Testing contingency hypotheses in budgetary research using moderated regression analysis: a second look

Frank G.H. Hartmann[a], Frank Moers[b],*

[a]*Amsterdam Graduate Business School, Universiteit van Amsterdam, The Netherlands*
[b]*Department of Accounting & Information Management, Maastricht University, The Netherlands*

Abstract

This paper addresses the use of moderated regression analysis (MRA) in contemporary management accounting and control research. It follows up on a discussion started by us [Hartmann & Moers, *Accounting, Organizations and Society* 24 (1999) 291–315] in this journal and provides a reaction to arguments put forward by Dunk (this issue). In doing so, this paper addresses the relationship between substantive theory and statistical test, emphasizes the need to distinguish between confirmatory and exploratory uses of MRA and argues that the importance of moderated hypotheses and tests may have been overstated in the management accounting literature under review.
© 2003 Elsevier Ltd. All rights reserved.

Introduction

The development of the behavioral management accounting literature central in Hartmann and Moers (1999) and Dunk (this issue) spans more than forty years. Argyris' (1952) investigation of budget-related behaviors is commonly seen as the starting point of a literature that addresses behavioral and attitudinal effects of management accounting systems in organizations. This literature received an impulse from an apparent conflict in findings of Hopwood (1972) and Otley (1978) that spurred the adoption of contingency perspectives that aimed to find a match between the use of budgets and the context in which they are used. Over time, several papers have provided critical evaluations of different aspects of this contingency literature on budgeting (e.g. Briers & Hirst, 1990; Fisher, 1995; Hartmann, 2000; Hartmann & Moers, 1999; Kren & Liao, 1988; Shields & Shields, 1998). We (Hartmann & Moers, 1999) focused on the statistical method used to test contingency hypotheses, and identified Moderated Regression Analysis (MRA) as the dominant statistical format to detect contingency effects. We analyzed 28 budgetary papers for several known problems of MRA and concluded that a large majority of papers showed errors in the application of MRA that may hinder the substantive interpretation of research findings.

In a recent reaction to our paper, Dunk (this issue) sketches a more positive picture of the application of MRA in the behavioral budgeting literature. Dunk (this issue) questions some of our analyses and findings, and argues that MRA-related problems are trivial and not material. In this paper, we provide a reaction to the points raised by Dunk (this issue). Where this reaction requires more in-depth discussions of MRA principles, we provide these in short form.

* Corresponding author.
E-mail addresses: hartmann@fee.uva.nl (F.G.H. Hartmann);
f.moers@AIM.unimaas.nl (F. Moers).

The paper proceeds as follows. In the following section, we discuss the motivation to use MRA when testing contingency hypotheses. This section also addresses key arguments in the paper by Dunk (this issue). In the next section, we provide a brief discussion of controversial issues in MRA. This section addresses the difference between form and strength effects, the different ways in which they can be statistically determined and, most importantly, their substantive relevance. Following this, we elaborate on the issue of treating management accounting systems as an independent variable and the endogeneity problem. Finally, we end the paper with our conclusions.

MRA and contingency theory

As a start, it is important to realize that no law exists stating that all use of statistics in academic research should be preceded by theory development or hypothesis formulation. Indeed, both the role of theory and the exploratory or confirmatory role of data analysis vary widely between literatures. The review of the budgetary literature in Hartmann and Moers (1999), however, explicitly analyzed theory-based empirical research. In this literature, it is paramount that hypotheses are clearly, logically and explicitly derived from theory, and subsequently tested using an appropriate technique. In this sense we agree completely with Dunk's (this issue) concern for papers that appear to speculate instead of theorize, or that seem to confuse predictor with moderator variables (cf. Luft & Shields, 2003). We are inclined to disagree, however, with Dunk's opinion on the necessity of stating *contingency* hypotheses, especially when this results in a plea for the use of—increasingly complex—interaction models. Instead, there are sufficient reasons to argue that relationships in accounting research should be singular and universal.

Generally, the reductionist principle from the philosophy of science, that underlies research as-we-know-it, demands us to distinguish between the complexity of the world and the complexity of theories *about* the world. Although the former is undeniably high, the latter should preferably be not. This not only holds for the social sciences, but also for the physical sciences. Different from what Dunk (this issue) seems to suggest, the physical sciences indeed do not settle for conditional knowledge, and are instead literally about establishing 'universal' relationships. It is thus not clear why the prediction and test of higher-order interactions are 'essential' in the budgeting literature (Dunk, this issue). Parsimony may simply require us to consider variables other than *RAPM*, *budgetary participation* and *task uncertainty* when we build our theories.

More specifically, promoting the use of three-way interactions because '…research increasingly shows that two-way interaction models fail to reflect robustly the many relations that occur in everyday life and in organizational settings' (Dunk, this issue) fails to address the question whether interaction models are appropriate at all. This question is even more relevant today. Given that economics has taken up a dominant position in management accounting research over the last decade, empirical relationships predicted may require somewhat different models than moderated regression analysis. The economic perspective has also drawn our attention to the endogeneity problem, which can result in path models being mistakenly presented as moderated relationships. Recently, Luft and Shields (2003) have argued that this is related to the question whether management accounting systems should be used as independent variable or as dependent variable. We elaborate on this issue in the section headed "MRA and the endogeneity problem". Apart from the noted upswing of economic theories structural equation models such as LISREL have become an important statistical tool in management accounting research. Such models allow testing alternative contingency forms of fit such as *selection* or *mediation* (Hartmann & Moers, 1999, p. 310). It is therefore important to reiterate here that the adoption of a contingency view does not imply the use of MRA. Nowhere have we made a claim otherwise (see e.g. Hartmann & Moers, 1999, Table A1).

For these reasons, it is not clear that formulating and testing interaction hypotheses follow some theoretical or methodological necessity. Theoretical progress may require simpler rather than more complex models, especially when new contingency variables are added without theoretical support (cf. Otley, 1980). The current status of the behavioral budgeting literature seems to support the presented arguments strongly (e.g. Shields & Shields, 1998; Hartmann, 2000), resulting in pleas for simpler theories and tests that lead to a solid understanding of essential relationships between management accounting system elements (e.g. Luft & Shields, 2003). If, however, the theory suggests moderated relationships, and MRA is selected for statistical analysis, the remaining issue is simply how to do this right. Here, many problems and pitfalls are still lurking, as is broadly recognized in the wider organizational research literature (see, e.g. Aguinis, 2002, Landis & Dunlap, 2000).

Hypothesis development and testing

The behavioral budgeting literature under scrutiny in Hartmann and Moers (1999) attempts to be theory based. This requires that arguments are proposed in a logical order, consistent with prior evidence, and that resulting theory is summarized in hypotheses that allow testing the theory's implications. This clearly excludes null hypotheses as an allowable format for hypothesis formulation, unless they are complemented with alternative hypotheses. We observed, however, that many budgeting papers do not summarize their theories in a directional hypothesis. It is further clear to us and to Dunk that the statistical form of hypothesis testing comes down to attempts to reject a null hypothesis and to accept the alternative hypothesis, and that accepting or proving a theory is formally impossible. Any discussion of philosophy of science semantics, however, should not obscure the fact that the link between theory and test should be as strict as possible, and that statistical procedures should be executed properly. The full argument about the relationship between theory, hypothesis and test is therefore a bit more complicated, than some of the arguments proposed by Dunk (this issue) may suggest.

Symmetry in MRA

One of the lasting complications is related to the issue of symmetry in MRA. Symmetry simply denotes the fact that the significance of β_3 in Eq. (1) indicates that X_2 moderates the relationship between X_1 and Y, and that at the same time X_1 moderates the relationship between X_2 and Y.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 \times X_2 + \varepsilon \qquad (1)$$

Therefore, theory rather than statistics should decide what variable is labeled 'moderator' (Hartmann & Moers, 1999, p. 249). The complication is that, given the fact that statistical symmetry is not equivalent to theoretical symmetry, a certain MRA model may be used to support various theories. Our concern related to the fact that this may provide incentives to researchers to indulge in an exploratory use of MRA resulting *in*, rather than *from* theory development, so that the statistical analysis appears to provide support for the theory, where it is really absent. Substantively, this may explain some idiosyncratic theories and findings in budgetary research, especially in small sample studies where statistical power is low. Dunk (this issue) does not seem to share our concern, but it is for exactly the same reasons that Luft and Shields (2003) recently suggested stringent rules and labels for moderators in the formulation of theory.

Form and strength hypotheses

Another complication relates to the substantive difference between moderating the *strength* of a relationship and moderating the *form* of a relationship. Taking a subgroup example, hypotheses of the *strength* type predict that correlation coefficients of a relationship between X_1 and Y will differ between subgroups. Hypotheses of the *form* type predict that the regression coefficients of the relationship between X_1 and Y will differ between subgroups. Both types of moderation may be meaningful, but of central importance is to establish what format is predicted by theory. Dunk (this issue) again seems to disagree with the precedence of theory, as he suggests that a clear focus on one format at a time would '...seriously limit the findings of a study'. However, this sounds

once more as a plea for the exploratory rather than confirmatory use of statistics in general, and MRA in particular.

Lower order effects and omission of variables

The complexity underlying MRA is also illustrated by the issue of the interpretation of main effects in a moderated regression equation (Dunk, this issue). In principle, the β_1 and β_2 coefficients obtained for the full model [Eq. (1)], will not be the same as those obtained for the reduced model [Eq. (2)] when β_3 is different from zero.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \qquad (2)$$

For ratio-scaled variables and interval-scaled variables that are centered, the β_1 and β_2 can be interpreted. This fact does not result from any 'appeal' made for it, nor from 'data manipulation' as Dunk (this issue) wants use to believe, but it simply follows from the basic properties of the types of measurement scales mentioned. Dunk (this issue) raises a more essential point as he questions the possible existence of a 'main' effect in the presence of an interaction. Dunk's answer is no, but it should be yes, as is illustrated in Fig. 1 below.

This figure presents an estimated relationship between Y and X_1 for different values of X_2 and illustrates the existence of a main effect in the presence of an interaction. In reaction to some of the other points mentioned by Dunk (this issue), it is easy to show that Fig. 1 is, per definition, not sensitive to scale manipulations of X_1 or X_2, and forcing the main effect coefficients to zero, by scale

manipulations, can never lead to the situation that '...the entire effect is carried by the estimate of the interaction term' (Dunk, this issue). The simple reason being that the significance of β_3 reflects the *incremental* variance explained by the interaction term, and not the *total* variance explained by the model. When establishing interaction, a more important issue is the 'reverse' question whether there exists a material or relevant interaction in the presence of main effects if the incremental variance explained by the interaction term is small compared to the total variance explained by the model. Also here, any discussion should not confuse parsimony of theories and the complexity of the world.

Parsimony is also important with regard to the inclusion of lower order variables and interactions in higher order equations. Earlier, we argued that some of the MRA-models we evaluated included variables that were theoretically unsupported, with the threat of the model becoming over-specified (Hartmann & Moers, 1999, p. 303). In the two budgetary papers that we used as examples (Imoisili, 1989; Harrison, 1992), higher-order interaction coefficients were initially found to be not significant, after which reduced models were tested involving the simultaneous test of multiple lower-order interactions. Dunk (this issues) unfortunately appears to still see no harm in such adhoc 'testing' of reduced models, even if such models have no bearing on the theory originally proposed. Fortunately, Dunk does seem to accept, however, our earlier conclusion about over-specification of regression models (Hartmann & Moers, 1999, p. 303) as he now acknowledges that
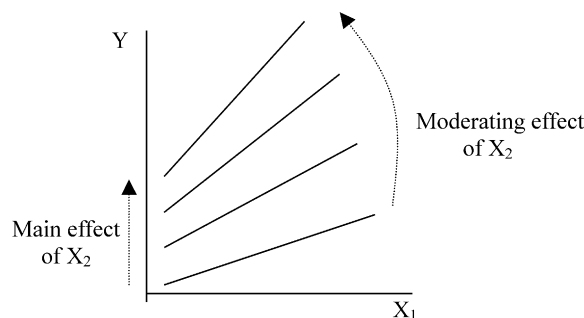

Fig. 1. Graphical representation of the main effect of X_2 on the dependent variable Y and the moderating effect of X_2 on the relationship between Y and X.

'...an overspecified model [...] makes it more difficult to reject the null hypotheses ... [which] ... has significant implications for research' (Dunk, this issue). This was precisely our point.

Effect size

Hartmann and Moers (1999) used the concept of 'effect size' (see, e.g. Pedhazur & Pedhazur-Schmelkin, 1991, p. 206) for demonstrating that the existence of an interactive effect of X_1 and X_2 on Y does not allow conclusions about what scores of X_1 and X_2 result in the highest score of Y. This is caused by a 'main effect' of the moderator (β_2), as is illustrated below by the difference between constants in Eqs (1a) and (1b).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 \times X_2 + \varepsilon \quad (1)$$

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$
(for $X_2 = 0$) $\quad (1a)$

$$Y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)X_1 + \varepsilon$$
(for $X_2 = 1$) $\quad (1b)$

Dunk (this issue) disagrees, and tries to demonstrate that the difference in intercepts is also caused by '...a difference in the constant term β_0 ...', but this is simply impossible since the constant term β_0 is what it is, constant. More important is that any properly formulated hypothesis clearly specifies ex-ante whether an interaction effect is predicted, or whether a certain score of the dependent variable is predicted. A clear focus on confirmatory, instead of exploratory, statistical analyses eliminates the need and possibility to restate hypotheses *ex-post*, such as now proposed by Dunk (this issue).

MRA and the endogeneity problem

It is clear from the previous discussions that much of the controversy surrounding MRA cannot be addressed in isolation from the theory that is being tested with MRA. On the one hand, as we argued before, there is reason to believe that the need to predict and test moderated relationships may be lower than argued by some (e.g. Dunk this issue). On the other hand, recently more attention is paid to the endogeneity problem that may occur when we treat management accounting systems as an independent variable, as is the case for the literature under scrutiny (Ittner & Larcker, 2001; Luft & Shields, 2003). The endogeneity problem can be illustrated with the example of the relationship between Job Related Tension (JRT), Budget Emphasis (BE) and Environmental Uncertainty (EU) that has received much attention in the budgetary literature (Hartmann, 2000).

A typical hypothesis would be that the impact of BE on JRT is higher for higher levels of EU. Such a hypothesis implicitly assumes that superiors do not adjust their use of budgets to the uncertainty in the environment, or, that BE is exogenously given and EU is a moderator, not an antecedent. When testing this hypothesis, researchers would typically use MRA (Hartmann & Moers, 1999), i.e.

$$\text{JRT} = \beta_0 + \beta_1 \text{BE} + \beta_2 \text{EU} + \beta_3 \text{BE} \times \text{EU} + \varepsilon \quad (3)$$

The question is what would be the proper interpretation of finding β_3 to be positive and significant. One interpretation is that, in line with the hypothesis stated, the higher the EU the greater the impact of BE on JRT. This would imply materially that superiors should put less emphasis on budgets when EU is high. Such a finding, however, constitutes an important paradox between the ex-ante assumption that superiors do not adjust their use of budgets to environmental uncertainty and the ex-post claim that they should.

What if superiors, on average, actually do behave in a way consistent with the researcher's claim, so that they do adjust their use of budgets to environmental uncertainty, such as depicted in Fig. 2, Panel A. In that case, EU in Eq. (3) can be rewritten into BE and, as a result, JRT becomes a quadratic, curvilinear function of BE. This is not extraordinary since, as Luft and Shields (2003, p. 189) recently documented, 'much of the theory underlying empirical management accounting

Panel A

BE

EU

Panel B

JRT

BE

BE = g(EU)
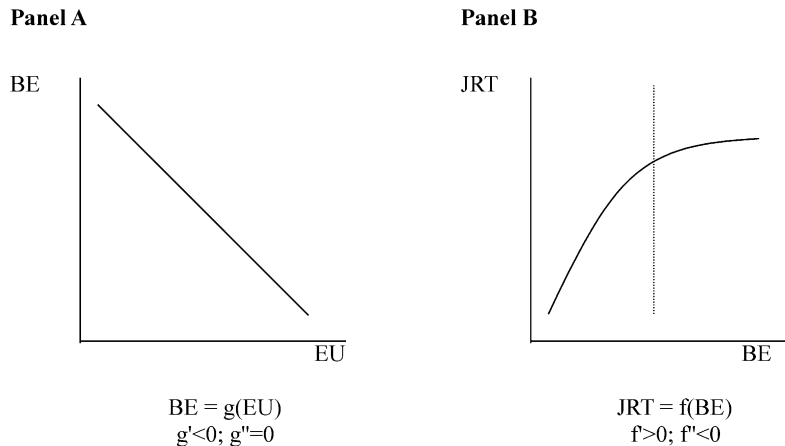g'<0; g"=0

JRT = f(BE)
f'>0; f"<0

Fig. 2. Hypothetical relationship between respectively Budget Emphasis (BE) and Environmental Uncertainty (EU) and Job Related Tension (JRT) and BE.

research predicts curvilinear relations'. If β_3 is found to be positive and significant in Eq. (3), the alternative interpretation therefore is that JRT and BE are directly (universally) related, as is depicted in Fig. 2, Panel B. Such a result, therefore, does not necessarily provide support for moderation, but may be due to an endogeneity problem.[1]

The fact that BE and EU are related, implies that certain values of BE are only observed when EU is low (high). More specifically, high (low) BE is observed when EU is low (high). Therefore, estimating the relationship between JRT and BE for different *subgroups* of EU is similar to estimating the relationship between JRT and BE for different *value ranges* of BE. Since the slope differs between high and low values of BE (see Fig. 2, Panel B), a positive and significant β_3 in Eq. (3) may cause MRA to mistakenly present a path model as a moderated relationship. We conclude that the endogeneity problem calls for more care when establishing the link between theory and test (cf. Ittner & Larcker, 2001, p. 397). Naturally, theoretical progress is all about coming up with the most likely story by eliminating alternative explanations.

---

[1] More formally, if JRT really is a quadratic function of BE and BE is a choice variable (partly) determined by EU, then MRA is subject to an endogeneity problem since the included interaction term (BE×EU) is correlated with the omitted quadratic term (BE×BE).

Conclusions

The discussion above and the discussion in Hartmann and Moers (1999) illustrate that the application of MRA is not straightforward and requires care. In contrast, Dunk (this issue) disagrees with many of the concerns we put forward stating that our concerns are overstated, incorrect, or immaterial. Although we firmly believe that reviews and critiques are important to scientific progress, we also believe that Dunk has not put forward a single argument that would require us to restate the general or partial conclusions drawn in Hartmann and Moers (1999). We acknowledge that these conclusions were strong, but both form and content of our previous review were genuine expressions of our surprise about the outcomes of our analyses.

It is important to note again that despite the concerns we have raised with respect to the application of MRA in the budgetary literature, MRA may be merciful as it may provide a robust form of testing that defies many of the errors found in its application. Moreover, even if MRA is not robust, it is still possible that researchers draw right substantive conclusions from analyses that are formally wrong. This is exactly why we noted earlier that we could not be sure that '…conclusions drawn and presented in these papers are not supported by the *data…*' (Hart-

mann & Moers, 1999, p.308, emphasis added). The only real test of substantive implications of flawed MRA would be to reanalyze original data, applying MRA properly. Since our attempts to do this remained fruitless for reasons beyond our control (Hartmann & Moers, 1999, footnote 27), it would have been interesting and valuable if Dunk would have addressed this issue by actually reanalyzing data in order to evaluate the concerns we have raised. The fact that he chose not to do so is troublesome.

The use of statistical techniques has, as elsewhere, become a cornerstone in management accounting research. Although management accounting research often encounters severe problems in the collection of data, no arguments can justify the improper application of well-described techniques such as MRA, even if these techniques are complex. Such arguments can also not be found by arguing for the futility of attempts to develop a contingency theory of management accounting (see, e.g. Chapman, 1997), by nitpicking about the use of the word 'methodology', especially not when it is used to express a 'body of methods used in a particular branch of activity' (Oxford Dictionary, p. 637), or by boldly claiming that the deficiencies in MRA are 'minor' and '...likely to be swamped by far greater issues relating to psychometrics and sampling methods' (Dunk, this issue, p. 18). Although these and other problems mentioned will continue to provide tough challenges to accounting researchers, no compensation should be sought in the statistical procedures applied, nor should they hinder any periodical, systematical and perhaps even critical evaluation of research's progress and doings.

References

Aguinis, H. (2002). Estimation of interaction effects in organization studies. *Organizational Research Methods*, *5*, 207–211.

Argyris, C. (1952). *The impact of budgets on people*. Ithaca, New York: The Controllership Foundation.

Briers, M., & Hirst, M. K. (1990). The role of budgetary information in performance evaluation. *Accounting, Organizations and Society*, *15*, 373–398.

Chapman, C. S. (1997). Reflections on a contingent view of accounting. *Accounting, Organizations and Society*, *22*, 189–205.

Dunk, A. S. (this issue). Moderated regression, constructs and measurement in management accounting: a reflection. Accounting, Organizations and Society, this issue, doi: 10.1016/S0361-3682(02)00057-0.

Fisher, J. (1995). Contingency-based research on management control systems: categorization by levels of complexity. *Journal of Accounting Literature*, *14*, 24–53.

Harrison, G. L. (1992). The cross-cultural generalizability of the relation between participation, budget emphasis and job-related attitudes. *Accounting, Organizations and Society*, *17*, 1–15.

Hartmann, F. G. H. (2000). The appropriateness of RAPM: toward the further development of theory. *Accounting, Organizations and Society*, *25*, 451–482.

Hartmann, F. G. H., & Moers, F. (1999). Testing contingency hypotheses in budgetary research: an evaluation of the use of moderated regression analysis. *Accounting, Organizations and Society*, *24*, 291–315.

Hopwood, A. G. (1972). An empirical study of the role of accounting data in performance evaluation. *Journal of Accounting Research*, *10*, 156–182.

Imoisili, O. A. (1989). The role of budget data in the evaluation of managerial performance. *Accounting, Organizations and Society*, *14*, 325–335.

Ittner, C. D., & Larcker, D. F. (2001). Assessing empirical research in managerial accounting: a value-based management perspective. *Journal of Accounting and Economics*, *32*, 349–410.

Kren, L., & Liao, W. M. (1988). The role of accounting information in the control of organizations: a review of the evidence. *Journal of Accounting Literature*, *7*, 280–309.

Landis, R. S., & Dunlap, W. P. (2000). Moderated multiple regression tests are criterion specific. *Organizational Research Methods*, *3*, 254–266.

Luft, J., & Shields, M. D. (2003). Mapping management accounting: graphics and guidelines for theory-consistent empirical research. *Accounting, Organizations and Society*, *28*, 169–249.

Otley, D. T. (1978). Budget use and managerial performance. *Journal of Accounting Research*, *16*, 122–149.

Otley, D. T. (1980). The contingency theory of management accounting: achievement and prognosis. *Accounting, Organizations and Society*, *5*, 413–428.

Oxford Dictionary, Oxford University Press, 1991.

Pedhazur, E. J., & Pedhazur-Schmelkin, L. (1991). *Measurement, design and analysis: an integrated approach*. Hillsdale: Erlbaum.

Shields, J. F., & Shields, M. D. (1998). Antecedents of participative budgeting. *Accounting, Organizations and Society*, *23*, 49–76.