

Multinational Finance Journal (2026) 30, 4: 42–50
doi: 10.67454/30-4-4
Copyright University of Vaasa
Published online: 03.10.2026
ISSN 3087-9132

Publishing in the Multinational Finance Journal, Part 2: Research Methods

Timothy King^{1,‡}

¹School of Accounting and Finance, University of Vaasa, Vaasa, Finland

[‡] Corresponding author e-mail: timothy.king@uwasa.fi

Abstract.

This second editorial in a series of four, written to mark the thirtieth volume of the Multinational Finance Journal, concerns research design. I discuss the problems of data and measurement that arise when tools developed in one country are applied in another, the identification of causal effects and its limits, the discretion researchers exercise over routine choices and what it implies for replication and transparency, the newer sources of evidence in text, machine learning, surveys and spatial data, and the reporting of results. Most points are illustrated with papers from the journal's archive, and all are set against the practices that the leading journals in finance and accounting now expect. I do not favour any one method. I ask that the design fit the question, that its assumptions be stated and defended, and that a reader could reproduce what was done.

Keywords: Research design, Identification, Measurement, Replication, Multinational finance

JEL Classification: C18, C50, G00, G15

1. Introduction

The first editorial in this series dealt with the choice of topic, and this one deals with execution. The Multinational Finance Journal has never had a house method. The archive I described in the companion editorial moves from the time-series econometrics of the late 1990s through panel regression to text, surveys and experiments (King, 2026), and I have no wish to narrow it. I look for a design that fits the question, for assumptions that are stated and defended, and for a description complete enough that a reader could reproduce the analysis. Leventis (2024b) observes of international accounting research that the designs used in earlier studies are not always applicable in a new setting. That observation is my starting point, because a multinational journal meets the problem in almost every paper it receives.

2. Data and measurement across borders

Most of the tools of empirical finance were developed on American data, and they do not always work elsewhere. Ooghe and Balcaen (2007) re-estimated seven published failure prediction models on Belgian accounts and found that some of the best known performed very poorly outside the samples on which they were built. The daily correlations between national stock markets that many studies of the time reported were biased by the fact that the markets are not open at the same hours, downward for contemporaneous correlations and upward for lagged ones, and Kahya (1997) provided a correction. McDonald, Michelfelder and Theodossiou (2009) showed that robust estimators built on symmetric distributions give biased intercepts when returns are skewed, so that an estimated alpha can be an artefact of the estimator. Papers about measurement are as welcome at this journal as papers that test hypotheses.

The same care is needed with the variables that describe countries. Isidro, Nanda and Wysocki (2020) show that the country attributes used in international accounting research are so highly correlated that four underlying factors summarise 72 of them, which means that a coefficient on investor protection, legal origin or trust may be capturing the effect of any of the others. Ratings of environmental, social and governance performance have a related weakness. Berg, Kölbel and Rigobon (2022) find that the ratings of different providers diverge mainly because the providers measure the same attribute differently, so that a result obtained with one rating should be shown to hold with another. Commercial databases also cover countries unevenly, and a sample that is broadly representative in the United States may consist of the largest and most international firms elsewhere. I ask authors who work with cross-country data to say how coverage varies, to validate the key measure in the setting where they use it, and to show that the results do not depend on a single provider.

A survey of 45 listed firms can be the right sample when the data can be obtained in no other way, as in the study of internal audit quality in Greece by Vadasi, Bekiaris and Andrikopoulos (2021), and a single cross-section is appropriate to an experiment, in which random assignment identifies the effect. Leventis (2024b) cautions that a single cross-section cannot identify change or establish cause, and I apply the same caution to archival cross-sections that contain no source of exogenous variation. Where the research question concerns the effect of one variable on another over time, I expect panel data of adequate length, and I expect authors working in small markets to be candid about what their sample can and cannot support.

3. Identification and its limits

Most empirical papers in finance and accounting make a causal claim, whether or not they use the word, and most therefore face the problems of omitted variables, simultaneity and measurement error that Roberts and Whited (2013) set out, often with a problem of selection besides. Each of the familiar remedies holds only under conditions that authors have to demonstrate. Fixed effects remove only the unobserved differences that are constant over time within the unit, and the common shortcuts of subtracting the industry mean from the dependent variable or adding the group mean of the dependent variable as a control give inconsistent estimates (Gormley and Matsa, 2014). Standard errors have to allow for the correlation of residuals within a firm over time and across firms at a point in time (Petersen, 2009), and in a cross-country study the number of countries, which is often small, limits what clustering can achieve. The exclusion restriction of an instrument cannot be tested directly and has to be defended by economic argument, and the conventional threshold for instrument strength is too lenient. Lee *et al.* (2022) show that, with a single instrument, the usual t-ratio has its stated size only when the first-stage F-statistic exceeds about 105, far above the value of ten that most authors cite, and Montiel Olea and Pflueger (2013) provide a test of instrument strength that remains valid when errors are heteroskedastic or clustered. Where selection on unobservables cannot be ruled out, the bounds proposed by Oster (2019) tell the reader how strong it would have to be, relative to selection on the observed controls, to overturn the result.

The multinational setting offers many events that help with identification. Countries change their rules and their currencies at different times and for reasons that have little to do with any one firm, and a crisis is seldom of one firm's making. Such events are not random, since governments legislate in response to economic conditions, and authors still have to show that nothing else changed at the same time. The introduction of the euro allowed Cheung and Westermann (2001) to study the effect of the exchange rate regime on equity markets, and the 1987 crash allowed Eun and Jang (2000) to infer, from the returns of firms with different bond ratings, that investors treat the bankruptcy costs borne by shareholders as significant. Papafilis, Psillaki and Margaritis (2019) applied linear and nonlinear tests of Granger causality and a difference-in-differences design to the restructuring of Greek sovereign debt and showed that the link between sovereign and bank credit risk weakened afterwards. Atanasov and Black (2016) review the literature on governance and firm value and conclude that the studies with credible causal designs rely on external shocks, most of them legal. Careful dating helps even where the event is chosen by the firm. Hertz, Lowengrub and Melvin (2000) separated the effects of an American listing on German firms by distinguishing three dates, the filing of accounts with the American regulator, the announcement of the depositary receipt programme and the first day of trading, and they recognised that the firms that choose to list may be those with nothing to hide.

In banking the standard was set by Khwaja and Mian (2008), who used loan-level data from Pakistan to compare how banks hit differently by a liquidity shock changed their lending to the same firm, so that the firm's demand for credit is held constant. Studies of credit supply have since been expected to separate it from credit demand in this way where the data allow. Credit registers that make the comparison possible now exist in many countries, although access to them is restricted and the comparison is confined to firms that borrow from more than one bank.

Designs that exploit the staggered adoption of a rule across countries or states need particular care. The standard two-way fixed effects estimator uses units that are already treated as controls for those treated later. When the effect of treatment changes over time, these comparisons distort the estimate, which can even take the opposite sign to the true effect (Goodman-Bacon, 2021). Baker, Larcker and Wang (2022) show that correcting for the problem changes the estimates and inferences of the published studies they re-examine, and estimators that compare newly treated units only with units not yet or never treated, such as that of Callaway and Sant'Anna (2021), are now the expected remedy. Conventional standard errors are also too small when the outcome is serially correlated, and the remedy is to cluster at the level at which treatment is assigned (Bertrand, Duflo and Mullainathan, 2004), which in a cross-country design is the country and brings back the problem of few clusters.

A paper need not contain a natural experiment. Armstrong *et al.* (2022) caution against confining research to the questions for which a quasi-experiment happens to exist, and Glaeser and Guay (2017) point out that the settings with the cleanest identification are often those whose results generalise least. Leuz (2022) argues for judging the strength of a design as a whole, which seems to me the right test. The journal has published valuable descriptive work, such as the account by Brown, Demetriou and Theodossiou (2018) of how the Cypriot banking system failed, and it will continue to do so. I cannot accept causal language where the design supports only an association. Authors who have a correlation should call it one, say which alternative explanations they have been able to exclude, and leave the reader to judge the rest. Longarela and Bjønnes (2024) set an example when they report that price discovery in currency markets is correlated with the bid-ask spread and then find, in two-stage regressions, no evidence that liquidity has a causal effect on it.

4. Discretion, replication and transparency

Every empirical paper is the result of many small choices, among them which observations to exclude, how to treat outliers, which controls to include and how to define the dependent variable. Mitton (2022) documents how widely these choices vary across published papers in corporate finance and shows that discretion over ten routine choices is enough to make most randomly generated variables appear significant. Menkveld *et al.* (2024) gave the same data and the same hypotheses to 164 research teams and found that the variation in their answers was sizeable, and larger than the teams themselves expected. In asset pricing, Harvey, Liu and Zhu (2016) argue that the profession has tested so many factors that a new one should be required to clear a t-statistic of about three.

For authors, the implication is that a result should be shown to survive the reasonable alternatives to each choice that produced it, and that the alternatives that fail should be reported alongside those that succeed. Variables should be defined in an appendix in enough detail for someone else to construct them, and sample selection should be set out step by step with the number of observations lost at each. I encourage authors to make their code available, and their data where licences permit, and I would welcome submissions that state their hypotheses and tests in advance of the analysis. Bloomfield, Rennekamp and Steenhoven (2018) find that editorial processes built on registration improve the planning of a study at some cost to the work done after the results are known, which suggests that registration suits confirmatory tests

better than exploratory ones. Replications belong to the same effort, and one that extends a well-known result to other markets, or fails to find it there, fits the purpose of this journal well.

5. Newer sources of evidence

Text was among the first of the newer sources of evidence to reach the journal, in the studies of news and sovereign spreads by Liu (2014) and of press coverage and stock returns in the United Kingdom by Ferguson *et al.* (2015), and more recently in the comparison of sentiment on social media and in the news by Garcia (2024). The central lesson of this literature is that measurement has to be fitted to the context. Loughran and McDonald (2011) showed that a general-purpose dictionary misclassifies most of the words it counts as negative in company filings. The same applies across languages and countries, since a word list or a language model built on American filings in English cannot be assumed to work on Finnish annual reports or Korean news.

Authors who use machine learning for prediction (Gu, Kelly and Xiu, 2020) or to extract data from documents and images (Dell, 2025) should report performance out of sample, explain what the model has learned in terms an economist can assess, and guard against look-ahead bias, in which information that was not available at the date of the prediction enters the training data. Large language models can make such tasks as coding, classifying text and extracting data from it considerably less costly (Korinek, 2023), and they introduce problems of their own. A model trained on text written after the sample period already contains information about what happened next. Its output also varies from one run to the next and from one version to the next, and commercial models are withdrawn, so the model and its version, the settings, the prompts and the date of use need to be recorded and the output archived if the work is to be reproducible.

Surveys obtain what archival data cannot, the beliefs and intentions behind decisions, and experiments add random assignment. Baker, Dutta and Saadi (2011) surveyed Canadian firms on their financial practices, Rieger, Wang and Hens (2013) drew on an international survey of time preferences, and Zhylyevskyy and Stover (2025) asked finance professionals in six countries how they would respond as depositors to a banking crisis. Krueger, Sautner and Starks (2020) used a survey to establish what institutional investors believe about climate risk, and Floyd and List (2016) explain how field experiments can be designed in our disciplines. The quality of a survey depends on the match between the respondents and the decision being studied, on the response rate, on whether those who responded differ from those who did not and on the wording of the instrument, all of which belong in the paper.

Spatial data have not yet appeared in the journal, and I expect that they will. Research on nature-related risk increasingly needs measures at the level of the individual asset and its location, which will require researchers in finance to work with satellite imagery and geocoded records of plants, mines and farms. Meta-analysis (Zabolotnyuk, 2018) and bibliometric mapping (Calosci *et al.*, 2026) have their own standards of search, inclusion and coding, which should be reported with the same care.

6. Reporting the results

On reporting I agree with Leventis (2024b) that a reader needs a correlation table and descriptive statistics, the latter in order to judge whether outliers that trimming or winsorising

has not removed could be driving a coefficient. I would add that the treatment of outliers should be described, that the tables should be readable without the text, and that the text should interpret the estimates instead of repeating them. I ask for economic magnitudes throughout. Mitton (2024) recommends expressing an effect relative to the standard deviation of the dependent variable and supplies benchmarks from several hundred published findings, which makes it possible to say whether an effect is large or small by the standards of the field. An effect that is statistically significant and economically trivial should be described as such.

7. Concluding remarks

The standards described here are those of the leading journals in finance and accounting, and they have risen quickly. They need not discourage authors whose data are limited or whose setting offers no experiment, and I would sooner publish a modest claim supported by a transparent design than an ambitious one that its design cannot support. The next editorial turns from the analysis to the writing, and in particular to the introduction, where most readers and most reviewers form their view of a paper.

Acknowledgements

I thank Zacharias Sautner for contributing his paper on nature risks as part of the journal's 30th anniversary celebration, my fellow co-editors Tatiana King, Vanja Piljak, Stephen Ferris and Anastasiya Shamshur, the associate editors and reviewers of the journal, and the Multinational Finance Society and the University of Vaasa for their continuing support of the journal. I used generative AI tools (Claude) to help me improve the writing and structuring of my arguments.

References

- Armstrong, C., Kepler, J. D., Samuels, D., and Taylor, D. (2022) Causality redux: The evolution of empirical methods in accounting research and the growth of quasi-experiments, *Journal of Accounting and Economics* 74(2–3), 101521.
- Atanasov, V., and Black, B. (2016) Shock-based causal inference in corporate finance and accounting research, *Critical Finance Review* 5(2), 207–304.
- Baker, A. C., Larcker, D. F., and Wang, C. C. Y. (2022) How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics* 144(2), 370–395.
- Baker, H. K., Dutta, S., and Saadi, S. (2011) Corporate finance practices in Canada: Where do we stand? *Multinational Finance Journal* 15(3/4), 157–192.
- Berg, F., Kölbel, J. F., and Rigobon, R. (2022) Aggregate confusion: The divergence of ESG ratings, *Review of Finance* 26(6), 1315–1344.
- Bertrand, M., Duflo, E., and Mullainathan, S. (2004) How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119(1), 249–275.
- Bloomfield, R., Rennekamp, K., and Steenhoven, B. (2018) No system is perfect: Understanding how registration-based editorial processes affect reproducibility and investment in research quality, *Journal of Accounting Research* 56(2), 313–362.
- Brown, S., Demetriou, D., and Theodossiou, P. (2018) Banking crisis in Cyprus: Causes, consequences and recent developments, *Multinational Finance Journal* 22(1/2), 63–118.
- Callaway, B., and Sant’Anna, P. H. C. (2021) Difference-in-differences with multiple time periods, *Journal of Econometrics* 225(2), 200–230.
- Calosci, M., Crisafulli, A., Giantomassi, M., and Giorgio, S. (2026) Central bank digital currencies: Bibliometric and thematic insights from the literature, *Multinational Finance Journal* 30(3), 1–40.
- Cheung, Y.-W., and Westermann, F. (2001) Equity price dynamics before and after the introduction of the euro: A note, *Multinational Finance Journal* 5(2), 113–128.
- Dell, M. (2025) Deep learning for economists, *Journal of Economic Literature* 63(1), 5–58.
- Eun, C. S., and Jang, H. J. (2000) Investor recognition of bankruptcy costs: Evidence from the 1987 market crash, *Multinational Finance Journal* 4(3/4), 221–245.
- Ferguson, N. J., Philip, D., Lam, H. Y. T., and Guo, J. M. (2015) Media content and stock returns: The predictive power of press, *Multinational Finance Journal* 19(1), 1–31.
- Floyd, E., and List, J. A. (2016) Using field experiments in accounting and finance, *Journal of Accounting Research* 54(2), 437–475.
- Garcia, J. (2024) Sentiment divergence and its impact on share liquidity, *Multinational Finance Journal* 28(1/2), 3–39.
- Glaeser, S., and Guay, W. R. (2017) Identification and generalizability in accounting research: A discussion of Christensen, Floyd, Liu, and Maffett (2017), *Journal of Accounting and Economics* 64(2–3), 305–312.
- Goodman-Bacon, A. (2021) Difference-in-differences with variation in treatment timing, *Journal of Econometrics* 225(2), 254–277.
- Gormley, T. A., and Matsa, D. A. (2014) Common errors: How to (and not to) control for unobserved heterogeneity, *Review of Financial Studies* 27(2), 617–661.

- Gu, S., Kelly, B., and Xiu, D. (2020) Empirical asset pricing via machine learning, *Review of Financial Studies* 33(5), 2223–2273.
- Harvey, C. R., Liu, Y., and Zhu, H. (2016) ... and the cross-section of expected returns, *Review of Financial Studies* 29(1), 5–68.
- Hertzel, M., Lowengrub, P., and Melvin, M. (2000) Information, announcement, and listing effects of ADR programs and German-U.S. stock market integration, *Multinational Finance Journal* 4(3/4), 181–200.
- Isidro, H., Nanda, D., and Wysocki, P. D. (2020) On the relation between financial reporting quality and country attributes: Research challenges and opportunities, *The Accounting Review* 95(3), 279–314.
- Kahya, E. (1997) Correlation of returns in non-contemporaneous markets, *Multinational Finance Journal* 1(2), 123–135.
- Khwaja, A. I., and Mian, A. (2008) Tracing the impact of bank liquidity shocks: Evidence from an emerging market, *American Economic Review* 98(4), 1413–1442.
- King, T. (2026) Thirty years of the Multinational Finance Journal: Research past, present and future, *Multinational Finance Journal* 30(4), 9–32.
- Korinek, A. (2023) Generative AI for economic research: Use cases and implications for economists, *Journal of Economic Literature* 61(4), 1281–1317.
- Krueger, P., Sautner, Z., and Starks, L. T. (2020) The importance of climate risks for institutional investors, *Review of Financial Studies* 33(3), 1067–1111.
- Lee, D. S., McCrary, J., Moreira, M. J., and Porter, J. (2022) Valid t-ratio inference for IV, *American Economic Review* 112(10), 3260–3290.
- Leuz, C. (2022) Towards a design-based approach to accounting research, *Journal of Accounting and Economics* 74(2–3), 101550.
- Leventis, S. (2024b) Editorial: Publishing in JIAAT Part 2 – Research methods, *Journal of International Accounting, Auditing and Taxation* 56, 100636.
- Liu, S. (2014) The impact of textual sentiment on sovereign bond yield spreads: Evidence from the Eurozone crisis, *Multinational Finance Journal* 18(3/4), 215–248.
- Longarela, I. R., and Bjønnes, G. H. (2024) Price discovery for competing currency numeraires, *Multinational Finance Journal* 28(3/4), 40–76.
- Loughran, T., and McDonald, B. (2011) When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *Journal of Finance* 66(1), 35–65.
- McDonald, J. B., Michelfelder, R. A., and Theodossiou, P. (2009) Robust regression estimation methods and intercept bias: A capital asset pricing model application, *Multinational Finance Journal* 13(3/4), 293–321.
- Menkveld, A. J., Dreber, A., Holzmeister, F., Huber, J., Johannesson, M., et al. (2024) Nonstandard errors, *Journal of Finance* 79(3), 2339–2390.
- Mitton, T. (2022) Methodological variation in empirical corporate finance, *Review of Financial Studies* 35(2), 527–575.
- Mitton, T. (2024) Economic significance in corporate finance, *Review of Corporate Finance Studies* 13(1), 38–79.
- Montiel Olea, J. L., and Pflueger, C. (2013) A robust test for weak instruments, *Journal of Business & Economic Statistics* 31(3), 358–369.

- Ooghe, H., and Balcaen, S. (2007) Are failure prediction models widely usable? An empirical study using a Belgian dataset, *Multinational Finance Journal* 11(1/2), 33–76.
- Oster, E. (2019) Unobservable selection and coefficient stability: Theory and evidence, *Journal of Business & Economic Statistics* 37(2), 187–204.
- Papafilis, M.-P., Psillaki, M., and Margaritis, D. (2019) The effect of the PSI in the relationship between sovereign and bank credit risk: Evidence from the euro area, *Multinational Finance Journal* 23(3/4), 211–272.
- Petersen, M. A. (2009) Estimating standard errors in finance panel data sets: Comparing approaches, *Review of Financial Studies* 22(1), 435–480.
- Rieger, M. O., Wang, M., and Hens, T. (2013) International evidence on the equity premium puzzle and time discounting, *Multinational Finance Journal* 17(3/4), 149–163.
- Roberts, M. R., and Whited, T. M. (2013) Endogeneity in empirical corporate finance, in: Constantinides, G. M., Harris, M., and Stulz, R. M. (eds.), *Handbook of the Economics of Finance*, Vol. 2A, Elsevier, 493–572.
- Vadasi, C., Bekiaris, M., and Andrikopoulos, A. (2021) Internal audit function quality and corporate governance: The case of Greece, *Multinational Finance Journal* 25(1/2), 1–61.
- Zabolotnyuk, Y. (2018) Wealth effects of bond rating announcements, *Multinational Finance Journal* 22(3/4), 211–254.
- Zhylyevskyy, O., and Stover, R. (2025) The interacting roles of bank supervisory effectiveness and depositor discipline in a banking crisis, *Multinational Finance Journal* 29(1/2), 1–75.