Economics Letters 107 (2010) 327–329

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Fixed, random, or something in between? A variant of Hausman's specification test for panel data estimators

Manuel Frondel *, Colin Vance

Ruhr-Universität Bochum (RUB), Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), Hohenzollernstr. 1-3, D-45128 Essen, Germany

ARTICLE INFO

Article history:
Received 28 November 2008
Received in revised form 20 January 2010
Accepted 24 February 2010
Available online 3 March 2010

Keywords:
Specification tests
Fuel price elasticity

JEL classification:
C12

ABSTRACT

This paper proposes a variant of the Hausman specification test for panel models that allows us to examine both the equality of the whole sets of coefficients of two alternative models as well as that of individual variables.

1. Introduction

The econometric modeling of panel data typically applies two principal 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 obtained from the fixed-effects estimation – taken as a group – should not differ systematically from the set derived via random-effects estimation under the null hypothesis that the unobservable, individual-specific effects and the regressors are orthogonal. If the test results suggest rejecting the equality of both coefficient sets, applied researchers generally proceed to draw conclusions based on the fixed-effects estimates. This course of action 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 of interest.

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 (Hausman and Taylor, 1981). Specifically, we show that using a straightforward model specification allows us to simultaneously estimate the fixed- and between-groups effects – either on the basis of Ordinary (OLS) or Generalized Least Squares (GLS) – and to test 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 for Germany. 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 focuses on fixed- and random effects. Yet, using the fact that testing the equality of fixed- and random effects is numerically equivalent to testing 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 approach.

Proposition 1. Departing from a standard panel data model,

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

* Corresponding author.
E-mail address: frondel@rwi-essen.de (M. Frondel).

0165-1765/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2010.02.007

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

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

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 $\mathbf{b}$.

Heuristic derivation. First, the between-groups estimator of parameter vector β emerging from model (1) can be obtained by averaging Eq. (1) over time and estimating the result via OLS:

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

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 β can be retrieved by subtracting Eq. (3) from Eq. (1) and estimating the result via OLS:

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

Third, instead of estimating either Eq. (3) or Eq. (4), we alternatively suggest in Proposition 1 estimating Eq. (2) via OLS to at once get both the fixed- and between-groups effects. This can be seen as follows: Upon averaging Eq. (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 Eq. (3):

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

Therefore, averaging either Eq. (1) or Eq. (2) over time and estimating the results via OLS must yield the same estimates, namely those of the between-groups effects.

Finally, subtracting Eq. (5) from Eq. (2) yields Eq. (4), with $\mathbf{w}$ instead of β as the parameter vector:

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

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

Proposition 2. Estimating Eq. (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

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

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 Eq. (2) yields:

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

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

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

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

3. Empirical example

To demonstrate the usefulness of the test, we employ household data drawn from the German Mobility Panel (MOP, 2010) and estimate fuel price elasticities using the following specification suggested by Proposition 1:

$$\begin{aligned} ln(e_{it}) = {} & \beta + b_p\cdot\overline{ln(p_i)} + w_p\cdot\left(ln(p_{it}) - \overline{ln(p_i)}\right) + \mathbf{b}_{\mathbf{x}}^T\cdot\ \bar{\mathbf{x}}_i \\ & + \mathbf{w}_{\mathbf{x}}^T\cdot\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right) + \xi_i + \nu_{it}, \end{aligned} \tag{9}$$

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 et al. (2008) or Frondel and Vance (2009).

The advantage of estimating Eq. (9), irrespective of whether 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 on the basis of ordinary t-tests, and the equality of the whole range of coefficients using an F-test[1]:

$$H_0 : b_p = w_p, \mathbf{b}_{\mathbf{x}} = \mathbf{w}_{\mathbf{x}}, \tag{10}$$

where w_p and $\mathbf{w}_{\mathbf{x}}$ designate the fixed effects and b_p and $\mathbf{b}_{\mathbf{x}}$ the between-groups effects, respectively. According to Hausman and 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.

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.

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 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.[2] It also bears noting that these OLS estimates are 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 Eq. (2) are equal.

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

[2] 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 Eq. (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.

Table 1
OLS estimates of Eq. (9) and test results.[3]

$\ln(e)$	Fixed effects		Between-groups effects		t-Test
	Coeff. s	Std. errors	Coeff. s	Std. errors	Statistics
$\ln(p)$	-0.569^{*}	(0.166)	-0.605^{**}	(0.188)	-0.21
Car age	-0.010	(0.007)	-0.015^{*}	(0.006)	-0.78
Household size	-0.002	(0.048)	0.016	(0.028)	0.38
Children	-0.020	(0.118)	0.045	(0.079)	0.66
# High school diploma	0.002	(0.055)	0.085	(0.036)	1.59
# Employed	-0.117^{**}	(0.044)	0.217^{**}	(0.037)	7.76^{**}
Job change	0.116^{**}	(0.070)	0.115	(0.077)	0.00
Vacation with car	0.254^{**}	(0.037)	0.435^{**}	(0.068)	3.55^{**}
Diesel car	-0.297	(0.196)	0.026	(0.089)	1.09
Premium car	0.149	(0.149)	0.326^{**}	(0.052)	1.90
Standard Hausman Test			$\chi^2\ (10) = 57.34^{**}$		

Note: * denotes significance at the 5%-level and ** at the 1%-level, respectively. Number of observations (households) used for the estimations: 1341 (530).

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 the correlation between the regressors and the unobservable, individual-specific effects is zero, 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 the equality of both the whole sets of coefficients and that of individual variables, an issue that cannot be addressed on the basis of the standard Hausman test.

Acknowledgements

We are grateful for invaluable comments and suggestions by Walter Krämer and Christoph M. Schmidt. This work has been supported in part by the Collaborative Research Center “Statistical Modelling of Nonlinear Dynamic Processes” (SFB 823) of the German Research Foundation (DFG).

References

Frondel, M., Vance, C., 2009. Do high oil prices matter? Evidence on the mobility behavior of German households. Environmental and Resource Economics 43 (1), 81–94.
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.
MOP (2010) 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.

[3] 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 statistics are calculated using $t = (b_{x_k} - w_{x_k}) / \sqrt{\widehat{\mathrm{Var}}(b_{x_k}) + \widehat{\mathrm{Var}}(w_{x_k})}$, where the covariance $\mathrm{Cov}(b_{x_k}, w_{x_k})$ vanishes due to the orthogonality of the vectors $\mathbf{x}_i$ and $(\mathbf{x}_{it} - \mathbf{x}_i)$ pertaining to $\mathbf{b_x}$ and $\mathbf{w_x}$, respectively.