The Davies Problem:
A New Test for Random Slope in the Hierarchical Linear Model

Rutger van Oest
Department of Marketing, BI Norwegian Business School, Oslo, Norway

Philip Hans Franses
Econometric Institute, Erasmus University Rotterdam, The Netherlands

Econometric Institute Report 2015-01

Abstract

Crucial inference for the hierarchical linear model concerns the null hypothesis of no random slope. We argue that the usually applied statistical test suffers from the so-called Davies problem, that is, a nuisance parameter is only identified under the alternative. We propose an easy-to-implement methodology that exploits this property. We provide the relevant critical values and demonstrate through simulations that our new methodology has better power properties.

Keywords: hierarchical linear model, random effects, slope variance, Davies problem

This version: January 5, 2015

Address for correspondence: Rutger van Oest, Department of Marketing, BI Norwegian Business School, Oslo, Nydalsveien 37, 0484 Oslo, Norway, rutger.d.oest@bi.no

1. Introduction

The hierarchical linear model is becoming increasingly popular in social sciences and business studies (Lahuis and Ferguson 2007; Oakley, Iacobucci and Duhachek 2005). It links a dependent variable with independent variables, where the data appear at multiple nested levels, that is, a macro level and a micro level. Examples include literature studies and cases within studies in a meta analysis, individuals and tasks within individuals in conjoint analysis, countries and respondents within countries in cross-national surveys, and entities (such as companies or individuals) and time within entities in longitudinal analysis.

The parameters in the hierarchical linear model need not be fixed but may vary at the macro level according to some statistical (e.g., normal) distribution. A key issue in empirical applications concerns whether the model's response coefficients, i.e., slopes, are fixed or are characterized by variance. While a random slope complicates parameter estimation and makes the model less parsimonious, erroneously ignoring it would lead to incorrect inference on the effect of the independent variable on the dependent variable. A significant random slope also implies unexplained variance that may be explained by moderators and may lead to new hypotheses of interest. It is important to be able to identify random slopes, although current tests possess generally low power (LaHuis and Ferguson 2007).

The statistical test for random slope involves testing whether the variance of its statistical distribution equals zero. However, when this variance is zero (i.e., when the random component is effectively removed), another parameter in the model (the correlation between the random slope and the random intercept) is no longer identified. This disappearance of a nuisance parameter under the null hypothesis is usually called the Davies problem (Andrews and Ploberger 1994; Davies 1987; 2002). It frequently appears in statistical and econometric models like Hidden Markov models (Carrasco 2002), smooth transition autoregressive models (Hansen 1996), the geometric lag or Koyck model (Franses and Van Oest 2007), and the GARCH model (Andrews 2001, Beg et al. 2001). The Davies problem makes the Likelihood Ratio (LR) test of random slope a test with a non-standard reference distribution, that is, critical values are not the usual chi (bar) square based critical values. This problem has been ignored in the literature on hierarchical linear models (e.g., Snijders and Bosker 2012) and therefore we do not know whether the commonly presumed standard distribution of the LR test is far off or is still approximately valid.

The current paper addresses this issue. First, we elaborate on the overlooked Davies problem and simulate the correct critical values. We do so for both the asymptotic case (i.e., when the number of "groups" at the macro level, i.e., N, is very large) and for more common cases where N is small, i.e., N = 50, 100 or 200. We show that the currently used reference distribution for slope variance is slightly too conservative asymptotically and the deviation increases when N becomes (moderately) small, the typical case in most model applications. For instance, the number of included studies in a meta analysis is often below 100, the number of respondents in a conjoint analysis is often a few hundreds at most, there are around 200 countries in the world of which most may not be included in cross-national studies, the number of participating companies in a longitudinal business-to-business study or number of kids in a longitudinal growth study may be as low as a few tens, and so on. In such cases, it is recommended to use exact (rather than asymptotic) critical values to avoid relying on an overly conservative test. We provide these critical values and correct for bias caused by the Davies problem.

Our second contribution is that we go beyond merely controlling for the Davies problem and exploit the presence of the unidentified correlation parameter to come up with a simple procedure for testing slope variance with improved power properties. The basic idea is that the correlation, the nuisance parameter, is not identified under the null hypothesis but conceptually it cannot be assumed zero if it is not zero. At the same time keeping it in the model if it is zero leads to unnecessarily low power. We combine the best of both worlds. First, we test for the data set and model at hand whether the correlation between slope and intercept can be set equal to zero. If not, we consider the LR test including the unidentified correlation parameter and use the corresponding simulated critical values. If the estimated correlation is not significantly different from zero, we set it equal to zero and run the LR test of slope variance without this correlation parameter.

The outline of our paper is as follows. In Section 2 we present the hierarchical linear model and briefly outline parameter estimation. Section 3 discusses LR based tests and elicits the key problematic issue for inference, which is widely overlooked in current applications of this model. In Section 4 we present the proper methodology for inference in the hierarchical linear model. We provide critical values and present the results of power simulation studies, highlighting that the proper method yields more power. Section 5 summarizes the results from our power analysis in a meta regression. Section 6 concludes this paper.

2. The hierarchical model

The hierarchical linear model, with random intercept $\beta_{0,i}$ and explanatory variable $x_{i,t}$ with random slope coefficient $\beta_{1,i}$, is given by[1]

(1)

$$y_{i,t} = \beta_{0,i} + \beta_{1,i}x_{i,t} + \varepsilon_{i,t}, \qquad \varepsilon_{i,t} \sim N(0,\sigma^2) \text{ i.i.d.},$$
$$\beta_{0,i} = \beta_0 + u_{0,i},$$
$$\beta_{1,i} = \beta_1 + u_{1,i},$$
$$\begin{pmatrix} u_{0,i} \\ u_{1,i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \rho\tau_0\tau_1 \\ \rho\tau_0\tau_1 & \tau_1^2 \end{pmatrix}\right),$$

where index *i* is at the macro level (e.g., individuals) and index *t* is at the micro level (e.g., time within individual), τ_0^2 and τ_1^2 are the variances of the intercept and the slope coefficient, respectively, and parameter ρ captures the correlation. It is common practice to keep ρ in the model (e.g., Snijders and Bosker 2012). In reduced form, (1) can be written as

(2)

$$y_{i,t} = \beta_0 + \beta_1 x_{i,t} + \xi_{i,t}, \qquad \xi_{i,t} = u_{0,i} + u_{1,i}x_{i,t} + \varepsilon_{i,t}$$
$$\begin{pmatrix} u_{0,i} \\ u_{1,i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \rho\tau_0\tau_1 \\ \rho\tau_0\tau_1 & \tau_1^2 \end{pmatrix}\right), \qquad \varepsilon_{i,t} \sim N(0,\sigma^2) \text{ i.i.d.}$$

Model parameters can be estimated by maximizing the log-likelihood function

(3)

$$\ln L\,(\theta, \tau_1^2, \rho) = -\frac{1}{2}\sum_{i=1}^{N}\left(T_i \ln(2\pi) + \ln|\Sigma_i| + \left(y_i - \beta_{0,i} - \beta_{1,i}x_i\right)' \Sigma_i^{-1}\left(y_i - \beta_{0,i} - \beta_{1,i}x_i\right)\right),$$

[1] The basic model can be generalized by adding other explanatory variables to the regression equation and putting group-level variables z_i into the random intercept and random slope. As long as $x_{i,t}$ is the only explanatory variable with a random slope, these extensions affect the fixed part of the model but not the random part. A common approach in a model with multiple explanatory variables is to test for one random slope at a time (e.g., Snijders and Bosker 2012; Stoel et al. 2012) and this strategy fits nicely within our framework.

where $\theta \equiv (\beta_0, \beta_1, \sigma^2, \tau_0^2)$ contains all model parameters except the slope variance parameter τ_1^2 and the correlation parameter ρ. In (3), N is the number of groups at the macro level, T_i is the number of observations within group i, $y_i \equiv (y_{i,1}, \dots, y_{i,T_i})$ contains all observations $y_{i,t}$ within group *i* and $x_i \equiv (x_{i,1}, \dots, x_{i,T_i})$ is defined similarly. It follows from (2) that the elements (t, s), $t = 1, \dots, T_i$, $s = 1, \dots, T_i$, of covariance matrix Σ_i are given by

(4)

$$(\Sigma_i)_{t,t} = \tau_0^2 + 2x_{i,t}\rho\tau_0\tau_1 + x_{i,t}^2\tau_1^2 + \sigma^2,$$
$$(\Sigma_i)_{t,s} = \tau_0^2 + (x_{i,t} + x_{i,s})\rho\tau_0\tau_1 + x_{i,t}x_{i,s}\tau_1^2, \quad t \neq s.$$

We obtain (asymptotic) standard errors by taking the square root of the diagonal elements of the estimated covariance matrix of the parameter estimates, which in turn can be computed as minus the inverse of the Hessian of (3) evaluated at the optimal parameter values.

3. The key issue

Testing whether the slope variance τ_1^2 equals zero may seem a rather straightforward task. It has been suggested that $\tau_1^2 = 0$ implies $\rho = 0$ (e.g., Lahuis and Ferguson 2009; Snijders and Bosker 2012; Stoel et al. 2006). A first conjecture is that the two restrictions can be tested by implementing an LR test and comparing its outcome to a chi square distribution with two degrees of freedom; the 5% critical value is 5.99.

However, this approach is incorrect for two reasons: (1) variances cannot be negative and hence the null hypothesis $\tau_1^2 = 0$ is located on the boundary of the parameter space, violating one of the regularity conditions of the LR test, and (2) the correlation parameter ρ for random intercept $u_{0,i}$ and random slope $u_{1,i}$ is not identified under the null hypothesis in which $u_{1,i}$ disappears from the model, hence the Davies problem. Extant studies have dealt with the first issue of non-negative variance (but not with the Davies problem) by replacing the χ_2^2 reference distribution by a chi bar square distribution, i.e., a mixture of a χ_1^2 distribution and a χ_2^2 distribution, both with 50% weight (e.g., Lahuis and Ferguson 2009; Snijders and Bosker 2012; Stoel et al. 2006; Stram and Lee 1994); the corresponding 5% critical value is 5.14. The intuition of this reference distribution is that under the null hypothesis the realized (unrestricted) variance τ_1^2 would be negative and hence take a value of zero in 50% of the cases. While the first degree of freedom from the correlation parameter ρ (not on the boundary of the parameter space) is always present, the second degree of freedom from the slope variance τ_1^2 (on the boundary of the parameter space) is only present in 50% of the cases (Self and Liang 1987; Stoel et al. 2006).

Though the chi bar square distribution is statistically correct if one wants to test the joint restriction of zero slope variance, $\tau_1^2 = 0$, and zero correlation, $\rho = 0$, this null hypothesis is conceptually incorrect. The reason is that only the slope variance is a parameter of interest and the correlation is merely a nuisance. Put differently, we want to test for $\tau_1^2 = 0$ and need to control for ρ. It is easy to see from (1) and (2) that the nuisance parameter ρ is not identified under the null hypothesis of interest, i.e., $\tau_1^2 = 0$.

4. The solution

Davies (1987) proposed a general solution to the problem of dealing with an unidentified nuisance parameter under the null hypothesis. The idea is to construct a new test statistic by evaluating the original test statistic for the hypothesis of interest over the entire range of the unidentified parameter. In our context and for given ρ, the LR statistic is twice the difference between the log-likelihood of the full model, $\ln L\,(\theta, \tau_1^2, \rho)$ in (3), maximized over $\theta \equiv (\beta_0, \beta_1, \sigma^2, \tau_0^2)$ and τ_1^2, and the log-likelihood under the null hypothesis $\tau_1^2 = 0$:

(5)

$$\ln L_0\,(\theta) = -\frac{1}{2}\sum_{i=1}^{N}\Big(T_i \ln(2\pi) + \ln|\Sigma_{0,i}| + (y_i - \beta_{0,i} - \beta_{1,i}x_i)'\,{\Sigma_{0,i}}^{-1}(y_i - \beta_{0,i} - \beta_{1,i}x_i)\Big),$$

with

(6)

$$(\Sigma_{0,i})_{t,t} = \tau_0^2 + \sigma^2,$$
$$(\Sigma_{0,i})_{t,s} = \tau_0^2, \quad t \neq s,$$

maximized over $\theta \equiv (\beta_0, \beta_1, \sigma^2, \tau_0^2)$. The corresponding test statistic for given ρ can be written as

(7)

$$\text{LR}(\rho) \equiv 2\ln L\left(\hat{\theta}(\rho), \hat{\tau}_1^2(\rho), \rho\right) - 2\ln L_0\left(\hat{\theta}_0\right),$$

where the "hats" denote parameter estimates and we make explicit the dependence of the parameter estimates in the full model on the nuisance parameter ρ (unidentified under the null hypothesis). Maximizing $\text{LR}(\rho)$ over the range of ρ results in the regular LR test statistic (e.g., Andrews and Ploberger 1994; Hansen 1996):

(8)

$$\text{LR} \equiv \max_{\rho} \text{LR}(\rho),$$

but it no longer has a standard distribution.[2]

Simulated critical values

The top part of Table 1 provides our simulated critical values at the 10% and 5% levels for number of groups N equal to 50, 100, 200 and 1000, where N = 1000 represents the asymptotic case. We use 40,000 draws, consider $T_i = 20$ observations within each group and verified that the critical values did not depend on T_i. The simulated asymptotic critical value of 5.04 at the 5% level is slightly more liberal than the corresponding critical value of 5.14 from the chi bar square distribution. Intuitively, the latter critical value corresponds to testing $\tau_1^2 = 0$ and $\rho = 0$ together, i.e., it treats the correlation parameter ρ as a parameter of interest, adding a full degree of freedom. On the other hand, it should be treated as a nuisance parameter that disappears under the null hypothesis and contributes *less* than a full degree of freedom; ρ is not identified and may be *non-zero* under the null hypothesis. For this reason the critical values from the chi bar square distribution (e.g., 5.14 at the 5% level) are slightly too conservative. Furthermore, these critical values are conservative because they still are asymptotic. Though the chi bar square distribution wrongly ignores the Davies problem, it turns out to be quite accurate (though not perfect) if the number of groups N is large, but it is quite far off in the more common situation in which the number of groups is limited. This makes it harder to detect a random slope when it is actually present in the data.

[2] An alternative is taking the average of $\text{LR}(\rho)$ over the range of ρ, resulting in the class of "ave test statistics" (Andrews and Ploberger 1994). However, this did not result in substantially higher power while requiring a computationally intensive grid search over ρ. We did not pursue this approach.

For later reference, the bottom part of Table 1 contains the simulated critical values when imposing $\rho = 0$ in (1) and (2), i.e., when we want to test the single restriction $\tau_1^2 = 0$ in the model without intercept-slope correlation.

Statistical power

We assess statistical power of the LR test with nuisance parameter by varying the slope standard deviation τ_1, the correlation ρ and the number of groups N in the data generating process. As before, we have $T_i = 20$ observations within each group. Table 2 provides the corresponding percentages of rejection of the null hypothesis $\tau_1^2 = 0$ using the appropriate 5% critical values in the top part of Table 1 and based on 1000 draws. Though it is not surprising that power increases substantially if the number of groups N is large and if the slope variance τ_1^2 increases, it is noteworthy that power is low if the actual correlation ρ is close to zero. This power quickly increases if ρ increases in an absolute sense, and it does not appear to depend on whether the correlation becomes more positive or negative; the sign of the correlation does not matter. The dependence of power on the absolute size of the correlation is strongest when the slope variance is large, i.e., when relatively far away from the null hypothesis $\tau_1^2 = 0$, and when the number of groups N is large. Taken together, our analysis suggests that keeping the correlation parameter ρ in the model when it is not strongly present in the data results in low power.

So, while it has been advocated and is common practice to keep ρ in the model (e.g., Snijders and Bosker 2012, p. 76-77), we argue that statistical significance of ρ needs to be tested first to come up with the appropriate LR test for $\tau_1^2 = 0$, i.e., either the test with or without the nuisance parameter ρ that is not identified under the null hypothesis. While the current approach for testing slope variance is to consider the joint (but conceptually incorrect) null hypothesis $\rho = 0$ and $\tau_1^2 = 0$, i.e., *simultaneous* testing of both the nuisance parameter and the parameter of interest, our new approach is *sequential*: it first considers statistical significance of the nuisance parameter and next tests the parameter of interest conditional on the outcome for the nuisance parameter. This procedure acknowledges that *conceptually* there may be a parameter in the model that is not identified under the null hypothesis, but removes this parameter from the model if it turns out not to be present *empirically*.

A simple test with better power properties

Based on the discussion above, we outline our simple procedure that requires the log-likelihood of the model without random slope, i.e., $\ln L_0(\hat{\theta})$ defined by (5) and (6), the log-likelihood $\ln L_{\text{no}\,\rho}(\hat{\theta}, \hat{\tau}_1^2)$ with random slope but without correlation parameter ρ, obtained from (3) and (4) after setting $\rho = 0$, and the log-likelihood of the full model, i.e., $\ln L(\hat{\theta}, \hat{\tau}_1^2, \hat{\rho})$ defined by (3) and (4). As the three model specifications are nested, it is efficient to estimate them sequentially. Two of the models are already needed for the regular LR test, and the extra model (with random slope but without correlation) is nested in between. Our procedure involves the LR statistics from each of the three pairwise combinations. We test for the unidentified nuisance parameter ρ using a regular LR test with one degree of freedom (5% critical value = 3.84), keep or remove ρ based on its statistical significance, and next apply the appropriate LR test for $\tau_1^2 = 0$. In brief:

1. Check whether $2\ln L(\hat{\theta}, \hat{\tau}_1^2, \hat{\rho}) - 2\ln L_{\text{no}\,\rho}(\hat{\theta}, \hat{\tau}_1^2) > 3.84$.
2. If yes, reject $\tau_1^2 = 0$ if $2\ln L(\hat{\theta}, \hat{\tau}_1^2, \hat{\rho}) - 2\ln L_0(\hat{\theta}) > \alpha_{0.05,\text{full}}$, where $\alpha_{0.05,\text{full}}$ is the appropriate 5% simulated critical value from the top part of Table 1.
3. If no, reject $\tau_1^2 = 0$ if $2\ln L_{\text{no}\,\rho}(\hat{\theta}, \hat{\tau}_1^2) - 2\ln L_0(\hat{\theta}) > \alpha_{0.05,\text{no}\,\rho}$, where $\alpha_{0.05,\text{no}\,\rho}$ is the appropriate 5% simulated critical value from the bottom part of Table 1.

Table 3 shows the power of our sequential LR test and is analogous to Table 2 for the regular LR test. Though power is still lowest if the actual correlation ρ is close to zero, the dependence of power on the correlation parameter is much weaker than it is in Table 2. Table 4 shows the percentage point difference in power between our sequential test and the regular LR test (with correct critical values from the top part of Table 1). It confirms the pattern: the largest improvement in power occurs when correlation ρ is close to zero, i.e., when increased power is needed most. The difference in power is quite substantial. For all reported numbers of groups *N*, it is possible to achieve improvements of 10 percentage points. For instance, if $\rho = 0$ and $\tau_1 = 0.10$, power increases from 16 percent to 26 percent for *N* = 50 and from 43 percent to 56 percent for *N* = 200, even though the sequential LR test is hardly more complicated or time consuming than the regular LR test.

5. Meta regression for power

To extract the general patterns from our power analysis, we run a linear regression on all 360 elements of Tables 2 and 3, where power (divided by 100) serves as the dependent variable. As explanatory variables we include a 0/1 dummy variable indicating whether the method is our sequential approach, the number of groups *N* (divided by 100), the slope standard deviation τ_1, the correlation ρ, the squared correlation ρ^2 to account for the suggested U-shaped relationship between correlation and power, interaction terms for *N*, τ_1, ρ and ρ^2, and we interact everything with the 0/1 sequential approach indicator to assess when extra power relative to the regular LR test is largest. The regression provides an excellent fit, with an R-square equal to 0.873 and an adjusted R-square of 0.866. Table 5 contains the parameter estimates.

The most important result in Table 5 is that the sequential LR test provides extra power over the regular LR test (with appropriate simulated critical values from Table 1), and the effect is significant at the 1% level. Furthermore, power is higher when the number of groups *N* is large, the slope standard deviation τ_1 is large, i.e., when we are far away from the null hypothesis, and when the correlation parameter is large in absolute sense. The significant interaction terms indicate that the positive effects of *N*, τ_1 and ρ^2 on power are reinforced when these variables take large values at the same time; all effects but one are significant at 1%, with the other one being significant at 5%. An important null finding is that the intercept-slope correlation itself has no effect on power, neither as a main effect nor as an interaction term. The second part of the table shows that one of the 9 interaction terms involving the sequential approach indicator (Seq) is significant at the 10% level, while the other 8 are not significant. The negative coefficient for Seq × ρ^2 confirms that the gain in power from our sequential approach is largest when ρ^2 is small, i.e., when the correlation parameter is close to zero and power is generally lowest.

6. Conclusion

Crucial inference for the hierarchical linear model concerning the null hypothesis of no random slope is hampered by the fact that there is a parameter that is only identified under the alternative, the so-called Davies problem. This calls for an alternative methodology, and this is what we have pursued in the present paper. The method is easy to implement and tackles the conceptual issue that the correlation between intercept and slope should not be tested as a parameter of interest; it is nuisance that does not show up in the null hypothesis but instead disappears. With simulations we have demonstrated that our test should be practically relevant. Future applications and case-specific illustrations shall show that our methodology matters.

References

Andrews, Donald W.K. (2001), “Testing When a Parameter is on the Boundary of the Maintained Hypothesis”, *Econometrica,* 69 (May), 683-734.

Andrews, Donald W. K. and Werner Ploberger (1994), “Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative”, *Econometrica*, 62 (November), 1383-1414.

Beg, A. B. M. Rabiul A., Mervyn J. Silvapulle and Paramsothy Silvapulle (2001), “Tests Against Inequality Constraints When Some Nuisance Parameters Are Present Only Under the Alternative: Test of ARCH in ARCH-M Models”, *Journal of Business & Economic Statistics*, 19 (April), 245-253.

Carrasco, Marine (2002), “Misspecified Structural Change, Threshold, and Markov-Switching Models”, *Journal of Econometrics*, 109, 239-273.

Davies, Robert B. (1987), “Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative”, *Biometrika,* 74 (March), 33-43.

Davies, Robert B. (2002), “Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative: Linear Model Case”, *Biometrika*, 89 (June), 484-489.

Franses, Philip Hans and Rutger van Oest (2007), “On the Econometrics of the Geometric Lag Model”, *Economics Letters*, 291-296.

Hansen, Bruce E. (1996), “Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis”, *Econometrica*, 64 (March), 413-430.

LaHuis, David M. and Matthew W. Ferguson (2009), “The Accuracy of Significance Tests for Slope Variance Components in Multilevel Random Coefficient Models”, *Organizational Research Methods*, 12 (July), 418-435.

Oakley, James L., Dawn Iacobucci and Adam Duhachek (2005), “Multilevel Hierarchical Linear Models and Marketing: This Is Not Your Adviser’s OLS Model”, in Naresh K. Malhotra (ed.), *Review of Marketing Research*, *Vol. 2* (203-227), M.E. Sharpe.

Self, Steven G. and Kung-Yee Liang (1987), “Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions”, *Journal of the American Statistical Association*, 82 (June), 605-610.

Snijders, Tom A. B. and Roel J. Bosker (2012), *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, second edition, Sage.

Stoel, Reinoud D., Francisca Galindo Garre, Conor Dolan and Godfried van den Wittenboer (2006), “On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints”, *Psychological Methods*, 11 (4), 439-455.

Stram, Daniel O. and Jae Won Lee (1994), “Variance Components Testing in the Longitudinal Mixed Effects Model”, *Biometrics*, 50 (December), 1171-1177.

Table 1: simulated critical values for different number of groups *N;* correlation ρ is a free nuisance parameter in the top part, it is assumed zero in the bottom part

		10% level	5% level
If ρ is free	$N = 50$	3.41	4.67
	$N = 100$	3.52	4.80
	$N = 200$	3.59	4.86
	$N = 1000$	3.72	5.04
If $\rho = 0$	$N = 50$	1.23	2.17
	$N = 100$	1.37	2.37
	$N = 200$	1.41	2.44
	$N = 1000$	1.59	2.65

Table 2: power of regular LR test for different standard deviations of the slope parameter, correlations and numbers of groups

N = 50									
	ρ=−0.8	ρ=−0.4	ρ=−0.2	ρ=−0.1	ρ=0.0	ρ=+0.1	ρ=+0.2	ρ=+0.4	ρ=+0.8
τ_1=0.02	6.6	5.6	5.4	5.0	5.0	4.9	4.8	5.0	6.5
τ_1=0.04	15.7	7.6	6.2	6.3	5.9	6.1	6.7	7.4	13.1
τ_1=0.06	28.2	12.9	8.6	8.4	7.9	8.1	8.6	11.2	25.9
τ_1=0.08	44.9	20.0	12.4	10.8	10.6	11.3	12.0	17.4	43.1
τ_1=0.10	64.1	29.4	19.2	17.0	16.2	16.0	18.0	27.3	62.8

N = 100									
	ρ=−0.8	ρ=−0.4	ρ=−0.2	ρ=−0.1	ρ=0.0	ρ=+0.1	ρ=+0.2	ρ=+0.4	ρ=+0.8
τ_1=0.02	10.1	7.4	6.4	6.3	6.9	6.8	7.1	8.4	11.7
τ_1=0.04	25.0	12.3	8.9	7.9	7.8	8.7	10.4	13.3	27.0
τ_1=0.06	48.0	19.4	12.9	11.5	10.5	11.7	13.8	21.6	47.0
τ_1=0.08	73.0	34.1	20.3	17.9	17.0	18.4	22.1	34.0	72.0
τ_1=0.10	88.7	49.9	33.0	28.6	26.6	29.7	33.9	48.1	88.0

N = 200									
	ρ=−0.8	ρ=−0.4	ρ=−0.2	ρ=−0.1	ρ=0.0	ρ=+0.1	ρ=+0.2	ρ=+0.4	ρ=+0.8
τ_1=0.02	15.1	7.4	6.0	5.5	5.6	5.7	5.8	7.5	13.3
τ_1=0.04	46.0	18.6	9.5	7.6	6.8	7.9	9.9	15.8	42.2
τ_1=0.06	79.0	34.8	18.6	13.6	11.4	13.5	16.8	32.0	75.3
τ_1=0.08	96.1	56.7	35.0	27.3	23.3	25.5	31.0	54.5	94.7
τ_1=0.10	99.3	77.5	54.7	46.7	42.7	43.1	52.4	74.9	99.6

N = 1000									
	ρ=−0.8	ρ=−0.4	ρ=−0.2	ρ=−0.1	ρ=0.0	ρ=+0.1	ρ=+0.2	ρ=+0.4	ρ=+0.8
τ_1=0.02	50.2	19.0	9.7	6.8	5.5	6.0	7.4	14.8	50.7
τ_1=0.04	98.5	55.5	25.0	16.6	11.7	12.2	23.2	56.4	97.2
τ_1=0.06	100.0	90.7	53.3	37.7	33.8	39.0	55.3	90.0	100.0
τ_1=0.08	100.0	99.5	88.2	77.3	71.6	78.1	90.0	99.5	100.0
τ_1=0.10	100.0	100.0	98.8	97.2	96.5	98.0	99.2	100.0	100.0

Table 3: power of sequential LR test for different standard deviations of the slope parameter, correlations and numbers of groups

N = 50									
	ρ=−0.8	ρ=−0.4	ρ=−0.2	ρ=−0.1	ρ=0.0	ρ=+0.1	ρ=+0.2	ρ=+0.4	ρ=+0.8
τ_1=0.02	9.1	7.6	7.5	7.4	7.3	7.5	7.6	7.7	9.1
τ_1=0.04	16.6	10.3	9.3	9.3	9.3	9.6	9.5	10.8	15.6
τ_1=0.06	29.3	16.0	12.5	12.6	12.2	12.6	13.3	15.4	27.8
τ_1=0.08	45.7	25.1	18.8	18.5	18.3	18.3	19.1	23.1	44.6
τ_1=0.10	64.2	35.3	28.2	26.6	25.9	26.3	28.1	33.0	63.2

N = 100									
	ρ=−0.8	ρ=−0.4	ρ=−0.2	ρ=−0.1	ρ=0.0	ρ=+0.1	ρ=+0.2	ρ=+0.4	ρ=+0.8
τ_1=0.02	13.3	10.4	9.0	8.7	9.2	9.4	9.8	11.0	14.5
τ_1=0.04	26.5	15.0	12.3	10.7	10.6	11.6	13.3	15.8	28.7
τ_1=0.06	48.1	23.6	18.3	16.8	16.8	18.1	19.5	25.5	47.4
τ_1=0.08	72.8	38.4	28.5	26.1	25.7	27.2	29.4	38.2	71.5
τ_1=0.10	88.7	54.2	42.6	38.6	37.8	39.5	42.4	51.5	87.9

N = 200									
	ρ=−0.8	ρ=−0.4	ρ=−0.2	ρ=−0.1	ρ=0.0	ρ=+0.1	ρ=+0.2	ρ=+0.4	ρ=+0.8
τ_1=0.02	16.9	10.1	8.3	7.9	7.9	8.0	8.3	9.7	14.7
τ_1=0.04	46.7	21.5	14.3	12.8	11.9	12.4	13.6	18.3	42.4
τ_1=0.06	78.8	39.2	26.6	22.4	21.4	22.0	24.0	36.3	75.2
τ_1=0.08	95.8	59.8	42.8	37.7	34.9	36.5	40.0	57.6	94.9
τ_1=0.10	99.2	79.5	64.4	59.0	56.4	57.7	62.5	77.3	99.5

N = 1000									
	ρ=−0.8	ρ=−0.4	ρ=−0.2	ρ=−0.1	ρ=0.0	ρ=+0.1	ρ=+0.2	ρ=+0.4	ρ=+0.8
τ_1=0.02	50.1	20.2	11.6	9.6	8.9	9.3	10.6	17.4	50.6
τ_1=0.04	98.5	56.0	29.8	22.8	20.0	21.5	29.4	57.5	97.2
τ_1=0.06	100.0	90.2	58.8	49.5	46.3	48.8	60.3	89.6	100.0
τ_1=0.08	100.0	99.6	89.9	85.5	84.1	86.1	92.6	99.5	100.0
τ_1=0.10	100.0	100.0	99.3	99.0	99.1	99.5	99.8	100.0	100.0

Table 4: difference in power between sequential LR test and regular LR test for different standard deviations of the slope parameter, correlations and numbers of groups

N = 50									
	ρ=−0.8	ρ=−0.4	ρ=−0.2	ρ=−0.1	ρ=0.0	ρ=+0.1	ρ=+0.2	ρ=+0.4	ρ=+0.8
τ_1=0.02	2.5	2.0	2.1	2.4	2.3	2.6	2.8	2.7	2.6
τ_1=0.04	0.9	2.7	3.1	3.0	3.4	3.5	2.8	3.4	2.5
τ_1=0.06	1.1	3.1	3.9	4.2	4.3	4.5	4.7	4.2	1.9
τ_1=0.08	0.8	5.1	6.4	7.7	7.7	7.0	7.1	5.7	1.5
τ_1=0.10	0.1	5.9	9.0	9.6	9.7	10.3	10.1	5.7	0.4

N = 100									
	ρ=−0.8	ρ=−0.4	ρ=−0.2	ρ=−0.1	ρ=0.0	ρ=+0.1	ρ=+0.2	ρ=+0.4	ρ=+0.8
τ_1=0.02	3.2	3.0	2.6	2.4	2.3	2.6	2.7	2.6	2.8
τ_1=0.04	1.5	2.7	3.4	2.8	2.8	2.9	2.9	2.5	1.7
τ_1=0.06	0.1	4.2	5.4	5.3	6.3	6.4	5.7	3.9	0.4
τ_1=0.08	−0.2	4.3	8.2	8.2	8.7	8.8	7.3	4.2	−0.5
τ_1=0.10	0.0	4.3	9.6	10.0	11.2	9.8	8.5	3.4	−0.1

N = 200									
	ρ=−0.8	ρ=−0.4	ρ=−0.2	ρ=−0.1	ρ=0.0	ρ=+0.1	ρ=+0.2	ρ=+0.4	ρ=+0.8
τ_1=0.02	1.8	2.7	2.3	2.4	2.3	2.3	2.5	2.2	1.4
τ_1=0.04	0.7	2.9	4.8	5.2	5.1	4.5	3.7	2.5	0.2
τ_1=0.06	–0.2	4.4	8.0	8.8	10.0	8.5	7.2	4.3	–0.1
τ_1=0.08	–0.3	3.1	7.8	10.4	11.6	11.0	9.0	3.1	0.2
τ_1=0.10	–0.1	2.0	9.7	12.3	13.7	14.6	10.1	2.4	–0.1

N = 1000									
	ρ=−0.8	ρ=−0.4	ρ=−0.2	ρ=−0.1	ρ=0.0	ρ=+0.1	ρ=+0.2	ρ=+0.4	ρ=+0.8
τ_1=0.02	−0.1	1.2	1.9	2.8	3.4	3.3	3.2	2.6	−0.1
τ_1=0.04	0.0	0.5	4.8	6.2	8.3	9.3	6.2	1.1	0.0
τ_1=0.06	0.0	−0.5	5.5	11.8	12.5	9.8	5.0	−0.4	0.0
τ_1=0.08	0.0	0.1	1.7	8.2	12.5	8.0	2.6	0.0	0.0
τ_1=0.10	0.0	0.0	0.5	1.8	2.6	1.5	0.6	0.0	0.0

Table 5: meta summary for power, obtained from linear regression on Tables 2 and 3

	Coefficient	Std. error
Intercept	0.368***	(0.006)
Seq (1 if sequential LR, 0 else)	0.041***	(0.012)
N (# groups / 100)	0.044***	(0.002)
τ_1 (slope standard deviation)	6.603***	(0.215)
ρ (correlation)	–0.006	(0.014)
ρ^2 (squared correlation)	0.519***	(0.025)
$N \times \tau_1$	0.586***	(0.056)
$N \times \rho$	0.000	(0.004)
$N \times \rho^2$	0.016**	(0.006)
$\tau_1 \times \rho$	–0.046	(0.495)
$\tau_1 \times \rho^2$	3.850***	(0.869)
Seq × N	–0.001	(0.003)
Seq × τ_1	0.392	(0.431)
Seq × ρ	0.001	(0.028)
Seq × ρ^2	–0.085*	(0.049)
Seq × $N \times \tau_1$	–0.078	(0.111)
Seq × $N \times \rho$	0.000	(0.007)
Seq × $N \times \rho^2$	–0.001	(0.013)
Seq × $\tau_1 \times \rho$	–0.005	(0.991)
Seq × $\tau_1 \times \rho^2$	–1.519	(1.738)

* $p < .10$;

** $p < .05$;

*** $p < .01$