Regressor and random-effects dependencies in multilevel models

Peter Ebbes*

Department of Economics, University of Groningen, P.O. Box 800, 9700AV, Groningen, NL

Ulf Böckenholt

Faculty of Management, McGill University, Montreal, PQ, H3A 1G5, Canada

Michel Wedel

Business School, University of Michigan, Ann Arbor, MI 48109-1234, USA

The objectives of this paper are (1) to review methods that can be used to test for different types of random effects and regressor dependencies, (2) to present results from Monte Carlo studies designed to investigate the performance of these methods, and (3) to discuss estimation methods that can be used when some but not all of the random effects and regressor independence assumptions, are violated. Because current methods are limited in various ways, we will also present a list of open problems and suggest solutions for some of them. As we will show, the issue of regressor random-effects independence has received some attention in the econometrics literature, but this important work has had little impact on current research practices in the social and behavioral sciences.

Key Words and Phrases: instrumental variables, linear models, exogeneity, endogenous variables.

1 Introduction

In many situations data have a hierarchical structure. For example, when one investigates how workplace characteristics affect worker productivity, both workers and firms are units in the analysis. Similarly, hierarchical data arise in the context of panel research, when multiple observations are available on the 'objects' under

study. Typically, these types of data are analyzed with multilevel or hierarchical linear models. The model we consider is given by

$$y_{ij} = X'_{ij}\boldsymbol{\beta} + Z'_i\boldsymbol{\gamma} + \alpha_i + \eta_{ij}, \tag{1}$$

where y_{ij} is the dependent variable, $X_{ij} \in \mathbb{R}^{k \times 1}$ are level-one or individual specific regressors, $Z_i \in \mathbb{R}^{l \times 1}$ contains level-two or group specific regressors, η_{ij} is a random (error) component with $\mathrm{E}(\eta_{ij}) = 0$ and $\mathrm{var}(\eta_{ij}) = \sigma_\eta^2$, and where $i = 1,...,n$ and $j = 1,...,n_i$. Throughout this article, matrices are printed in capitals and scalars and vectors as lowercase. Greek symbols denote unobserved parameters that are to be estimated. The unit-specific intercept α_i may be specified to be random (with $\mathrm{E}(\alpha_i) = 0$ and $\mathrm{var}(\alpha_i) = \sigma_\alpha^2$) or fixed depending on the context of the study and the types of inferences that can be drawn (VERBEEK, 2000; JUDGE *et al.*, 1985; WOOLD-RIDGE, 2002; BRYK and RAUDENBUSH, 1992; SNIJDERS and BOSKER, 1999).

Frequently, in the modeling of hierarchical data structures it is assumed that the explanatory variables X and Z are independent of the random (error) components. If independence holds, the regressors are said to be 'exogenous' (or determined outside the model). However, in many applications it is unrealistic to assume that regressors and random components are independent. For the model given in (1) we consider two types of independence:

1. level-2 independence or $X\alpha$- and $Z\alpha$-independence, and
2. level-1 independence or $X\eta$- and $Z\eta$- independence.

This article shows that even in the presence of modest dependencies, regression effects can be biased substantially. Different approaches for testing the independence assumption are presented and illustrated with the help of simulation studies.

Importantly, the independence assumptions can be easily violated. Examples include (1) relevant omitted variables (CARD, 1999, 2001; UUSITALO, 1999; SPENCER and FIELDING, 1998a, 1998b), (2) measurement error in the regressors (PLAT, 1988; BAGOZZI *et al.*, 1999; WANSBEEK and MEIJER, 2000; CARROLL *et al.*, 1995), (3) self-selection (HAMILTON and NICKERSON, 2003; VELLA and VERBEEK, 1998; ANGRIST *et al.*, 1996), (4) simultaneity (WHITE, 2001; GREENE, 1997), and (5) serially correlated regressors in the presence of lagged dependent variables (WHITE, 2001; RUUD, 2000). In the standard (single level) regression model, the ordinary least squares (OLS) estimator can be written as $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} = (X'X)^{-1}X'y = \boldsymbol{\beta} + (X'X)^{-1}X'\epsilon$, where $\mathrm{E}(\epsilon) = 0$. When the assumption of independence of regressors and errors does not hold (i.e. when $\mathrm{E}(\epsilon|X) \neq 0$), it follows immediately that the OLS estimator is biased. This bias can be reduced, at least in large samples, by using instrumental variables estimation techniques (BOWDEN and TURKINGTON, 1984; WHITE, 2001; WOOLDRIDGE, 2002). Instrumental variables (IVs) should be uncorrelated with the error term ϵ, and should explain part of the variability in the endogenous regressors X. Once instruments are available, unbiased estimates for the regression parameters can be obtained. Furthermore, Hausman-like tests (HAUSMAN, 1978) can be used to test for regressor error dependencies in this standard linear regression model. The general

idea of this approach is to compare two estimators, one that is consistent under both the null hypothesis of regressor–error independence and the alternative hypothesis, and one that is only consistent under the null hypothesis. The null hypothesis is rejected once a significant difference between these two estimators is found (cf. VERBEEK, 2000). For a more detailed explanation of instrumental variables techniques and the Hausman test, we refer to Appendix 1.

In multilevel models additional random components reflect the nesting structure in the data. Henceforth, an investigation of independence of explanatory variables and random terms becomes even more important. Because of the potentially severe consequences when these independence assumptions are violated, they need to be tested for explicitly in any application of multilevel models. The literature suggest performing the following diagnostic steps when endogeneity is suspected, which serves as a roadmap to the remainder of this paper. First, a diagnostic check to examine $X\alpha$-independence is readily available for multilevel models based on the work by HAUSMAN and TAYLOR (1981). Fixed-effects (FE) estimation gives an unbiased estimate for $\boldsymbol{\beta}$ in model (1) regardless of violation of $X\alpha$-independence, whereas random-effects (RE) estimation yields biased estimates (see Section 2). If the test, which is based on the Hausman test (HAUSMAN, 1978), proposed by HAUSMAN and TAYLOR (1981) (which we denote by the H_α-test) does not reject the independence hypothesis, both fixed- and random-effects estimation for $\boldsymbol{\beta}$ can be used. Once rejected, only fixed-effects estimation yields consistent results, provided the regressors are independent of level-1 random components. We show how the inclusion of group means can be used to examine $X\alpha$-dependencies (MUND-LAK, 1978). We present the Hausman–Taylor (HT) estimator as an alternative to fixed-effects estimation, which is potentially more efficient and which, in contrast to the fixed-effects estimator, can be used to estimate level-2 effects. The H_α-test, Mundlak's π approach and the Hausman–Taylor estimator are discussed in Section 3.

However, these above-mentioned steps should be considered with caution. As will be shown in Section 4, the performance of the H_α-test relies on the independence of regressors and level-1 random components. Unfortunately, endogeneity at this level can often not be ruled out a priori. Although this type of endogeneity is often ignored, it is a crucial assumption in using standard multilevel estimators. As a first diagnostic check for it, one should carefully consider whether or not, based on theoretical grounds, level-1 independence can be assumed. If not, IV estimation techniques can be adopted to estimate regression parameters in model (1) (BOWDEN and TURKING-TON, 1984; WOOLDRIDGE, 2002). Several different multilevel IV estimators can be derived to estimate the regression parameters in model (1), depending on the specific assumptions about the exact form of the endogeneity problem (see Appendix 2). This approach is illustrated and discussed in Section 5. To test for level-1 independence, another test based on the general approach of HAUSMAN (1978) can be constructed. We will refer to this test as the H_η-test and illustrate its usefulness in Section 5.

The diagnostics steps for investigating independence assumptions in two-level multilevel models are presented in Table 1. For the sake of simplicity, no distinction is

Table 1.   Overview of diagnostic tests to determine independence between regressors and random effects in a linear two-level regression model, where 'yes' ('no') means that the specific independence assumption is (not) satisfied.

Case	$EX\alpha = 0$	$EX\eta = 0$	Section	Table	Test	Estimators
(i)	Yes	Yes	2, 5	2, 7	H_α or Mundlak's π, and H_η	FE or RE
(ii)	No	Yes	2, 3	2, 3, 4	H_α or Mundlak's π, and H_η	FE or HT
(iii)	Yes	No	2, 4, 5	2, 5, 6, 7	H_η	External IV
(iv)	No	No	2, 4, 5	2, 5, 7	H_η	External IV

made between level-1 and level-2 regressors. Now three types (cases (ii)–(iv)) of violations of regressor–error dependencies can be distinguished. This table specifies the various sections in which tests and estimators for each case are discussed in more detail.

2   Biases caused by level-1 ($X\alpha$)- and level-2 ($X\eta$)-dependencies

The parameters $\boldsymbol{\beta}$ in the multilevel model given in (1), can be estimated by fixed- or random-effects methods (VERBEEK, 2000; BALTAGI, 2001; GOLDSTEIN, 1995; LONG-FORD, 1993). We do not discuss the estimators here, but details can be found in Appendix 2. To illustrate the effects of $X\alpha$- and $X\eta$-dependencies under fixed-effects and random-effects estimation, consider Table 2 which summarizes the simulation results for the model:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \alpha_i + \eta_{ij}, \tag{2}$$

where $i = 1,\ldots,150$, $j = 1,\ldots,10$, $\alpha_i \sim N(0, \sigma_\alpha^2)$ and $\eta_{ij} \sim N(0, \sigma_\eta^2)$, and the following four cases are specified: (i) $\rho(x, \alpha) = \rho(x, \eta) = 0$, (ii) $\rho(x, \alpha) = 0.3$ and $\rho(x, \eta) = 0$, (iii) $\rho(x, \alpha) = 0$ and $\rho(x, \eta) = 0.3$, and (iv) $\rho(x, \alpha) = \rho(x, \eta) = 0.3$. The table presents means and standard deviations computed across 250 replications. As expected, both the fixed-effects and random-effects estimator yield unbiased results for β_1 and unbiased estimates for the variances when the regressor is truly exogenous (case (i)). The fixed-effects estimator cannot estimate the constant β_0 (nor the effects of other level-2 variables). Unbiased results for these parameters are obtained with the random-effects estimator. When $\rho(x, \alpha) = 0.3$ and $\rho(x, \eta) = 0$ (case (ii)), the random-effects estimator

Table 2.   Results of simulation study to examine bias fixed-effects (FE) and random-effects (RE) estimator for level-1 and level-2 endogeneity. True values: $\beta_0 = 10$, $\beta_1 = 2$, $\sigma_\eta^2 = 1$ and $\sigma_\alpha^2 = 1$.

		Case			
		(i)	(ii)	(iii)	(iv)
FE	β_0	–	–	–	–
	β_1	1.99 (0.04)	2.00 (0.04)	2.43 (0.04)	2.42 (0.04)
RE	β_0	10.04 (0.21)	8.87 (0.29)	7.88 (0.20)	5.79 (0.30)
	β_1	1.99 (0.04)	2.23 (0.05)	2.42 (0.04)	2.84 (0.06)
	σ_α^2	1.01 (0.14)	0.10 (0.02)	0.99 (0.13)	0.00 (0.00)
	σ_η^2	1.00 (0.04)	1.00 (0.04)	0.90 (0.03)	0.91 (0.04)

is biased upward and σ_α^2 exhibits a severe downward bias, but fixed-effects estimation is possible for β_1 and an unbiased estimate for σ_η^2 can be obtained. When $\rho(x,\alpha) = 0$ and $\rho(x,\eta) = 0.3$ (case (iii)), both the fixed-effects and the random-effects estimator yield biased results for the regression parameters and similar conclusions hold for σ_η^2. However, σ_α^2 can be estimated consistently in this case. Finally, when all independence assumptions are violated (case (iv)), it can be seen that both fixed-effects and random-effects estimation yields biased results for the regression parameters. The bias in the fixed-effects estimator for case (iv) is independent of the presence of $X\alpha$-dependency. It can be seen that random-effects estimation yields an even larger bias in this case. In all replications, the estimate of σ_α^2 was negative, and therefore set to 0. The bias in the estimate of σ_η^2 is approximately equal to its bias for case (iii).

These results indicate clearly that one should consider carefully whether to use random-effects estimation when there are reasons to assume that independence may not hold. Even a moderate (positive) correlation between x and α in model (2) induces in this case an (upward) bias of approximately 10% in the random-effects estimator for β_1 and an approximately 90% downward bias in $\hat{\sigma}_\alpha^2$. Dependencies between the regressors and α can be accommodated by using a fixed-effects estimation. However, failure to correct for dependencies between regressors and η leads to biases in *both* the random-effects and the fixed-effects estimator. A moderate positive correlation between the regressor and η induces a significant upward bias in both the fixed- and random-effects estimate for β_1. Finally, when the regressor is correlated with *both* α_i and η_{ij}, the bias in the random-effects estimator for the regression parameters is even larger and under case (iv) it would be concluded incorrectly that no random effects are present in the data. The following sections focus on the case when only $X\alpha$- but no $X\eta$-dependencies are present.

3 The case of level-2 ($X\alpha$) dependencies only

3.1 *Testing for $X\alpha$-dependencies*
In this section we first discuss two test statistics to examine $X\alpha$-dependencies. It is assumed that no $X\eta$-dependency is present. In case this type of dependency cannot be rejected, we present and illustrate alternative estimators.

HAUSMAN and TAYLOR (1981) show that the multilevel structure of the data and the presence of a consistent estimator regardless of the correlation between regressors and α_i (but with X and η independent), facilitate tests for this type of endogeneity in model (1) using the general idea of a Hausman test (HAUSMAN, 1978). This Hausman test statistic can be computed as follows:

$$H_\alpha = (\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE})' \hat{\Sigma}^{-1} (\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}), \qquad (3)$$

where $\hat{\Sigma}$ is an estimate of the covariance matrix of $\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}$ and computed as $\widehat{\text{cov}}(\hat{\boldsymbol{\beta}}_{FE}) - \widehat{\text{cov}}(\hat{\boldsymbol{\beta}}_{RE})$. The resulting test statistic H_α can be shown to have a chi-square distribution under the null hypothesis of independence of X, Z and α_i. If the

Table 3. Results of H_α-test and Mundlak's approach ($\alpha_i = \pi \bar{x}_i + \xi_i$). True values: $\beta_0 = 10$, $\beta_1 = 2$, $\sigma_\eta^2 = 1$ and $\sigma_\alpha^2 = 1$.

		Case	
		(i)	(ii)
H_α-test		4%	100%
Mundlak	β_0	9.90 (1.85)	−11.17 (0.86)
	β_1	1.99 (0.04)	2.00 (0.04)
	π	0.03 (0.38)	4.23 (0.18)
	σ_α^2	1.01 (0.14)	0.10 (0.02)
	σ_η^2	1.00 (0.04)	1.00 (0.04)

null hypothesis is rejected, the fixed-effects estimator should be used. A great advantage of multilevel over single level applications is the possibility to test for regressor–error disturbances of this type. This is not possible in single-level applications, as there is no estimator that is consistent under both the null hypothesis and alternative hypothesis when IVs are not available.

3.2 Mundlak's approach for $X\alpha$-dependencies

One approach for investigating potential correlations between X and the random effects α_i is to model the dependence between α_i and the regressors explicitly. MUNDLAK (1978) suggests the inclusion of group means by estimating $\alpha_i = \bar{X}_i \pi + \xi_i$. SNIJDERS and BOSKER (1999) argue that the inclusion of group means as explanatory variables in multilevel models can yield interesting substantive results. It can be shown that the test proposed by HAUSMAN and TAYLOR (1981) and Mundlak's approach, are closely related and, in fact, yield numerical identical results (BALTAGI, 2001, p. 65–72). Modeling this dependence explicitly allows for unbiased random-effects estimation for $\boldsymbol{\beta}$, regardless of whether X and α are independent or not. This approach is attractive when fixed-effects estimation is undesirable, but X_i and α_i cannot be assumed independent. However, this procedure does not yield unbiased estimates for level-2 effects/parameters (γ and σ_α^2).

These methods are illustrated in Table 3 where we present the results for the H_α-test and Mundlak's approach. The data were simulated according to the same design as in the previous simulation study (model (2)). It can be seen that when there is no regressor–error dependency (case (i)), the proportion of replication in which the H_α-test rejects the null hypothesis is very close to the nominal P-value of 5%. With a correlation between x and α of 0.3, the null hypothesis of no level-2 ($X\alpha$) dependency is rejected in all replications. The same conclusions follow from Mundlak's π, which is significantly different from zero for $\rho_{x,\alpha} = 0.3$ but not for $\rho_{x,\alpha} = 0$. Furthermore, random-effects estimation in Mundlak's model allows for unbiased estimates of the level-1 predictor, but the constant (and other potential level-2 predictors) cannot be estimated unbiasedly. The same holds for the variance σ_α^2, but σ_η^2 can be estimated unbiasedly. In the next section we present a more satisfying solution to the problem when $X\alpha$-dependencies, but no $X\eta$-dependencies, are present.

Table 4. Results of Hausman–Taylor (HT) estimator for $\rho_{x_2\alpha} = 0.3$ and $\rho_{z_2\alpha} = 0.3$, but no level-1 dependencies (case (ii)). True values: $\beta_0 = 10$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 2$, $\sigma_\eta^2 = 1$ and $\sigma_\alpha^2 = 1$.

	FE	RE	HT
β_0	–	9.25 (0.34)	10.06 (0.97)
β_1	2.00 (0.05)	1.59 (0.08)	2.00 (0.05)
β_2	2.00 (0.04)	2.38 (0.07)	2.00 (0.04)
γ_1	–	1.22 (0.17)	2.02 (0.44)
γ_2	–	2.40 (0.11)	1.97 (0.44)
σ_η^2	1.00 (0.04)	1.00 (0.04)	1.00 (0.04)
σ_α^2	–	0.01 (0.01)	1.13 (0.36)

3.3 *The Hausman–Taylor estimator under Xα-dependencies*

Although Mundlak's approach allows for random-effects estimation, no unbiased results can be obtained for the level-2 (group-specific) variables. As a solution, HAUSMAN and TAYLOR (1981) suggested an estimator that consistently and efficiently estimates both level-1 and level-2 parameters. It requires a priori knowledge about which of the level-1 and level-2 regressors are uncorrelated with the random components. Let $X_{ij} = [X_{1ij} : X_{2ij}]$ and $Z_i = [Z_{1i} : Z_{2i}]$, where the variables in sets X_1 and Z_1 are assumed to be uncorrelated with α_i and all regressors are assumed to be independent of η_{ij}. The idea is that X_{1ij} and Z_{1i} serve as their own instruments; $X_{2ij} - \bar{X}_{2i}$ can be used as instruments for X_{2ij} (as in the fixed-effects approach), and $\bar{X}_{1i}$ serves as instrument for Z_{2i}. To identify all the regression parameters, the number of variables contained in set X_1 needs to be at least as large as the number of variables in set Z_2. An attractive feature of the Hausman and Taylor estimator is that no external instruments (i.e. variables that are not included in the main regression equation) are needed, as this estimator constructs instrument from available data ('internal' instruments). More recent studies suggest modifications (to improve efficiency) of the Hausman–Taylor estimator, see ARELLANO and BOVER (1995).

Table 4 illustrates the Hausman–Taylor estimator. The previously considered model to generate the data is extended as follows:

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \gamma_1 z_{1i} + \gamma_2 z_{2i} + \alpha_i + \eta_{ij} \tag{4}$$

for $i = 1,...,150$ and $j = 1,...,10$. We specify x_1 and z_1 to be independent of the random components. x_2 and z_2 are related to α ($\rho_{x_2\alpha} = \rho_{z_2\alpha} = 0.3$), but independent of η_{ij} (i.e. case (ii)). Table 4 contains the means and standard deviations of the estimated parameters computed across 250 simulation replications. As can be seen, the fixed-effects (FE) estimator yields consistent results for level-1 effects, but no estimator for level-2 effects can be obtained. The random-effects (RE) estimator yields biased results for all regression parameters and σ_α^2, which is in agreement with the results in Table 2. The Hausman–Taylor estimator uses the additional information that x_1 and z_1 are exogenous. These 'internal' instruments can be used to estimate the effects of all regression parameters consistently. Furthermore, an approximately unbiased estimate for σ_α^2 can be obtained. In all cases, σ_η^2 can be estimated unbiasedly.

The Hausman–Taylor estimator is very powerful as it does not require external instruments. We agree with VERBEEK (2000) that despite this obvious advantage, the method has played a surprisingly minor role in empirical work. In practice one does not know which X and Zs are independent of the α, but it is possible to test for this assumption (HAUSMAN and TAYLOR, 1981).

In this section we assumed independence of regressors and η_{ij}. Unfortunately, the methods presented in this section become unreliable and yield incorrect conclusions in the presence of $X\eta$-dependencies. Similar observations were made for the fixed- and random-effects estimators in Section 2. This is illustrated and discussed in the following section.

4   Limitations in the presence of level-1 ($X\eta$)-dependencies

This section considers two problems in using the methods discussed so far. First, as noted when discussing the results in Table 2, both random-effects and fixed-effects estimation fails when endogeneity arises from level-1 dependencies (case (iii) and (iv)). Second, although successful in testing and solving for $X\alpha$-dependencies, we will show that the H_α-test, Mundlak's approach, and the Hausman–Taylor estimator also break down in this case.

In Section 2 we illustrated the consequences of using the fixed-effects and the random-effects estimator when regressors are correlated with the lowest level error term η_{ij}. It was illustrated that even a small correlation between x and η in model (2) induced biases in both the fixed- and random-effects estimators. Similar limitations apply to the H_α-test and Mundlak's approach discussed in Sections 3.1 and 3.2. Based on model (2), the simulation results in Table 5 illustrate this situation. First, it can be seen that a situation with endogeneity at the first level ($X\eta$-dependency) but no $X\alpha$-dependency, cannot be detected by the H_α-test and Mundlak's approach (case (iii)). This is not surprising, as the test is not designed for investigating this hypothesis. However, the estimates for both β_1 and σ_η^2 are still significantly biased due to $X\eta$-dependencies. Researchers who are not aware of potential endogeneity problems at the first level may incorrectly conclude from these tests that either

Table 5.   Results of H_α test and Mundlak's approach ($\alpha_i = \pi\bar{x}_{i.} + \xi_i$).
True values: $\beta_0 = 10$, $\beta_1 = 2$, $\sigma_\eta^2 = 1$ and $\sigma_\alpha^2 = 1$.

		Case	
		(iii)	(iv)
H_α-test		6%	100%
Mundlak	β_0	8.00 (1.95)	−13.34 (0.20)
	β_1	2.42 (0.04)	2.42 (0.04)
	π	−0.03 (0.39)	4.25 (0.05)
	σ_α^2	0.99 (0.13)	0.00 (0.00)
	σ_η^2	0.90 (0.03)	0.91 (0.04)

Table 6. Results of Hausman–Taylor (HT) estimator for $\rho_{x_2\eta} = 0.3$ and $\rho_{z_2\eta} = 0.3$. True values: $\beta_0 = 10$, $\beta_1 = \beta_2 = \gamma_1 = \gamma_2 = 2$, $\sigma_\eta^2 = 1$ and $\sigma_\alpha^2 = 1$.

	FE	RE	HT
β_0	–	8.48 (0.54)	9.64 (1.27)
β_1	1.58 (0.06)	1.57 (0.06)	1.58 (0.06)
β_2	2.21 (0.03)	2.21 (0.02)	2.21 (0.03)
γ_1	–	1.59 (0.13)	1.78 (0.23)
γ_2	–	2.19 (0.08)	1.99 (0.21)
σ_η^2	0.95 (0.04)	0.95 (0.04)	0.95 (0.04)
σ_α^2	–	0.95 (0.13)	1.06 (0.19)
H_α-test		15%	

fixed- or random-effects estimation can be used, although, in fact, both methods yield biased results. When $X\alpha$-dependencies and $X\eta$-dependencies are present (case (iv)), the H_α-test and Mundlak's π diagnose the $X\alpha$-dependency. Given that $X\alpha$-dependency is detected, one should now use fixed-effects estimation (or the Hausman–Taylor estimator). However, it was seen in Table 2 that in the presence of $X\eta$-dependencies fixed-effects estimates for β are biased. The researcher in this case correctly concludes that $X\alpha$-dependencies are present, but misses the $X\eta$-dependencies and, henceforth, still uses biased estimates.

The same fallacious conclusion follows from the Hausman–Taylor estimator based on internal instrumental variables, as can be seen from Table 6. These results are based on model (4), where x_1 and z_1 are specified to be independent of all random components, but x_2 and z_2 are correlated with η_{ij} (but not with α_i). The table shows that both the fixed-effects and the random-effects estimators yield biased results. The H_α-test does not diagnose this type of endogeneity and rejects the null hypothesis in 15% of the cases, when the nominal rate is at 5%. Thus, importantly, this test indicates too often that there is a $X\alpha$-dependency whereas in fact there is none, as the dependency is caused by correlation between X and η. The Hausman–Taylor estimator is also biased in general, but because x_1 is truly exogenous it is a valid instrument for z_2, and the Hausman–Taylor estimate for γ_2 is unbiased. The bias in the estimate for σ_η^2 is small, as is the one observed for σ_α^2.

We conclude that when endogenous regressors are present at the lowest level of the hierarchical model, caused by correlations between X and η, all available tests and estimators presented in Section 3 yield invalid inferences. In the next section we discuss possible solutions to this problem.

5 Testing and solving for *Xη*-dependencies

5.1 *External instruments*
We consider potential remedies to the situation where $X\eta$-dependencies are present in the form of 'classical' IV methods. These methods are similar to the Hausman–Taylor estimator, but require the availability of 'external' instruments.

External instrumental variables are desirable for an unbiased and consistent estimation when $X\eta$-dependencies are present in the data (BOWDEN and TURKINGTON, 1984; WOOLDRIDGE, 2002). The main ideas behind these estimators are similar to the ones of classical IV estimators developed for cross-sectional situations, with an additional step to account for nonspherical disturbances due to the hierarchical structure (WOOLDRIDGE, 2002). Two multilevel IV estimators that yield unbiased estimation of the parameters in model (1) in the presence of $X\eta$-dependencies are the (multilevel) two- and three-stage least-squares (SLS) estimators (see Appendix 2), where the latter estimator takes the random error component structure into account yielding a potential more efficient estimator (IM *et al.*, 1999, WOOLDRIDGE, 2002; BOWDEN and TURKINGTON, 1984). In the following, we will use the multilevel 2SLS estimator to illustrate the usefulness of external IV estimators when the $X\eta$-independence assumption is violated. We also show how this estimate can be used to construct another Hausman-based test (H_η-test) to test for $X\eta$-independencies. The results of this test can be used to decide whether fixed-effects, random-effects or the Hausman–Taylor estimator, or multilevel (external) IV estimators should be used.

Using model (2) we illustrate the multilevel IV estimator with one level-1 instrument. The endogenous regressor is now simulated as $x_{ij} = c + v_{ij} + \phi_{ij}$, where $\rho_{\phi,\eta} = 0.3$, c is a constant, and v_{ij} is the instrument generated independent of all error terms. In addition, a Hausman-based test is computed that compares the multilevel IV estimate for (β_0, β_1) with the random-effects estimate for (β_0, β_1) (or the fixed-effects estimate for β_1). The results are presented in Table 7. Note that in Table 7 we estimate σ_ϵ^2, which is the variance of $\epsilon_{ij} = \alpha_i + \eta_{ij}$, from the residuals computed from the IV regression. The table shows that once valid external instruments are available, we obtain approximately unbiased estimates for the model parameters. Furthermore, these estimates are unbiased regardless of $X\alpha$-independence (case (iii) vs. (iv)). The H_η-test based on these estimates detects both case (iii) and case (iv) endogeneity, indicating that the multilevel (external) IV estimators should be used. A disadvantage of this method is that it is less efficient than fixed- and random-effects estimators. Furthermore, valid instruments that have no direct effect on y and explain a substantial part of the variance in x, have to be available, which is often difficult in empirical work.

Although external IVs can be useful for dealing with $X\eta$-dependencies, it should be noted that IV estimators can be seriously biased in small samples and may exhibit

Table 7.   Results for multilevel IV for case (iii) and case (iv) violations. True values: $\beta_0 = 10$, $\beta_1 = 2$, $\sigma_\eta^2 = 1$ and $\sigma_\alpha^2 = 1$.

		Case		
		(i)	(iii)	(iv)
H_η-test		3.2%	96.4%	100%
Multilevel IV	β_0	10.00 (0.21)	9.98 (0.14)	9.98 (0.20)
	β_1	1.99 (0.07)	2.00 (0.07)	2.01 (0.07)
	σ_ϵ^2	1.99 (0.12)	1.98 (0.14)	1.99 (0.15)

poor asymptotic properties when weak instruments are applied. An instrument is said to be 'weak' when it explains none or only a small part of the variance in the endogenous regressor (i.e. it is only weakly correlated). There is a considerable literature that investigates the potential pitfalls in IV estimation when weak instruments are used and several recommendations to deal with these problems are made (STAIGER and STOCK, 1997; BOUND *et al.*, 1995; NELSON and STARTZ, 1990; KLEIBERGEN and ZIVOT, 2003).

To address the problem of weak instruments, HAHN and HAUSMAN (2002) recently developed a test for the validity of instruments. Their approach is also based on the general Hausman specification test approach (HAUSMAN, 1978). The test statistic is fairly simple to compute and is shown to have a *t*-distribution under the null hypothesis. Rejection of the null hypothesis might indicate a failure of the orthogonality assumption of the instruments or that the instruments are weak. HAHN and HAUSMAN (2002) suggested a two step approach, based on this test, to decide which IV estimator, or none, should be used. This approach may provide a helpful guide in guarding against weak instruments. Furthermore, it is relatively straightforward to use and it could prevent the researcher from relying on results obtained with weak instruments.

However, although IV methods are attractive in theory, they can be difficult to apply in practice because it may prove difficult to locate 'good' IVs as indicated by the Hanh and Hausman test. As a possible solution, we next consider LEWBEL's (1997) method for computing instrumental variables from the data at hand and demonstrate that this method could potentially be extended to multilevel models with general $X\eta$-dependencies.

5.2 *Internal instruments: Lewbel's approach*

LEWBEL (1997) provides a method for constructing internal instruments when $X\eta$-dependencies exist. This approach has been proposed originally in the context of measurement error models, but we argue that it is also useful in the context of general correlated-regressor error. To the best of our knowledge, the issue of constructing internal instruments from available data in multilevel models where $X\eta$-dependencies are present has not been addressed before. LEWBEL's (1997) idea is based on the observation that when the endogenous regressor in model (2) has a skewed distribution, the following transformations of the available data may yield valid instruments:

$$\tilde{v}_{1ij} = (y_{ij} - \bar{y})(x_{ij} - \bar{x})$$
$$\tilde{v}_{2ij} = (y_{ij} - \bar{y})^2 \qquad\qquad (5)$$
$$\tilde{v}_{3ij} = (x_{ij} - \bar{x})^2$$

The results in Table 8 illustrate the internal instrumental variable approach for model (2) and compare it with the external instrumental variable approach in the previous section. The same simulation data as in Table 7 was used, where the endogenous regressor was generated as $x_{ij} = c + v_{ij} + \phi_{ij}$, with $\rho_{\phi,\eta} = 0.3$ and v_{ij} is

Table 8. Results of multilevel and Lewbel's internal IVs for cases (iii) and (iv) compared with (i). True values: $\beta_0 = 10$, $\beta_1 = 2$ and $\sigma_\alpha^2 = \sigma_\eta^2 = 1$.

		Case		
		(i)	(iii)	(iv)
Multilevel IV	β_0	10.00 (0.21)	9.98 (0.14)	9.98 (0.20)
	β_1	1.99 (0.07)	2.00 (0.07)	2.01 (0.07)
	σ_ϵ^2	1.99 (0.12)	1.98 (0.14)	1.99 (0.15)
Lewbel IV	β_0	10.05 (0.75)	9.86 (0.61)	9.81 (0.58)
	β_1	1.98 (0.28)	2.05 (0.23)	2.07 (0.22)
	σ_ϵ^2	2.04 (0.16)	2.00 (0.17)	1.97 (0.21)

the exogenous instrument. We compare Lewbel's approach with the benchmark, where we assume that the v_{ij} are observed instruments. Thus the results from the multilevel IV estimator in Table 8 are the same as in Table 7, and were obtained by using v_{ij} as 'observed' instruments, whereas the Lewbel approach uses the constructed instruments in (5) instead. Table 8 indicates that the Lewbel IVs may yield approximately unbiased results. Using these IVs is less efficient than using the true observed IVs, which is not surprising as the former uses less information. Nevertheless, the Lewbel approach appears to be quite promising since it provides a method to construct instruments from the available data. These instruments can either be used alone, or to augment a set of existing instruments in order to improve efficiency.

6 Discussion and future research

Although the previous discussion may suggest that regressor and random components dependencies can be adequately addressed in multilevel models, much care is required in using these methods in actual applications. First, the estimation methods and test procedures to solve and test for $X\alpha$-dependencies rely critically on the independence of X and η. Second, methods that rely on IVs are known to be biased in small samples and standard asymptotic results break down when instruments are weak (i.e. they are poorly correlated with X). This holds in particular for the IV-based methods to solve for $X\eta$-dependencies and for the Hausman–Taylor estimator to solve for $X\alpha$-dependencies.

Although the issues about the validity and the number of instrumental variables have primarily been investigated in cross-sectional applications, it is clear that they are relevant for multilevel applications as well. For instance, when for the simulation study in Table 4 the instrument $\bar{x}_{1i.}$ is weakly correlated with the endogenous regressor z_2 as $z_{2i} = 0.01 \times \bar{x}_{1i.} + 0.01 \times z_{1i} + \zeta_i$, where ζ_i is a random component correlated with α_i, and with all other input parameters unchanged, the Hausman–Taylor estimator yields $\hat{\gamma}_2 = 3.45 (21.72)$ and $\hat{\sigma}_\alpha^2 = 238.01 (2611.22)$. Similar observations can be made for the 'external' multilevel IV estimates concerning Table 7. To deal with these problems, BOUND *et al.* (1995) suggest that the R^2 or the F-statistic of the regression of the endogenous regressors on the instruments serve as rough guides

to the quality of the instruments and should routinely be reported. The HAHN and HAUSMAN (2002) test or the method suggested by DONALD and NEWEY (2001) to choose the number of instruments could potentially be extended to serve as a guide for identifying and selecting 'valid' instruments for the Hausman–Taylor estimator or multilevel IV estimators.

Further, it is often suggested in cross-sectional applications to use the 'limited information maximum likelihood' (LIML) estimator instead of least squares estimators, since it is found to be less sensitive to 'weak' instruments (e.g. DAVIDSON and MACKINNON, 1993; STAIGER and STOCK, 1997). To the best of our knowledge, this issue has not been addressed for multilevel models, but perhaps it should be because it may lead to improved results for the Hausman–Taylor estimator or the multilevel IV estimators discussed in Section 5.

Finally, the Lewbel approach has been shown to yield consistent results for a simple multilevel model with $X\eta$-dependency. This method deserves more attention and could potentially be powerful in situations where no or weak instruments are available. The performance of this method depends critically on its underlying assumptions as is shown in LEWBEL (1997) and WANSBEEK and MEIJER (2000). Most importantly, the method may be sensitive to outliers as it relies on third-order moments. Furthermore, the constructed instruments are weak when the distribution of the endogenous regressor is not strongly skewed. It is well known that in this case IV estimators can be seriously biased (STAIGER and STOCK, 1997; BOUND *et al.*, 1995; WANSBEEK and MEIJER, 2000). As a result, additional work is needed to determine the exact conditions under which this approach can be used effectively in multilevel applications.

In some applications where endogeneity arises, however, the nature of the data generating process itself suggests suitable instruments. This holds in particular for measurement error models, autoregressive models, and simultaneous equation models. Possible approaches for measurement error models are discussed by WANSBEEK and MEIJER (2000), CARROLL *et al.* (1995) and BOWDEN and TURKINGTON (1984). These models can be estimated using IV techniques, for instance by using other (potentially) mismeasured variables (see WHITE, 2001). Another method is based on WALD (1940), which assumes that the observations can be divided into groups. This classification should be independent of the error terms and discriminate between high and low values of the unobservable true construct (see also MADANSKY, 1959). LEWBEL's (1997) idea presented in Section 5.2 was originally proposed to solve for measurement error problems. We showed however that that approach can be fruitfully applied in the analysis of the general IV problems as well and deserves more attention. In autoregressive models one can often use lagged dependent or independent variables as instruments (see for instance WHITE, 2001; WOOLDRIDGE, 2002). Similarly, in simultaneous equations models instruments for each equation can be obtained from the set of excluded exogenous variable for that equation (see for instance GREENE, 1997).

Our discussion of various methods did not address estimation methods in (general) random coefficient and non-linear models (like probit- or logit models)

having endogenous regressors. In both cases, however, a similar reasoning applies as for linear (random intercept) models. BOWDEN and TURKINGTON (1984) discuss IV approaches for nonlinear models with additive disturbances (i.e. $y = g(\theta, x) + \epsilon$). Techniques developed for linear models, in particular (generalized) method of moments (GMM) techniques, can be used to estimate these models. BLUNDELL and POWELL (2000) investigate endogeneity issues in several generalizations of the linear model. These authors discuss the extent to which commonly used methods in linear models can be applied to the generalized models and show that the methods' applicability depends on the structural form.

Random coefficient models assume that differences between the level-2 objects are not only reflected by different intercepts as in model (1) but also by different slope coefficients. These models can be written as $y_i = X_i\boldsymbol{\beta}_i + \eta_i$, where $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\mu}_{\beta,i}$, with $\boldsymbol{\mu}_{\beta,i}$ a random component having mean zero, $\mathrm{E}(\boldsymbol{\mu}_{\beta,i}\boldsymbol{\mu}'_{\beta,i}) = \Delta$ and $E(\boldsymbol{\mu}_{\beta,i}\boldsymbol{\mu}'_{\beta,j}) = 0$ for $i = 1,\dots,n$, and $j \neq i$. As for random intercept models, the question whether to use a fixed-effects approach (in fact a seemingly unrelated regression framework), or a random-effects approach (a random-effects framework), depends on potential correlation between the random coefficients and the explanatory variables. If dependencies are present, which is sometimes referred to as 'heterogeneity bias', the random-effects estimator of $\boldsymbol{\beta}$ is biased and a fixed-effects approach should be used. PUDNEY (1978) provides a test for the null hypothesis that the explanatory variables are not correlated with the random coefficients. This test is based on the sample covariance between the (standard) least-squares estimators for $\boldsymbol{\beta}_i$ and the means of the explanatory variables for each individual (see also CHAMBERLAIN, 1982).

In general, we conclude that much needs to be done before problems of endogeneity in multilevel models can be adequately addressed. We showed that even small violations of the independence structure result in biased estimates for parameters of interest. In Table 1 we distinguished four cases of (in)dependence relations among the level-1 regressors and the random components. No distinction was made between level-1 and level-2 regressors. If this distinction is introduced, fifteen instead of three possible cases of violations of the independence assumptions emerge. Each of these combinations could lead to different biases in the estimators discussed in this article. Although it is possible to apply the methods presented here to address the various cases, detailed studies are necessary to assess their performance in practice. Clearly, endogeneity problems require much more attention than they receive in current applications of multilevel models.

Appendix 1: Classical instrumental variables estimation

The (single level) standard linear regression model for n observations is given by

$$y = X\boldsymbol{\beta} + \epsilon, \tag{6}$$

where $X \in \mathbb{R}^{n \times k}$ are the regressors, $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$ are the (unobserved) and identically

independently distributed errors with mean 0 and variance $\sigma_\epsilon^2 I_n$, and y is an $n \times 1$ vector of dependent variables. The ordinary least-squares (OLS) estimator is BLUE and is given by $\hat{\boldsymbol{\beta}}_{\text{OLS}} = (X'X)^{-1}X'y$. If $\text{E}(\epsilon|X) = 0$, $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ is unbiased (e.g. WHITE, 2001).

In large samples instrumental variables techniques can be used when this assumption is not met. Instrumental variables (IVs), collected in matrix $V \in \mathbb{R}^{n \times m}$, should be uncorrelated with the error term ϵ, i.e. $\text{E}(\epsilon|V) = 0$, meaning that the instruments cannot have a direct effect on y (external instruments). Furthermore, the instruments should explain part of the variability in the endogenous regressors. Once instruments are available (and $m \geq k$), two-stage least-squares techniques, for example, can be used to obtain better estimates of β. The 'classical' IV estimator for model (6) is computed as $\hat{\boldsymbol{\beta}}_{\text{IV}} = (X'P_V X)^{-1} X' P_V y$, where $P_V = V(V'V)^{-1}V'$ (BOWDEN and TURKINGTON, 1984; WHITE, 2001; GREENE, 1997).

When 'valid' instruments are available, a Hausman test (HAUSMAN, 1978) can be used to test for regressor error dependencies in model (6). Under $H_0 : \text{E}(\epsilon|X) = 0$, the Hausman test-statistic computed as $H = (\hat{\boldsymbol{\beta}}_{\text{IV}} - \hat{\boldsymbol{\beta}}_{\text{OLS}})' \Sigma^{-1} (\hat{\boldsymbol{\beta}}_{\text{IV}} - \hat{\boldsymbol{\beta}}_{\text{OLS}})$, where $\Sigma = \text{var}(\hat{\boldsymbol{\beta}}_{\text{IV}}) - \text{var}(\hat{\boldsymbol{\beta}}_{\text{OLS}})$, has a χ^2 distribution. For a more detailed discussion on how to obtain an estimate for Σ and how to determine the degrees of freedom (d.f.), see e.g. GREENE (1997).

Appendix 2: Estimation for the hierarchical linear model

The parameters $\boldsymbol{\beta}$ in the multilevel model given in (1), can be estimated by either fixed-effects (assume α_i to be fixed parameters for $i = 1,\dots,n$) or random-effects (assume the α_i to be drawn from a distribution) methods. The fixed-effects estimator, also known as the within-groups- or the covariance-estimator, for $\boldsymbol{\beta}$ can be computed as a simple regression on the transformed equation (7) which is obtained by averaging (1) across j for every i, and subtracting the result from (1), resulting in

$$y_{ij} - \bar{y}_i = (X_{ij} - \bar{X}_i)\boldsymbol{\beta} + (\eta_{ij} - \bar{\eta}_i), \tag{7}$$

where $\bar{y}_i = (1/n_j) \sum_j y_{ij}$ and similarly for $\bar{X}_i$ and $\bar{\eta}_i$. Now α_i and $Z_i\gamma$ drop out, and thus γ is not identifiable from (7). An alternative would be to replace all group variables by dummy variables and apply OLS on the equation $y_{ij} = \sum_i \alpha_i d_{ij} + X_{ij}\boldsymbol{\beta} + \eta_{ij}$, where $d_{ij} = 1$ if $i = j$ and 0 otherwise. The resulting estimator for $\boldsymbol{\beta}$ is known as the least-squares dummy variable (LSDV) estimator and is exactly identical to the fixed-effects estimator for $\boldsymbol{\beta}$ from (7). For consistent and unbiased estimation in (7) the ordinary least-squares (OLS) estimator can be used, if the constructed regressors $X_{ij} - \bar{X}_i$ are independent of the constructed error $\eta_{ij} - \bar{\eta}_i$. This implies that $\text{E}(X_{ij}\eta_{il}) = 0$ for all i, j, l.

The random-effects estimator provides an important alternative under the assumption that the α_is are i.i.d. random variables. Now $\epsilon_{ij} = \alpha_i + \eta_{ij}$ in (1) is the

composite (random) error term. The OLS estimator for β and γ is consistent and unbiased, but not fully efficient. Combining all observations we can rewrite model (1) as

$$y = X\beta + Z\gamma + \epsilon = W\delta + \epsilon, \tag{8}$$

where $W = [X : Z]$ and $\delta = (\beta', \gamma')'$, and the other symbols defined accordingly to stacking. For known Ω, where $\Omega = \text{var}(\epsilon_i)$, $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})'$, the generalized least-squares (GLS) estimator for β and γ, given by $\hat{\delta}_{GLS} = \left(W'(I_n \otimes \Omega^{-1})W\right)^{-1} W'(I_n \otimes \Omega^{-1})y$ is efficient. However, when Ω is not known, it needs to be estimated, yielding a feasible GLS estimator. A feasible GLS estimator can be obtained in several ways. We use the method explained in VERBEEK (2000, p. 317). The GLS estimator is shown to be equal to a weighted average between the fixed-effects estimator computed from (7) and the so-called- between estimator, which is the OLS estimator in the model

$$\bar{y}_i = \bar{X}_i\beta + Z_i\gamma + \alpha_i + \bar{\eta}_i \tag{9}$$

for $i = 1, \ldots, n$. The latter estimator ignores the within-group information and exploits only differences between groups. For more details on the computation of the weighting matrix, see VERBEEK (2000), HSIAO (1986) and BALTAGI (2001). Several other random-effects estimation procedures for model (1) are available that include the iterative GLS (IGLS) approach, (restricted) maximum likelihood (REML), or Bayesian procedures (see e.g. GOLDSTEIN, 1995; LONGFORD, 1993).

From standard OLS results, it follows that the between estimator for β and γ from (9) is consistent and unbiased when the constructed regressors $\bar{X}_i$ and Z_i are independent of α_i and $\bar{\eta}_i$. The fixed-effects estimator from (7) is consistent and unbiased when $E(X_{ij}\eta_{il}) = 0$ for all i, j, l. When both conditions hold, the random-effects estimator for β and γ is consistent and unbiased.

In the simulations studies, the variance for η_{ij} was estimated as the sum of the squared residuals from model (7) divided by $n(m - 1) - (k + l)$, where $n_j = m$ for all j in our case. The variance for α_i is estimated as $\hat{\sigma}_\alpha^2 = \hat{\sigma}_B^2 - \frac{1}{m}\hat{\sigma}_\eta^2$, where $\hat{\sigma}_B^2$ is estimated from the squared residuals divided by n from (9).

Multilevel instrumental variables estimators
Two IV estimators that yield unbiased estimation of the parameters in model (8) in the presence of $X\eta$-dependencies, are the multilevel 2SLS estimator, given by

$$\delta_{2SLS} = (W'P_V W)^{-1} W'P_V y, \tag{10}$$

where $P_V = V(V'V)^{-1}V$, and the multilevel 3SLS estimator, given by

$$\delta_{3SLS} = (W'\tilde{P}_V W)^{-1} W'\tilde{P}_V y, \tag{11}$$

with $\tilde{P}_V = V(V'(I_n \otimes \hat{\Omega})V)^{-1}V'$ and where $\hat{\Omega}$ can be estimated from the residuals from a 2SLS estimation. As in appendix 1, V is a set of (external) instruments. For more details, see e.g. WOOLDRIDGE (2002), IM *et al.* (1999) and BOWDEN and TURKINGTON (1984).

References

ANGRIST, J. D., G. W. IMBENS and D. B. RUBIN (1996), Identification of causal effects using instrumental variables, *Journal of the American Statistical Association* **91**, 444–455.

ARELLANO, M. and O. BOVER (1995), Another look at the instrumental variables estimation of error-components models, *Journal of Econometrics* **68**, 29–51.

BAGOZZI, R. P., Y. YI and K. D. NASSEN (1999), Representation of measurement error in marketing variables: review of approaches and extension to three-faced designs, *Journal of Econometrics* **89**, 393–421.

BALTAGI, B. H. (2001), *Econometric analysis of panel data*, John Wiley & Sons Ltd., Chichester.

BLUNDELL, R. and J. L. POWELL (2000), *Endogeneity in nonparametric and semiparametric regression models*, Working paper.

BOUND, J., D. A. JAEGER and R. M. BAKER (1995), Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak, *Journal of the American Statistical Association* **90**, 443–450.

BOWDEN, R. J. and D. A. TURKINGTON (1984), *Instrumental variables*, Cambridge University Press, New York.

BRYK, A. S. and S. W. RAUDENBUSH (1992), *Hierarchical linear models, applications and data analysis methods*, Sage Publications, Newbury Park, CA.

CARD, D. (1999), The causal effect of education on earnings, in: O. C. ASHENFELTER and D. CARD (eds), *Handbook of labor economics*, volume 3A, Elsevier Science B.V., North-Holland, 1801–1863.

CARD, D. (2001), Estimating the return to schooling: Progress on some persistent econometric problems, *Econometrica* **69**, 1127–1160.

CARROLL, R. J., D. RUPPERT and L. A. STEFANSKI (1995), *Measurement error in nonlinear models*, Chapman & Hall, London.

CHAMBERLAIN, G. (1982), Multivariate regression models for panel data, *Journal of Econometrics* **18**, 5–46.

DAVIDSON, R. and J. G. MACKINNON (1993), *Estimation and inference in econometrics*, Oxford University Press, New York.

DONALD, S. G. and W. K. NEWEY (2001), Choosing the number of instruments, *Econometrica* **69**, 1161–1191.

GOLDSTEIN, H. (1995), *Multilevel statistical models*, John Wiley & Sons Ltd., New York.

GREENE, W. H. (1997), *Econometric analysis*, Prentice Hall, New Jersey.

HAHN, J. and J. HAUSMAN (2002), A new specification test for the validity of instrumental variables, *Econometrica* **70**, 163–189.

HAMILTON, B. H. and J. A. NICKERSON (2003), Correcting for endogeneity in strategic management research, *Strategic organization* **1**, 51–78.

HAUSMAN, J. A. (1978), Specification tests for econometrics, *Econometrica* **46**, 1251–1271.

HAUSMAN, J. A. and W. E. TAYLOR (1981), Panel data and unobservable individual effects, *Econometrica* **49**, 1377–1398.

HSIAO, C. (1986), *Analysis of panel data*, Cambridge University Press, New York.

IM, K. S., S. C. AHN, P. SCHMIDT and J. M. WOOLDRIDGE (1999), Efficient estimation of panel data models with strictly exogenous explanatory variables, *Journal of Econometrics* **93**, 177–201.

JUDGE, G. G., W. E. GRIFFITHS, R. C. HILL, H. LÜTKEPOHL and T.-C. LEE (1985), *The theory and practice of econometrics*, John Wiley & Sons Inc., New York.

KLEIBERGEN, F. and E. ZIVOT (2003), Bayesian and classical approaches to instrumental variables regression, *Journal of Econometrics* **114**, 29–72.

LEWBEL, A. (1997), Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D, *Econometrica* **65**, 1201–1213.

LONGFORD, N. T. (1993), *Random coefficient models*, Oxford University Press, New York.

MADANSKY, A. (1959), The fitting of straight lines when both variables are subject to error, *Journal of the American Statistical Association* **54**, 173–205.

MUNDLAK, Y. (1978), On the pooling of time-series and cross section data, *Econometrica* **46**, 69–85.

NELSON, C. R. and R. STARTZ (1990), Some further results on the exact small sample properties of the instrumental variable estimator, Econometrica **58**, 967–976.

PLAT, F. W. (1988), *Modelling for Markets: Applications of advanced models and methods for data analysis*, PhD thesis, University of Groningen.

PUDNEY, S. E. (1978), *The estimation and testing of some error components models*, Technical report, London School of Economics.

RUUD, P. A. (2000), *An introduction to classical econometric theory*, Oxford University Press, New York.

SNIJDERS, T. A. B. and R. J. BOSKER (1999), *Multilevel analysis*, SAGE Publications, London.

SPENCER, N. H. and A. FIELDING (1998a), *A comparison of modelling strategies for value-added analyses of educational data*, University of Hertfordshire.

SPENCER, N. H. and A. FIELDING (1998b), *An instrumental variable consistent estimation procedure to overcome the problem of endogenous variables in multilevel models*, University of Hertfordshire.

STAIGER, D. and J. H. STOCK (1997), Instrumental variables regression with weak instruments, *Econometrica* **65**, 557–586.

UUSITALO, R. (1999), *Essays in economics of education*, PhD thesis, University of Helsinki.

VELLA, F. and M. VERBEEK (1998), Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men, *Journal of Applied Econometrics* **13**, 163–183.

VERBEEK, M. (2000), *A guide to modern econometrics*, John Wiley & Sons Ltd., Chichester.

WALD, A. (1940), The fitting of straight lines if both variables are subject to error, *The Annals of Mathematical Statistics* **11**, 284–300.

WANSBEEK, T. and E. MEIJER (2000), *Measurement error and latent variables in econometrics*, Elsevier, Amsterdam.

WHITE, H. (2001), *Asymptotic theory for econometricians*, Academic Press, New York.

WOOLDRIDGE, J. M. (2002), *Econometric analysis of cross section and panel data*, Massachusetts Institute of Technology, Cambridge.