Title
Making Theoretically Informed Choices in Specifying Panel-Data Models

Journal
Production and Operations Management, 30(7)

ISSN
1059-1478

Authors
Ketokivi, Mikko
Bromiley, Philip
Awaysheh, Amrou

Publication Date
2021-07-01

DOI
10.1111/poms.13347

Peer reviewed

Making Theoretically Informed Choices in Specifying Panel-Data Models

Mikko Ketokivi (corresponding author)
IE Business School, IE University
Maria de Molina 12-5, 28006 Madrid, Spain
mikko.ketokivi@ie.edu
+34-684-124-516

Philip Bromiley
UCI Paul Merage School of Business
University of California - Irvine, CA 92697-3125
bromiley@uci.edu
+1-949-824-6657

Amrou Awaysheh
Indiana University Kelley School of Business
Indianapolis, IN 46202-5151
awaysheh@iu.edu
+1-317-274-0955

Making Theoretically Informed Choices in Specifying Panel-Data Models

ABSTRACT

We argue that in analyzing panel-data econometric models, researchers rely excessively on statistical criteria to determine model specification, treating it primarily as a matter of statistical inference. This inferential emphasis is most obvious in the common practice of using statistical tests (e.g., the Hausman test) to choose between fixed- and random-effects specifications, often ignoring the assumptions underpinning these tests. For instance, the Hausman test depends on the true within-panel (longitudinal) and between-panel (cross-sectional) parameters being equal. This assumption is often not justified, because longitudinal and cross-sectional variances and covariances may manifest different underpinning mechanisms. In addition to different *mechanisms* often resulting in different variables determining within and between effects, within and between *variables* may also have different meanings. To help researchers make theoretically informed choices, we formulate five questions that can help researchers think of model specification in a theoretically rigorous way. We examine these issues with examples from both the general management and operations management research. Importantly, we argue that addressing the questions regarding model specification must involve primarily theoretical and contextual judgement, not statistical tests.

Keywords: Panel data, level of analysis, model specification, within/between effects, fixed/random effects

History: Received June 2020; accepted December 2020 by Kalyan Singhal after two revisions.

1. INTRODUCTION

Consider the cost of housing. Cross-sectional differences across households depend on factors such as mortgage payments and property taxes. In stark contrast, month-to-month variation in a household's housing costs depends on time-varying factors such as heating and cooling expenses, seasonal variation in maintenance such as landscaping and snow removal, various non-routine costs such as maintenance and repair costs, et cetera. The factors that drive the differences in housing costs across households have little to do with the factors that drive cost variation over time for individual households. In panel-data econometric terms, not only does the *within model* of housing costs manifest a different causal logic than the *between model*, but also the two involve different explanatory variables.

Similar phenomena can be found in many domains of operations management research, and management research more generally (e.g., Karatzas, Papadopolous, & Godsell, 2020; Shockley, Plummer, Roth, & Fredendall, 2014). For example, cross-sectional differences in operational efficiency link to how operational systems are *designed* (e.g., their scale and technology), whereas variation in efficiency over time tends to be associated with how the systems are *managed*. Therefore, both the antecedents and the consequences of an increase in, say, plant productivity from one year to the next (the within effect) may differ greatly from the antecedents and the consequences of two operational systems differing in productivity by that same amount (the between effect). The phenomenon is obviously not limited to operational systems: The effect of a one-year bonus on employee behavior (the within effect) may differ from the effect of a stable difference in compensation between employees (the between effect).

In this paper, we argue that while those seeking to understand housing costs will not confound the between and within models, management researchers often fail to see the fundamental differences between the two; this may cause a disconnect between theory and empirical analysis. In part, this reflects treating model choice as a matter of consistency and precision of estimation (i.e., inference) instead of treating it as a theoretical matter (Bell, Fairbrother, & Jones, 2019; Certo, Withers, & Semadeni, 2017). The problem appears often in the analysis of (multilevel) panel data on firms, where the between effect links to inter-firm differences and the within effect to variation within firms over time.

Confounding of within and between models can lead to both the *ecological* fallacy and the *atomistic fallacy.* Robinson (1950) noted that analyzing data at the aggregate level but drawing conclusions at the individual level constitutes an ecological fallacy. Ecological fallacies have been found in management research as well (e.g., Brewer & Venaik, 2014; Certo et al., 2017). In the context of panel-data analysis, the ecological fallacy occurs when one draws *within conclusions* from *between data*: It is an ecological fallacy to conclude that a positive cross-sectional correlation between a specific operational practice and performance implies that implementing said practice will improve performance. The opposite, drawing between conclusions from within data, is the atomistic fallacy (Bliese, Schepker, Essmann, & Ployhart, 2020). For example, one might incorrectly assume that something that helps a firm improve its performance will also explain why some firms perform better than others.

Management researchers often ignore the substantive (theoretical) differences of the within and between models, and instead, treat the question as a matter of statistical inference. Our central argument is that the choice between these models must depend first on theoretical, not inferential considerations. This involves asking questions such as the following: What is the research question of interest? What causal mechanisms link the variables of interest to one another? Do the mechanisms address inter-firm differences (a between effect) or are they about what happens in individual firms over time (a within effect), or both? Only after addressing these questions should we think of model specification, estimator choice, and inferential tests.

This paper is structured as follows. To complement the housing cost example, we first discuss examples that emphasize the importance of making the within-between distinction in management research. Because methodological texts extensively cover the inferential (statistical) issues, we focus on an examination of theoretical criteria. As for the inferential issues, we clarify how the within-between distinction relates to the more commonly made distinction of random and fixed effects (see the Technical Appendix at the end of the paper for details), however, our main focus is on theoretical criteria in model specification. To this end, we formulate five questions to help researchers gauge the theoretical alignment of their panel-data models.

2. WITHIN AND BETWEEN EXPLANATIONS: THEORY AND INFERENCE

Researchers need to distinguish the between ("What is the expected difference between Mary and Joe's s values of y if they differ in x_1 by 1?") and within ("What is the expected change in Mary's value of y if her x_1 increases by 1?") effects from one another. In practice, researchers often implicitly assume that the two true effects are identical. This problem occurs, for example, when researchers draw within conclusions from cross-sectional differences: A common way of misinterpreting the positive association between practice X and performance in a cross-section is to conclude that adopting X improves performance; similarly, a zero cross-sectional correlation between practice X and performance is interpreted as evidence that adopting X does not improve performance. Many empirical articles based on cross-sectional research fall prey to this misinterpretation in the discussion of managerial implications in particular.

A less conspicuous but perhaps even more common manifestation of the equality assumption comes in choosing between the fixed-effects (FE) and the random-effects (RE) specifications in panel-data analysis. In the Technical Appendix to this paper, we show how the commonly used Hausman's (1978) specification test effectively rests on the assumption that the true within and between effects are equal.

Should we generally assume within and between effects to be equal? To establish that they may *not* be equal, consider factory scale and productivity (Hayes, Pisano, Upton, & Wheelwright, 2005, Chapter 3). Design decisions about product mix and production volumes occur before construction when a firm chooses production technology, plant layout, and other relevant factors. These choices depend on production cost estimates, demand projections, projected transportation costs, and the like. These design differences lead to a between effect: Larger factories may perform better due to economies of scale and scope. After a plant starts production, product mix and volume often vary over time. However, rather than a general plant redesign and *capacity change*, over-time variation usually reflects changes in *capacity utilization*, driven by seasonal variation in demand, product lifecycles, production shutdowns, and the like. In sum, the factors that influence plant average scale and output (largely determined in the design stage) differ from those that influence variation over time; the within and between models have different underpinning theoretical mechanisms.

In addition to potential substantive differences in the theoretical mechanisms at different levels, the within and the between effects observed in empirical research often differ not only in *magnitude* (e.g., Certo et al., 2017) but also in their *interpretation* (Snijders & Berkhof, 2008). Interpretational differences arise from two sources. First, the theoretical mechanisms that link the independent variables to the dependent variables and produce the observed within and between effects can differ. Second, even when within and between analyses involve the same variables, the interpretation of these variables may be differ across levels. Specifically, the average level of a variable—the *between variable* used in the between model—and the deviation of the same variable from its mean—the *within variable* used in the within model—may have different interpretations (Muthén, 1994; Snijders & Berkhof, 2008). Comparing the magnitudes of the within and between effects makes little sense if the interpretations of the two effects differ.

To illustrate how within and between effects of the same variables can have different interpretations, consider the production volume example above. A stable difference in production volumes of two production lines (between-line variation) is reasonably interpreted as a difference in *scale*. In stark contrast, within-line variation in volume over time may simply indicate changes in *capacity utilization*. Similarly, if a person's salary increases by $10,000, we describe this (within phenomenon) either as a *raise* if it is permanent, or a *bonus* if it is a one-time payment. However, these are not the words we use to interpret the between phenomenon of one person making $10,000 more than another.

2.1. Theoretical and Inferential Criteria in Model Specification

In specifying and analyzing statistical models, researchers rely on both theoretical and inferential criteria. The theoretical criterion pertains to whether the model aligns with the substantive theory used to derive one's hypotheses and knowledge about the phenomenon and context. We label this *theoretical alignment*.

Theoretical alignment has many facets, but we focus here specifically on *levels*: Do the model parameters that represent the relations between the independent and dependent variables reside in the model at a level consistent with theory? For example, a model of competitive advantage—often operationalized as abnormal returns—must involve between-level variables and parameters (Bou & Satorra, 2007: 707). Bliese et al. (2020: 24) noted that "researchers who study between-firm theories may implicitly commit the atomistic fallacy by believing results based on fixed effects [the within model] support theory involving between-firm relationships."

In addition to theoretical concerns, use of models in empirical research also involves numerous inferential criteria. For statistical models, this includes the use of statistical tests that address model assumptions. We label this *inferential alignment*. The important distinction between theoretical and inferential alignment can be expressed in terms of model parameters: Theoretical alignment ensures that the model contains the parameters that operationalize the substantive hypotheses, and inferential alignment ensures the consistent estimation of these parameters. The ecological and atomistic fallacies reflect theoretical misalignment; unmodeled serial correlation or heteroskedasticity of errors are examples of inferential misalignment. Whether errors are serially correlated and heteroskedastic is obviously an inferential question, usually unrelated to substantive theory. Conversely, no inferential procedure can tell us whether the model aligns with substantive theory.

3. ENSURING THEORETICAL ALIGNMENT

Problems with theoretical alignment occur when theoretical logic does not clearly establish whether a hypothesis is about the between or the within effect. Certo et al. (2017: 1546) noted that hypotheses in management research often exhibit "lack of theoretical specificity regarding within- versus between-firm relationships." For instance, Chen and Miller's (2007) hypothesis "R&D search intensity increases with firms' slack resources" can be read both as a within and a between hypothesis (Certo et al., 2017: 1546): "As a firm increases its slack resources, its R&D search intensity increases" is the within reading; "Firms with more slack resources have higher R&D search intensity" is the between reading. Both readings are reasonable: R&D search intensity and slack can both vary across firms and within firms over time, making both within and between hypotheses empirically relevant. Many theoretical concepts including diversification, structures, performance exhibit similar duality.

The crucial questions are whether the same theoretical logic applies to the within and between models, and whether the logic predicts equal parameter sizes in the within and between cases. Put managerially, if we find diversified firms have higher profits than undiversified firms, can Firm A expect increasing diversification to improve its performance? Alternatively, is the mechanism by which diversification influences performance within firms over time the same as the mechanism that produces cross-sectional differences? The answer is far from obvious and requires a careful comparison of both within and between explanations of how and why performance and diversification relate to one another.

It is easy to think of research situations where within and between mechanisms differ. In his classic treatment of process and variance theories, Mohr (1982: 13) noted that "an explanation of the process of acquiring power within an organization—how this process works—is not the same as an explanation for the distribution of power—why one person or unit has more power than another." Distribution of power (the between phenomenon) may be explained by demographic variables such as age, gender, or ethnicity, but these variables may have little to do with how an individual increases his or her power (the within phenomenon); indeed, some of the variables in the between model are fixed over time, that is, they have no within variance.

Given that within and between mechanisms often differ, it is disconcerting that management researchers are not explicit about whether the theoretical mechanisms they propose are of within or between variety (Certo et al., 2017). To help researchers think about the key issues more explicitly, we present below five questions. We invite all researchers who analyze panel data to consider these questions in their own research. The questions are designed specifically to establish theoretical alignment. Although the answers will vary from one research context to another, thus precluding a general substantive prescription, we believe the questions themselves are general enough to relate to almost all types of management research.

3.1. Question 1: What Is the Level of the *Phenomenon* of Interest?

Prior to addressing any theoretical or inferential questions, researchers must first clarify what phenomenon or managerial issue they seek to understand: Is the interest in how something varies within entities (e.g., firms) over time, how entities differ from one another, or both? Many organizational phenomena exhibit both within and between characteristics. For example, the general manager of an individual manufacturing plant focuses on improving plant productivity (a within phenomenon), whereas a corporate-level Vice President of Operations may care more about why some plants perform better than others (a between phenomenon). In the context of operations management research, within and between effects are relevant in the management of *multi-unit* or *multi-site* operations in particular (e.g., Dreyfus, Nair, & Talluri, 2020; Metters, Frey, & Vargas, 1999). Obviously, even though between and within levels are distinct, they can also be complementary: "[B]etween effects in longitudinal studies are often equally illuminating [as within effects], despite being by definition non-changing… [I]n cross-sectional studies, the effects of wider social contexts on individuals can also be extremely relevant. Social science is concerned with understanding the world as it exists, not just dynamic changes within it" (Bell et al., 2019).

We emphasize focusing first on the phenomenon, because in the statistical methods literature, one sometimes encounters general claims such as the following: "In multilevel research, the interest often lies on the level-two variables, and between-cluster effects of the level-one variables are level-two effects" (Schunck & Perales, 2017); and "In many applications the whole point of using panel data is to allow [unobserved heterogeneity] to be arbitrarily correlated with the [regressors]. A fixed effects analysis achieves this purpose explicitly" (Wooldridge, 2002: 265). These kinds of statements—abundant in the methods literature—must never be taken out of context and read as prescriptive. We do not know whether Schunck and Wooldridge sought prescription, but researchers often read methodological texts as prescriptive. As to the first claim, researchers must use their own judgment and contextual consideration in deciding whether level-one (within) or level-two (between) variables are of interest. As to the second, choosing the fixed-effects specification is fundamentally misguided if one is interested in a between phenomenon. While econometricians may be primarily concerned with the quality of the parameter estimates, empirical researchers must first consider whether the parameters appropriately operationalize the theory being tested.

3.2. Question 2: What Are the Theoretical Mechanisms that *Explain* the Phenomenon?

After considering the phenomenon of interest, researchers need to provide an account of what exactly is driving what, how, and why. Consider "What drives productivity improvements at the plant?" versus "Why are some plants more productive than others?" These are questions of theoretical logic, because they are about the theoretical mechanisms that link the predictor variables to the outcomes.

Again, a detailed examination and explication of the theoretical mechanisms should reveal the level, or levels, at which the mechanisms operate. Or put differently, a theory that is not explicit about levels is probably too vague to be tested. In panel-data models, the obvious mechanisms are (1) within variation in the regressor drives within variation in the outcome, (2) between variation in the regressor drives between variation in the outcome, or (3) both. Within and between mechanisms may share the same general theoretical foundation but differ in specifics. In Mohr's (1982) example of an individual acquiring power in an organization versus the distribution of power across individuals in an organization, both mechanisms may be described theoretically by invoking *hierarchical position power* (Galbraith, 1974), but the specific mechanisms that link the drivers to the outcomes at the two levels likely differ.

The answer to the question "Is it a within or a between mechanism, or both?" depends on the context and the research question. In asking how risk results in some firms performing better than others, Bliese et al. (2020: 24) noted that one can "articulate theoretical mechanisms that operate both within firm and between firms." In other contexts, it may be that only one of the mechanisms is relevant. For example, aspirations models (e.g., Bromiley & Harris, 2014) address how aspirations and performance vary over time within firms, suggesting an exclusively within mechanism. More generally, theories of decision-making and resource allocation often emphasize the longitudinal aspects over the cross-sectional.

3.3. Question 3: Is the Mechanism Empirically Accessible?

If one's theoretical logic unambiguously specifies the levels at which the propositions operate, theoretical alignment of the model is possible. However, several obstacles remain to the adequate estimation of model parameters. Here, we are particularly concerned with whether the available data can be used to empirically address the question of theoretical interest.

Estimating a within parameter obviously requires within variance, which means that the within effects of variables that do not vary over time cannot be estimated. For example, economies of scale and scope easily lend themselves to between analysis (because firms differ in scale and scope), but longitudinal datasets where firms change scale and scope significantly over time are scarce, and in some industries, non-existent. For example, commercial airlines tend to be either generalists, such as American Airlines, or specialists, such as Southwest Airlines (Lapré

& Scudder, 2004). Southwest serves only 14 international destinations and its fleet consists of only Boeing 737 aircraft; American Airlines serves 95 international destinations and operates 15 different types of aircraft. We are not aware of a single specialist airline that has developed into a generalist airline over time, which means that the data to examine what happens if a specialist airline expands its scope to become a generalist do not exist. Assuming it becomes like the other generalists is an ecological fallacy.

3.4. Question 4: What Do the Variables Mean?

Does the between variable $\bar{x}_i$ (panel mean) have the same interpretation as the within variable $x_{it} - \bar{x}_i$ (deviation from panel mean)? Whether this *within transformation* (Wooldridge, 2002: 267) performed as part of the FE specification maintains the meaning of the variable must be explicitly established: The empirical transformation of a variable may change the interpretation of the variable, and consequently, the interpretation of model parameters.

Consider the case where x_{it} measures sales for firm *i* in month *t*. The between variable $\bar{x}_i$ is a measure of average firm sales over all the months observed, which is plausibly interpreted as representing firm scale or size. But the within variable $x_{it} - \bar{x}_i$ cannot represent size because subtracting $\bar{x}_i$ removes the size effect. The within variable is positive when sales are above average and negative when they are below average. Assuming no trends in the within data, within variation may simply reflect differences in monthly demand, not that the firm changes size. More generally, most between variables reflect complex interactions of systems such that the firm does not (and probably cannot) simply scale these up or down with annual variation. For example, the administrative and managerial side of firm expenses does not change scale automatically to adjust to annual variation in sales, whereas purchase of raw materials does.

Even where within and between variables are calculated from the same raw data and have identical measurement units, they may have different meanings in the model; the meaning of the variable in the context of *the data* is not necessarily the same as its meaning in the context of *the model*. In our sales example, between variation in sales may be interpreted as differences in size (and scale), but within variation plausibly reflects differences in demand (and operationally, capacity utilization) over time. On observing that sales in November were higher than in October, we probably would not conclude that the firm has grown in size.

An important aspect of model context is the level at which the variable resides. This may reasonably be taken as a general rule: "In the practice of multilevel analysis, it is known that the cluster means often have a substantively meaningful interpretation, different from the level one variables from which they are calculated" (Snijders & Berkhof, 2008: 146).

In sum, in establishing theoretical alignment, researcher must consider the possibility that within and between interpretations of variables differ. Assuming without demonstration that a variable has the same meaning at different levels of the model can lead to an ecological or atomistic fallacy. Our prescription here is straightforward: Assume *a priori* that within and between meanings differ and conclude otherwise only if an explicit theoretical and substantive analysis supports it.

3.5. Question 5: Do Within and Between Effects Differ?

If the meanings of the within and between variables differ, comparing the within and between effects makes little sense. However, if we can establish that the within and between variables have the same meaning, we might want to examine whether the within and the between effects are equal. Thus, we might ask: When would an examination of a firm at different points in time provide the same empirical insight as a cross-sectional analysis of firms? We are hard pressed to think of any examples where the answer would be an unambiguous *yes*.

Whether what we see cross-sectionally differs from what we see longitudinally must start at the theoretical question: Do the hypothesized within and between mechanisms differ? We can always conduct an inferential test of equality, but a statistical test should be deployed only after formulating a hypothesis from theory. As the examples discussed earlier suggest, the within and between mechanisms are likely to differ.

The equality of the within and between effects has received little attention in the management literature. However, the psychology literature addresses this topic in some detail, and could provide some insight and guidance (Fisher, Medaglia, & Jeronimus, 2019; Molenaar, 2004). Researchers in the psychological sciences and clinical psychology must constantly address the question of *group-to-individual generalizability*: Do studies of inter-individual variation in a group give information about intra-individual variation over time? They may or may not: Fisher, Madaglia, and Jeronimus (2019: 6107) described the consequences of using cross-sectional differences uncritically to draw conclusions about within effects as potentially "catastrophic." In management research, the consequences may not be catastrophic, but fallacious conclusions are always a cause for concern.

In addressing whether within and between effects are equal, psychologists invoke the notion of *ergodicity* from probability theory. A stochastic process is ergodic if we can draw conclusions from its statistical properties by examining a sufficiently long random sample of it. Importantly, ergodicity assumes that observing any two instances of the process (e.g., by following two different firms), we would make the same inferences. If the process is ergodic, then interindividual and intraindividual variation are asymptotically equivalent (Molenaar, 2004:

206). Consequently, inferences based on group-level data can be applied at the individual level as well (Fisher et al., 2019). The critical question is: When can ergodicity be assumed?

Borsboom et al. (2009: 78) concluded in the context of psychological phenomena that "ergodicity should be viewed as an esoteric condition, that is, we should normally work from the hypothesis that ergodicity does not hold." This is ultimately an empirical question, but we propose that in management research, the mechanisms that produce the within and the between effects likely differ. Further, as several of our examples show, sometimes the question whether within and between effects are the same is nonsensical due to different interpretations of the variables at different levels: It makes little sense to ask whether production scale (the between interpretation of production volume) and capacity utilization (the within interpretation of production volume) have identical effects on productivity.

4. CONCLUSION

Despite the decline of purely cross-sectional studies and increased use of longitudinal data, researchers still need to address both cross-sectional and longitudinal research questions, because the research questions of interest may reside both at the between and within levels. Specifically, while some phenomena of interest may be *between phenomena* and others *within phenomena*, many exhibit duality being relevant both at within and between levels. Productivity in a multi-factory industrial firm is a good example: "Why do some factories perform better than others?" (the between question) and "How can a factory improve its performance?" (the within question) are both theoretically and practically relevant. In all empirical research, researchers need to be clear about the level of the phenomenon, and statistical models must be specified accordingly. Recent methodological texts suggest that this is often overlooked.

Addressing model specification in its entirety involves both theoretical and inferential considerations. Researchers have become increasingly skilled in examining the inferential aspects of their models: How does the within-panel disturbance behave? How do I ensure I model it correctly? How should I address unobserved panel-level heterogeneity? What can I do about endogeneity at within and between levels? The inferential tools in statistical software packages have likewise made impressive progress possible over the past forty years.

In this paper, we have sought to complement the remarkable inferential progress by directing attention to a more fundamental issue in panel-data models: Is the model consistent with the level of analysis of the substantive theory? The question is more profound than whether the estimator one chooses provides consistent estimates; the question is whether the model contains the parameters that operationalize the hypotheses of interest. We hope this paper demonstrates the answer to this seemingly simple question is both important and complex.

Finally, whereas inferential issues and statistical methodology are in some sense context free, addressing theoretical alignment must incorporate the specific phenomenon of interest and the specific theoretical mechanisms that link model variables to one another. Further, because theoretical alignment depends on context, there is no textbook or technical procedure that provides the sort of general guidance methodological texts provide on inferential matters. At the same time, systematic attention to theoretical alignment is necessary. We hope that this paper establishes that some general guidance and prescription is possible. But instead of decision rules and formalized procedures, our prescription takes the form of questions. By addressing these questions in their specific, idiosyncratic research context, researchers can work toward ensuring theoretical alignment of their statistical models.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for helping us crystallize our argument. We also thank Editor Kalyan Singhal for his guidance throughout the review process. In addition, we acknowledge the useful comments given by our colleagues on earlier versions of this article. In particular, we thank the participants of the Paper Session at the 2019 Academy of Management Annual Meeting and our colleagues at Universidad Carlos III de Madrid. This article has also benefited greatly from numerous conversations with our colleagues. All errors and omissions are our own.

REFERENCES

Bell, A., M. Fairbrother, K. Jones. 2019. Fixed and random effects models: Making an informed choice. *Quality & Quantity* **53**(2) 1051-1074.

Bliese, P.D., D.J. Schepker, S.M. Essmann. 2020. Bridging methodological divides between macro- and microresearch: Endogeneity and methods for panel data. *Journal of Management* **46**(1) 70-99.

Borsboom, D., R.A. Kievit, D. Cervone, B.S. Hood. 2009. The two disciplines of scientific psychology, or: The disunity of psychology as a working hypothesis. J. Valsiner, P.C.M. Molenaar, M.C.D.P. Lyra, N. Chaudhary, eds. *Dynamic Process Methodology in the Social and Developmental Sciences*. Springer, Dordrecht, The Netherlands, 67-97.

Bou, J.C., A. Satorra. 2007. The persistence of abnormal returns at industry and firm levels: Evidence from Spain. *Strategic Management Journal* **28**(7) 707-722.

Brewer, P., S. Venaik. 2014. The ecological fallacy in national culture research. *Organization Studies* **35**(7) 1063-1086.

Bromiley, P., J.D. Harris. 2014. A comparison of alternative measures of organizational aspirations. *Strategic Management Journal* **35**(3) 338-357.

Certo, S.T., M.C. Withers, M. Semadeni. 2017. A tale of two effects: Using longitudinal data to compare within- and between-firm effects. *Strategic Management Journal* **38**(7) 1536-1556.

Chen, W., K.D. Miller. 2007. Situational and institutional determinants of firms' R&D search intensity. *Strategic Management Journal* **28**(4) 369-381.

Dreyfus, D., A. Nair, S. Talluri. 2020. The impact of chain organization size on efficiency and quality of affiliated facilities—Implications for multi-unit organizational forms. *Production and Operations Management* **29**(7), 1605-1623.

Fisher, A.J., J.D. Medaglia, B.F. Jeronimus. 2019. Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences* **117**(27) E6106-E6115.

Galbraith, J.R. 1974. Organization design: An information processing view. *Interfaces* **4**(3) 28-36.

Hausman, J.A. 1978. Specification tests in econometrics. *Econometrica* **46**(6) 1251-1271.

Hayes, R.H., G.P. Pisano, D.M. Upton, S.C. Wheelwright. 2005. *Operations, Strategy, and Technology: Pursuing the Competitive Edge*. Wiley, New York.

Karatzas, A., G. Papadopoulos, J. Godsell. 2020. Servitization and the effect of training on service delivery system performance. *Production and Operations Management* **29**(5) 1101-1121.

Lapré, M.A., G.D. Scudder. 2004. Performance improvement paths in the U.S. airline industry: Linking trade-offs to asset frontiers. *Production and Operations Management* **13**(2) 123-134.

Maddala, G.S. 1971. The use of variance components models in pooling cross section and time series data. *Econometrica* **39**(2) 341-358.

Metters, R.D., F.X. Frei, V.A. Vargas. 1999. Measurement of multiple sites in service firms with data envelopment analysis. *Production and Operations Management* **8**(3) 264-281.

Mohr, L.B. 1982. *Explaining Organizational Behavior*. Jossey-Bass, San Francisco, CA.

Molenaar, P.C.M. 2004. A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement* **2**(4) 201-218.

Muthén, B.O. 1994. Multilevel covariance structure analysis. *Sociological Methods & Research* **22**(3) 376-398.

Robinson, W.S. 1950. Ecological correlations and the behavior of individuals. *American Sociological Review* **15**(3) 351-357.

Shockley, J., L.A. Plummer, A.V. Roth, L.D. Fredendall. 2015. Strategic design responsiveness: An empirical analysis of US retail store networks. *Production and Operations Management* **24**(3) 451-468.

Schunck, R., F. Perales. 2017. Within- and between-cluster effects in generalized linear mixed models: A discussion of approaches and the xthybrid command. *Stata Journal* 17(1) 89-115.

Snijders, T.A.B., J. Berkhof. 2008. Diagnostic checks for multilevel models. J. de Leeuw, E. Meijer, eds. *Handbook of Multilevel Analysis*. Springer, Berlin, 141-175.

Wooldridge, J.M. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

TECHNICAL APPENDIX

The purpose of this technical appendix is to show how the assumption of equality of the within and between effects inevitably enters into our reasoning whenever we use Hausman's specification test to choose between the fixed-effects (FE) and the random-effect (RE) estimator, labeled $\hat{\beta}_{FE}$ and $\hat{\beta}_{GLS}$, respectively (Hausman, 1978: 1263). Consider these estimators in the context of estimating the familiar regression model $y_{it} = X_{it}\beta + \mu_i + \varepsilon_{it}$.

In choosing the estimator for β, one is effectively choosing how the statistical information in the sums of squares and sums of products is used in estimation. Maddala (1971: 343) gives a convenient expression for the general estimator: $\hat{\beta} = [W_{xx} + \theta B_{xx}]^{-1}[W_{xy} + \theta B_{xy}]$, where W_{xx} and B_{xx} are the within and between sums of squares of x, respectively, and W_{xy} and B_{xy} are the within and between sums of products, respectively. Note that both within and between variances and covariances are assumed to supply information on the *same* vector of parameters, β; there are no separate within and between parameters.

It is straightforward to use Maddala's expression to describe the common panel-data estimators. Specifically, $\hat{\beta}_{FE}$ is obtained by setting $\theta = 0$: The FE estimator makes use of only within variance, which is why $\hat{\beta}_{FE}$ is called the *within estimator*. $\hat{\beta}_{GLS}$ is obtained when $\theta > 0$, making $\hat{\beta}_{GLS}$ a "weighted average of $\hat{\beta}_{FE}$ (the within group estimator) and the between group estimator" (Hausman, 1978: 1263).

If we now define $\hat{q} = \hat{\beta}_{FE} - \hat{\beta}_{GLS}$ (Hausman, 1978: 1263), the structure of Hausman's specification test can be expressed effectively as follows: $H_0\colon \hat{q} = 0$, $H_A\colon \hat{q} \neq 0$. In this context, a useful way to interpret the test is to ask whether an estimate that is calculated exclusively from within information (W_{xy}) changes if between information (B_{xy}) is incorporated? A statistically significant Hausman test statistic means that the estimate indeed changes, that is, B_{xy} supplies statistical information that is different from W_{xy}. If this is the case, Hausman's (1978: 1269) prescription is to seek an understanding of why this happens.

There are two reasons for why incorporating B_{xy} changes the estimate. One is that $\hat{\beta}_{GLS}$ is inconsistent, because the RE assumption that unobserved level-2 heterogeneity does not correlate with the regressors is violated. $\hat{\beta}_{FE}$ is in turn unaffected by endogenous level-2 heterogeneity. The other explanation, often overlooked, is that the mechanism that links *x* and *y* to one another within firms over time (the within effect) is different from the mechanism that links them cross-sectionally across firms (the between effect). In Maddala's (1971: 347) terminology, the between-group estimate $B_{xx}^{-1}B_{xy}$ and the within-group estimate $W_{xx}^{-1}W_{xy}$ may estimate *different* parameters, with different underpinning theoretical logic and different interpretations.

The first explanation is the one to which panel-data researchers default. But this default assumes that the second can be eliminated as an alternative explanation. The chronic problem in research practice is that researchers are not aware of the second alternative, and instead, default to the logic that a significant Hausman statistic means that $\hat{\beta}_{FE}$ is consistent but $\hat{\beta}_{GLS}$ is not.

There is of course a potential third reason for a significant Hausman test statistic: Both $\hat{\beta}_{FE}$ and $\hat{\beta}_{GLS}$ are inconsistent, because the assumption that unobserved level-1 heterogeneity (the disturbance term) does not correlate with the independent variables is violated. In this case the entire test is unreliable; Hausman test requires the assumption that $\hat{\beta}_{FE}$ is consistent.

In sum, unless one assumes the true within and between effects to be equal, the interpretation of a significant Hausman statistic is ambiguous at best and misleading at worst. Therefore, we conclude that anyone who wishes to use the Hausman test to choose between the FE and the RE specification in an unambiguous and non-misleading way must assume both that that $\hat{\beta}_{FE}$ is consistent and that the true within and between parameters are equal.