonically with δ, and show that LM_g and LMS_g tests are more powerful relative to their respective counterparts LM and LMS for this particular

Table 2
Empirical rejection frequencies for LM score tests.

N	*T*	*LMS*				LMS_g			
		Normal	t(2)	t(3)	Chi	Normal	t(2)	t(3)	Chi
30	5	0.0619	0.0393	0.0478	0.0679	0.0789	0.0353	0.0520	0.0750
30	7	0.0584	0.0358	0.0442	0.0612	0.0691	0.0253	0.0443	0.0648
30	10	0.0577	0.0285	0.0375	0.0593	0.0588	0.0220	0.0415	0.0605
30	20	0.0520	0.0256	0.0401	0.0542	0.0586	0.0188	0.0356	0.0576
30	30	0.0553	0.0222	0.0338	0.0529	0.0546	0.0209	0.0346	0.0518
30	50	0.0509	0.0200	0.0366	0.0489	0.0554	0.0178	0.0330	0.0450
50	5	0.0628	0.0468	0.0530	0.0692	0.0780	0.0356	0.0511	0.0752
50	7	0.0574	0.0357	0.0477	0.0650	0.0672	0.0260	0.0430	0.0684
50	10	0.0552	0.0317	0.0373	0.0662	0.0615	0.0231	0.0374	0.0586
50	20	0.0531	0.0260	0.0381	0.0542	0.0517	0.0194	0.0350	0.0521
50	30	0.0502	0.0247	0.0389	0.0503	0.0521	0.0215	0.0370	0.0530
50	50	0.0511	0.0214	0.0343	0.0500	0.0527	0.0183	0.0309	0.0520
100	5	0.0620	0.0463	0.0543	0.0709	0.0785	0.0322	0.0534	0.0739
100	7	0.0578	0.0335	0.0454	0.0658	0.0684	0.0273	0.0433	0.0643
100	10	0.0589	0.0316	0.0425	0.0541	0.0643	0.0224	0.0385	0.0554
100	20	0.0588	0.0254	0.037	0.0544	0.0554	0.0199	0.0339	0.0586
100	30	0.0557	0.0241	0.0374	0.0547	0.0518	0.0216	0.0343	0.0539
100	50	0.0516	0.0236	0.0334	0.0533	0.0515	0.0199	0.0338	0.0479
200	5	0.0661	0.0495	0.0601	0.0756	0.0780	0.0320	0.0553	0.0799
200	7	0.0598	0.0420	0.0494	0.0624	0.0693	0.0228	0.0474	0.0674
200	10	0.0580	0.0319	0.0480	0.0595	0.0658	0.0205	0.0401	0.0582
200	20	0.0487	0.0267	0.0393	0.0525	0.0563	0.0226	0.0353	0.0538
200	30	0.0523	0.0222	0.0378	0.0526	0.0542	0.0179	0.0333	0.0527
200	50	0.0560	0.0227	0.0391	0.0503	0.0534	0.0202	0.0379	0.0539
300	5	0.0640	0.0470	0.0623	0.0770	0.0755	0.0315	0.0482	0.0746
300	7	0.0621	0.0381	0.0532	0.0611	0.0675	0.0215	0.0514	0.0654
300	10	0.0563	0.0299	0.0450	0.0577	0.0628	0.0221	0.0406	0.0618
300	20	0.0517	0.0235	0.0388	0.0523	0.0549	0.0192	0.0368	0.0566
300	30	0.0525	0.0239	0.0392	0.0545	0.0511	0.0229	0.0362	0.0569
300	50	0.0517	0.0239	0.0386	0.0547	0.0526	0.0169	0.0377	0.0524
400	5	0.0664	0.0449	0.0620	0.0751	0.0776	0.0308	0.0554	0.0773
400	7	0.0582	0.0404	0.0549	0.0681	0.0660	0.0268	0.0506	0.0679
400	10	0.0548	0.0365	0.0477	0.0620	0.0626	0.0222	0.0431	0.0628
400	20	0.0544	0.0243	0.0400	0.0501	0.0563	0.0205	0.0379	0.0492
400	30	0.0534	0.0248	0.0394	0.0531	0.0515	0.0233	0.0375	0.0527
400	50	0.0521	0.0230	0.0378	0.0540	0.0521	0.0213	0.0371	0.0570

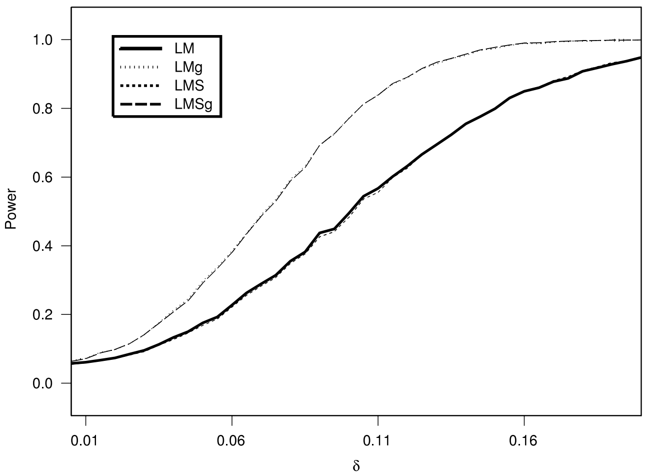

Fig. 3. $H_1 : \upsilon_{it} = \theta_{it}(1 + \delta u_{it})$.

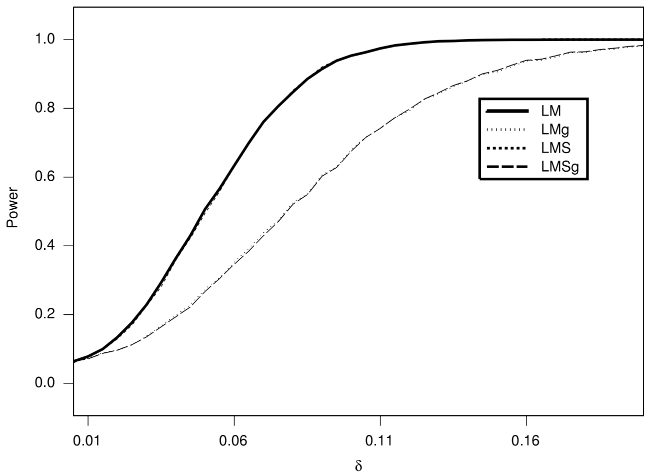

Fig. 4. $H_2 : \upsilon_{it} = \theta_{it}(1 + \delta x_{it})$.

experiment. Intuitively, LM_g and LMS_g focus on the only present source of heteroskedasticity (the heteroskedasticity arising from variation within x_{it}) which explains the power gain.

Next, we consider H_2 where $\upsilon_{it} = \theta_{it}(1 + \delta x_{it})$ with $\theta_{it} \sim N(0, 1)$. This particular experiment illustrates the local power results from Section 2. That is, for the type of alternative that is a function of the x_{it} variable itself (as opposed to u_{it} in H_1), the theorems suggested that *LM* and *LMS* should have higher local power relative to LM_g and LMS_g respectively. The power results from our experiment confirm this result in Fig. 4.

In addition, we have considered H_3, with $\upsilon_{it} = \theta_{it}(1 + \delta\alpha_i)$ and $\theta_{it} \sim N(0, 1)$, so the only variation in conditional variance comes from the different cross-sectional units from the variables α_i. Given this alternative, each group has a different variance, yet there is no within group heteroskedasticity. As expected and designed, we see that only the *LM* and *LMS* tests reject for this type of alternative, while the LM_g and LMS_g tests have correct power, equal to size. This is illustrated in Fig. 5.

As predicted by our analytical local power results, the LM_g and LMS_g tests are more powerful than the non-adjusted versions of

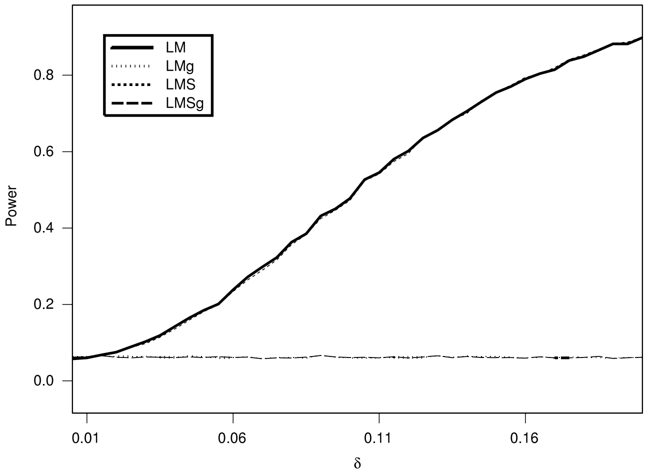

Fig. 5. $H_3 : v_{it} = \theta_{it}(1 + \delta\alpha_{it})$.

the test for the alternatives of the form H_1, since all of the heteroskedastic variation is within groups. Moreover, the *LM* and *LMS* versions of the test are more powerful when the heteroskedastic variation is both within and between groups.

Finally, we explore the consequences of heterokurtosis on the proposed tests. To this end, consider the data generating process for x_{it} and v_{it} where v_{it} is drawn from a t-distribution with ν degrees of freedom divided by $[\nu/(\nu-2)]^{1/2}$ so that the variance is one. The excess kurtosis is $6/(\nu-4)$ when $\nu > 4$ (and infinite for $\nu < 4$), and is related to the regressor x_{it} by $\nu_{it} = x_{it} + 8$. Due to the link between x_{it} and ν_{it}, the process exhibits conditional heterokurtosis. We generate $x_{it} = \alpha_i + u_{it}$, where α_i is again drawn from a normal distribution with mean one and variance one.

The results from the heterokurtosis experiment appear in Table 3. The score based tests *LMS* and LMS_g are slightly more robust to heterokurtosis than their counterparts *LM* and LM_g, although the differences are very small. In general, all of the tests continue to have reasonable size properties under heterokurtosis.

We also revisited the power analysis for our tests when the innovations are governed by heterokurtosis. The power curves are similar to the case of homokurtosis, and we present the results for H_1 under heterokurtosis in Fig. 6. The other cases are similar and are omitted to conserve space. The findings suggest that, although the score based tests *LMS* and LMS_g are robust to heterokurtosis, its presence does not seem to affect *LM* and LM_g significantly. Montes-Rojas and Sosa-Escudero (2011) obtain similar results for their robust tests under heterokurtosis, in the random effects case.

The results of the experiment suggest that using these tests provides a strategy for narrowing down the source of heteroskedasticity. If the *LM* or *LMS* tests reject but the LM_g or LMS_g tests do not, we can conclude that the source of heteroskedasticity is at the between level, and not within groups. If both types of tests reject, heteroskedasticity at the within level is relevant.

In general, the robustness of LM_g and LMS_g to individual level heteroskedasticity (where σ_i^2's are different) has a cost of less power in certain alternatives. However, the combination of tests is very informative toward the structure of heteroskedasticity in fixed effects models.

5. Final remarks

This paper proposes simple tests for heteroskedasticity in linear panels using residuals from fixed effects estimation. The incidental parameters problem is circumvented by using a pseudo likelihood function, conditional on a sufficient statistic for these parameters. Two types of tests are derived in order to help distinguish whether heteroskedasticity is present at the between level or some combination of the between and within levels. Both tests are distribution free and can be readily implemented after fixed effects estimation. A Monte Carlo experiment suggests that our proposed tests have a very good performance in small samples similar to the ones used in practice. There are several extensions that can be explored. Two-way models and unbalanced panels can easily be accommodated in our framework, since our tests can be implemented through simple artificial regressions. Extending our framework to handle dynamic panels or endogenous variables in general is a relevant challenge for further work.

Table 3
Size under heterokurtosis.

N	*T*	*LM*	LM_g	*LMS*	LMS_g
30	5	0.0677	0.0768	0.0562	0.0722
30	7	0.0610	0.0636	0.0513	0.057
30	10	0.0598	0.0623	0.0541	0.0562
30	20	0.0609	0.0582	0.0524	0.0532
30	30	0.0547	0.0546	0.0506	0.0485
30	50	0.0602	0.0570	0.0560	0.0532
50	5	0.0723	0.0796	0.0610	0.0728
50	7	0.0612	0.0655	0.0553	0.0601
50	10	0.0604	0.0667	0.0556	0.0574
50	20	0.0615	0.0591	0.0525	0.0559
50	30	0.0570	0.0566	0.0500	0.0519
50	50	0.0562	0.0572	0.0485	0.0520
100	5	0.0720	0.0828	0.0668	0.0788
100	7	0.0644	0.0704	0.0579	0.0656
100	10	0.0633	0.0647	0.0525	0.0592
100	20	0.0655	0.0600	0.0560	0.0560
100	30	0.0632	0.0624	0.0579	0.0583
100	50	0.0635	0.0536	0.0558	0.0496
200	5	0.0750	0.0869	0.0668	0.0812
200	7	0.0704	0.0725	0.0634	0.0682
200	10	0.0664	0.0655	0.0618	0.0629
200	20	0.0629	0.0597	0.0554	0.0541
200	30	0.0577	0.0580	0.0502	0.0542
200	50	0.0603	0.0547	0.0523	0.0500
300	5	0.0702	0.0791	0.0642	0.0760
300	7	0.0688	0.0732	0.0630	0.0686
300	10	0.0634	0.0658	0.0565	0.0610
300	20	0.0635	0.0604	0.0559	0.0554
300	30	0.0586	0.0550	0.0483	0.0503
300	50	0.0660	0.0601	0.0530	0.0535
400	5	0.0774	0.0859	0.0687	0.0819
400	7	0.0698	0.0726	0.0605	0.0687
400	10	0.0687	0.0662	0.0583	0.0637
400	20	0.0628	0.0616	0.0543	0.0562
400	30	0.0663	0.0576	0.0567	0.0521
400	50	0.0655	0.0574	0.0551	0.0530

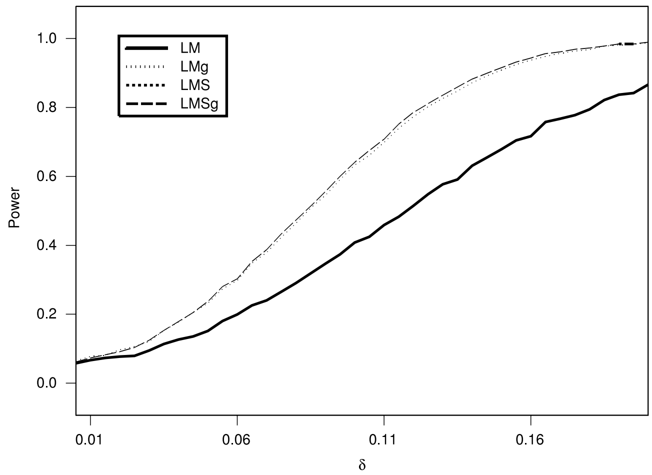

Fig. 6. Heterokurtosis $H_1 : v_{it} = \theta_{it}(1 + \delta u_{it})$.

Acknowledgments

We thank the Editor, an Associate Editor, three anonymous referees, Roger Koenker, and participants at New York Camp Econometrics for comments on earlier versions of the paper.

Appendix. Proofs

Proof of Theorem 2.1.

$$\ell_C(\beta,\gamma) = \sum_{i=1}^{N}\left[-\frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}\ln(\iota_T^\top \Sigma_i^{-1}\iota_T)\right] - \frac{1}{2}\sum_{i=1}^{N} v_i^\top \Sigma_i^{-1} v_i + \frac{1}{2}\sum_{i=1}^{N} v_i^\top\left(\frac{\Sigma_i^{-1}\iota_T\iota_T^\top\Sigma_i^{-1}}{\iota_T^\top\Sigma_i^{-1}\iota_T}\right)v_i = \sum_{i=1}^{N}(A_i + B_i + C_i + D_i).$$

In order to find a score vector that can be used to construct a test for heteroskedasticity, we take derivatives with respect to γ for each of the terms A_i, B_i, C_i, and D_i:

$$\frac{\partial vec(A_i)}{\partial vec(\gamma)^\top} = -\frac{1}{2}\frac{\partial \ln|\Sigma_i|}{\partial vec(\Sigma_i)^\top}\frac{\partial vec(\Sigma_i)}{\partial\gamma^\top}.$$

From Magnus and Nuedecker (1999), we have

$$\frac{\partial|\Sigma_i|}{\partial vec(\Sigma_i)^\top} = |\Sigma_i| vec(\Sigma_i^{-1})^\top$$

so that

$$\frac{\partial \ln|\Sigma_i|}{\partial vec(\Sigma_i)^\top} = vec(\Sigma_i^{-1})^\top.$$

Moreover, it is straightforward to see that

$$\frac{\partial vec(\Sigma_i)}{\partial\gamma^\top} = \begin{pmatrix} h'(\gamma^\top z_{i1})z_{i1}^\top \otimes e_1 \\ h'(\gamma^\top z_{i2})z_{i2}^\top \otimes e_2 \\ \vdots \\ h'(\gamma^\top z_{iT})z_{iT}^\top \otimes e_T \end{pmatrix}$$

where e_j is a $T \times 1$ vector with a one in the jth entry and zeros elsewhere. Evaluating these derivatives at $\gamma = 0$ (under the null hypothesis where $h(0) = \sigma_v^2$) gives

$$\left.\frac{\partial vec(A_i)}{\partial vec(\gamma)^\top}\right|_{\gamma=0} = -\frac{1}{2\sigma_v^2} vec(I_T)^\top \omega_i$$

where

$$\omega_i = h'(0)\begin{pmatrix} z_{i1}^\top \otimes e_1 \\ z_{i2}^\top \otimes e_2 \\ \vdots \\ z_{iT}^\top \otimes e_T \end{pmatrix}$$

$$\frac{\partial vec(B_i)}{\partial\gamma^\top} = -\frac{1}{2}\frac{\partial \ln(\iota_T^\top\Sigma_i^{-1}\iota_T)}{\partial vec(\Sigma_i^{-1})^\top}\frac{\partial vec(\Sigma_i^{-1})}{\partial vec(\Sigma_i)^\top}\frac{\partial vec(\Sigma_i)}{\partial\gamma^\top}.$$

We have

$$\frac{\partial \iota_T^\top\Sigma_i^{-1}\iota_T}{\partial vec(\Sigma_i^{-1})^\top} = \iota_T^\top \otimes \iota_T^\top.$$

Similarly, we have

$$\frac{\partial vec(\Sigma_i^{-1})}{\partial vec(\Sigma_i)^\top} = -\Sigma_i^{-1}\otimes\Sigma_i^{-1}$$

and the derivative

$$\frac{\partial vec(\Sigma_i)}{\partial\gamma^\top}$$

is as before. Combining these derivatives and evaluating at $\gamma = 0$ results in

$$\left.\frac{\partial vec(B_i)}{\partial\gamma^\top}\right|_{\gamma=0} = \frac{1}{2T\sigma_v^2}(\iota_T^\top \otimes \iota_T^\top)\omega_i.$$

Next

$$\frac{\partial vec(C_i)}{\partial\gamma^\top} = -\frac{1}{2}\frac{\partial v_i^\top\Sigma_i^{-1}v_i}{\partial vec(\Sigma_i^{-1})^\top}\frac{\partial vec(\Sigma_i^{-1})}{\partial vec(\Sigma_i)^\top}\frac{\partial vec(\Sigma_i)}{\partial\gamma^\top} = -\frac{1}{2}(v_i\otimes v_i)(-\Sigma_i^{-1}\otimes\Sigma_i^{-1})\frac{\partial vec(\Sigma_i)}{\partial\gamma^\top}.$$

Evaluating under the null, we have

$$\left.\frac{\partial vec(C_i)}{\partial\gamma^\top}\right|_{\gamma=0} = \frac{1}{2\sigma_v^4}(v_i\otimes v_i)\omega_i.$$

Now for D_i, let

$$F(\gamma) = \frac{I_T}{\iota_T^\top\Sigma_i^{-1}\iota_T}$$

$$G(\gamma) = \Sigma_i^{-1}\iota_T\iota_T^\top\Sigma_i^{-1}.$$

Using Magnus (2010), we have

$$\frac{\partial vec\,(F(\gamma)G(\gamma))}{\partial\gamma} = (G(\gamma)^\top \otimes I_T)\frac{\partial vec(F(\gamma))}{\partial\gamma} + (I_T\otimes F(\gamma))\frac{\partial vec(G(\gamma))}{\partial\gamma}.$$

In addition,

$$\frac{\partial vec(F(\gamma))}{\partial\gamma^\top} = vec(I_T)\frac{\partial(\iota_T^\top\Sigma_i^{-1}\iota_T)^{-1}}{\partial\iota_T^\top\Sigma_i^{-1}\iota_T}\frac{\partial\iota_T^\top\Sigma_i^{-1}\iota_T}{\partial vec(\Sigma_i^{-1})^\top}\frac{\partial vec(\Sigma_i^{-1})}{\partial\gamma} = -\frac{1}{(\iota_T^\top\Sigma_i^{-1}\iota_T)^2}(\iota_T^\top\otimes\iota_T^\top)(-\Sigma_i^{-1}\otimes\Sigma_i^{-1})\frac{\partial vec(\Sigma_i)}{\partial\gamma}.$$

For the derivative of $G(\gamma)$, we have

$$d(\Sigma_i^{-1}\iota_T\iota_T^\top\Sigma_i^{-1}) = d(\Sigma_i^{-1}\iota_T\iota_T^\top)\Sigma_i^{-1} + \Sigma_i^{-1}\iota_T\iota_T^\top d(\Sigma_i^{-1});$$

hence

$$vec\,d(\Sigma_i^{-1}\iota_T\iota_T^\top\Sigma_i^{-1}) = vec\left[d(\Sigma_i^{-1}\iota_T\iota_T^\top)\Sigma_i^{-1} + \Sigma_i^{-1}\iota_T\iota_T^\top d(\Sigma_i^{-1})\right] = (\Sigma_i^{-1}\iota_T\iota_T^\top\otimes I_T)d\,vec(\Sigma_i^{-1}) + (I_T\otimes\Sigma_i^{-1}\iota_T\iota_T^\top)d\,vec(\Sigma_i^{-1}).$$

This implies that

$$\frac{\partial vec(\Sigma_i^{-1}\iota_T\iota_T^\top\Sigma_i^{-1})}{\partial vec(\Sigma_i^{-1})^\top} = (\Sigma_i^{-1}\iota_T\iota_T^\top\otimes I_T) + (I_T\otimes\Sigma_i^{-1}\iota_T\iota_T^\top),$$

consequently

$$\frac{\partial vec(G(\gamma))}{\partial\gamma^\top} = \frac{\partial vec(\Sigma_i^{-1}\iota_T\iota_T^\top\Sigma_i^{-1})}{\partial vec(\Sigma_i^{-1})^\top}\frac{\partial vec(\Sigma_i^{-1})}{\partial vec(\Sigma_i)^\top}\frac{\partial vec(\Sigma_i)}{\partial\gamma^\top} = \left[(\Sigma_i^{-1}\iota_T\iota_T^\top\otimes I_T) + (I_T\otimes\Sigma_i^{-1}\iota_T\iota_T^\top)\right]\times(-\Sigma_i^{-1}\otimes\Sigma_i^{-1})\frac{\partial\Sigma_i}{\partial\gamma^\top}.$$

We can state the derivative under the null of $\gamma = 0$ as

$$\left.\frac{\partial vec\left(\frac{\Sigma_i^{-1}\iota_T\iota_T^\top\Sigma_i^{-1}}{\iota_T^\top\Sigma_i^{-1}\iota_T}\right)}{\partial\gamma^\top}\right|_{\gamma=0}$$
$$= \left.\left[(G(\gamma)^\top \otimes I_T)\frac{\partial vec(F(\gamma))}{\partial\gamma} + (I_T \otimes F(\gamma))\frac{\partial vec(G(\gamma))}{\partial\gamma}\right]\right|_{\gamma=0}$$
$$= \frac{1}{T^2\sigma_v^4}(\iota_T\iota_T^\top \otimes I_T)vec(I_T)(\iota_T^\top \otimes \iota_T^\top)\omega_i$$
$$- \left[\left(\iota_T\iota_T^\top \otimes \frac{I_T}{T\sigma_v^4}\right) + \left(I_T \otimes \frac{\iota_T\iota_T^\top}{T\sigma_v^4}\right)\right]\omega_i.$$

Using these results, we have

$$\frac{\partial D_i}{\gamma} = \frac{1}{2T^2\sigma_v^4}(v_i^\top\iota_T\iota_T^\top v_i)(\iota_T^\top \otimes \iota_T^\top)\omega_i$$
$$- \frac{1}{2T\sigma_v^4}\left[\left(v_i\iota_T\iota_T^\top \otimes v_i\right) + \left(v_i \otimes v_i\iota_T\iota_T^\top\right)\right]\omega_i$$

where the first line uses $vec(ABC) = (C^\top \otimes A)vec(B)$; in our case $A = v_i, B = I_T, C = \iota_T\iota_T^\top v_i$. Combining the terms gives

$$\left.\frac{\partial\ell_C(\beta,\gamma)}{\partial\gamma}\right|_{\gamma=0} = -\frac{1}{2}\sum_{i=1}^N\frac{1}{\sigma_v^2}vec(I_T)^\top\omega_i \tag{A.1}$$
$$+\frac{1}{2}\sum_{i=1}^N\frac{1}{T\sigma_v^2}(\iota_T^\top \otimes \iota_T^\top)\omega_i \tag{A.2}$$
$$+\frac{1}{2}\sum_{i=1}^N\frac{1}{\sigma_v^4}(v_i \otimes v_i)\omega_i \tag{A.3}$$
$$+\frac{1}{2}\sum_{i=1}^N\frac{1}{T^2\sigma_v^4}(v_i^\top\iota_T\iota_T^\top v_i)(\iota_T^\top \otimes \iota_T^\top)\omega_i \tag{A.4}$$
$$-\frac{1}{2}\sum_{i=1}^N\frac{1}{T\sigma_v^4}\left[\left(v_i\iota_T\iota_T^\top \otimes v_i\right) + \left(v_i \otimes v_i\iota_T\iota_T^\top\right)\right]\omega_i. \tag{A.5}$$

Term (A.1) simplifies to

$$-\frac{h'(0)}{2\sigma_v^4}\sum_{i=1}^N T\bar{z}_{i\cdot}\sigma_v^2$$

and (A.2) is

$$\frac{h'(0)}{2\sigma_v^4}\sum_{i=1}^N\bar{z}_{i\cdot}\sigma_v^2.$$

(A.3) becomes

$$\frac{h'(0)}{2\sigma_v^4}\sum_{i=1}^N\sum_{t=1}^T v_{it}^2 z_{it};$$

(A.4) is

$$\frac{h'(0)}{2\sigma_v^4}\sum_{i=1}^N\bar{v}_{i\cdot}^2\bar{z}_{i\cdot}.$$

The last terms are

$$-\frac{h'(0)}{\sigma_v^4}\sum_{i=1}^N\sum_{t=1}^T\bar{v}_{i\cdot}v_{it}z_{it}.$$

Combining all terms, the score for γ is given by

$$\frac{h'(0)}{2\sigma_v^4}\sum_{i=1}^N\sum_{t=1}^T\left[(v_{it}-\bar{v}_{i\cdot})^2 - \left(1-\frac{1}{T}\right)\sigma_v^2\right]z_{it},$$

which is a $p \times 1$.

Proof of Theorem 2.2. Using the Frisch–Waugh–Lovell Theorem, NTR^2 can be expressed as follows:

$$NTR^2 = NT\frac{\hat{W}^{2\top}MZ(Z^\top MZ)^{-1}Z^\top M\hat{W}^2}{\hat{W}^{2\top}M\hat{W}^2}$$

where $\hat{W}^2$ is an $NT \times 1$ vector with typical element $\hat{w}_{it}^2$, Z is an $NT \times K$ matrix where each column is formed with the observations of the kth explanatory variable sorted first by individuals and then by periods, M is the centering matrix $M \equiv I_{NT} - \iota_{NT}(\iota_{NT}^\top\iota_{NT})\iota_{NT}^\top$, and ι_{NT} is an $NT \times 1$ vector of ones. Moreover, let M_0 be defined in a similar manner using ι_T, a $T \times 1$ vector of ones. First, we deal with the estimated v_{it} using the relation

$$\hat{w}_i = M_0(y_i - X_i\hat{\beta}_{FE})$$
$$= M_0(\alpha_i\iota_T + X_i\beta + v_i - X_i\hat{\beta}_{FE})$$
$$= M_0 v_i + M_0 X_i(\beta - \hat{\beta}_{FE})$$

so that each element can be written as

$$(\widehat{v_{it}-\bar{v}_{i\cdot}}) = (v_{it}-\bar{v}_{i\cdot}) + (x_{it}-\bar{x}_{i\cdot})^\top(\beta-\hat{\beta}_{FE}).$$

Then we can write

$$\hat{W}^{2\top}MZ = \sum_{i=1}^N\sum_{t=1}^T\left[\widehat{v_{it}-\bar{v}_{i\cdot}}\right]^2(z_{it}-\bar{z}_{\cdot\cdot})$$

$$\bar{z}_{\cdot\cdot} = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T z_{it}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N\sum_{t=1}^T(\widehat{v_{it}-\bar{v}_{i\cdot}})^2(z_{it}-\bar{z}_{\cdot\cdot})$$
$$= \frac{1}{\sqrt{N}}\sum_{i=1}^N\sum_{t=1}^T(v_{it}-\bar{v}_{i\cdot})^2(z_{it}-\bar{z}_{\cdot\cdot})$$
$$+\frac{1}{\sqrt{N}}\sum_{i=1}^N\sum_{t=1}^T(\beta-\hat{\beta}_{FE})^\top(x_{it}-\bar{x}_{i\cdot})$$
$$\times(x_{it}-\bar{x}_{i\cdot})^\top(\beta-\hat{\beta}_{FE})(z_{it}-\bar{z}_{\cdot\cdot})$$
$$+\frac{2}{\sqrt{N}}\sum_{i=1}^N\sum_{t=1}^T(v_{it}-\bar{v}_{i\cdot})(x_{it}-\bar{x}_{i\cdot})^\top(\beta-\hat{\beta}_{FE})(z_{it}-\bar{z}_{\cdot\cdot}).$$

The second term is $O_p(N^{-1/2})$ since $(\beta-\hat{\beta}_{FE}) = O_p(N^{-1/2})$ and the third is also $O_p(N^{-1/2})$ since we can apply a central limit theorem so that

$$\frac{1}{\sqrt{N}}\hat{W}^{2\top}MZ = \frac{1}{\sqrt{N}}W^{2\top}MZ + o_p(1)$$

and hence

$$NTR^2 = NT\frac{W^{2\top}MZ(Z^\top MZ)^{-1}Z^\top MW^2}{W^{2\top}MW^2} + o_p(1).$$

It will be convenient to rewrite it as

$$NTR^2 = T\left(\frac{Z_0^\top W_0^2}{\sqrt{N}}\right)^\top\left[\left(\frac{W_0^{2\top}W_0^2}{N}\right)\frac{Z_0^\top Z_0}{N}\right]^{-1}\left(\frac{Z_0^\top W_0^2}{\sqrt{N}}\right) + o_p(1)$$

where $Z_0 \equiv MZ$ and $W_0^2 = MW^2$. First we will establish the asymptotic normality of

$$\frac{Z_0^\top W_0^2}{\sqrt{N}} = \sqrt{N}\frac{Z_0^\top W_0^2}{N}.$$

By the Cramer–Wold Theorem, it is equivalent to establishing the asymptotic normality of

$$\sqrt{N}c^\top \frac{Z_0^\top W_0^2}{N} = \sqrt{N}\frac{\sum_{i=1}^N c^\top Z_{0i}^\top w_{0i}^2}{N} = \sqrt{N}\frac{\sum_{i=1}^N m_i}{N}$$

for any K vector c, where Z_{0i} is a $T \times p$ matrix with rows equal to the T observations of the variables for individual i, w_{0i}^2 is a $T \times 1$ vector with element w_{0it}^2, and $m_i \equiv c' Z_{0i}^\top w_{0i}^2$. We can write

$$\begin{aligned}\frac{1}{\sqrt{N}}\sum_{i=1}^N m_i &= \frac{1}{\sqrt{N}}\sum_{i=1}^N c^\top Z_{0i}^\top \left[\left(w_i^2 - \left(1-\frac{1}{T}\right)\sigma_v^2 \iota_T\right)\right.\\ &\quad \left. + \left(\left(1-\frac{1}{T}\right)\sigma_v^2\iota_T - \bar{w}^2\iota_T\right)\right]\\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^N c^\top Z_{0i}^\top\left(w_i^2 - \left(1-\frac{1}{T}\right)\sigma_v^2\iota_T\right)\end{aligned}$$

since $\sum_{i=1}^N \iota_T z_{0i} = 0$. It is easy to check that $E[c^\top z_{0i}^\top (w_i^2 - \sigma_v^2)] = 0$ by the assumptions, and that $Var(m_i) = \sigma_{w^2}^2 c^\top D_{0i} c < \infty$ with $D_{0i} \equiv E(Z_{0i}^\top Z_{0i}) < \infty$. Then, by the Lindeberg–Feller Central Limit Theorem

$$\sqrt{N}\frac{\sum_{i=1}^N m_i}{N} = c^\top \frac{Z_0' V}{\sqrt{N}} \xrightarrow{d} N(0, \sigma_{w^2}^2 c^\top D_0 c)$$

($D_0 = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N D_{0i}$), which by the Cramer–Wold Theorem is equivalent to

$$\frac{Z_0'V}{\sqrt{N}} \xrightarrow{d} N(0, \sigma_{w^2}^2 D_0).$$

Now note that $(1/N)\ W^{2\top}MW^2 = (1/N)\sum_{i=1}^N w_{0i}^{2\top} w_{0i}^2 \xrightarrow{p} T\sigma_{w^2}^2$ by Khinchine's Law of Large Numbers under H_0 and under the homokurtosis assumption. Additionally,

$$\frac{Z^\top MZ}{N} = \frac{Z_0^\top Z_0}{N} = (1/N)\sum_{i=1}^N Z_{0i}^\top Z_{0i} \xrightarrow{p} D_0$$

by Chebychev's LLN element-by-element in the matrix $(1/N)\ Z^\top MZ$. Then

$$\left[\left(\frac{W^{2\top}MW^2}{N}\right)\left(\frac{Z_0^\top Z_0}{N}\right)\right]^{-1} \xrightarrow{p} \left(\sigma_{w^2}^2 T D_0\right)^{-1};$$

then, collecting all previous results,

$$NTR^2 \xrightarrow{d} \chi_p^2(0)$$

where $\chi_p^2(0)$ denotes a central chi-squared law with p degrees of freedom.

We derive the asymptotic distribution of the test statistic under the sequence of local alternatives $H_A : \gamma = \delta/\sqrt{N}$ where δ is a finite p-vector. Now, we have

$$\begin{aligned}E(Z^\top MW^2) &= \sum_{i=1}^N\sum_{t=1}^T E\left[z_{0it}(v_{it} - \bar{v}_{i\cdot})^2\right]\\ &= \sum_{i=1}^N\sum_{t=1}^T E\left[z_{0it}\left(v_{it}^2 - \frac{2}{T}\sum_{s=1}^T v_{it}v_{is} + \frac{1}{T^2}\sum_{s=1}^T\sum_{s'=1}^T v_{is}v_{is'}\right)\right]\\ &= \sum_{i=1}^N\sum_{t=1}^T E\left(z_{0it}\left[h(z_{it}^\top\gamma) - \frac{2}{T}h(z_{it}^\top\gamma) + \frac{1}{T^2}\sum_{s=1}^T h(z_{is}^\top\gamma)\right]\right).\end{aligned}$$

Expanding around zero gives

$$h(z_{it}^\top\gamma) = h(0) + h'(0)z_{it}^\top\gamma + \frac{1}{2}h''(c^*)(z_{it}^\top\gamma)^2$$

where $c^* \in [0, \gamma^\top z_{it}]$. Using $\gamma = \delta/\sqrt{N}$, we have

$$\begin{aligned}E\left(\frac{Z^\top MW^2}{\sqrt{N}}\right) &= \frac{1}{\sqrt{N}}\sum_{i=1}^N\sum_{t=1}^T E\left(z_{0it}\left[\left(1-\frac{2}{T}\right)\frac{h'(0)}{\sqrt{N}}z_{it}^\top\delta\right.\right.\\ &\quad\left.\left. + \frac{1}{T^2}\frac{h'(0)}{\sqrt{N}}\sum_{s=1}^T z_{is}^\top\delta\right]\right) + o(1)\\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^N\sum_{t=1}^T E\left(z_{0it}\left[\left(1-\frac{2}{T}\right)\frac{h'(0)}{\sqrt{N}}z_{it}^\top\delta\right.\right.\\ &\quad\left.\left. + \frac{h'(0)}{\sqrt{N}T}\bar{z}_{i\cdot}^\top\delta\right]\right) + o(1)\\ &= G + H + o(1).\end{aligned}$$

First, because $\sum_{i=1}^N\sum_{t=1}^T z_{0it} = 0$, we have,

$$\begin{aligned}G &= \frac{h'(0)}{N}\left(1-\frac{2}{T}\right)\sum_{i=1}^N\sum_{t=1}^T E\left[z_{0it}(z_{it} - \bar{z}_{\cdot\cdot})^\top\right]\delta\\ &= \frac{h'(0)}{N}\left(1-\frac{2}{T}\right)\sum_{i=1}^N\sum_{t=1}^T E\left(z_{0it}z_{0it}^\top\right)\delta\\ &= \frac{h'(0)}{N}\left(1-\frac{2}{T}\right)\sum_{i=1}^N E(Z_{0i}^\top Z_{0i})\delta.\end{aligned}$$

Next, we have

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T z_{0it}\bar{z}_{i\cdot}^\top &= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T (z_{it} - \bar{z}_{\cdot\cdot})\bar{z}_{i\cdot}^\top\\ &= \frac{1}{N}\sum_{i=1}^N \bar{z}_{i\cdot}\bar{z}_{i\cdot}^\top - \frac{1}{T}\sum_{t=1}^T \bar{z}_{\cdot\cdot}\bar{z}_{\cdot\cdot}^\top\\ &= \frac{1}{N}\sum_{i=1}^N (\bar{z}_{i\cdot} - \bar{z}_{\cdot\cdot})(\bar{z}_{i\cdot} - \bar{z}_{\cdot\cdot})^\top\end{aligned}$$

so that

$$H = \frac{h'(0)}{N}\sum_{i=1}^N E\left[(\bar{z}_{i\cdot} - \bar{z}_{\cdot\cdot})(\bar{z}_{i\cdot} - \bar{z}_{\cdot\cdot})^\top\right]\delta.$$

Combining these results, we see that the limiting non-centrality parameter is given by

$$\frac{[h'(0)]^2}{\sigma_{w^2}^2}\left(1-\frac{2}{T}\right)^2\delta^\top D_0\delta + 2\frac{[h'(0)]^2}{\sigma_{w^2}^2}\left(1-\frac{2}{T}\right)\delta^\top D_B\delta + \frac{[h'(0)]^2}{\sigma_{w^2}^2}\delta^\top D_B D_0^{-1} D_B\delta$$

where

$$D_B = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N E(\bar{z}_{i\cdot} - \bar{z}_{\cdot\cdot})(\bar{z}_{i\cdot} - \bar{z}_{\cdot\cdot})^\top.$$

Proof of Theorem 2.3. The proof is similar but with M replaced with $M_G = I_N \otimes M_0$ where M_0 is defined as in Assumption 2. However, for the non-centrality parameter, we have

$$\begin{aligned}E\left(\frac{1}{\sqrt{N}}Z^\top M_G W^2\right) &= \frac{1}{\sqrt{N}}\sum_{i=1}^N\sum_{t=1}^T E\left(z_{Git}\left[\left(1-\frac{2}{T}\right)\frac{h'(0)}{\sqrt{N}}z_{it}^\top\delta\right.\right.\\ &\quad\left.\left. + \frac{1}{T^2}\frac{h'(0)}{\sqrt{N}}\sum_{s=1}^T z_{is}^\top\delta\right]\right) + o(1)\end{aligned}$$

where $z_{Git} = z_{it} - \bar{z}_{i\cdot}$. Using the fact that $\sum_{t=1}^{T} z_{Git} = 0$, we have

$$E\left(\frac{1}{\sqrt{N}} Z^\top M_G W^2\right)$$
$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} E\left(z_{Git}\left[\left(1-\frac{2}{T}\right)\frac{h'(0)}{\sqrt{N}} z_{it}^\top \delta\right]\right) + o(1)$$
$$= \left(1-\frac{2}{T}\right) h'(0)\delta^\top \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} E(z_{Git} z_{Git}^\top) + o(1)$$
$$= \left(1-\frac{2}{T}\right) h'(0)\delta^\top \frac{1}{N} \sum_{i=1}^{N} E(Z_i^\top M_0 Z_i)$$

so that the non-centrality parameter becomes

$$\frac{[h'(0)]^2}{\sigma^2_{w^2}} \left(1-\frac{2}{T}\right)^2 \delta^\top D_G \delta.$$

References

Angrist, J., Pischke, J., 2008. Mostly Harmless Econometrics: An Empiricist's Companion. Oxford University Press, Oxford.
Baltagi, B.H., Bresson, G., Pirotte, A., 2006. Joint LM test for homoskedasticity in a one-way error component model. Journal of Econometrics 134, 401–417.
Baltagi, B.H., Bresson, G., Pirotte, A., 2005. Adaptive estimation of heteroskedastic error component models. Econometric Reviews 24, 39–58.
Baltagi, B.H., Jung, B.C., Song, S.H., 2010. Testing for heteroskedasticity and serial correlation in a random effects panel data model. Journal of Econometrics 154, 122–124.
Bera, A., Jarque, C., 1982. Model specification tests. a simultaneous approach. Journal of Econometrics 20, 59–82.
Breusch, T.S., Pagan, A.R., 1979. A simple test for heteroskedasticity and random coefficient variation. Econometrica 47, 1287–1294.
Bresson, G., Hsiao, C., Pirotte, A., 2007. Assessing the contribution of R&D to total factor productivity. A Bayesian approach to account for heterogeneity and heteroskedasticity. Mimeo.
Chamberlain, G., 1980. Analysis of covariance with qualitative data. Review of Economic Studies 146, 225–238.
Davidson, R., MacKinnon, J., 1993. Estimation and Inference in Econometrics. Oxford University Press, Oxford.
Davidson, R., MacKinnon, J., 2003. Econometric Theory and Methods. Oxford University Press, Oxford.
Evans, M., 1992. Robustness of size of tests of autocorrelation and heteroskedasticity to nonnormality. Journal of Econometrics 51, 7–24.
Inoue, A., Solon, G., 2006. A Portmanteau test for serially correlated errors in fixed effects models. In: Econometric Theory, vol. 22. Cambridge University Press, pp. 835–851.
Koenker, R., 1981. A note on studentizing a test for heteroskedasticity. Journal of Econometrics 17, 107–112.
Koenker, R., Bassett, G.S., 1982. Robust tests for heteroscedasticity based on regression quantiles. Econometrica 50, 43–61.
Lejeune, B., 2006. A full heteroscedastic one-way error components model allowing for unbalanced panel: pseudo-maximum likelihood estimation and specification testing. In: Baltagi, B.H. (Ed.), Panel Data Econometrics: Theoretical Contributions and Empirical Applications. Elsevier, Amsterdam, pp. 31–66 (Chapter 2).
Li, Q., Stengos, T., 1994. Adaptive estimation in the panel data error component model with heteroskedasticity of unknown form. International Economic Review 35 (4), 981–1000.
Magnus, J., Nuedecker, H., 1999. Matrix Differential Calculus with Applications in Statistics and Econometrics, second ed. Wiley, New York.
Magnus, J., 2010. On the concept of matrix derivative, manuscript, Department of Econometrics and OR, Tilburg University.
Montes-Rojas, G., Sosa-Escudero, W., 2011. Robust tests for heteroskedasticity in the one-way error components model. Journal of Econometrics 160, 300–310.
Newey, W., 1985. Maximum likelihood specification testing and conditional moment tests. Econometrica 53, 1047–1070.
Pesaran, H.M., Yamagata, T., 2008. Testing slope heterogeneity in large panels. Journal of Econometrics 142, 50–93.
Roy, N., 2002. Is adaptative estimation useful for panel models with heteroskedasticity in the individual specification error component: some monte carlo evidence. Econometric Reviews 21, 189–203.
Savin, N.E., 1984. Multiple hypothesis tsting. In: Griliches, Z., Intriligator, M.D. (Eds.), Handbook of Econometrics, vol. II. Elsevier, Amsterdam, pp. 827–879. Chapter 14.
Tauchen, G., 1985. Diagnostic testing and evaluation of maximum likelihood models. Journal of Econometrics 30, 415–443.
White, H., 1987. Specification testing in dynamic models. In: Bewley, T.F. (Ed.), Advances in Econometrics — Fifth World Congress, vol. 1. Cambridge University Press, Cambridge, pp. 1–58 (Chapter 1).
Wooldridge, J.M., 1990. A unified approach to robust, regression-based specification tests. Econometric Theory 6, 17–43.
Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data, 2nd Ed.. Cambridge University Press, Cambridge.
Zietz, J., 2001. Heteroskedasticity and neglected parameter heterogeneity. Oxford Bulletin of Economics and Statistics 63, 263–273.