Fixed, Random, or Something in Between? A Variant of HAUSMAN's Specification Test for Panel Data Estimators

Manuel Frondel, Colin Vance, RWI Essen

Abstract: This paper proposes a variant of the classical HAUSMAN specification test commonly employed to distinguish between fixed- and random effects. The proposed variant substitutes the classical test for the equality of fixed- and random effects with a numerically identical test for the equality of between-groups and fixed effects. Using the panel model specification suggested here allows us to simultaneously estimate the fixed- and between-groups effects so that we are able to examine both the equality of the whole range of coefficients as well as that of individual variables, an issue that cannot be addressed on the basis of the standard HAUSMAN test. The usefulness of the test variant is illustrated using a panel of household travel diary data collected in Germany between 1997 and 2005.

JEL classification: C12.

Key words: Specification tests, fuel price elasticity.

Correspondence: Dr. Manuel Frondel, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI Essen), Hohenzollernstr. 1-3, D-45128 Essen. E-mail: frondel@rwi-essen.de.

Acknowledgements: We are grateful for valuable comments and suggestions by Walter KRÄMER.

1 Introduction

The econometric modeling of panel data typically applies two principle approaches, fixed- and random-effects estimators. In the fixed-effects approach, time-invariant, unobservable factors for each observation unit are either explicitly captured by dummy variables or wiped out through time-demeaning. In contrast, these time-invariant unobservables are treated as part of the disturbances in the random-effects model, thereby assuming that their correlation with the regressors is zero. If this assumption is met, the random-effects estimator confers the advantage of greater efficiency over the fixed-effects estimator. Violation of the assumption, however, implies biased estimates.

To investigate the appropriateness of either of these two approaches, HAUSMAN's (1978) specification test is commonly employed. It is based on the idea that the set of coefficient estimates gleaned from the fixed-effects estimation should not differ systematically from that of the random-effects estimation if the orthogonality assumption of the unobservable, individual-specific effects and the regressors is correct. If the test results vote against the equality of both coefficient sets, applied researchers generally draw conclusions based on the fixed-effects estimates, a course of action that effectively results in the wholesale discarding of the random-effects estimates. In such cases, however, it would be frequently interesting to know whether the inequality holds for the complete set of coefficients or whether there are exemptions for specific variables. This issue cannot be addressed on the basis of the standard HAUSMAN test.

To examine the equivalence of the coefficients for individual variables, this paper suggests a variant of the HAUSMAN test that is based on the fact that testing the equality of fixed- and random effects is numerically identical to testing the equality of between-groups and fixed effects – see HAUSMAN and TAYLOR (1981). Specifically, we prove that using an appropriate specification allows us to simultaneously estimate the fixed- and between-groups effects – either on the basis of OLS or GLS – and to examine both the equality of coefficients for individual variables as well as that of the whole range of coefficients.

The following section presents the theoretical basis of the test variant. Its usefulness is illustrated in Section 3 using a panel of household travel diary data collected in Germany between 1997 and 2005. The last section summarizes and concludes.

2 Methodology

Most applied panel analyses eschew the between-groups estimator, as it fails to capture inter-temporal information, and instead focus on fixed- and random effects. Yet, using the fact that the null of the HAUSMAN test is equivalent to the hypothesis that the set of fixed effects $\mathbf{w}$ equals the set of between-groups effects $\mathbf{b}$, the between-groups estimator plays a major role in our test variant.

Proposition 1: Departing from a standard panel data model,

$$y_{it} = \beta_0 + \boldsymbol{\beta}^T \mathbf{x}_{it} + \xi_i + \nu_{it}, \quad i = 1, ..., N, t = 1, ..., T, \tag{1}$$

where ξ_i denotes an unknown individual-specific term and ν_{it} is a random error component that varies over individuals i and time t, and estimating the specification

$$y_{it} = \beta_0 + \mathbf{w}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + \mathbf{b}^T \bar{\mathbf{x}}_i + \xi_i + \nu_{it}, \tag{2}$$

via OLS simultaneously yields estimates of the between-groups and fixed effects, where the OLS estimator of $\mathbf{w}$ provides for the fixed-effects estimates and the estimates of the between-groups effects are given by the OLS estimator of $\mathbf{b}$.

Proof: First, the between-groups estimator of parameter vector $\boldsymbol{\beta}$ emerging from model (1) can be obtained by averaging (1) over time and estimating the result via OLS:

$$\bar{y}_i = \beta_0 + \boldsymbol{\beta}^T \bar{\mathbf{x}}_i + \xi_i + \bar{\nu}_i, \tag{3}$$

where $\bar{y}_i, \bar{\mathbf{x}}_i$ and $\bar{\nu}_i$ denote the time means of $y_{it}, \mathbf{x}_{it}$ and ν_{it}, respectively. Second, the fixed-effects estimates of $\boldsymbol{\beta}$ can be retrieved by subtracting (3) from (1) and estimating the result via OLS:

$$y_{it} - \bar{y}_i = \boldsymbol{\beta}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + \nu_{it} - \bar{\nu}_i. \tag{4}$$

Third, instead of estimating either (3) or (4), we alternatively suggest in Proposition 1 estimating (2) via OLS to at once get both the fixed- and between-groups effects. This can be seen as follows: Upon averaging (2) over time, the term related to $\mathbf{w}$ washes out so that the result is, aside from the notation of the parameter vector, identical to (3):

$$\bar{y}_i = \beta_0 + \mathbf{w}^T \cdot \mathbf{0} + \mathbf{b}^T \bar{\mathbf{x}}_i + \xi_i + \bar{\nu}_i = \beta_0 + \mathbf{b}^T \bar{\mathbf{x}}_i + \xi_i + \bar{\nu}_i. \tag{5}$$

Subtracting (5) from (2) yields (4), with $\mathbf{w}$ instead of $\boldsymbol{\beta}$ as parameter vector:

$$y_{it} - \bar{y}_i = \mathbf{w}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + \nu_{it} - \bar{\nu}_i. \tag{6}$$

In other words, either demeaning (1) or (2) and estimating the result via OLS provides for the fixed-effects estimates.

Finally, inserting (6) into (2) shows that (2) is equivalent to (5). Therefore, estimating either (2) or (5) using OLS provides for the between-groups effects $\mathbf{b}$. To put it alternatively, whether averaging either (1) or (2) over time and estimating the results via OLS must yield the same estimates, namely those of the between-groups effects.

Proposition 2: Estimating specification (2) via GLS yields exactly the same results as the OLS estimation of this specification.

Proof: First, estimating panel model (1) via GLS is equivalent to estimating

$$y_{it} - \lambda \cdot \bar{y}_i = \beta_0 \cdot (1 - \lambda) + \boldsymbol{\beta}^T \cdot (\mathbf{x}_{it} - \lambda \cdot \bar{\mathbf{x}}_i) + \xi_i - \lambda \cdot \xi_i + \nu_{it} - \lambda \cdot \bar{\nu}_i \tag{7}$$

via OLS (see e.g. WOOLDRIDGE 2008:490), where λ is a parameter that is determined by the time horizon and the variances of the error terms ξ_i and ν_{it}. Second, employing the same transformation to the modified specification (2) yields:

$$y_{it} - \lambda \cdot \bar{y}_i = \beta_0 \cdot (1-\lambda) + \mathbf{w}^T(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) - \lambda \cdot \mathbf{w}^T \cdot \overline{(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)} + \mathbf{b}^T \cdot \bar{\mathbf{x}}_i - \lambda \cdot \mathbf{b}^T \cdot \bar{\bar{\mathbf{x}}}_i + \xi_i - \lambda \cdot \xi_i + \nu_{it} - \lambda \cdot \bar{\nu}_i. \tag{8}$$

Recognizing that $\overline{(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)} = 0$ and $\bar{\bar{\mathbf{x}}}_i = \bar{\mathbf{x}}_i$ and rearranging gives:

$$y_{it} = \beta_0 + \mathbf{w}^T(\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + \mathbf{b}^T \bar{\mathbf{x}}_i + \xi_i + \nu_{it} - \lambda \cdot \underbrace{(\bar{y}_i - \beta_0 - \mathbf{b}^T \bar{\mathbf{x}}_i - \xi_i - \bar{\nu}_i)}_{=0},$$

where the last bracket vanishes because of equation (3). In short, both transformation (8) and specification (2) are identical and, hence, the OLS estimation of specification (2) delivers the same results as estimating transformation (8) via OLS, which is in turn equivalent to estimating specification (2) via GLS.

3 Empirical Example

To demonstrate the usefulness of the test variant, we employ household data drawn from the German Mobility Panel (MOP 2007) and estimate fuel-price elasticities using the following specification suggested by Proposition 1:

$$\ln(e_{it}) = \beta + b_p \cdot \overline{\ln(p_i)} + w_p \cdot (\ln(p_{it}) - \overline{\ln(p_i)}) + \mathbf{b_x}^T \cdot \bar{\mathbf{x}}_i + \mathbf{w_x}^T \cdot (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + \xi_i + \nu_{it} \;, \quad (9)$$

where $\ln(e)$ is the logged monthly fuel consumption, $\ln(p)$ denotes logged real fuel price per liter and $\mathbf{x}$ designates a vector of control variables such as age of the car and whether it is a premium make. A detailed data description can be found in FRONDEL, PETERS, VANCE (2008).

The advantage of estimating specification (9), irrespective of whether either OLS or GLS is used, is that it allows us to at once retrieve the entire set of between-groups and fixed effects and, hence, to easily examine both the equality of the coefficients for individual variables, such as $\ln(p)$, on the basis of ordinary t-tests, as well as the equality of the whole range of coefficients using an F-test[1]:

$$H_0 : b_p = w_p, \mathbf{b_x} = \mathbf{w_x}, \quad (10)$$

where w_p and $\mathbf{w_x}$ designate the fixed effects and b_p and $\mathbf{b_x}$ the between-groups effects, respectively. According to HAUSMAN, TAYLOR (1981), any rejection of the null hypothesis H_0 also implies that the fixed- and random effects are different and, hence, that the fixed effects should be preferred over the random effects.

[1] In a different context, KRÄMER and REHBERGER (1986:98-99) show for the classical regression model $y = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}$, where $\boldsymbol{\gamma} = 0$ under H_0 and $\boldsymbol{\gamma} \neq 0$ under the alternative, that the F-test is identical to the HAUSMAN test if the number of variables included in $\mathbf{Z}$ is smaller than those included in $\mathbf{X}$.

In our empirical example, the result of the standard HAUSMAN test reported in Table 1 indicates that the orthogonality hypothesis of the unobservable individual-specific effects and the regressors is rejected. From the t-test results reported in the last column, it becomes obvious that the reason for the rejection of the null is primarily due to the difference in the estimates of just two variables: the number of employed household members and the indicator for whether a vacation was taken during the survey period.

Table 1: OLS Estimates of Specification (9) and Test Results.[2]

	Fixed Effects		Between-Groups Effects		t-Test
$\ln(e)$	Coeff.s	Std. Errors	Coeff.s	Std. Errors	Results
$\ln(p)$	* -0.569	(0.166)	**-0.605	(0.188)	-0.15
car age	-0.010	(0.007)	* -0.015	(0.006)	-0.55
household size	-0.002	(0.048)	0.016	(0.028)	0.33
children	-0.020	(0.118)	0.045	(0.079)	0.46
# high school diploma	0.002	(0.055)	0.085	(0.036)	1.28
# employed	** -0.117	(0.044)	** 0.217	(0.037)	** 5.80
job change	** 0.116	(0.070)	0.115	(0.077)	0.00
vacation with car	** 0.254	(0.037)	** 0.435	(0.068)	* 2.36
diesel car	-0.297	(0.196)	0.026	(0.089)	1.51
premium car	0.149	(0.149)	** 0.326	(0.052)	1.12
Standard HAUSMAN Test:			χ^2 (10) = ** 57.34		

Note: * denotes significance at the 5 %-level and ** at the 1 %-level, respectively. Number of observations (households) used for the estimations: 1,341 (530).

In contrast, the between-groups and fixed-effects estimates of the key variable, $\ln(p)$, do not significantly differ from each other. This also suggests that the respective fixed- and random-effects estimates are equal in statistical terms, a suggestion that is

[2] To correct for the non-independence of repeated observations from the same households over the years of the survey, the regression disturbance terms are clustered at the household level. The presented measures of statistical significance are robust to this survey design feature. The t-test results are calculated using $t = (b_{x_k} - w_{x_k})/\sqrt{\widehat{Var(b_{x_k})} + \widehat{Var(w_{x_k})}}$, where the covariance $Cov(b_{x_k}, w_{x_k})$ vanishes due to the orthogonality of $\mathbf{b_x}$ and $\mathbf{w_x}$.

substantiated by the closeness of the fixed-effects estimate of -0.569 and the (unreported) random-effects estimate of -0.579. Furthermore, it can be empirically demonstrated that using STATA's fixed- and between-groups effects estimation options precisely reproduces the OLS estimates displayed in Table 1, as is claimed in Proposition 1.[3] It also bears noting that these OLS estimates are perfectly identical to those obtained when using STATA's random-effects estimation option, which is in line with Proposition 2's claim that both the OLS and GLS estimates of specification (2) are equal.

4 Summary and Conclusion

The HAUSMAN (1978) specification test is commonly employed for selecting between the fixed- and random-effects estimators for panel data. The random-effects estimator is based on the assumption that there is zero correlation between the regressors and the error term, a situation that should be considered the exception rather than the rule (WOOLDRIDGE 2008:493). It is therefore not surprising that this null hypothesis is frequently not found to withstand empirical scrutiny. If the test statistic, which contrasts the fixed- and random-effects estimates, rejects the null, applied researchers generally discard the random effects and base their conclusions on the fixed-effects estimates.

This all-or-nothing choice prompted HAUSMAN and TAYLOR (1981) to propose a model that introduces an instrumental variable estimator using both between- and within-groups variation to correct for the correlation of selected regressors with the individual effect. Using a straightforward model specification that also draws on the between- and within-groups variation, we suggest a test variant that is based on the contrast of between-groups and fixed effects and allows us to examine both the equality of the whole set of coefficients as well as that of individual variables, an issue that cannot be addressed on the basis of the standard HAUSMAN test.

[3]The employed data set and code is available from the authors upon request. It should be noted that STATA's between-groups estimates and the OLS estimates of $\mathbf{b}$ of specification (2) are not perfectly identical if the employed panel is unbalanced, as in our example. In this case, one has to use weighted least squares (WLS) in order to correct for the frequency of a household's occurrence in the panel.

References

FRONDEL, M. , PETERS, J. , VANCE, C. (2008) Identifying the Rebound: Evidence from a German Household Panel *The Energy Journal*, **29 (4)**, 154-163.

KRÄMER, W. , SONNBERGER, H. (1986) *The Linear Regression Model under Test*. Physica, Heidelberg, Wien.

MOP (2008) German Mobility Panel. http://www.ifv.uni-karlsruhe.de/MOP.html

HAUSMAN, J. A. (1978) Specification Tests in Econometrics *Econometrica* **46**, 1251-1271.

HAUSMAN, J. A. , TAYLOR, W. E. (1981) Panel Data and Unobservable Individual Effects *Econometrica* **49**, 1377-1398.

WOOLDRIDGE, J. M. (2008) *Introductory Econometrics*. Fourth Edition, South Western.