Private returns to R&D in the presence of spillovers, revisited

Giovanni Millo

Group Insurance Research at Generali S.p.A., Trieste, Italy

Correspondence
Giovanni Millo, Group Insurance Research at Generali S.p.A., via Machiavelli 3, 34123 Trieste, Italy.
Email: millogiovanni@gmail.com

Summary

This is both a replication of Eberhardt et al. (*Review of Economics and Statistics*, 2013, *95*(2), 436–448) using different software, and a critical extension and diagnostic reassessment of the original results. The main findings of the paper are confirmed and sometimes reinforced. We point out some inconsistencies, in particular in the calculation of standard errors for the common correlated effects pooled model; we extend the diagnostic checks; lastly, in the spirit of the original contribution, we show how local cross-sectional dependence diagnostics can be used to provide a first assessment of the direction of spillovers. We provide complete replication code in open source R.

1 | INTRODUCTION

The principle of common correlated effects (CCE) augmentation of Pesaran provides a simple and powerful way of controlling for common factors in panel regressions. While the original CCE paper (Pesaran, 2006) had already spanned a large literature, Eberhardt, Helmers, and Strauss (2013; henceforth EHS) was among the first applications to make it into a general-interest top journal.[1] They applied standard and CCE methods to R&D-augmented production functions in an unbalanced panel of up to 12 manufacturing subsectors in 10 OECD countries, observed over a maximum of 26 years, showing how controlling for common effects and spillovers did away with the econometric significance of own R&D productivity, and concluding that spillover effects of R&D dominated direct effects, thus highlighting the inadequacy of the standard approach and setting an agenda for future work.

EHS estimated static and dynamic versions of the production function: $y_{it} = \alpha_i l_{it} + \beta_i k_{it} + \gamma_i r_{it} + \lambda_{it} + \psi_i + \epsilon_{it}$, where y, l, k and r are the logs of production and of the three inputs: labor (L), capital (K), and R&D (RD). The estimators employed can be categorized in two ways: according to the hypotheses on the technology parameters (homogeneous or heterogeneous: that is, assuming $\alpha_i = \alpha$, $\beta_i = \beta$ and $\gamma_i = \gamma$ across all i or, respectively, leaving them free to vary); and, analogously, according to those on the impact of unobservables, as summarized in EHS, Table 4. In particular, in the pooled models with year dummies $\lambda_{it} = \lambda_t$ is assumed homogeneous across i, whereas in the CCE models time common factors f_t enter the model through individual-specific, time-invariant factor loadings, so that $\lambda_{it} = \lambda_i' f_t$.

Together with CCE, EHS showcased most of the standard estimation methods for panels of time series and a wealth of panel diagnostics. As panel time series econometrics is a relatively recent field and still under development, some aspects of their approach, while state of the art at the time the paper appeared, lend themselves to criticism when seen with today's eyes. Moreover, the lack of well-tested and user-friendly software is reflected in some inconsistencies, and in their use of computational shortcuts which are not entirely justified. We replicate their results—and our suggested modifications—in open source R (R Core Team, 2014) using user-level features, mainly from the “plm” package (Croissant & Millo, 2008).

The views expressed are solely the author's own and do not necessarily reflect those of his employer.

[1]The very first mention of CCE in a major journal pre-dates publication of the original contribution: Imbs, Mumtaz, Ravn, and Rey (2005) employed it as a robustness check, both in the homogeneous and in the heterogeneous version (see their footnote 29).

This paper and its online Appendix (provided as Supporting Information; henceforth OA) therefore double up as a review of panel time series techniques in R (see also Millo, 2015).

Our replication, while generally upholding the results of EHS, is not completely successful. There are issues in EHS's standard errors for the pooled models: those of the pooled (POLS), 2FE (fixed effects) and first difference (FD) models do not account for clustering; those of the CCEP models, due to the peculiar estimation method employed, are seriously underestimated, resulting in spurious significance; employing, respectively, clustered and nonparametric standard errors yields consistent results across the different specifications. Some of the diagnostic tests, especially the CD test, are known to be biased in the context of data such as that in EHS, so that they are compared and contrasted with results for tests which do not suffer this shortcoming; as a result some puzzles regarding pooled models with favorable CD test statistics are solved. Moreover, there are issues with applying the CCE approach to a dynamic specification that were not known at the time EHS was published. In particular, the time dimension of the EHS data set is too short for dynamic CCE to be reliable.[2]

2 | COMPUTING STANDARD ERRORS FOR CCEP ESTIMATES

According to Pesaran (2006, Eqs. 15, 27, 65) CCEP estimates ($\hat{\beta}_{\text{CCEP}}$) are computed applying ordinary least squares (OLS) to data transformed by projecting them onto the matrix of cross-sectional averages and deterministic components. Since this last estimator has few user-friendly implementations, EHS use a workaround: they obtain $\hat{\beta}_{\text{CCEP}}$ using standard regression software by adding the interaction terms between cross-sectional averages and individual dummies. The problem with this procedure is that it will not produce the correct standard errors.[3]

Pesaran (2006) gives two main methods to estimate the variance of CCEP coefficients. The *nonparametric* estimator in Equation 67, based on the empirical covariance of the individual CCE estimates $\hat{\beta}_{\text{CCE},i}$, is appropriate (Pesaran, 2006, Th. 3) for the case of heterogeneous parameters β_i. For the special case of homogeneity ($\beta_i = \beta$), and for $T/N \to 0$ as N and T diverge jointly, a *sandwich* estimator is justified (Th. 4, see Eq. 74).[4] In particular, the Newey–West procedure is suggested (Eqs. 50–52). Pesaran (2006, p 988) advocates the use of the nonparametric estimator, which is consistent under a wider set of assumptions (most notably, in "longer" panels) and performs well in simulations. The EHS data, where $N = 119$ and $T = 11 - 26$, are problematic from the viewpoint of the sandwich estimators; they rather comply with the assumptions underlying the nonparametric version.

Bootstrapping has also been recommended (Sarafidis & Wansbeek, 2012), in particular cluster-bootstrapping, whereby entire cross-sectional units are resampled in order to preserve within-individual correlation properties (see Cameron, Gelbach, & Miller, 2008; henceforth CGM).[5] It should nevertheless be noted that cluster-bootstrapping in general complies with the assumptions underlying the parametric clustering/sandwich estimator,[6] which it can be expected to outperform as regards finite-sample properties; in the EHS case, where the number of clusters is sizable, it is more likely to share the sandwich estimator's weaknesses than it is to yield significant improvements. In Table 1 we compare SEs from all the methods mentioned above; as expected, cluster bootstrapping methods and sandwich estimators produce similar results, leading to qualitatively identical conclusions.

Controlling for clustering, both homogeneous models accounting for individual heterogeneity (2FE, FD) now again have SEs of magnitude comparable to that of the CCEP, provided the latter are computed according to Pesaran (2006). The main economic result is that own R&D expenditure is consistently insignificant when controlling for individual heterogeneity. This reconciles the findings of EHS, Table 5, with those of the dynamic and/or heterogeneous models in EHS, Tables 6–8. Table 2 is our modified version of EHS, Table 5.

---

[2]While the former two points are addressed below, providing alternative results, there seems to be no solution available for this last issue: The discussion is expanded upon in the OA. We thank an anonymous reviewer for helping us single out the relevant contributions of this relatively large replication exercise, and for suggesting the last point.

[3]It is also potentially prone to computational problems because of the proliferation of regressors. On this point, see the OA.

[4]This can be computed applying the standard methods for robust coefficient covariance estimation in panels (see Millo, 2017a, for a computationally oriented review) to the projected data and residuals.

[5]Computationally, cluster-bootstrapped CCEP SEs can be easily obtained through the same transformation framework as for the sandwich estimator: With reference to the classification in CGM, once the model is expressed as pooled OLS on transformed data it is easy to resample either $(\tilde{y}_i, \tilde{\mathbf{x}}_i)$ pairs (*pairs bootstrap*) or transformed residuals $\tilde{e}_i$ (*wild bootstrap*) to obtain a bootstrap sample of parameter estimates $\hat{\beta}^*$ from which in turn var($\hat{\beta}$) is estimated; asymptotic refinement for both methods, translating into better finite-sample properties especially in the case of a small number of clusters, can be gained along the same lines by directly bootstrapping (robust) t tests (t-bootstrap) which, unlike the $\hat{\beta}^*$s, are pivotal statistics (CGM, III).

[6]An exception is the *residual-se* method, which assumes i.i.d. clusters and is therefore actually more restrictive than the sandwich estimator (CGM, p. 418).

TABLE 1 Original (EHS) versus alternative standard errors for the CCEP model from Table 5 in EHS, with corresponding significance diagnostics

	ln *L*		**ln *K***		**ln *RD***	
EHS	0.027	0***	0.036	0***	0.017	0***
Nonparametric	0.088	0***	0.161	0.07148	0.068	0.21526
HC (cluster by firm)	0.045	0***	0.077	0.00017***	0.033	0.01023**
Newey–West	0.031	0***	0.045	0***	0.02	0.00003***
Driscoll–Kraay	0.042	0***	0.076	0.00015***	0.019	0.00001***
Wild bootstrap	0.045	0***	0.077	0.00018***	0.034	0.01285**
Pairs bootstrap	0.047	0***	0.078	0.00021***	0.035	0.01473**
Wild *t*-bootstrap (HC)		0***		0.002***		0.020**
Pairs *t*-bootstrap (HC)		0***		0***		0.016**

Note. Nonparametric: as in Pesaran 2006, Th. 3; HC, NW and DK: *sandwich* estimates as in Pesaran 2006, Th. 4, with the *meat* calculated by the respective methods; bootstrap ($M = 999$): *pairs* or *wild* bootstrap estimates, clustered by firm, as defined in CGM; *t*-bootstrap: *pairs* or *wild* resampling of cluster-robust *t*-statistics. For each variable: column 1, standard error; column 2, *p*-value for the *t*-test. Asterisks indicate significance at: ***1%; **5%; *10%.

TABLE 2 Static homogeneous models, selected results from modified EHS, Table 5

	POLS	**2FE**	**FD**	**CCEP**
ln *L*	0.464	0.608	0.646	0.562
	(11.8399)***	(5.567)***	(16.543)***	(6.379)***
ln *K*	0.465	0.487	0.262	0.289
	(11.294)***	(3.057)***	(2.771)***	(1.802)*
ln *RD*	0.096	0.063	0.045	0.084
	(6.704)***	(1.351)	(1.447)	(1.239)
CRS	0.068*	0.338	0.674	0.675

Note. POLS, pooled OLS with time FEs; 2FE, two-way fixed effects; FD, first differences with time FEs; CCEP, pooled common correlated effects. *t*-statistics in parentheses, constructed from clustered (POLS, 2FE, FD), respectively nonparametric (CCEP) standard errors. Asterisks indicate significance at: ***1%; **5%; *10%. CRS, Wald test for H_0 of constant returns to scale (*p*-values).

3 | CCE, DETERMINISTICS, AND CROSS-SECTIONAL DEPENDENCE TESTING

CCEP models are a generalization of time fixed-effects ones, which last assume $\lambda_i = 1 \forall i$ (see above). Therefore, augmenting a CCEP specification with time dummies (as in EHS, Table 5, column 5) is redundant from a theoretical viewpoint, while in practice it leads to exact collinearity between the time dummies and the cross-sectional averages of response and regressors. Deterministic trends are instead a legitimate addition to CCE models (Pesaran, 2006, Remark 2).

Adding deterministics to the models can also have noteworthy side effects on cross-sectional dependence diagnostics. In particular, the CD test has no power when $E(\lambda) = 0$ or, equivalently, when the estimated model contains time fixed effects or the data have been cross-sectionally demeaned (Sarafidis & Wansbeek, 2012). Therefore the CD test cannot detect cross-sectional dependence in pooled models with either time or two-way fixed effects (EHS, Tables 5 and 6). In this sense, the addition of cross-sectional averages in the CCEP estimator has the same effect as adding time dummies: it too does away with the power of the CD test. In a situation where substantially cross-sectional dependent data have been centered on zero by data transformation or—similarly—model augmentation, the average (pairwise) cross-correlation coefficient $\bar{\hat{\rho}}$ will be near zero and the CD test statistic correspondingly small,[7] whereas the average absolute cross-correlation coefficient $|\bar{\hat{\rho}}|$ will be large; the difference between these two statistics can be telling in this respect. As for deterministic

[7]For example, compare the statistics in Holly, Pesaran, and Yamagata (2010, Tables 7 and 9) with the results in their Section 5.5 (see also Baltagi & Li, 2014, Tables 4, 5, and A5–7).

TABLE 3 Comparison of cross-section and spatial dependence diagnostics for the models in EHS, Table 5

	POLS	2FE	FD	CCEP
Global CD	−1.570	−1.464	−1.598	2.588
	0.116	*0.143*	*0.110*	*0.010*
$\bar{\hat{\rho}}$	−0.004	−0.005	−0.005	0.005
$\lvert\bar{\hat{\rho}}\rvert$	0.503	0.502	0.218	0.263
Frees rank test	21.959	22.575	−8.308	−2.013
	0.000	*0.000*	*0.000*	*0.000*
CD(1), sector	8.631	8.677	2.910	3.937
	0.000	*0.000*	*0.004*	*0.000*
CD(1), country	24.584	22.579	16.409	20.092
	0.000	*0.000*	*0.000*	*0.000*
RW, sector	*0.002*	*0.002*	*0.008*	*0.026*
RW, country	*0.002*	*0.002*	*0.002*	*0.002*

Note. Global CD is Pesaran (2015) test; $\bar{\hat{\rho}}$ is the mean of pair-wise correlation coefficients; $\lvert\bar{\hat{\rho}}\rvert$ the mean of their absolute values; Frees (1995) is the rank test for cross-sectional dependence; CD(1) *sector* and, respectively, *country* are the local CDp tests (Pesaran, 2004) when defining neighborhood as sharing either the same industrial sector or the same country. RW, again *sector* or *country*, is Millo's (2017b) randomization test for spatial correlation robust to common factors (symmetric pseudo-*p*-values from 999 random draws are reported, hence 0.002 is the minimum). *p*-values in italics.

trends, omitting them from a specification can result in serial error correlation, which in turn biases the CD test (violating Assumption 1 in Pesaran, 2015): The (otherwise harmless) addition of a trend can therefore help avoid false positives.

As a first step towards reassessing cross-sectional dependence in EHS's models, we compare $\bar{\hat{\rho}}$ with $\lvert\bar{\hat{\rho}}\rvert$. Then we perform the rank-based sibling of the CD test—the Frees test (Frees, 1995), which does not share its weakness. The $\bar{\hat{\rho}}$ is near zero for all models, which, as expected, affects the power of the CD test. By contrast, $\lvert\bar{\hat{\rho}}\rvert$, although bigger for POLS and 2FE, is of sizable magnitude for all models, hinting at substantial cross-sectional dependence, which is actually detected by the Frees test. Unlike EHS, we conclude that cross-sectional dependence is present in all models (see Table 3).

The latter can, a priori, be of either *global* (pervasive, factor-type) or *local* (decaying, spatial-type) nature. Note that the CCEP estimator is designed to control for common factor dependence, and it is consistent in the presence of spatial dependence. It is therefore no wonder, and no problem, that cross-sectional dependence be left in the CCEP residuals ($\hat{\epsilon}_{\mathrm{CCEP}}$), as long as it is due to local correlation and not to a pervasive factor structure.[8] We check by applying the exponent of cross-sectional dependence (Bailey, Kapetanios, & Pesaran, 2016) to $\hat{\epsilon}_{\mathrm{CCEP}}$. $\mathring{\alpha}$ is estimated at 0.616 (95% CI: 0.574–0.658): The evidence of pervasive factors in the residuals is weak, if any.[9] The residual spatial dependence is instead discussed in the following.

4 | SPATIAL DEPENDENCE AND THE DIRECTION OF SPILLOVERS

In light of the conclusions of EHS on the importance of identifying the direction of spillovers from R&D, we assess the degree of *local* cross-sectional dependence between "neighboring" observations with respect to two naturally defined "proximity" dimensions in the EHS data: country and industry. As a first step, we apply the local, or spatial, CDp test (Pesaran, 2004, Section 7); the null hypothesis of independence is consistently rejected for either definition of space, and across all specifications. Nevertheless, global dependence can cause the CDp test to reject as well (Holly et al., 2010;

[8]We thank an anonymous reviewer for raising this point.

[9]Strictly speaking, $\hat{\mathring{\alpha}}$ in excess of 0.5 is not fully compatible with weak dependence. Yet, $\mathring{\alpha}$ is not estimable if lower than 0.5, and is usually overestimated in this region (Bailey et al., 2016): Hence we conclude that a factor structure is really not relevant. The values reported are relative to the bias-corrected version of $\mathring{\alpha}$ (Bailey et al., 2016, Eq. 13), allowing for persistent factors but no spatial correlation. All other versions are presented in the OA; results are largely similar.

Millo, 2017b; Moscone & Tosetti, 2010); therefore the latter can be safely used only if controlling for common factors. Moreover, neither the CD nor the CDp test tolerates serial correlation, which seems present in all static models from EHS. Factor-robust testing procedures are compared and discussed at length in Millo (2017b), where a new test (RW) based on randomizing the proximity matrix is proposed, which is both factor robust and—interestingly, in view of the above concerns—also insensitive to serial correlation. Unlike the CDp, the RW test can be safely applied to any of the specifications considered by EHS. Cross-section and spatial dependence diagnostics are reported in Table 3.

From the evidence presented, we conclude that all models have *locally* correlated residuals within single countries, and, although the evidence is weaker, also within industrial sectors.

REFERENCES

Bailey, N., Kapetanios, G., & Pesaran, M. H. (2016). Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics*, *31*(6), 929–960.

Baltagi, B. H., & Li, J. (2014). Further evidence on the spatio-temporal model of house prices in the United States. *Journal of Applied Econometrics*, *29*(3), 515–522.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, *90*(3), 414–427.

Croissant, Y., & Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, *27*(2), 1–43.

Eberhardt, M., Helmers, C., & Strauss, H. (2013). Do spillovers matter when estimating private returns to R&D? *Review of Economics and Statistics*, *95*(2), 436–448.

Frees, E. W. (1995). Assessing cross-sectional correlation in panel data. *Journal of Econometrics*, *69*(2), 393—414.

Holly, S., Pesaran, M. H., & Yamagata, T. (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics*, *158*(1), 160–173.

Imbs, J., Mumtaz, H., Ravn, M., & Rey, H. (2005). PPP strikes back: aggregation and the real exchange rate. *Quarterly Journal of Economics*, *120*(1), 1–43.

Millo, G. (2015). Narrow replication of "A spatio-temporal model of house prices in the USA" using R. *Journal of Applied Econometrics*, *30*(4), 703–704.

Millo, G. (2017a). Robust standard error estimators for panel models: A unifying approach. *Journal of Statistical Software*, *82*(1), 1–27.

Millo, G. (2017b). A simple randomization test for spatial correlation in the presence of common factors and serial correlation. *Regional Science and Urban Economics*, *66*, 28–38.

Moscone, F., & Tosetti, E. (2010). Testing for error cross section independence with an application to US health expenditure. *Regional Science and Urban Economics*, *40*(5), 283–291.

Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. (*Cambridge Working Papers in Economics*). Cambridge, UK: University of Cambridge.

Pesaran, M. H (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, *74*(4), 967–1012.

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews*, *34*(6–10), 1089–1117.

R Core Team (2014). R: A language and environment for statistical computing. Vienna: Austria: R Foundation for Statistical Computing.

Sarafidis, V., & Wansbeek, T. (2012). Cross-sectional dependence in panel data analysis. *Econometric Reviews*, *31*(5), 483–531.