**Working Paper Series 624**
(ISSN 1211-3298)

# Time-invariant Regressors under Fixed Effects: Identification via a Proxy Variable

Matěj Bělín

CERGE-EI
Prague, September 2018

**ISBN 978-80-7343-431-1 (Univerzita Karlova, Centrum pro ekonomický výzkum a doktorské studium)**
**ISBN 978-80-7344-474-7 (Národohospodářský ústav AV ČR, v. v. i.)**

# Time-invariant Regressors under Fixed Effects: Identification via a Proxy Variable

Matěj Bělín*

September 11, 2018

**Abstract**

Identification of a coefficient associated with a time-invariant regressor (TIR) often relies on the assumption that the TIR is uncorrelated with the unobserved heterogeneity across panel units. We derive an estimator which avoids the random-effects assumption by employing a proxy for the unobserved heterogeneity thus extending the existing results on proxy variables from the cross-sectional literature. In addition, we quantify the sensitivity of the estimates to potential violations of the random-effects assumption when no proxy is available. The utility of this approach is illustrated on the problem of implausibly high distance elasticity produced by gravity models of international trade.

***Key words:*** Identification; Model specification; Omitted variable bias; Panel data; Variable addition

***JEL Classification:*** C01, C18, C33

*CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Politických vězňů 7, 111 21 Prague, Czech Republic. Email: matej.belin@cerge-ei.cz. Citigroup endowment scholarship is gratefully acknowledged. This paper benefited greatly from helpful discussions with Štěpán Jurajda and Nikolas Mittag. Any remaining errors are my own.

# 1 Introduction

The estimation of parameters associated with time-invariant regressors (TIRs) in panel data is often based on strong assumptions. This is because separating the effect of TIRs from the unobserved, time-invariant heterogeneity places either high demands on the data in terms of the availability of instruments (Hausman & Taylor, 1981) or high demands on the restrictiveness of the model in terms of the assumed lack of correlation between TIRs and individual-specific effects (Krishnakumar, 2006; Plümper & Troeger, 2011; Woodcock, 2015).

In this paper, we show that assumptions commonly imposed on proxy variables in cross-sectional settings (e.g. Lewbel, 1997; Lubotsky & Wittenberg, 2006; Bollinger & Minier, 2015) lead readily to the identification of the TIR coefficient. Thus, when a proxy variable for the unobserved heterogeneity is available, it allows indentification of the TIR coefficient under arbitrary correlation between the TIR and the latent confounder.

Further, when no such proxy is observed, we derive an expression for the bias in the random-effects coefficient that allows the researcher to quantify the sensitivity of the regression coefficient to potential violations in the random-effects assumption. The intuition behind this approach is the same as in Altonji, Elder, and Taber (2005), but the panel data allow us to impose fewer assumptions on the latent variable and hence the sensitivity measure becomes more informative.

To illustrate the utility of the proposed technique, we study the estimates of the elasticity of international trade flows with respect to the distance between trading countries. As documented by Balistreri and Hillberry (2006), Leamer (2007), Coe, Subramanian, and Tamirisa (2007), Disdier and Head (2008), and recently re-emphasised by Head and Mayer (2014), standard gravity models of international trade imply very large trade costs that are highly persistent over time. This has led to suspicions that the geographic distance reflects not only the transportation costs, but also other factors which dampen trade flows, e.g. cultural or institutional dissimilarity, which are likely to correlate with distance and may suppress trade activities (e.g. Blum & Goldfarb, 2006; Guiso et al., 2009; Head & Mayer, 2013;

Lendle et al., 2016). Proxying these latent differences with an index of institutional similarity reveals a substantial bias in the naive distance elasticity estimate.

The remainder of this paper is organised as follows: first, it specifies the model under consideration, and then studies its properties under the inclusion of a proxy variable. Further, we extend the discussion to cases when no proxy is available. Thereafter, we apply the proposed approach on trade data from OECD countries. The paper closes with brief concluding remarks.

# 2 Structural model

## 2.1 Basic framework

Consider a generic panel model:

$$Y_{it} = U_i + Z_i\beta + \varepsilon_{it}, \tag{1}$$

where $Y_{it}$ is the outcome variable, $U_i$ is an individual-specific intercept, and $Z_i$ is the TIR of interest. To the extent that the model includes other variables, assume that they can be purged by the Frisch-Waugh theorem. The objective is to obtain a consistent estimate of $\beta$. To allow arbitrary covariance between $Z_i$ and $U_i$, Assumption 1 specifies a linear relationship between these variables.
**Assumption 1.** *Let the data-generating process in (1) be governed by:*

$$U_i = \alpha + Z_i\lambda + \xi_i \tag{2}$$
$$0 = \mathbb{E}[\xi_i|Z_i] = \mathbb{E}[\xi_i] \tag{3}$$
$$0 = \mathbb{E}[\varepsilon_{it}|Z_i, \xi_i]. \tag{4}$$

*Assume further that finite unconditional variances of $\varepsilon_{it}$, $Z_i$, and $\xi_i$ exist, denoted by* $\mathrm{var}\,(\varepsilon), \mathrm{var}\,(Z)$, *and* $\mathrm{var}\,(\xi)$ *respectively, which are estimable from the data.*

Condition (3) is fairly standard as it requires strict exogeneity of $Z_i$ and $\xi_i$ with respect to the idiosyncratic disturbances. Condition (4) requires that (2) is correctly specified. Except for pathological cases, estimating variances is relatively straightforward, although note that the standard estimator for $\mathrm{var}(\xi)$ rules out autocorrelations in $\varepsilon_{it}$ (Swamy & Arora, 1972). Therefore, it is advisable to check if allowing these autocorrelations changes the estimate substantially.

Plugging (2) into (1) shows that under Assumption 1, running a regression

$$Y_{it} = b_0 + Z_i b_1 + e_{it} \tag{5}$$

leads in population[1] to $b_1 = \beta + \lambda$. In cases when an instrumental variable can be obtained (Hausman & Taylor, 1981), the bias term $\lambda$ can be eliminated easily. However, the approach taken here is to utilise information contained in an imperfect *proxy* for $U_i$, thus avoiding the need to defend the proper exclusion of an instrument.

## 2.2 Identification

Following the literature on imperfect control variables (Lewbel, 1997; Lubotsky & Wittenberg, 2006; Bollinger & Minier, 2015), let us assume that a proxy variable for $U_i$ is available, which is subject to classical measurement error.

**Assumption 2.** *Suppose that a proxy variable for $U_i$ is available, and obeys the following relationship:*

$$U_i^* = \phi_0 + U_i \phi_1 + \eta_i, \tag{6}$$

*such that:*

$$0 = \mathbb{E}[\eta_i Z_i \xi_i] = \mathbb{E}[\eta_i]\mathbb{E}[\xi_i]\mathbb{E}[Z_i] = \mathbb{E}[\eta_i]. \tag{7}$$

*Assume further that a finite unconditional variance of $\eta_i$ exists, denoted by* $\mathrm{var}(\eta)$.

Condition (6) specifies a linear projection between $U_i^*$ and $U_i$, and (7) guaran-

[1] All estimated quantities in this paper will be treated as population estimates to simplify the notation by suppressing the probability limits.

tees that the measurement error is classical, i.e. uncorrelated with regressor $Z_i$. This additional requirement is mild. By Assumption 1, $\text{cov}\,(Z, \xi) = 0$ and the linear projection in (6) renders $\eta_i$ uncorrelated with $U_i$. It therefore follows that $\text{cov}\,(\xi, \eta) = -\lambda \text{cov}\,(Z, \eta)$. This equality is naturally satisfied when the measurement error $\eta_i$ does not correlate with either $Z_i$ or $\xi_i$. Hence (7) only rules out the knife-edge cases when the two covariances happen to offset each other precisely.

With Assumptions 1 and 2 in place, it is tempting to solve the identification problem by including the proxy variable $U^*$ on the right-hand side of the regression equation and to fit an augmented model:

$$Y_{it} = b_0^* + Z_i b_1^* + U_i^* b_2^* + e_{it}^*. \tag{8}$$

Note that the regression with a mismeasured control introduces bias as well. Asymptotically, the coefficient of interest becomes a weighted average of $(\beta + \lambda)$ and $\beta$ itself:

$$b_1^* = \frac{(\lambda + \beta)\text{var}\,(\eta) + \beta \phi_1^2 \text{var}\,(\xi)}{\text{var}\,(\eta) + \phi_1^2 \text{var}\,(\xi)}. \tag{9}$$

Appendix A.1 provides the derivation. Intuitively, the poorer proxy for $U_i$ is available, the smaller proportion of omitted variable bias is eliminated, which is one of the reasons Pei, Pischke, and Schwandt (2018) argue against using $U_i^*$ as an additional control. Nevertheless, here $b_1^*$ leads to the identification of $\beta$. Observe first that running a regression:

$$U_i^* = c_0 + Z_i c_1 + v_i \tag{10}$$

yields the following estimates:

$$c_1 = \phi_1 \lambda \tag{11}$$

$$s_v^2 = \phi_1^2 \text{var}\,(\xi) + \text{var}\,(\eta), \tag{12}$$

where $s_v^2$ is the estimated variance of the error terms $v_i$. Since $\text{var}\,(\xi)$ is estimable from (5), there are only four unknowns remaining: $\beta, \lambda, \phi_1$, and $\mathrm{var}(\eta)$. The following proposition supplies the formula for the key parameter $\lambda$.

**Proposition 1.** *Under Assumptions 1 and 2, the bias term $\lambda$ can be recovered from regressions (5), (8), and (10) as:*

$$\lambda = \frac{c_1^2 s_\xi^2}{(b_1 - b_1^*) s_v^2}, \tag{13}$$

*where $s_\xi^2$ is the estimated variance of the random effects.*

*Proof:* see Appendix A.1. Having identified $\lambda$, it remains to compute the unbiased estimate of $\beta$ as $b_1 - \widehat{\lambda}$. The reason identification is feasible here and not in the cross-sectional cases studied by Lubotsky and Wittenberg (2006), Bollinger and Minier (2015), and Oster (2016) is that panel data furnish an additional piece of information about the latent variable, $\mathrm{var}(\xi)$, which is unavailable in cross-sections.

It bears noting that in practical contexts, $b_1 - b_1^*$ might be close to zero, so using (13) to compute the bias term may lead to a very noisy estimate of $\lambda$.[2] Hence, care needs to be taken to verify that including the proxy variable on the right-hand side changes the TIR parameter of interest by a statistically significant margin.

# 3 Correlation-based robustness measure

When Assumption 2 is not justifiable, a more agnostic approach is needed. After estimating model (5), one may follow the intuition of Altonji et al. (2005) to quantify how strong the endogeneity has to be to produce bias $\lambda$ large enough to invalidate the conclusions drawn from $b_1$. Measuring endogeneity of $Z_i$ as its

[2]See Lewbel (1997) for a discussion of identification under classical measurement error sans external instruments. His estimator suffers from a similar problem as (13), as it depends on the skewness of the proxy (performing worst for symmetrically distributed variables).

correlation with $U_i$ leads to the following expression (derived in Appendix A.1):

$$\text{corr}\,(U, Z) = \lambda \times \sqrt{\frac{\text{var}\,(Z)}{\lambda^2 \text{var}\,(Z) + \text{var}\,(\xi)}}. \tag{14}$$

Plugging $b_1$ from regression (5) as $\lambda$ into (14) can be used to calculate how far $\text{corr}\,(U, Z)$ has to deviate from the assumed value of zero to account for the estimated coefficient in its entirety. Should even small values of $\text{corr}\,(U, Z)$ lead to high biases, then one would be ill-advised to rely on the regression results. Panel data again provide a significant advantage here, since the calculation in (14) needs no further assumptions, while in the cross-sectional settings considered by Altonji et al. (2005), an additional assumption is needed on how much residual variance in regression (5) is attributable to the latent variable (cf. Oster, 2016, sec. 3.3.1 for discussion).

# 4 Empirical application

We estimate a simple gravity model of international trade between OECD countries for the period 2005–2014. The parameter of interest, $b_1$, is the elasticity of international trade flows with respect to the distance between the trading partners. To obtain $b_1^*$, an index of institutional similarity is used as a proxy for the latent cross-country differences. Regressions (5), (8), and (10) were estimated using three separate OLS models. Since computation of $\widehat{\lambda}$ requires combining parameters from multiple regressions, bootstrapping was used to obtain standard errors. In addition, the whole system was estimated jointly as a single MLE model for comparison (see Appendix A.2 for details). Table 1 reports the results. The model estimates trade elasticity at roughly negative unity, which is well in line with the published literature (see Disdier & Head, 2008, for a survey). While adding the proxy variable leads to an almost imperceptible change in the estimated elasticity, once (13) is used to account for the measurement error, the estimated $\lambda$ becomes much more pronounced (albeit rather imprecisely estimated). Hence $b_1 - b_1^*$ on its own severely understates the omitted variable bias. The large negative bias in the estimated

**Table 1:** Estimated distance elasticity of international trade flows.

| | | Bootstrap OLS | | MLE | |
|---|---|---|---|---|---|
| Param. | Description | Est. | SE | Est. | SE |
| $b_1$ | Distance elasticity (Eq. 5) | -1.055 | 0.035 | -1.058 | 0.034 |
| $b_1^*$ | Distance elasticity (Eq. 8) | -1.040 | 0.035 | -1.044 | 0.033 |
| $\lambda$ | Bias in $b_1$ | -0.501 | 0.246 | -0.533 | 0.194 |
| $b_1 - \lambda$ | Distance elasticity (unbiased) | -0.554$^\dagger$ | 0.246 | -0.525 | 0.187 |
| $\rho^0_{U,Z}$ | corr $(U, Z)$ if $b_1 = \lambda$ | -0.826 | 0.012 | -0.828 | 0.014 |

*Note:* All results are significant at 1% unless significance at 10% is indicated by $^\dagger$.

distance elasticity is also consistent with the empirical literature (Blum & Goldfarb, 2006; Head & Mayer, 2013; Lendle et al., 2016). However, it is unlikely that the entirety of the estimated parameter is attributable to omitted variable bias, as $b_1 = \lambda$ implies a correlation between distance and the latent confounder ($\rho^0_{U,Z}$) of more than –0.8. Thus, the unobserved variable would have to be a "clone" of distance, which is scarcely plausible.

# 5 Concluding remarks

This paper offers a method for accommodating time-invariant regressors into linear panel models when the presence of fixed effects is suspected. To this end, it shows that it is possible to identify the relevant TIR coefficient, provided that a proxy for the latent confounder can be procured.

In addition, the paper proposes a technique for measuring the robustness of results from random-effects models, which is attractive as it avoids the need to assume exogeneity of the TIR in question, or to use instruments, or indeed to find a proxy for the omitted variable.

# A Appendices

## A.1 Proofs

*Proof of Equation (9).* Assume for simplicity that $Y_{it}, U_i$, and $Z_i$ have zero means. OLS therefore estimates parameters in (8) as:

$$\begin{bmatrix} b_1^* \\ b_2^* \end{bmatrix} = \begin{bmatrix} \text{var}\,(Z) & \text{cov}\,(Z, U^*) \\ \text{cov}\,(Z, U^*) & \text{var}\,(U^*) \end{bmatrix}^{-1} \begin{bmatrix} \text{cov}\,(Z, Y) \\ \text{cov}\,(U^*, Y) \end{bmatrix}. \tag{15}$$

Carrying out the matrix multiplication yields:

$$b_1^* = \frac{\text{cov}\,(Z, Y)\text{var}\,(U^*) - \text{cov}\,(U^*, Y)\text{cov}\,(Z, U^*)}{\text{var}\,(U^*)\text{var}\,(Z) - \text{cov}\,^2(Z, U^*)}. \tag{16}$$

Under Assumptions 1 amd 2, the requisite variances and covariances can be expressed in terms of the structural parameters as:

$$\begin{aligned} \text{cov}\,(Z, Y) &= (\beta + \lambda)\text{var}\,(Z) && (17) \\ \text{cov}\,(Z, U^*) &= \lambda\phi_1\text{var}\,(Z) && (18) \\ \text{cov}\,(U^*, Y) &= \phi_1\left[\text{var}\,(\xi) + \lambda(\beta + \lambda)\text{var}\,(Z)\right] && (19) \\ \text{var}\,(U^*) &= \phi_1^2\lambda^2\text{var}\,(Z) + \phi_1^2\text{var}\,(\xi) + \text{var}\,(\eta). && (20) \end{aligned}$$

Plugging (17)–(20) into (16) yields:

$$b_1^* = \frac{\text{var}\,(Z)[(\beta + \lambda)\text{var}\,(\eta) + \beta\phi_1^2\text{var}\,(\xi)]}{\text{var}\,(Z)[\text{var}\,(\eta) + \phi_1^2\text{var}\,(\xi)]}. \tag{21}$$

Cancelling out var $(Z)$ delivers Equation (9). □

Note that when there is no measurement error, (i.e. $\text{var}\,(\eta) = 0$), then $b_1^* = \beta$ since in that case $U_i^*$ properly controls for the omitted variable irrespective of the scaling factor $\phi_1$. An alternative derivation via the formula for omitted variable bias can be found in Oster (2016), and Pei et al. (2018).

*Proof of Proposition (1).* Recall that regression (5) produces $b_1 = \beta + \lambda$ and the augmented model (8) yields $b_1^*$ given in (9). Therefore:

$$b_1 - b_1^* = \frac{\lambda \phi_1^2 \operatorname{var}(\xi)}{\phi_1^2 \operatorname{var}(\xi) + \operatorname{var}(\eta)}. \tag{22}$$

Since (22) eliminates $\beta$, what remains is a system of three simultaneous equations (11), (12), and (22) with three unknowns $\lambda, \phi_1$, and $\operatorname{var}(\eta)$. The solution for $\lambda$ is given in Proposition 1 after replacing $\operatorname{var}(\xi)$ with its estimate $s_\xi^2$. □

*Proof of Equation (14).* Note that $U_i = \lambda Z_i + \xi_i$ as above. Hence the correlation between $Z_i$ and $U_i$ is:

$$\operatorname{corr}(U, Z) = \frac{\operatorname{cov}(U, Z)}{\sqrt{\operatorname{var}(U)\operatorname{var}(Z)}} = \frac{\lambda \operatorname{var}(Z)}{\sqrt{\lambda^2 \operatorname{var}^2(Z) + \operatorname{var}(\xi)\operatorname{var}(Z)}}, \tag{23}$$

where the second equality holds due to assumed uncorrelatedness between $Z_i$ and $\xi_i$. Cancelling out $\sqrt{\operatorname{var}(Z)}$ yields the required result. □

## A.2 Data and empirical specification

Trade flows of goods were extracted from the OECD *Structural Analysis Database* (STAN). A dataset by Gurevich and Herman (2018) was used as a source of contextual variables. The proxy for the unobserved cross-country differences was constructed as a Koghut-Singh index of the Worldwide Governance Indicators maintained by the World Bank (see Möhlmann et al., 2010, for a similar approach).

The model was specified as:

$$\begin{aligned}
\ln(\text{Trade flows})_{ijt} = & b_0 + b_1 \ln(\text{Distance})_{ij} + a_1 \ln(\text{GDP})_{it} + a_2 \ln(\text{GDP})_{jt} + \\
& + a_3 \ln(\text{Population})_{it} + a_4 \ln(\text{Population})_{jt} + \\
& + a_5 \overline{\ln(\text{GDP})}_i + a_6 \overline{\ln(\text{GDP})}_j + a_7 \overline{\ln(\text{Population})}_i \\
& + a_8 \overline{\ln(\text{Population})}_j + \sum_{s=2006}^{2014} a_s \mathbf{1}[t = s] + e_{ijt}.
\end{aligned} \tag{24}$$

Subscripts $i$, $j$, and $t$ index exporter, importer, and time respectively. The overline specifies the mean within a country dyad. Following Mundlak (1978), we assume that $Y_{ijt} = U_{ij} + Z_{ij}\beta + \mathbf{X}_{ijt}\boldsymbol{\gamma} + \varepsilon_{ijt}$ and $U_{ij} = Z_{ij}\lambda + \overline{\mathbf{X}}_{ij}\boldsymbol{\delta} + \xi_{ij}$, which is why within means were included in the specification. Due to collinearity, within means of year dummies were omitted. In addition, a second regression was fitted where (24) was augmented by the inclusion of the proxy variable described above. The third regession projected the proxy on ln(Distance) and the within means. To measure sensitivity of $b_1$ to endogeneity by (14), we obtained the conditional variance of ln(Distance) by regressing it on within means and isolating the residuals. Variance of the random effects, $\operatorname{var}(\xi)$, was calculated by the Swamy and Arora (1972) method since allowing autocorrelated idiosyncratic disturbances has small effect on the resulting estimate. After estimating these four models, the results were plugged into (13) to calculate $\lambda$. Standard errors were obtained by bootstrapping with 1000 replications clustered by the country dyads.

In addition, the system of four regressions was fitted in a single MLE model which avoids the need to bootstrap, but instead it assumes that the random effects are normally distributed. Nevertheless, both point estimates and standard errors match closely between the MLE and bootstrapped OLS.

# References


Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, *113*(1), 151-184.

Balistreri, E. J., & Hillberry, R. H. (2006). Trade frictions and welfare in the gravity model: How much of the iceberg melts? *The Canadian Journal of Economics / Revue canadienne d'Economique*, *39*(1), 247-265.

Blum, B. S., & Goldfarb, A. (2006). Does the internet defy the law of gravity? *Journal of International Economics*, *70*(2), 384-405.

Bollinger, C. R., & Minier, J. (2015). On the robustness of coefficient estimates to the inclusion of proxy variables. *Journal of Econometric Methods*, *4*(1), 101.

Coe, D. T., Subramanian, A., & Tamirisa, N. T. (2007). The missing globalization puzzle: Evidence of the declining importance of distance. *IMF Staff Papers*, *54*(1), 34-58.

Disdier, A.-C., & Head, K. (2008). The puzzling persistence of the distance effect on bilateral trade. *The Review of Economics and Statistics*, *90*(1), 37-48.

Guiso, L., Sapienza, P., & Zingales, L. (2009). Cultural biases in economic exchange? *The Quarterly Journal of Economics*, *124*(3), 1095-1131.

Gurevich, T., & Herman, P. (2018). The dynamic gravity dataset: 1948-2016. *USITC Working Paper 2018-02-A.*.

Hausman, J. A., & Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econometrica*, *49*(6), 1377-1398.

Head, K., & Mayer, T. (2013). What separates us? Sources of resistance to globalization. *Canadian Journal of Economics/Revue canadienne d'économique*, *46*(4), 1196-1231.

Head, K., & Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In E. Helpman, G. Gopinath, & K. Rogoff (Eds.), *Handbook of international economics, vol. 4* (p. 131-195). Amsterdam: North-Holland.

Krishnakumar, J. (2006). Time invariant variables and panel data models : A generalised Frisch Waugh theorem and its implications. In B. H. Baltagi

(Ed.), *Panel data econometrics — theoretical contributions and empirical applications.* Elsevier.

Leamer, E. E. (2007). A flat world, a level playing field, a small world after all, or none of the above? A review of Thomas L Friedman's The World is Flat. *Journal of Economic Literature*, *45*(1), 83-126.

Lendle, A., Olarreaga, M., Schropp, S., & Vézina, P.-L. (2016). There goes gravity: eBay and the death of distance. *The Economic Journal*, *126*(591), 406-441.

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*(5), 1201-1213.

Lubotsky, D., & Wittenberg, M. (2006). Interpretation of regressions with multiple proxies. *The Review of Economics and Statistics*, *88*(3), 549-562.

Möhlmann, J., Ederveen, S., de Groot, H. L., & Linders, G.-J. M. (2010). Intangible barriers to international trade: A sectoral approach. In P. A. G. van Bergeijk & S. Brakman (Eds.), *The gravity model in international trade: Advances and applications* (p. 224-251). Cambridge: Cambridge University Press.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, *46*(1), 69-85.

Oster, E. (2016). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 1-18.

Pei, Z., Pischke, J.-S., & Schwandt, H. (2018). Poorly measured confounders are more useful on the left than on the right. *Journal of Business & Economic Statistics*, 1-12.

Plümper, T., & Troeger, V. E. (2011). Fixed-effects vector decomposition: Properties, reliability, and instruments. *Political Analysis*, *19*(2), 147-164.

Swamy, P. A. V. B., & Arora, S. S. (1972). The exact finite sample properties of the estimators of coefficients in the error components regression models. *Econometrica*, *40*(2), 261-275.

Woodcock, S. D. (2015). Match effects. *Research in Economics*, *69*(1), 100-121.

**Abstrakt**

Identifikace koeficientu pro časově invariantní regresor se často opírá o předpoklad, že tento regresor není korelován s nepozorovanou heterogenitou mezi panelovými jednotkami. Odvozujeme estimátor, který se tomuto předpokladu vyhýbá tím, že využije proxy pro nepozorovanou heterogenitu, čímž rozšiřujeme stávající výsledky o proxy proměnných z literatury o průřezových datech. Dále kvantifikujeme citlivost odhadů na potenciální porušení předpokladu náhodných efektů, není-li žádné proxy k dispozici. Užitnost tohoto přístupu je ilustrována na problému vysokých odhadů elasticity mezinárodních obchodních toků vzhledem k vzdálenosti pomocí gravitačních modelů.

Working Paper Series
ISSN 1211-3298
Registration No. (Ministry of Culture): E 19443

Individual researchers, as well as the on-line and printed versions of the CERGE-EI Working Papers (including their dissemination) were supported from institutional support RVO 67985998 from Economics Institute of the CAS, v. v. i.

Specific research support and/or other grants the researchers/publications benefited from are acknowledged at the beginning of the Paper.



Published by
Charles University, Center for Economic Research and Graduate Education (CERGE)
and
Economics Institute of the CAS, v. v. i. (EI)
CERGE-EI, Politických vězňů 7, 111 21 Prague 1, tel.: +420 224 005 153, Czech Republic.
Printed by CERGE-EI, Prague
Subscription: CERGE-EI homepage: http://www.cerge-ei.cz

Phone: + 420 224 005 153
Email: office@cerge-ei.cz
Web: http://www.cerge-ei.cz

Editor: Byeongju Jeong

The paper is available online at http://www.cerge-ei.cz/publications/working_papers/.

ISBN 978-80-7343-431-1 (Univerzita Karlova, Centrum pro ekonomický výzkum a doktorské studium)
ISBN 978-80-7344-474-7 (Národohospodářský ústav AV ČR, v. v. i.)