Exact inference for the linear model with groupwise heteroscedasticity

Paul Bekker and Erik Leertouwer*

January, 2000

Abstract

Exact inference on a single coefficient in a linear regression model, as introduced by Bekker (1997), is elaborated for the case of normally distributed heteroscedastic disturbances. Instead of approximate inference based on feasible generalized least squares, exact confidence sets are formulated based on partial rotational invariance of the distribution of the vector of disturbances. The approach is applied to the random-effects and fixed-effects models for panel data.

Keywords: *exact inference, panels, weighted least squares, assumption equivalent inference, heteroskedasticity.*

JEL classifications: *C19, C23.*

---

*Department of Economics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands. E-mail: P.A.Bekker@eco.rug.nl, E.C.Leertouwer@eco.rug.nl. We would like to thank Tom Wansbeek for helpful comments.

1 Introduction

In this paper we will elaborate a nonparametric approach for exact inference in the linear model for normally distributed heteroskedastic disturbances. Bekker (1997) describes exact inference for the linear model for cases where the distribution of the vector of disturbances is invariant under a group of linear transformations. For disturbances with spherical distributions (cf. Chmielewski, 1981), which are invariant under rotations, this approach results in classical exact inference with confidence intervals based on t-distributions. Here, we will apply this approach in a context of groupwise heteroscedasticity. We derive a family of exact inferences for cases where the disturbance vector shows a WLS-structure, such that its distribution is invariant under groupwise rotations. Optimal inference is then based on an intuitively appealing criterion. Thus, we provide exact inference as an alternative to approximate inference based on the asymptotics related to feasible WLS.

Although it is closely related to Bekker (1997), this paper is self-contained. First we describe, similar to the earlier paper, exact inference for the case where the disturbance vector has a spherical distribution. For cases where the disturbances are assumed to be independent, this would amount to Gaussianity. Consequently, Gaussian distributions are the most interesting spherical distributions in practice. However, in the derivation we do not need this additional assumption. For example, we will not assume the existence of moments.

Our aim is to derive exact inference on a single parameter. The inference will take the form of a random function $F(\beta)$, of a scalar β, which is uniformly distributed when evaluated at the true value β_0, say, of the regression coefficient of interest. Exact confidence sets can then easily be formulated as sets $\{\beta \mid F(\beta) \in S\}$, where the Lebesgue measure of S equals the probability of coverage $\beta_0 \in \{\beta \mid F(\beta) \in S\}$. In case of a spherical distribution of the disturbance vector, $F(\beta)$ will be nondecreasing and ranging from 0 to 1.

Next, we will describe a similar approach in a WLS-context. Here we consider subgroups of the rotation group, or the orthonormal group, in order to describe partially spherical or elliptical distributions. This approach leads to many possibilities for functions $F(\beta)$. Based on a minimum (conditional) variance argument we formulate a random function $F(\beta)$ that ranges from 0, as $\beta \to -\infty$, to 1, as $\beta \to \infty$, but does not necessarily increase monotonically with probability 1. Thus, two-sided confidence sets will be bounded, but they need not be convex.

Finally, this approach will be applied to the fixed effects and random effects panel models. In particular, the results allow for exact inference on a coefficient in an error-components or random-effects model, an issue which has not yet been resolved satisfactorily (cf. Taylor (1977), Park and Simar (1994)). The results also allow for new insight in the relation between the random-effects and fixed-effects model (cf. Mundlak (1978)).

We use the following notation. A vector of n ones will be indicated by ι_n, the symbol I_n is used for the $n \times n$ identity matrix. Let A be an $n \times m$ matrix of rank m. Then P_A indicates the projection matrix $A(A'A)^{-1}A'$ and L_A is an $n \times (n-m)$ matrix indicating an orthonormal complement of A, i.e. $A'L_A = 0$ and $L'_A L_A = I_{n-m}$. Notice that $L_A L'_A = I_n - P_A$. For square orthonormal matrices we use the symbol R, i.e. $R' = R^{-1}$.

2 Exact inference based on rotational invariance

To introduce our contributions in the following sections, consider, similar to Bekker (1997), a regression model

$$y = x\beta + X_2\delta + u.$$

Here, y, x, and u are n-vectors and X_2 is an $n \times m$ matrix such that (x, X_2) has full column rank. Conditional on (x, X_2) the distribution of u is assumed to be spherical: u can be rotated without affecting its distribution, i.e. the distribution depends only on $u'u$, and if R is an orthonormal matrix, u and Ru have the same distribution: $u \sim Ru$.

Let

$$\epsilon = X_2\delta + u,$$

and consider orthonormal matrices R that affect the u part of ϵ, not the X_2 part, i.e. $R\epsilon = RX_2\delta + Ru = X_2\delta + Ru$. That is, let P_{X_2} and L_{X_2} be the projection matrix and the orthonormal complement, respectively, related to X_2, as defined in Section 1, and consider the group of linear transformations

$$\mathcal{R} = \{R \mid R = P_{X_2} + L_{X_2}\tilde{R}L'_{X_2};\ \tilde{R}'\tilde{R} = I_{n-m}\}. \tag{1}$$

For an n-vector ϵ^*, the vector $R\epsilon^*$, with $R \in \mathcal{R}$, will be located on the sphere in $\mathbb{R}^n$ with radius $\|\epsilon^*\|$, and the set of vectors $L'_{X_2}R\epsilon^*$ form the sphere in $\mathbb{R}^{n-m}$ with squared radius equal to $\epsilon^{*\prime}(I_n - P_{X_2})\epsilon^*$. Thus we find, conditional on (x, X_2) and $R \in \mathcal{R}$, that ϵ and $R\epsilon$ have identical distributions:

$$\epsilon \sim R\epsilon.$$

Let $\{R_1, \cdots, R_N\}$ be a random sample, independent of ϵ, drawn from $\mathcal{R}$ and let $R_0 = I_n$. That is, the elements of $\{L'_{X_2} R_1 \epsilon^*, \cdots, L'_{X_2} R_N \epsilon^*\}$ are assumed to be independently uniformly distributed over the sphere with squared radius $\epsilon^{*\prime}(I_n - P_{X_2})\epsilon^*$. Thus, the vectors $R_i \epsilon,\ i = 0, \cdots, N$, have identical distributions, but they need not be independent since their shared length may be random.

To achieve independence, consider a set

$$C_{\epsilon^*} = \{R\epsilon^* \mid R \in \mathcal{R}\},$$

which describes an equivalence class related to the group $\mathcal{R}$. Notice that $\epsilon \in C_{\epsilon^*}$ amounts to $P_{X_2}\epsilon = P_{X_2}\epsilon^*$ and $\epsilon'(I_n - P_{X_2})\epsilon = \epsilon^{*\prime}(I_n - P_{X_2})\epsilon^*$. Consequently, conditional on (x, X_2) and $\epsilon \in C_{\epsilon^*}$, the vectors $R_i\epsilon,\ i = 0, \cdots, N$ are independent and identically distributed (i.i.d.) with a uniform distribution over the equivalence class on which we condition.

Now consider a possibly random n-vector z that is a function of (x, X_2) and C_ϵ so that, conditional on (x, X_2) and C_ϵ, the scalars $z' R_i \epsilon,\ i = 0, \cdots, N$, will be i.i.d. For example this holds for $z = x$. Based on this conditioning we find that

$$\#\{i \mid z'\epsilon < z' R_i \epsilon\}/N$$

will be uniformly distributed over $\{0, 1/N, 2/N, \cdots, 1\}$. If z also satisfies the condition

$$z'(I_n - R_i)x > 0, \tag{2}$$

for almost all $R_i \in \mathcal{R}$, we find, similar to Bekker (1997), that

$$F(\beta, z) = \#\{i \mid \frac{z'(I_n - R_i)y}{z'(I_n - R_i)x} < \beta\}/N$$

is uniformly distributed over $\{0, 1/N, 2/N, \cdots, 1\}$. This distribution does not depend on (x, X_2) or C_{ϵ^*}, so it also holds unconditionally.

The confidence limits,

$$c_i(z) = \frac{z'(I_n - R_i)y}{z'(I_n - R_i)x}, \qquad i = 1, \cdots, N, \tag{3}$$

describe the boundaries of $N+1$ elementary $\frac{100}{N+1}\%$—confidence sets that can be combined to form relevant confidence sets. The number N does not depend on the sample size n, and can be increased at will.

For the present group of transformations $\mathcal{R}$, as given in (1), we find that for $R \in \mathcal{R}$

$$(I_n - R) = L_{X_2}(I_{n-m} - \tilde{R})L'_{X_2}.$$

Consequently, a necessary and sufficient condition for (2) to hold is that $L'_{X_2}z$ is a scalar multiple of $L'_{X_2}x$, with positive inner product. That is, if the vectors are not scalar multiples, we could find, on the sphere in $\mathbb{R}^{n-m}$ with squared radius $x'(I_n - P_{X_2})x$, a set of vectors $\tilde{R}L'_{X_2}x$, with positive Lebesgue measure, that are closer, in the Eucledian metric, to $L'_{X_2}z$ than $L'_{X_2}x$ is, so that condition (2) would not be satisfied. On the other hand, for $L'_{X_2}x \neq 0$ and almost all orthonormal $\tilde{R}$ it holds that $x'L_{X_2}\tilde{R}L'_{X_2}x < x'L_{X_2}L'_{X_2}x$ by the Cauchy-

Schwarz inequality. So the confidence limits are in fact given by

$$c_i(x) = \frac{x'(I_n - R_i)y}{x'(I_n - R_i)x}, \qquad i = 1, \cdots, N.$$

However, in the next section we will consider subgroups of $\mathcal{R}$ for which the distinction between z and x becomes relevant. Such vectors z, which are fixed conditional on $\epsilon \in C_{\epsilon^*}$, will be referred to as instruments and, similar to Bekker (1997), as monotonic instruments if they satisfy condition (2) as well. The computation of the confidence limits is briefly discussed in the Appendix.

Bekker shows that inference about β based on the $N + 1$ elementary confidence sets, converges to classical inference based on the t_{n-m-1}-distribution if $N \to \infty$. This is in agreement with Efron (1969) who shows that Student's t distribution remains unchanged if we assume a spherical distribution rather than a Gaussian distribution. Thus, the median of the N confidence limits $c_i(x)$ converges to the OLS-estimator of β.

In the next section we will show that the present approach of exact inference can be generalized to a heteroscedastic context. The usual asymptotic approach in this generalized context provides only approximate confidence sets.

3 Exact inference based on partial rotational invariance

Consider the regression model of Section 2 while the distribution of u is not assumed to be invariant under general rotations. Instead, let

$$u = \sum_{j=1}^{k} A_j v_j,$$

where the A_j are known $n \times n_j$-matrices, where $n_j > m + 1$ and $\sum_{j=1}^{k} n_j = n$. We will assume that the distribution of the n-vector $v = (v_1', \cdots, v_k')'$ is invariant under rotations of its subvectors $v_j, \ j = 1, \cdots, k$.

We consider cases where $A = (A_1, \cdots, A_k)$ is nonsingular. Hence,

$$\begin{aligned} A^{-1}y &= A^{-1}x\beta + A^{-1}\epsilon, \\ A^{-1}\epsilon &= A^{-1}X_2\delta + v. \end{aligned}$$

So, without loss of generality, A can be assumed to equal the identity matrix. However, for the sake of later applications, we will assume that A is orthonormal: $A'A = I_n$. Thus, we find

$$\begin{aligned} A_j'y &= A_j'x\beta + A_j'\epsilon, \\ A_j'\epsilon &= A_j'X_2\delta + v_j. \end{aligned}$$

The distribution of u is, therefore, assumed to be invariant under transformations taken from

$$\mathcal{R}_u = \{R \mid R = \sum_{j=1}^{k} A_j R_j^* A_j';\ R^{*\prime}_j R_j^* = I_{n_j},\ j = 1, \cdots, k\}.$$

In that case, $\epsilon \sim R\epsilon$ if $R \in \mathcal{R}_\epsilon$, where

$$\begin{aligned} \mathcal{R}_\epsilon &= \{R \mid R = \sum_{j=1}^{k} A_j R_j^* A_j'; \\ &\quad R_j^* = P_{A_j' X_2} + L_{A_j' X_2} \tilde{R}_j L'_{A_j' X_2}, \\ &\quad \tilde{R}_j' \tilde{R}_j = I_{n_j - m},\ j = 1, \cdots, k\}. \end{aligned}$$

Notice that $\mathcal{R}_\epsilon$ is a subgroup of $\mathcal{R}$, as given in (1). That is, if $R \in \mathcal{R}_\epsilon$, then R is orthonormal and $RX_2 = X_2$, since $A_i' A_j = 0$ if $i \neq j$ and $A_i' A_j = I_{n_j}$ if $i = j$. Furthermore, $\sum_{j=1}^{k} A_j A_j' = I_n$. Intuitively, the matrices R only rotate "within" the groups.

For a random sample $\{R_1, \cdots, R_N\}$ taken from $\mathcal{R}_\epsilon$ we find confidence limits

$$c_i(z) = \frac{z'(I_n - R_i)y}{z'(I_n - R_i)x}, \qquad i = 1, \cdots, N,$$

as in (3), where now condition (2) should be satisfied with respect to $\mathcal{R}_\epsilon$. However, we can now also consider instruments where the confidence limits $c_i(z)$ are different from $c_i(x)$. That is, for $R \in \mathcal{R}_\epsilon$ we find

$$(I_n - R) = \sum_{j=1}^{k} A_j L_{A_j' X_2} (I_{n_j - m} - \tilde{R}_j) L'_{A_j' X_2} A_j'.$$

Consequently, condition (2) is satisfied if and only if

$$zA_jL_{A_j'X_2}(I_{n_j-m} - \tilde{R}_j)L'_{A_j'X_2}A_j'x > 0$$

for almost all $\tilde{R}_j$, and $j = 1, \cdots, k$. That is, $L'_{A_j'X_2}A_j'z$ and $L'_{A_j'X_2}A_j'x$ should be scalar multiples, with nonnegative inner-product, for $j = 1, \cdots, k$. Consequently, all confidence limits different from $c_i(x)$ can be generated by choosing monotonic instruments of the form

$$z(\lambda) = \sum_{j=1}^{k} \lambda_j A_j A_j' x, \tag{4}$$

where $\lambda = (\lambda_1, \cdots, \lambda_k)' \neq 0$ and $\lambda_j \geq 0$, $j = 1, \cdots, k$.

For any such choice of λ, confidence limits are given by

$$c_i(z(\lambda)) = \frac{\sum_{j=1}^{k} \lambda_j x' A_j (I_{n_j} - R_{ij}^*) A_j' y}{\sum_{j=1}^{k} \lambda_j x' A_j (I_{n_j} - R_{ij}^*) A_j' x}, \tag{5}$$

and

$$F(\beta, z(\lambda)) = \#\{i \mid c_i(z(\lambda)) < \beta\}/N \tag{6}$$

will be uniformly distributed over $\{0, 1/N, 2/N, \cdots, 1\}$.

Would there be an 'optimal' choice for the vector λ? In order to answer this question affirmatively, let the equivalence class C_{ϵ^*} now be defined with respect to the group $\mathcal{R}_\epsilon$,

$$C_{\epsilon^*} = \{R\epsilon^* \mid R \in \mathcal{R}_\epsilon\},$$

and consider the following result.

Lemma 1 *Let, conditional on (x, X_2), $\epsilon \sim R\epsilon$ for $R \in \mathcal{R}_\epsilon$, then, conditional on $\epsilon \in C_{\epsilon^*}$, the expectation and covariance matrix of ϵ are given by*

$$\begin{aligned} E(\epsilon \mid (x, X_2), \epsilon \in C_{\epsilon^*}) &= \sum_{j=1}^{k} A_j P_{A'_j X_2} A'_j \epsilon^*, \\ \mathrm{Var}(\epsilon \mid (x, X_2), \epsilon \in C_{\epsilon^*}) &= \sum_{j=1}^{k} \sigma_j^{*2} A_j (I_{n_j} - P_{A'_j X_2}) A'_j, \end{aligned}$$

where

$$\sigma_j^{*2} = (n_j - m)^{-1} \epsilon^{*\prime} A_j (I_{n_j} - P_{A'_j X_2}) A'_j \epsilon^*.$$

The proof is given in the Appendix.

The optimal choice for λ can now be based on the following result.

Theorem 1 *Let, conditional on (x, X_2), $\epsilon \sim R\epsilon$ for $R \in \mathcal{R}_\epsilon$, then*

$$E(c_i(z(\lambda)) \mid R_i, (x, X_2), \epsilon \in C_{\epsilon^*}) = \beta$$

and the minimum of

$$\mathrm{Var}(c_i(z(\lambda)) \mid R_i, (x, X_2), \epsilon \in C_{\epsilon^*})$$

is found for $\lambda_j = 1/\sigma_j^{*2}$, $j = 1, \cdots, k$, which does not depend on R_i.

The proof is given in the Appendix.

Let $\lambda_j^* = 1/\sigma_j^{*2}$, and $\lambda^* = (\lambda_1^*, \cdots, \lambda_k^*)'$. An optimal choice for the weights λ_j would be given by λ_j^*, $j = 1, \cdots, k$, and $F(\beta, z(\lambda^*))$, as defined in (6), would provide optimal inference about the true value of β. However, λ^* is unknown.

The solution to this problem will be based on a vector of weights $\lambda(\beta)$ as a function of β, which will coincide with the optimal weight λ^* for the true value of β. Notice that the value of σ_j^{*2}, as defined in the lemma, does not depend on the choice of element ϵ^* from C_{ϵ^*}. That is, if $\epsilon \in C_{\epsilon^*}$ then $C_{\epsilon^*} = C_\epsilon$ and

$$\sigma_j^{*2} = \frac{\epsilon' A_j (I_{n_j} - P_{A_j' X_2}) A_j' \epsilon}{n_j - m}.$$

Therefore, we consider weights

$$\lambda_j(\beta) = \frac{n_j - m}{(y - x\beta)' A_j (I_{n_j} - P_{A_j' X_2}) A_j' (y - x\beta)}. \tag{7}$$

These weights depend on the unknown true value of β, but they are constant, for the true value, over an equivalence class. So, using (4), $z(\lambda(\beta))$ provides a legitimate, albeit unknown, monotonic instrument. Notice that replacing β by an estimate $\hat{\beta}$ would make $z(\lambda(\hat{\beta}))$ vary as ϵ varies within an equivalence class C_{ϵ^*}. This would violate an essential requirement for our exact inference procedure.

If we let β_0 denote the true value, optimal, but unknown, inference would be described by the function $F(\beta, z(\lambda(\beta_0))) = F(\beta, z(\lambda^*))$. Our computable exact inference is described by the function $F(\beta, z(\lambda(\beta))$, which will simply be indicated by

$$F(\beta) = \#\{i \mid \frac{\sum_{j=1}^k \lambda_j(\beta) x' A_j (I_{n_j} - R_{ij}^*) A_j' y}{\sum_{j=1}^k \lambda_j(\beta) x' A_j (I_{n_j} - R_{ij}^*) A_j' x} < \beta\}/N, \tag{8}$$

where $\lambda_j(\beta),\ j = 1, \cdots, k$, is defined in (7). When evaluated at β_0, $F(\beta)$ will be uniformly distributed over $\{0, 1/N, 2/N, \cdots, 1\}$.

Notice that if

$$\text{rank}\{(y, x)'(I_{n_j} - P_{A'_j X_2})(y, x)\} = 2, \;\; j = 1, \cdots, k,$$

almost surely, which is possible since $n_j > m + 1$, then

$$\frac{(y - x\beta)' A_j (I_{n_j} - P_{A'_j X_2}) A'_j (y - x\beta)}{(y - x\beta)' A_1 (I_{n_1} - P_{A'_1 X_2}) A'_1 (y - x\beta)}$$

has a minimum and a maximum a.s. Consequently, $F(-\infty) = 0$ and $F(\infty) = 1$. However, contrary to the case $k = 1$, the function $F(\beta)$ may also decrease. Therefore, two-sided confidence sets

$$S_{1-\alpha} = \{\beta \mid F(\beta) \in [\alpha/2, 1 - \alpha/2]\},$$

evaluated for sufficiently large N, will be bounded, but they may be non-convex.

Another difference relates to the computation of the inference. For $k = 1$ the confidence limits can easily be computed since they do not depend on β. For the present case $k > 1$, the function $F(\beta)$ can only be computed based on a grid for β. That is, for $j = 1, \cdots, k$ we need to compute, as described in Section 2, N points given by

$$((x' A_j (I_{n_j} - R^*_{ij}) A'_j y), \;\; (x' A_j (I_{n_j} - R^*_{ij}) A'_j x)),$$

for $i = 1, \cdots, N$. Subsequently, we need to compute for each grid-point $\beta_l, \;\; l = 1, \cdots, L$, and for $j = 1, \cdots, k$, the values

$$(y - x\beta_l)' A_j (I_{n_j} - P_{A'_j X_2}) A'_j (y - x\beta_l).$$

Thus we can compute $F(\beta)$ for L values β_l.

4 An application in a panel context

Now, consider a panel data model where subjects $s = 1, \cdots, S$ are observed at times $t = 1, \cdots, T$:

$$
\begin{aligned}
y_{st} &= x_{st}\beta + \epsilon_{st}, \\
\epsilon_{st} &= \tilde{x}'_{2st}\delta + u_{st}.
\end{aligned}
$$

The scalar β is the parameter of interest and the explanatory variables are given by the vectors $(x_{st},\ \tilde{x}'_{2st})$. In many panel data applications a one-way error components model is used for the disturbances with

$$u_{st} = \mu_s + w_{st},$$

where μ_s denotes the unobservable subject effect and w_{st} denotes the remainder disturbance, see Baltagi (1995).

Let $y_s = (y_{s1}, \cdots, y_{sT})'$ and $y = (y'_1, \cdots, y'_S)$ with similar definitions for x, $\tilde{X}_2$ and w. Then, in matrix notation the model can be formulated as

$$y = x\beta + \tilde{X}_2\delta + (I_S \otimes \iota_T)\mu + w.$$

For the fixed effects model the S-vector μ is assumed to be nonrandom. If we assume that

$w \sim \mathcal{N}(0, \sigma_w^2 I_{ST})$, then the analysis of Section 2 applies where

$$X_2 = (\tilde{X}_2,\ I_S \otimes \iota_T),$$
$$u = w.$$

In that case exact inference on β amounts to the classical inference based on t-distributions.

For the random effects model μ is assumed to be random. In addition to the assumption on w, we assume that $\mu \sim \mathcal{N}(0, \sigma_u^2 I_S)$ and that μ and w are independent. Then

$$\begin{aligned} y = x\beta &+ \tilde{X}_2\delta + (I_S \otimes \iota_T)(\mu + T^{-1}(I_S \otimes \iota_T')w) \\ &+ (I_S \otimes L_{\iota_T})(I_S \otimes L_{\iota_T}')w. \end{aligned}$$

As $(I_S \otimes \iota_T')w$ and $(I_S \otimes L_{\iota_T}')w$ are independent, this model fits within the framework of Section 3. We find

$$y = x\beta + \epsilon,$$

where the number of observations equals $n = ST$,

$$\epsilon = X_2\delta + u,$$

where $X_2 = \tilde{X}_2$ is an $n \times m$-matrix, and

$$u = \sum_{j=1}^{k} A_j v_j,$$

with $k = 2$, $A_1 = I_S \otimes T^{-1/2}\iota_T$, of order $n \times n_1$ with $n_1 = S$, and $A_2 = I_S \otimes L_{\iota_T}$, which is of order $n \times n_2$ with $n_2 = S(T-1)$. Furthermore, $v_1 = \mu T^{1/2} + (I_S \otimes T^{-1/2}\iota_T')w$, which can be rotated independently of the rotations of $v_2 = (I_S \otimes L_{\iota_T}')w$. Exact inference on β, based on the uniform distribution of $F(\beta)$, as given in (8), now immediately applies.

The orthonormal subgroup $\mathcal{R}_\epsilon$, as defined in section 3, is given here by

$$
\begin{aligned}
\mathcal{R}_\epsilon \quad = \quad & \{R \mid R = (R_1^* \otimes P_{\iota_T}) + (I_S \otimes L_{\iota_T})R_2^*(I_S \otimes L_{\iota_T}') \\
& R_j^* = P_{A_j' X_2} + L_{A_j' X_2}\tilde{R}_j L_{A_j' X_2}', \\
& \tilde{R}_j' \tilde{R}_j = I_{n_j - m},\ j = 1, 2\}.
\end{aligned}
$$

Interestingly, the orthonormal subgroup $\mathcal{R}_\epsilon$ for the fixed effects model is simply found by fixing $\tilde{R}_1$ to the identity matrix, which is also a group, so that $R_1^* = I_S$. Consequently, we find that the function $F(\beta)$, which has been used to describe exact inference for the random effects model, also applies to the fixed effects model, if we fix the matrices $R_{i1}^* = I_S$.

Appendix

1. The computation of the confidence limits

The actual computation of the confidence limits, as discussed in Section 2, may take place as follows. Let $w_i = \tilde{R}_i' L_{X_2}' x$, then

$$
\begin{aligned}
c_i(x) &= \frac{x'(I_n - R_i)y}{x'(I_n - R_i)x} = \frac{x' L_{X_2}(I_{n-m} - \tilde{R}_i) L_{X_2}' y}{x' L_{X_2}(I_{n-m} - \tilde{R}_i) L_{X_2}' x} \\
&= \frac{x'(I_n - P_{X_2})y - w_i' L_{X_2}' y}{x'(I_n - P_{X_2})x - w_i' L_{X_2}' x}.
\end{aligned}
$$

As w_i is uniformly distributed over the sphere in $\mathbb{R}^{n-m}$ with squared radius $x'(I_n - P_{X_2})x$, it can be generated by a random drawing $v_i \sim \mathcal{N}(0, I_{n-m})$ which is subsequently renormalized:

$$
w_i = \left(\frac{x'(I_n - P_{X_2})x}{v_i' v_i} \right)^{1/2} v_i .
$$

Furthermore, for the distribution of $w_i' L_{X_2}' y$ and $w_i' L_{X_2}' x$ the only relevant aspects are the lengths of the vectors $L_{X_2}' y$ and $L_{X_2}' x$, and the angle between them. To reduce the dimension of the problem, let

$$
(\tilde{y}, \tilde{x}) = \{(y, x)'(I_n - P_{X_2})(y, x)\}^{1/2},
$$

and let $\tilde{w}_i$ consist of the first two elements of w_i, then the confidence limits can be generated by

$$
c_i(x) = \frac{(\tilde{x} - \tilde{w}_i)' \tilde{y}}{(\tilde{x} - \tilde{w}_i)' \tilde{x}}.
$$

2. The proof of Lemma 1

Proof: We use the fact that $E(v) = 0$ and $\text{Var}(v) = (v'v/l)I_l$ if the l-vector v is uniformly distributed over the sphere in $\mathbb{R}^l$ with fixed radius $(v'v)^{1/2}$. Therefore,

$$\begin{aligned} E(\tilde{R}_j L'_{A'_j X_2})A_j\epsilon^* \mid (x, X_2)) &= 0, \\ \text{Var}(\tilde{R}_j L'_{A'_j X_2})A_j\epsilon^* \mid (x, X_2)) &= \sigma_j^{*2} I_{n_j - m}. \end{aligned}$$

For a random drawing $R\epsilon^*$ from C_{ϵ^*} we find

$$\begin{aligned} R\epsilon^* &= \sum_{j=1}^{k} A_j R_j^* A'_j \epsilon^* \\ &= \sum_{j=1}^{k} A_j (P_{A'_j X_2} + L_{A'_j X_2} \tilde{R}_j L'_{A'_j X_2}) A_j \epsilon^*, \end{aligned}$$

so that

$$\begin{aligned} E(\epsilon \mid (x, X_2), \epsilon \in C_{\epsilon^*}) &= E(R\epsilon^* \mid (x, X_2)), \\ \text{Var}(\epsilon \mid (x, X_2), \epsilon \in C_{\epsilon^*}) &= \text{Var}(R\epsilon^* \mid (x, X_2)), \end{aligned}$$

equal the expressions as given in the lemma. □

3. The proof of Theorem 1

Proof: From (5), we find

$$c_i(z(\lambda)) = \beta + \frac{\sum_{j=1}^{k} \lambda_j x' A_j (I_{n_j} - R_{ij}^*) A'_j \epsilon}{\sum_{j=1}^{k} \lambda_j x' A_j (I_{n_j} - R_{ij}^*) A'_j x}.$$

Using the lemma, we have

$$E(c_i(z(\lambda)) \mid R_i, (x, X_2), \epsilon \in C_{\epsilon^*}) =$$

$$\beta + \frac{(\sum_{j=1}^k \lambda_j x' A_j (I_{n_j} - R_{ij}^*) A_j')(\sum_{j=1}^k A_j P_{A_j' X_2} A_j' \epsilon^*)}{\sum_{j=1}^k \lambda_j x' A_j (I_{n_j} - R_{ij}^*) A_j' x}.$$

Furthermore,

$$(\sum_{j=1}^k \lambda_j x' A_j (I_{n_j} - R_{ij}^*) A_j')(\sum_{j=1}^k A_j P_{A_j' X_2} A_j' \epsilon^*) =$$
$$\sum_{j=1}^k \lambda_j x' A_j (I_{n_j} - R_{ij}^*) A_j' A_j P_{A_j' X_2} A_j' \epsilon^* =$$
$$\sum_{j=1}^k \lambda_j x' A_j L_{A_j' X_2} (I_{n_j - m} - \tilde{R}_{ij}) L_{A_j' X_2}' P_{A_j' X_2} A_j' \epsilon^* = 0,$$

since

$$L_{A_j' X_2}' P_{A_j' X_2} = 0,$$

so

$$E(c_i(z(\lambda)) \mid R_i, (x, X_2), \epsilon \in C_{\epsilon^*}) = \beta.$$

Similarly,

$$\mathrm{Var}(c_i(z(\lambda)) \mid R_i, (x, X_2), \epsilon \in C_{\epsilon^*}) =$$

$$\frac{\sum_{j=1}^k \lambda_j^2 \sigma_j^{*2} x' A_j (I_{n_j} - R_{ij}^*)(I_{n_j} - R_{ij}^*)' A_j' x}{(\sum_{j=1}^k \lambda_j x' A_j (I_{n_j} - R_{ij}^*) A_j' x)^2} =$$

$$\frac{2\sum_{j=1}^{k}\lambda_j^2\sigma_j^{*2}x'A_j(I_{n_j}-R_{ij}^*)A_j'x}{(\sum_{j=1}^{k}\lambda_j x'A_j(I_{n_j}-R_{ij}^*)A_j'x)^2},$$

whose minimum is found for $\lambda_j = 1/\sigma_j^{*2}$. □

References

Baltagi, B.H. (1995), *Econometric analysis of panel data*, Wiley, Chichester.

Bekker, P.A. (1997), “Nonparametric assumption-equivalent inference for the linear model in an instrumental variable setting”.

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

Park, B.U. and L. Simar (1994), “Efficient semiparametric estimation in a stochastic frontier model”, *Journal of the American Statistical Association*, **89**, 929–936.

Taylor, W.E. (1977), “Small sample properties of a class of two stage Aitken estimators”, *Econometrica*, **45**, 497–508.