Teaching Corner

Gueorgui I. Kolev* and Helmuts Āzacis

On the Use of the Helmert Transformation, and its Applications in Panel Data Econometrics

Abstract: We revisit the Helmert transformation, and provide a useful and simple derivation of the joint distribution of the sample mean and the sample variance in samples from independently and identically distributed normal random variables. Our derivation is distinguished by concreteness, very little abstractness, and should be appealing to beginning students of statistics, and to both beginning and advanced students of econometrics. We also highlight one fruitful application of the Helmert transformation in panel data econometrics. The Helmert transformation can be used to eliminate the fixed effects in the estimation of fixed effects models, and we briefly review this application of the transformation in the panel data context.

Keywords: Helmert transformation, sample mean, sample variance, panel data econometrics, fixed effects model

JEL Classifications: B23, C16, C20

1 Introduction

The Helmert transformation is named after the German geodesist Friedrich Robert Helmert (1876) and has a long history of use in statistics (Sawkins 1940; Cramér 1946, p. 116; Kruskal 1946; Weatherburn 1961, p. 164; Brownlee 1965, p. 271; Rao 1973, p. 182 among others). One application (*the application* from now on for brevity) of the Helmert transformation in statistics is to find the joint distribution of the sample mean, and the sample variance, calculated from a sample from a normal population. Although the cited references are authoritative, the presentations there of the application are in our view often rather complicated, which inhibits readers' understating of the topic. We present our version of how the Helmert transformation works in the application, and our version should be particularly suitable and accessible to either beginning undergraduate university students in statistics, or students at both undergraduate and postgraduate level in econometrics, or quantitative social and political sciences which use econometrics.

For example, Cramér (1946, p. 116) and Weatherburn (1961, p. 164) both follow Sawkins (1940), and introduce in abstract terms an orthogonal linear transformation having certain properties. Instead, we think that explicitly stating the transformation, and then directly establishing its key properties would have pedagogical advantages. Rao (1973, p. 182) similarly uses an orthogonal matrix, i.e. a Helmert matrix, which he defines in terms of certain abstract properties. We think there would be pedagogical advantage to instead explicitly display one such Helmert matrix, so that the reader can visualize what is going on, and then proceed to establishing its properties and what it exactly does in the context of the application.

*Corresponding author: Gueorgui I. Kolev, Department of Finance, Tilburg University, Tilburg, Netherlands,
E-mail: joro.kolev@gmail.com. https://orcid.org/0000-0002-9250-9917
Helmuts Āzacis, Cardiff Business School, Cardiff University, Cardiff, UK, E-mail: azacish@cardiff.ac.uk

Upon presenting our treatment of the application, which of course uses the vary same ideas of the Helmert transformation and the associated Helmert matrix as the sources mentioned above, we also briefly discuss the use of the Helmert transformation in panel data econometrics. In particular, the Helmert transformation can be employed to eliminate fixed effects in the estimation of fixed effects models. Therefore, our treatment has the advantage that it is very concrete and explicit compared to previous literature, and advocates for more extensive use of the Helmert transformation. Unlike previous authors, we firstly explicitly state the Helmert transformation and display the Helmert matrix, and then proceed to prove the implications of applying the Helmert transformation in the application.

Our proof is closest to Brownlee (1965, p. 271). Brownlee (1965, p. 271) starts with a formula for the sample variance and using it, derives new variables that possess certain desired properties. This derivation of the new variables is fairly involved. We instead start by defining the new variables and then using a simple induction argument, prove that these variables have the same variance as the original variables.

We use mathematical induction in our proof, and Stigler (1984) also uses mathematical induction in his proof, but the argument there is different. He first establishes the joint distribution of the sample mean and variance for a sample of just two observations and then via the induction argument, proves that this joint distribution extends to samples of larger size. Zehna (1991) asserts that Stigler's proof is not completely rigorous, and relies three times throughout the proof on a faulty argument.

This article proceeds as follows. In Section 2 we introduce the Helmert transformation and derive the joint distribution of the sample mean and the sample variance. In Section 3 we introduce the Helmert matrix associated with the transformation, and provide some remarks on its properties. In Section 4 we suggest that the use of the Helmert transformation has been somewhat overlooked in within estimation of one way error component fixed effects models in panel data econometrics, and we briefly survey the extant applications of the Helmert transformation in panel data context. In the last section we conclude.

2 The Helmert Transformation and the Joint Distribution of the Sample Mean and the Sample Variance in Samples from a Normal Population

We have a set of T random variables $x_t, t = 1, 2, \ldots, T$ independently and identically distributed (i.i.d. from now on) as Normal(μ, σ^2), and we want to find the joint distribution of the sample mean $\bar{x}_T = \sum_{t=1}^{T} x_t / T$ and the sample variance $s^2 = \sum_{t=1}^{T} (x_t - \bar{x}_T)^2 / (T - 1)$.

Consider the Helmert transformation, which takes the set $x_t, t = 1, 2, \ldots, T$ and produces a new set of variables $z_t, t = 1, 2, \ldots, T$ as follows:

$$z_1 = \bar{x}_T, \quad z_2 = (x_2 - x_1)\sqrt{1/2}, \quad z_3 = x_3 - [(x_2 + x_1)/2]\sqrt{2/3}, \ldots,$$

$$z_t = (x_t - \bar{x}_{t-1})\sqrt{(t-1)/t}, \ldots, z_T = (x_T - \bar{x}_{T-1})\sqrt{(T-1)/T}, \tag{1}$$

where $\bar{x}_{t-1} = \sum_{\tau=1}^{t-1} x_\tau / (t - 1)$.

The new set of variables z_t have convenient properties. (In what follows, we liberally use the properties of the expectation $E(\cdot)$, variance $\text{Var}(\cdot)$, and covariance $\text{Cov}(\cdot, \cdot)$ operators, and we assume that the reader is familiar with these properties.)

Properties of the Helmert transformation:

1. $z_t, t = 1, 2, \ldots, T$ are mutually uncorrelated: $\text{Cov}(z_1, z_t) = \text{Cov}\left(\bar{x}_T, (x_t - \bar{x}_{t-1})\sqrt{(t-1)/t}\right)$ $= (1/T)\sqrt{(t-1)/t} \cdot \text{Cov}\left(\sum_{\tau=1}^{T} x_\tau, x_t - \sum_{\tau=1}^{t-1} x_\tau / (t-1)\right) = (1/T)\sqrt{(t-1)/t}[\sigma^2 - (t-1)\sigma^2/(t-1)] = 0$ for $t = 2, \ldots, T$. And, $\text{Cov}(z_t, z_\theta) = \text{Cov}\left(x_t - \sum_{\tau=1}^{t-1} x_\tau / (t-1), x_\theta - \sum_{\tau=1}^{\theta-1} x_\tau / (\theta-1)\right) = -\sigma^2 / (\theta - 1)$

$+ (t-1)\sigma^2/[(t-1)(\theta-1)] = 0$, where without loss of generality we take $\theta > t > 1$. (Property 1 does not require normality; an i.i.d. distribution of the variables would be sufficient.)

2. $z_t, t = 1, 2, \ldots, T$ are linear combinations of $x_t, t = 1, 2, \ldots, T$ (which we have assumed, are jointly normally distributed) and hence $z_t, t = 1, 2, \ldots, T$ are jointly normally distributed as well.

3. $E z_1 = E\bar{x}_T = \mu$; $E z_t = E[(x_t - \bar{x}_{t-1})\sqrt{(t-1)/t}] = \sqrt{(t-1)/t} \cdot (\mu - \mu) = 0$, for $t = 2, 3, \ldots, T$.

4. $\mathrm{Var}\, z_1 = \sigma^2/T$, $\mathrm{Var}\, z_t = [\sigma^2 + \sigma^2/(t-1)](t-1)/t = \sigma^2, t = 2, \ldots, T$. So $z_2, \ldots, z_T$ are homoskedastic.

5. $\sum_{t=1}^{T}(x_t - \bar{x}_T)^2 = \sum_{t=2}^{T} z_t^2 = \sum_{t=2}^{T}(x_t - \bar{x}_{t-1})^2(t-1)/t$, which we prove in the theorem below.

An Auxiliary Theorem. Let x_t be a scalar quantity observed over $t = 1, 2, \ldots, T$ periods. Then,

$$\sum_{t=1}^{T}(x_t - \bar{x}_T)^2 = \sum_{t=2}^{T}(x_t - \bar{x}_{t-1})^2(t-1)/t, \qquad (2)$$

where $\bar{x}_T = \sum_{t=1}^{T} x_t/T$ and $\bar{x}_{t-1} = \sum_{\tau=1}^{t-1} x_\tau/(t-1)$.

Proof of the Auxiliary Theorem by induction. For $T = 2$ we have

$$\sum_{t=1}^{2}(x_t - \bar{x}_T)^2 = (x_1 - x_1/2 - x_2/2)^2 + (x_2 - x_1/2 - x_2/2)^2$$

$$= (x_2 - x_1)^2/2 = \sum_{t=2}^{2}(x_t - \bar{x}_{t-1})^2(t-1)/t,$$

so for $T = 2$ the relationship holds indeed.

Now we assume that the relationship holds for T, and demonstrate that if it holds for T, then it holds for $T + 1$ too. First,

$$\sum_{t=2}^{T+1}(x_t - \bar{x}_{t-1})^2(t-1)/t = \sum_{t=2}^{T}(x_t - \bar{x}_{t-1})^2(t-1)/t + (x_{T+1} - \bar{x}_T)^2 T/(T+1).$$

Second,

$$\sum_{t=1}^{T+1}(x_t - \bar{x}_{T+1})^2 = \sum_{t=1}^{T+1} x_t^2 - (T+1)\bar{x}_{T+1}^2 = \sum_{t=1}^{T} x_t^2 + x_{T+1}^2 - (T+1)\bar{x}_{T+1}^2$$

$$= \sum_{t=1}^{T}(x_t - \bar{x}_T)^2 + T\bar{x}_T^2 + x_{T+1}^2 - (T+1)\bar{x}_{T+1}^2.$$

By the induction hypothesis

$$\sum_{t=1}^{T}(x_t - \bar{x}_T)^2 = \sum_{t=2}^{T}(x_t - \bar{x}_{t-1})^2(t-1)/t.$$

Therefore, what remains to be shown is that

$$(x_{T+1} - \bar{x}_T)^2 T/(T+1) = T\bar{x}_T^2 + x_{T+1}^2 - (T+1)\bar{x}_{T+1}^2.$$

Rewrite

$$(T+1)\bar{x}_{T+1}^2 = \left(\sum_{t=1}^{T} x_t + x_{T+1}\right)^2 /(T+1) = (T\bar{x}_T + x_{T+1})^2/(T+1)$$

$$= [T^2/(T+1)]\bar{x}_T^2 + 2[T/(T+1)]\bar{x}_T x_{T+1} + x_{T+1}^2/(T+1).$$

Hence,

$$T\bar{x}_T^2 + x_{T+1}^2 - (T+1)\bar{x}_{T+1}^2 = T\bar{x}_T^2 + x_{T+1}^2 - [T^2/(T+1)]\bar{x}_T^2 - 2[T/(T+1)]\bar{x}_T x_{T+1} - x_{T+1}^2/(T+1)$$
$$= [T/(T+1)]\bar{x}_T^2 - 2[T/(T+1)]\bar{x}_T x_{T+1} + [T/(T+1)]x_{T+1}^2,$$

which is indeed equal to $(x_{T+1} - \bar{x}_T)^2 T/(T+1)$. $\qquad\square$

We will now use the properties of the Helmert transformation to deduce the joint distribution of the sample mean and the sample variance in samples from a normal population. We use the following three facts. First, linear combinations of jointly normal random variables are themselves jointly normally distributed (Cramér 1946, p. 213; Weatherburn 1961, p. 57). Second, for a set of jointly normal random variables, if they are uncorrelated, they are independent as well (e.g. Lancaster 1959; David 2009). There is a subtle point here, the set of variables need to be jointly normal, as one can construct counter examples where the marginal distributions are normal but the joint distribution is not normal, variables are uncorrelated, and yet they are not independent. Lancaster (1959) presents such a counter example, and precisely states the conditions in his Theorem 1. Third, the sum of k independent, squared, standard normal variables is distributed as $\chi^2(k)$, which is a distribution discovered and described by Helmert (1876). (Helmert (1876) discovered the χ^2 distribution, however he did not observe that the sample mean and the sample variance are independent, see David (2009).)

Main Theorem. The joint distribution of the sample mean and the sample variance in a sample of T i.i.d. random variables $x_t, t = 1, 2, \ldots, T$, each x_t distributed as Normal(μ, σ^2), has the following properties:
- The sample mean $\bar{x}_T$ and the sample variance s^2 are independently distributed.
- The sample mean $\bar{x}_T$ is distributed as Normal($\mu, \sigma^2/T$).
- $(T-1)s^2/\sigma^2$ is distributed as $\chi^2(T-1)$.

Proof of the Main Theorem. In the proof we will refer to the listed **Properties of the Helmert transformation** as Property plus the number under which the property was listed.
- The sample mean $\bar{x}_T$ is a function of z_1 only, and by Property 5, the sample variance s^2 is a function of $z_t, t = 2, \ldots, T$ only. By Property 1, $z_t, t = 1, 2, \ldots, T$ are uncorrelated, and by Property 2, $z_t, t = 1, 2, \ldots, T$ are jointly normally distributed. Because $z_t, t = 1, 2, \ldots, T$ are a set of uncorrelated and jointly normal random variables, they are a set of independent random variables as well (Lancaster 1959, Theorem 1). Therefore the independence of $\bar{x}_T$ (a function of z_1 only) and s^2 (a function of $z_2, z_3, \ldots, z_T$ only) follows because z_1 is independent of $z_2, z_3, \ldots, z_T$.
- The sample mean $\bar{x}_T$ is normally distributed by Property 2. Its mean is given in Property 3, and its variance is given in Property 4. Overall $\bar{x}_T$ is distributed as Normal($\mu, \sigma^2/T$).
- $(T-1)s^2/\sigma^2 = \sum_{t=1}^{T}(x_t - \bar{x}_T)^2/\sigma^2 = \sum_{t=2}^{T}z_t^2/\sigma^2$, where the last equality follows by dividing both sides of Property 5 by σ^2. However, $\sum_{t=2}^{T}z_t^2/\sigma^2$ is the sum of $T-1$ squared standard normal variables, and hence is distributed as $\chi^2(T-1)$.

The Main Theorem was first proved by Fisher (1915, 1925), but Fisher's "geometric arguments" are difficult to follow. What we have presented, albeit somewhat lengthy, is an elementary and concrete proof which requires very little abstract thinking. Our proof should be accessible to students with any moderately quantitative background, and should be much preferable to students who are not used to elaborate abstract mathematical thinking, such as beginning undergraduates in statistics, or both undergraduate and graduate students in econometrics. Overall, from the transformed set of variables z_t it is very easy to deduce the joint distribution of the sample mean and the sample variance.

We finish this section by stating another property of the Helmert transformation, which we will briefly use in Section 4. Suppose there is another set of variables $y_t, t = 1, \ldots, T$. Then,

6. $\sum_{t=1}^{T}(x_t - \bar{x}_T)(y_t - \bar{y}_T) = \sum_{t=2}^{T}(x_t - \bar{x}_{t-1})(y_t - \bar{y}_{t-1})(t-1)/t,$

The proof of this property follows the same steps as the proof of the Auxiliary Theorem and, therefore, is omitted. Note that if $y_t = x_t$ for all t, this property reduces to Property 5.

3 The Helmert Matrix

The previous section is self contained, and using the Helmert transformation to deduce the joint distribution of the sample mean and the sample variance does not require any matrix algebra. However if there is need, or desire to do so, one can also relate the Helmert transformation from the previous section to what Lancaster (1965) calls a Helmert matrix in the strict sense.

Consider the following matrix

$$\text{Ho}_{T \times T} \equiv \begin{pmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ -1 & 1 & 0 & 0 & \dots & 0 \\ -1/2 & -1/2 & 1 & 0 & \dots & 0 \\ -1/3 & -1/3 & -1/3 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ -1/(T-1) & -1/(T-1) & -1/(T-1) & -1/(T-1) & \dots & 1 \end{pmatrix}$$

We can verify by direct multiplication that the rows of this matrix are orthogonal, that is, HoHo' results in a diagonal matrix. We can also consider the rescaled version of Ho

$$\text{Hm}_{T \times T} \equiv \text{diag}\left(1/T \quad \sqrt{1/2} \quad \sqrt{2/3} \quad \sqrt{3/4} \quad \dots \quad \sqrt{(T-1)/T}\right) \cdot \text{Ho},$$

where $\text{diag}(v)$ is an operator that transforms a vector v into a diagonal matrix, and overall the operation $\text{diag}(v) \cdot \text{Ho}$ results in multiplying the ith row of the matrix Ho by the ith element of the vector v. With the choice of the first element in v as $1/T$, we can see by direct multiplication that $\text{Hm}'\text{Hm}$ is symmetric with one element repeated on the main diagonal and another element repeated everywhere off the main diagonal. HmHm' is diagonal and is almost the identity matrix, only the upper left element is different from 1, the rest of HmHm' coincides with the identity matrix.

If we instead choose the first element in v to be $1/\sqrt{T}$,

$$\text{Hn}_{T \times T} \equiv \text{diag}\left(1/\sqrt{T} \quad \sqrt{1/2} \quad \sqrt{2/3} \quad \sqrt{3/4} \quad \dots \quad \sqrt{(T-1)/T}\right) \cdot \text{Ho},$$

direct multiplication shows that $\text{Hn}'\text{Hn}$ and HnHn' are both the identity matrix, and we would call such a matrix Hn an orthonormal matrix.

We see that if we arrange the set x_t, $t = 1, 2, \dots, T$ into a column vector $x \equiv \begin{pmatrix} x_1 & x_2 & \dots & x_T \end{pmatrix}'$, and choose the first element of v to be $1/T$, we will obtain the Helmert transformation from the previous section, $z = \text{Hm} \cdot \text{x}$, where $z \equiv \begin{pmatrix} z_1 & z_2 & \dots & z_T \end{pmatrix}'$. Because of the appealing aesthetics of Hn with the first element in v chosen to be $1/\sqrt{T}$, i.e. a choice resulting in an orthonormal matrix $\text{Hn}'\text{Hn} = \text{HnHn}' = \text{Identity}$, all authors we are aware of use this version of the Helmert matrix. However for our application all we need is that HmHm' be a diagonal matrix with all the elements on the main diagonal below the first being equal to 1. In this situation, the elements of the vector $z = \text{Hm} \cdot x$ will be mutually uncorrelated and each element z_t, $t = 2, \dots, T$ will be homoskedastic. Therefore, for our application choosing the first element in the vector v as $1/T$ serves perfectly fine.

4 The Helmert Transformation in Panel Data Models

We have derived the joint distribution of the sample mean and the sample variance with particular focus on the simplicity and concreteness of the derivation, and particular focus on the use of the Helmert transformation.

The Helmert transformation has found its application in the "fixed effects" panel data model too. Consider the standard one way "fixed effects" panel data model (e.g. Wooldridge 2010, Ch.10; Hsiao 2014, Ch.3)

$$y_{it} = x'_{it}\beta + \mu_i + \varepsilon_{it}, \quad i = 1, 2, \ldots, I, \quad t = 1, 2, \ldots, T \tag{3}$$

where the regressand y_{it} is a scalar, the regressor vector x_{it} is $K \times 1$, (y_{it}, x'_{it}) are i.i.d. for $i = 1, 2, \ldots, I$, i.e. the variables constitute an i.i.d. random sample in the cross section. The individual "fixed effects" (so called by convention) are time constant, random and potentially correlated with the regressor vector. The idiosyncratic error term ε_{it} is uncorrelated with the regressor $x_{j\tau}$ for all t, τ, i, j, i.e. the regressor x_{it} is strictly exogenous with respect to ε_{it} conditional on the fixed effects, and ε_{it} is i.i.d. both in the cross section and in the time series dimensions.

Consistent estimation of the parameter vector β in Eq. (3) under the assumptions that the fixed effects can be arbitrarily correlated with the regressors, and the regressors are strictly exogenous with respect to ε_{it}, conditional on the fixed effects, proceeds by eliminating the fixed effects. The conventional way of eliminating the fixed effects is by the within transformation, firstly averaging Eq. (3) across time, to obtain $\bar{y}_{iT} = \bar{x}'_{iT}\beta + \mu_i + \bar{\varepsilon}_{iT}$, where $\bar{y}_{iT} = \sum_{t=1}^{T} y_{it}/T$, $\bar{x}_{iT} = \sum_{t=1}^{T} x_{it}/T$ and $\bar{\varepsilon}_{iT} = \sum_{t=1}^{T} \varepsilon_{it}/T$. Then we subtract this averaged equation from Eq. (3) to eliminate the fixed effects and to obtain the estimating equation

$$y_{it} - \bar{y}_{iT} = (x_{it} - \bar{x}_{iT})'\beta + (\varepsilon_{it} - \bar{\varepsilon}_{iT}), \quad i = 1, 2, \ldots, I, \quad t = 1, 2, \ldots, T. \tag{4}$$

Finally, we estimate Eq. (4) by ordinary least squares (OLS) over the $I \cdot T$ pooled observations to obtain the within estimator

$$\hat{\beta} = \left(\sum_{i=1}^{I} \sum_{t=1}^{T} (x_{it} - \bar{x}_{iT})(x_{it} - \bar{x}_{iT})' \right)^{-1} \left(\sum_{i=1}^{I} \sum_{t=1}^{T} (x_{it} - \bar{x}_{iT})(y_{it} - \bar{y}_{iT}) \right). \tag{5}$$

The within estimator is well studied, well understood and a basic building block in the panel data econometrics literature. However the within transformation introduces strong correlation between the transformed errors in the estimating Eq. (4), because all the transformed errors $(\varepsilon_{it} - \bar{\varepsilon}_{iT})$ for a cross sectional unit i share the same $\bar{\varepsilon}_{iT}$. This makes residual diagnostic checks and residual analysis awkward and difficult, for example, if we wanted to test that the errors ε_{it} are indeed i.i.d.

On the other hand if we apply the Helmert transformation in Eq. (1) on each variable in the fixed effects model Eq. (3) and for each cross sectional unit i separately, to construct transformed variables corresponding to $z_2, z_3, \ldots, z_T$ which have mean 0, then we again eliminate the fixed effects μ_i:

$$(y_{it} - \bar{y}_{it-1})\sqrt{(t-1)/t} = \left[(x_{it} - \bar{x}_{it-1})'\sqrt{(t-1)/t} \right]\beta + (\varepsilon_{it} - \bar{\varepsilon}_{it-1})\sqrt{(t-1)/t}, \tag{6}$$

$i = 1, 2, \ldots, I, t = 2, \ldots, T$, where $\bar{y}_{it-1} = \sum_{\tau=1}^{t-1} y_{i\tau}/(t-1)$, $\bar{x}_{it-1} = \sum_{\tau=1}^{t-1} x_{i\tau}/(t-1)$, $\bar{\varepsilon}_{it-1} = \sum_{\tau=1}^{t-1} \varepsilon_{i\tau}/(t-1)$.

We can proceed with the OLS estimation of the Helmert-transformed estimating Eq. (6),

$$\hat{\beta} = \left(\sum_{i=1}^{I} \sum_{t=2}^{T} (x_{it} - \bar{x}_{it-1})(x_{it} - \bar{x}_{it-1})' \right)^{-1} \left(\sum_{i=1}^{I} \sum_{t=2}^{T} (x_{it} - \bar{x}_{it-1})(y_{it} - \bar{y}_{it-1}) \right), \tag{7}$$

which is the best linear unbiased estimator, because the errors in the Helmert-transformed estimating equation $(\varepsilon_{it} - \bar{\varepsilon}_{it-1})\sqrt{(t-1)/t}$ are uncorrelated and homoskedastic, (and normal if the original ε_{it} were normal to start with). By invoking Property 6 of the Helmert transformation, one can verify that the estimator in Eq. (7) coincides with the within estimator in Eq. (4). However, as an added benefit, we can apply any residual diagnostics and checks that we might have in mind, because the Helmert-transformed errors in the estimating equation have the same stochastic properties as the original errors in the structural model. To the best of our knowledge, the existing literature on panel data models has not exploited this simplification in the analysis afforded by the Helmert transformation.

The use of the Helmert transformation in the context of panel models and, in particular, dynamic panel models has been popularized by Arellano and Bover (1995) and Arellano (2003). (See also Alvarez and Arellano 2003.) It is also described in Hansen (2022, Section 17.43). A dynamic panel model is like the one given in Eq. (3), but it additionally contains the lagged values of the dependent variable as regressors. For example, if y_{it} (directly) depends only on its own value in the previous period, then the model is

$$y_{it} = y_{it-1}\alpha + x'_{it}\beta + \mu_i + \varepsilon_{it}, \quad i = 1, 2, \ldots, I, \quad t = 2, \ldots, T. \tag{8}$$

Arellano and Bover (1995) suggest to transform the variables in the following way:

$$(y_{it} - \bar{y}_{it+1})\sqrt{(T-t)/(T-t+1)}, \quad (x_{it} - \bar{x}_{it+1})'\sqrt{(T-t)/(T-t+1)}, \tag{9}$$

$i = 1, 2, \ldots, I, t = 1, \ldots, T-1$, where $\bar{y}_{it+1} = \sum_{\tau=t+1}^{T} y_{i\tau}/(T-t)$, $\bar{x}_{it+1} = \sum_{\tau=t+1}^{T} x_{i\tau}/(T-t)$. Observe that this is the same Helmert transformation in Eq. (1) but with the variables ordered in the reverse order according to the time index. Arellano and Bover (1995) refer to this transformation as "the forward orthogonal deviation" as opposed to "the backward orthogonal deviation" which is the one displayed in Eq. (6).

Both forward and backward orthogonal deviations produce uncorrelated and homoskedastic errors in the transformed model, provided that the original errors ε_{it} are i.i.d. However, the forward orthogonal deviation has the following advantage in the dynamic panel model. When transforming the variables to remove the fixed effects, it introduces correlation between the transformed lagged dependent variable and the transformed error term. Therefore, one needs to use an instrumental variables estimator to obtain consistent estimates of α and β. With the forward orthogonal deviation, the past values of the dependent variable $y_{i1}, \ldots, y_{it-1}$ are valid instruments for $(y_{it-1} - \bar{y}_{it})\sqrt{(T-t-1)/(T-t)}$, which is the new lagged dependent variable after the variable transformation. Furthermore, Hayakawa (2009a, 2009b) shows that the instrumental variables estimator is more efficient if the instruments themselves are constructed using backward orthogonal deviations.

5 Conclusion

We revisit the Helmert transformation, and we provide a simple and useful induction-based derivation of the joint distribution of the sample mean and the sample variance in samples from independently and identically distributed normal random variables. Our derivation is concrete and should be appealing to students of statistics and econometrics. We also suggest one fruitful application of the Helmert transformation in panel data econometrics – residual based tests in fixed effects models. We briefly review the applications of the Helmert transformation in panel data context, where the transformation is more commonly known as "the forward/backward orthogonal deviations operator".

References

Alvarez, J., and M. Arellano. 2003. "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators." *Econometrica* 71 (4): 1121−59.

Arellano, M. 2003. *Panel Data Econometrics*. Oxford: Oxford University Press.

Arellano, M., and O. Bover. 1995. "Another Look at the Instrumental Variable Estimation of Error-Components Models." *Journal of Econometrics* 68 (1): 29−51.

Brownlee, K. A. 1965. *Statistical Theory and Methodology in Science and Engineering*. New York: A Wiley Publication in Applied Statistics.

Cramér, H. 1946. *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

David, H. A. 2009. "A Historical Note on Zero Correlation and Independence." *The American Statistician* 63 (2): 185−6.

Fisher, R. A. 1915. "Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population." *Biometrika* 10 (4): 507−21.

Fisher, R. A. 1925. "Applications of "Student's" Distribution." *Metron* 5: 90−104.

Hansen, B. 2022. *Econometrics*. Princeton: Princeton University Press.

Hayakawa, K. 2009a. "A Simple Efficient Instrumental Variable Estimator for Panel AR(p) Models when Both N and T Are Large." *Econometric Theory* 25 (3): 873−90.

Hayakawa, K. 2009b. "First Difference or Forward Orthogonal Deviation- Which Transformation Should be Used in Dynamic Panel Data Models?: A Simulation Study." *Economics Bulletin* 29 (3): 2008−17.

Helmert, F. R. 1876. "Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers director Beobachtungen gleicher Genauigkeit." *Astronomische Nachrichten* 88: 113.

Hsiao, C. 2014. *Analysis of Panel Data*, 3rd ed. Cambridge: Cambridge University Press.

Kruskal, W. 1946. "Helmert's Distribution." *The American Mathematical Monthly* 53 (8): 435−8.

Lancaster, H. O. 1959. "Zero Correlation and Independence." *Australian Journal of Statistics* 1 (2): 53−6.

Lancaster, H. O. 1965. "The Helmert Matrices." *The American Mathematical Monthly* 72 (1): 4−12.

Rao, C. R. 1973. *Linear Statistical Inference and its Applications*, 2nd ed. New York: Wiley.

Sawkins, D. T. 1940. "Elementary Presentation of the Frequency Distribution of Certain Statistical Populations." *Journal and Proceedings of the Royal Society of New South Wales* 74: 209−39.

Stigler, S. M. 1984. "Kruskal's Proof of the Joint Distribution of $\bar{X}$ and S^2." *The American Statistician* 38 (2): 134−5.

Weatherburn, C. E. 1961. *A First Course in Mathematical Statistics*. Cambridge: The English Language Book Society and Cambridge University Press.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT press.

Zehna, P. W. 1991. "On Proving that $\bar{X}$ and S^2 are Independent." *The American Statistician* 45 (2): 121−2.