Testing for Complementarities between Accounting Practices

Stijn Masschelein[a,*], Frank Moers[b,*]

[a] *University of Western Australia Business School, Department of Accounting and Finance, 35 Stirling Highway, Crawley 6009, Australia.*
[b] *Maastricht University School of Business and Economics, Department of Accounting & Information Management, P.O. Box 616, 6200 MD Maastricht, The Netherlands.*

Abstract

Following the theoretical innovations of complementarity theory, management control studies have investigated interdependencies between different management control practices. In this paper, we compare the two dominant statistical specifications to test for the presence of an interdependency. We show theoretically how the power of the demand and the performance specification varies with the level of optimality in the sample and how those specifications are vulnerable to correlated omitted variable bias. Our simulation results reveal that the demand specification is more robust to variations in optimality and correlated omitted variables than the performance specification. We use these results to formulate recommendations for future research into management control interdependencies.

Keywords: Complementarity Theory, Management Control, Statistical Methods

*The authors are grateful for comments by two anonymous reviewers, David Bedford, seminar participants at Leuven University, Australian National University, the University of Technology Sydney, the University of New South Wales, the University of Western Australia, the Vienna University of Economics and Business, the 2018 AFAANZ conference, and the 2018 AOS conference on Management Control as System or Package in Maastricht.

1. Introduction

The management accounting literature has investigated the fit between accounting practices and the firms' environment (Chenhall, 2003; Otley, 2016), as well as the firms' choices of interdependent practices such as delegation and incentives (Bouwens & Van Lent, 2007; Indjejikian & Matejka, 2012; Moers, 2006), or the levers of control (Simons, 1994, 2000; Widener, 2007). The interdependencies between different practices are the reason a collection of practices form an accounting system (Grabner & Moers, 2013; Milgrom & Roberts, 1995). When two practices positively reinforce or complement each other, the firm benefits from using them together, as a system. When two practices negatively reinforce each other, they act as substitutes. In this paper, we compare the two most common specifications in the empirical literature to test whether practices are complements (or substitutes), i.e., the performance specification and the demand specification (Grabner & Moers, 2013), and examine the vulnerability of these specifications to their underlying assumptions.

The performance specification tests whether the interaction between two practices is postively correlated with performance (Athey & Stern, 1998; Carree et al., 2011; Grabner & Moers, 2013; Hofmann & van Lent, 2017). For instance, the interaction between delegation and accounting based incentives is positively related to business unit performance. The performance specification assumes that there are a "sufficient number" of firms that deviate from the optimal level for the practices which allows researchers to detect performance differences between optimal and suboptimal accounting systems. The extreme version of this assumption is that all firms make random choices. The demand specification on the other hand tests whether two practices are positively correlated with each other after controlling for environmental factors (Arora, 1996; Grabner & Moers, 2013; Johansson, 2018; Hofmann & van Lent, 2017). For instance, delegation and accounting based incentives are correlated after controlling for environmental factors. The demand specification assumes that there are a "sufficient number" of firms that simultaneously choose the optimal level of the practices taking into account the interdependency and the firm's environment. The extreme version of this assumption is that all firms make optimal choices.

In observational samples neither the assumption of completely randomly chosen practices nor the assumption of completely optimal choices will hold (Brynjolfsson & Milgrom, 2013). While an individual decision is either optimal or not, a non-optimal decision can be closer or further from the optimal level. As a result, a sample of those individual decisions can exhibit different levels of optimality where higher levels of optimality imply more observations that are close to the optimal decision. The methodology literature has treated the assumption about the level of optimality in a sample as an untestable assumption and recommends that researchers argue whether their chosen specification is appropriate for their research setting. The demand specification is often seen as more appropriate than the performance specification when one decision maker has designed the entire accounting system, when the optimal design is not too complicated, and when the decision maker has had sufficient time and incentives to choose the optimal system (Grabner & Moers, 2013; Hofmann & van Lent, 2017; Carree et al., 2011; Johansson, 2018). The performance specification is often deemed more appropriate when it involves relatively new practices and technologies that require experimentation (Carree et al., 2011; Bedford et al., 2016). While these recommendations are intuitive, their empirical validity is not clear. That is, it is not clear how vulnerable each specification is to deviations from its underlying assumption of (lack of) optimality and whether this vulnerability differs between the two specifications. The key question we address in this paper is whether the typical arguments for choosing a specification are necessary and/or sufficient.

In addressing this question, we first show how the demand and performance specification arise from the same underlying and unobserved objective function. The objective function formalises the performance effects of management accounting practices as well as how those performance effects depend on other practices and contingency factors. As a result, the objective function captures the main insight of complementarity theory (Milgrom & Roberts, 1995; Grabner & Moers, 2013) and contingency theory (Chenhall, 2003; Otley, 2016) and is thus the theoretical foundation for a hypothesis test regarding interdependencies. We then show the statistical problems that arise with the demand and performance specification when the assumptions underlying these specifications are not satisfied. First, the demand specification has more power to detect an interdependency when management control practices are closer to optimal while the performance specification is more likely to detect an interdependency when the control practices are further from optimal. But again, how close to or how far from optimal the practices need to be to have sufficient power is an open question. Second, while the correlated omitted variable problem is widely recognised for the demand specification (Grabner & Moers, 2013; Arora, 1996; Hofmann & van Lent, 2017), it is largely ignored (Grabner & Moers, 2013; Hofmann & van Lent, 2017) or thought of as non-existing for the performance specification (Carree et al., 2011). We show that both specifications are vulnerable to the same omitted variable bias when they do not appropriately control for a contingency factor that affects both practices. When all relevant practices and contingency factors are observed, this bias can be easily addressed, although the most commonly used variant of the performance specification unfortunately does not address the bias. However, when not all contingency factors are observed, it is an open question whether the two specifications are equally vulnerable to correlated omitted variables.

To investigate the robustness of both specifications to variations in the level of optimality, we use a simulation approach. In doing this, we focus on Type I errors (rejection of a true null hypothesis of no complementarity) and power (or Type II errors, i.e., failure to reject a false null hypothesis of no complementarity). The results show that the demand specification has appropriate Type I error rates while maintaining power to detect a real complementary effect, even at low levels of optimality. In contrast, the theoretically derived performance specification suffers from elevated Type I error rates at all levels of optimality and loses power to detect true interdependencies at higher levels of optimality. In addition, the performance specification as it is typically implemented in the literature is biased when the practices are contingent on the same environmental factor. Finally, in contrast to common assumptions, even in the presence of correlated omitted variables, the demand specification fares better than the performance specification.

This study contributes to the literature by providing three recommendations for studies that test for interdependencies between management control practices. First of all, the demand specification should generally be preferred over the performance specification in an observational sample of firms. While researchers are typically recommended to argue whether their chosen specification is appropriate for their research setting, our findings indicate that the use of the performance specification needs to be especially well-substantiated. Even when performance data is available, it is still advisable to report the demand functions before reporting the performance specification (Aral et al., 2012; Cassiman & Veugelers, 2006). Second, when theory and prior research indicate that the practices are contingent on environmental factors, studies should appropriately control for these factors. Because the management control literature has a rich history of studying contingency effects, we believe most management control settings will warrant appropriate controls for contingency factors. Importantly, these controls are as relevant for the performance specification as they are for the demand specification. In the demand specification, controlling for contingency factors means including the contingency variables as separate independent variables. In the performance specification, controlling for contingency effects requires including the interaction of the contingency factors with the management control practices. Of the published studies using the performance specification following Grabner & Moers (2013), only Bedford et al. (2016, 2019) appropriately control for contingency factors. Third, because the performance specification is vulnerable to Type I errors, we advise to use more robust estimation techniques than ordinary least squares.

The remainder of this paper is structured as follows. First, we derive the demand and performance specifications from a common objective function. Second, we explain and calibrate the simulation approach. Third, we compare the robustness of the different specifications in a simulation study. Fourth, we provide guidance for researchers who estimate interdependencies. Last, we summarise the simulation results and recommendations, and discuss how future research can improve the estimation of interdependencies.

2. Model and formal analysis

In this section, we present a firm's objective function to model the essential elements of both complementarity theory and contingency theory. According to complementarity theory, the performance effect of a management control practice depends on the use of another practice. According to contingency theory, the performance effect of the practice depends on the environment. The theoretical model helps to make explicit the assumptions in the two statistical specifications. We start with the assumption that a firm has to decide on the level of two management accounting practices which both depend on one environmental factor. We represent the levels of the management accounting practices as x_1 and x_2 and the level of the environmental factor as z. We further assume that performance, y, depends on a factor, ν while the performance effects of each management control practice further depend on a factor ϵ_1 and ϵ_2 respectively. That is, there is variation among firms in the extent to which each management control practice affects performance. x_1, x_2 and z are observed by the researcher but ϵ_1, ϵ_2, and ν are not.

We illustrate the model with an example from the management accounting literature where z stands for environmental uncertainty, one of the most commonly studied environmental factor in the contingency literature. x_1 and x_2 are two commonly studied management control decisions, the extent to which decisions are delegated to middle level managers (henceforth delegation) and the extent to which accounting measures are used to evaluate and reward performance of the middle level managers (henceforth accounting incentives). Finally, y is the financial performance of the business unit. These four constructs have received considerable attention in both the contingency and the complementarity literature (Grabner & Moers, 2013; Chenhall, 2003; Otley, 2016). In the example, we cannot do justice to the rich theories and measurement developed in the literature and therefore the example should only be seen as an illustration of the objective function. In addition, the linear form is a special case, but its simplicity allows us to make our point without loss of generality.

$$y = \beta_0 + (\beta_1 + \gamma_1 z + \epsilon_1)x_1 + (\beta_2 + \gamma_2 z + \epsilon_2)x_2 + \beta_{12}x_1x_2 - \frac{1}{2}\delta_1 x_1^2 - \frac{1}{2}\delta_2 x_2^2 + \nu \quad (1)$$

The objective function (1) shows how profit, y, depends on delegation and accounting incentives, x_1 and x_2, environmental uncertainty z and the unobservable factors. The parameter β_{12} represents the complementarity between delegation and accounting incentives (i.e. $\frac{\delta^2 y}{\delta x_1 \delta x_2} = \beta_{12}$, see Grabner & Moers (2013)). Irrespective of which empirical test will be performed, interdependence implies that $\beta_{12} \neq 0$ and this is what needs to be theoretically substantiated. In this paper, we focus on specifications to test the hypothesis that $\beta_{12} \neq 0$.[1] An example of such a test is whether the effectiveness of incentive contracts depends on the level of delegation for managers (Moers, 2006; Indjejikian & Matejka, 2012). The parameters γ_1 and γ_2 represent the contingency effect of environmental uncertainty, z, for delegation, x_1, and accounting incentives, x_2. For instance, $\gamma_1 > 0$ implies that delegation is more valuable for the firm with higher environmental uncertainty while $\gamma_2 < 0$ implies that accounting incentives are less valuable with higher environmental uncertainty (Chenhall, 2003). δ_1 and δ_2, which we assume to be positive, represent increasing marginal costs to the practices.[2]

[1] Initially, we limit the paper to two-way complementarities for two reasons. First, theory in management control typically does not predict higher order interdependencies (for an exception outside of management accounting see Aral et al. (2012)). Second, the paper's main focus is on the consequences of deviations from completely optimal and completely random practices for hypothesis tests. The problems we identify apply to more complex hypothesis tests for higher order interdependencies. For simplicity of exposition, we avoid the additional complications of testing for higher order interdependencies (Carree et al., 2011). In section 4.5, we consider contingency effects on the interdependencies (Grabner, 2014; Grabner & Speckbacher, 2016; Matejka & Ray, 2017) and in section 4.4, we investigate an objective function with three practices and two-way interdependencies (Indjejikian & Matejka, 2012).

[2] δ_1 and δ_2 are the second derivatives of the performance effects of the practices. In general, $\delta > 0$ is a mathematical representation of decreasing marginal returns, increasing marginal

The objective function (1) formalises management accounting theory. The key assumptions of complementarity theory and contingency theory are captured by the parameters β_{12} and γ_1, γ_2 respectively. In order to test for a complementarity between delegation and accounting incentives, we need to make additional assumptions about the statistical model, which we discuss below. The theoretical assumptions discussed above and the statistical assumptions discussed below are the same for both the demand specification and the performance specification.

We assume that ϵ_1 and ϵ_2 and ν are independent which implies that the model contains all contingency variables and all practices that affect the performance of the two practices x_1 and x_2. The objective function (1) is similar to the the objective function in Kretschmer et al. (2012) with two exceptions. First, we assume that the values of x_1 and x_2 are continuous while Kretschmer et al. (2012) allow for binary practices. We return to the issue of binary practices as an extension to the current model in section 4.2.3 and 4.4. Second, we assume independence of the unobserved factors. We will relax this assumption, when we introduce the problem of correlated omitted variables further in section 4.3.[3] In the remainder of this section, we will work with the simple objective function with two practices and one environmental factor.

2.1. Optimal level of practices

Profit maximising firms try to adopt the optimal level for each practice. As a benchmark, we derive the optimal level for all practices by setting the first derivative of the objective function (1) to each practice equal to 0 and then solve for each practice. This results in:

---

costs, or a combination of both. For simplicity and in line with the specification in Grabner & Moers (2013), we interpret δ as the parameter of increasing marginal costs.

[3]Athey & Stern (1998) discuss the implications of relaxing the independence assumption in more detail.

$$
\begin{aligned}
x_1^* &= \frac{\beta_1 + \gamma_1 z + \beta_{12} x_2^* + \epsilon_1}{\delta_1} = \frac{\delta_2(\beta_1 + \gamma_1 z + \epsilon_1) + \beta_{12}(\beta_2 + \gamma_2 z + \epsilon_2)}{\delta_1 \delta_2 - \beta_{12}^2} \\
x_2^* &= \frac{\beta_2 + \gamma_2 z + \beta_{12} x_1^* + \epsilon_2}{\delta_2} = \frac{\delta_1(\beta_2 + \gamma_2 z + \epsilon_2) + \beta_{12}(\beta_1 + \gamma_1 z + \epsilon_1)}{\delta_1 \delta_2 - \beta_{12}^2}
\end{aligned} \tag{2}
$$

A number of conclusions can be drawn from the optimality conditions (2). First, in the absence of complementarity ($\beta_{12} = 0$), the optimal level of delegation, x_1^* and the optimal level of accounting incentives, x_2^* are still correlated. The contingency effects of environmental uncertainty introduce a relation between the optimal level of delegation, x_1^*, the optimal level of accounting incentives, x_2^* and environmental uncertainty, z. Second, in the absence of a complementarity ($\beta_{12} = 0$) and after controlling for environmental uncertainty, there is no relation between the optimal level of delegation and accounting incentives, i.e. the conditional correlation, $cor(x_1^*|z, x_2^*|z) = 0$. In contrast, in the presence of a complementarity, $\beta_{12} > 0$, the optimal level of delegation positively correlates with the optimal level of accounting incentives after controlling for z (Arora, 1996).

In what follows, we will refer to the observed correlations between the practices and the environmental variable. We define r_{12}, r_{1z}, and r_{2z} as the observed sample correlations between delegation, x_1, and accounting incentives, x_2, between delegation, x_1, and environmental uncertainty, z, and between accounting incentives, x_2, and environmental uncertainty, z, respectively. In samples with higher levels of optimality, the observed correlations are determined by the equations (2). In samples where management control practices are randomly chosen, the correlations are 0.

Finally, the second order condition for the optimality conditions (2) equals $\delta_1 \delta_2 - \beta_{12}^2 > 0$. The intuition behind this condition is that the increase in marginal costs to delegation and accounting incentives needs to be relatively large so that the interdependency does not dominate the optimal solution. That is, it avoids corner solutions where the optimal use of the management control practices is to use them to their full extent or not at all.[4]

2.2. Performance Specification

The performance specification estimates the objective function (1) directly. In essence, the stochastic form of function (1) is a regression equation. With cross-sectional data, the closest approximation to the objective function is the first of the following three regression models.

$$y = \beta_0^{p1} + (\beta_1^{p1} + \gamma_1^{p1} z)x_1 + (\beta_2^{p1} + \gamma_2 z)x_2 + \beta_{12}^{p1} x_1 x_2 + \delta_1^{p1} x_1^2 + \delta_2^{p1} x_2^2 + \alpha^{p1} z + \nu^{p1}$$
$$y = \beta_0^{p2} + (\beta_1^{p2} + \gamma_1^{p2} z)x_1 + (\beta_2^{p2} + \gamma_2 z)x_2 + \beta_{12}^{p2} x_1 x_2 + \alpha^{p2} z + \nu^{p2}$$
$$y = \beta_0^{p3} + \beta_1^{p3} x_1 + \beta_2^{p3} x_2 + \beta_{12}^{p3} x_1 x_2 + \alpha^{p3} z + \nu^{p3}$$

The first specification captures all features of the objective function (1) with the exception of the unobservable contingency effects on delegation and accounting incentives, ϵ_1 and ϵ_2. Cross-sectional data does not permit estimating the unobserved heterogeneity. In this specification, β_{12}^{p1} tests for the interdependency between the management accounting practices. We are not aware of any accounting studies using this specification, which we label the *performance 1* specification.

The second specification, which we label the *performance 2* specification, is the correct specification for binary practices, i.e., when practices are either absent or present. In this case, the quadratic terms, x_i^2, automatically drop out of the equation. While investigating continuous practices, Bedford et al. (2016, 2019) follow this specification as they control appropriately for contingency factors and drop the quadratic terms in the specification. The majority of the literature follows the third performance specification, which also drops the controls for the contingency factors or assumes that $\gamma_1 = \gamma_1 = 0$. This specification thus ignores the insights from contingency theory. We call this

[4] In the main analysis of this paper, we will assume that the second-order condition holds when the control practices are continuous. In Section 4.4 we relax this assumption.

specification *performance 3*. In the remainder of this section, we explain the potential problems with these three specifications.

2.2.1. Lack of Power

All three specifications will suffer form the well known problem that performance is no longer a function of the management practices when firms optimally adopt interdependent practices (Grabner & Moers, 2013). This can be illustrated using objective function (1) and the optimality conditions (2). When we set $x_1 = x_1^*$ and $x_2 = x_2^*$ in the objective function (1), i.e., the firm makes optimal decisions, we find that profit, y, is fully determined by the unknown parameters and environmental uncertainty, z, and profit no longer depends on the value of the practices. Thus, *observed* profit is not a function of *observed* delegation and *observed* accounting incentives when all firms adopt the optimal system. With optimal levels of delegation and accounting incentives, business unit profit is only a function of environmental uncertainty and there is no longer information about the practices in the performance variable. This is what economists are referring to when they say that one cannot examine performance effects of choices. The point is not that there are no performance effects of adopting management control practices, but rather that such effects cannot be empirically detected.

The loss of power of the performance specification depends on the level of optimality in the sample. A higher level of optimality leads to stronger correlations between the control practices and the contingency factor, and less independent information in the *observed* practices about the *observed* profit. For a performance specification to have sufficient power, there needs to be a sufficient number of firms that do not adopt the optimal level of the practices (Bedford et al., 2016; Carree et al., 2011; Hofmann & van Lent, 2017). However, the literature has not gone beyond this rule of thumb and provides no guidance on how large the deviations from optimality need to be. Our simulation study will provide an answer to this question.

2.2.2. Correlated Omitted Variable

In this paper, we identify a second problem with the most popular specification in the literature, *performance 3*, which follows directly from the contingency effects. The specification omits the terms x_1z, x_2z, x_1^2, and x_2^2. Although a full treatment of the omitted variable bias is beyond the scope of this study, we illustrate the problem for the special case where x_1, x_2, and z follow a multivariate standard normal distribution. Appendix Appendix A shows that the bias for omitting x_1z, x_2z, x_1^2, and x_2^2 in the estimate β_{12}^{p3} is proportional to the following components:

$$\begin{array}{ll} \gamma_1 \dfrac{cov(x_1x_2, x_1x_2)}{var(x_1x_2)} = \gamma_1 \dfrac{r_{2z} + r_{12}r_{1z}}{1 + r_{12}^2} & \gamma_2 \dfrac{cov(x_1x_2, x_2z)}{var(x_1x_2)} = \gamma_2 \dfrac{r_{1z} + r_{12}r_{2z}}{1 + r_{12}^2} \\ \delta_1 \dfrac{cov(x_1x_2, x_1^2)}{var(x_1x_2)} = \delta_1 \dfrac{r_{12}}{1 + r_{12}^2} & \delta_2 \dfrac{cov(x_1x_2, x_2^2)}{var(x_2x_2)} = \delta_2 \dfrac{r_{12}}{1 + r_{12}^2} \end{array} \tag{3}$$

where r_{12}, r_{1z}, and r_{2z} are the observed sample correlations, as defined before. Let us assume that there is no interdependency, i.e., $\beta_{12} = 0$. The estimate β_{12}^{p3} , which equals $\beta_{12} + bias$, is then affected by the four components specified in (3). $\beta_{12}^{p3} = \beta_{12} = 0$ when the observed correlations r_{12}, r_{1z}, and r_{2z} are all zero, or in other words, when firms just randomly pick their management control practices. However, as explained in the previous section, contingency theory implies that those empirical correlations between delegation, accounting incentives, and environmental uncertainty are different from 0 when firms are not completely ignorant of the optimal levels. As a result, when environmental uncertainty is not appropriately controlled for, the *performance 3* specification is vulnerable to an omitted variable bias when testing for a complementarity between delegation and accounting incentives. The intuition behind the bias is as follows. If delegation, x_1, and environmental uncertainty, z, are correlated, the interaction between delegation and accounting incentives, x_1x_2, and the interaction between accounting incentives and environmental uncertainty, x_2z, are also correlated. As a result, when the contingency effect, x_2z, is omitted from the performance specification, the complementarity test, x_1x_2, will be confounded by the contingency effect, even when $\beta_{12} = 0$. That is, a Type I error occurs. Note that this problem extends to the *performance 1* and *performance 2* specifications when factors unobservable to the researcher affect both practices, which we will address later.

Because the bias does not require perfectly optimal decisions and follows directly from the contingency effect of environmental uncertainty on delegation ($\gamma_1 \neq 0$) and accounting incentives ($\gamma_2 \neq 0$), this problem cannot be easily ignored in a typical management accounting study. A special case of the omitted variable bias is the omission of the quadratic terms $\delta_1 x_1^2$ and $\delta_2 x_2^2$ in *performance 2*. Omission of these terms will bias the estimate of the β_{12}^{p2} with a factor proportional to the correlation between delegation and incentives.

The empirical and methodology literature has focused on this bias in the demand specification (see also section 2.3) but has largely ignored the problem of correlated omitted environmental factors in performance specifications. However, in the above, we showed that the performance specifications suffer from the exact same bias in contrast to what has been argued in the methodology literature (Carree et al., 2011).

2.3. Demand Specification

The *demand specification* can take two forms, the regression approach and the conditional correlation approach. The first approach regresses one practice (e.g. delegation) on the other (e.g. accounting incentives) and controls for environmental uncertainty.

$$x_1 = \beta_1^d + \beta_{12}^d x_2 + \gamma_1^d z + \epsilon^d$$

The regression specification approximates the optimality condition (2) where β_{12}^d is the parameter that estimates the complementarity effect. An alternative and equivalent approach is to estimate the conditional correlation between delegation and incentives. Prior research has used seemingly unrelated regressions with x_1 and x_2 as dependent variables or seperate regressions to condition on environmental practices (Indjejikian & Matejka, 2012; Matejka & Ray, 2017).[5] In the remainder of this section, we will explain the problems with the demand specification from the perspective of the regression approach. Those problems equally apply to the conditional correlation approach.

2.3.1. Lack of Power

When $\beta_{12} \neq 0$ but firms do not take these interdependencies and the contingency effects into account, the empirical correlations are expected to be 0 and so is the regression estimate β_{12}^d. In the remainder of this study, we consider this an unreasonable assumption unless researchers can identify a natural experiment for the adoption of the practices, x_1 and x_2. In the simulation study, we will investigate what happens when we vary the extent to which firms adopt the optimal level of delegation and accounting incentives. When firms' accounting systems deviate strongly from the optimal accounting system, the empirical correlations will be small and the demand specification will lack power to detect a real interdependency.

2.3.2. Correlated Omitted Variable

The omitted variable bias for the demand function is a well known problem (Arora, 1996; Grabner & Moers, 2013; Hofmann & van Lent, 2017). When testing for the complementarity between delegation and accounting incentives without controlling for environmental uncertainty, the estimate of the interdependency, β_{12}^d, will be biased by $\gamma_1^d \frac{cov(\mathbf{z},\mathbf{x_2})}{var(\mathbf{x_2})}$. For instance, if environmental uncertainty is positively associated with delegation ($\gamma_1^d > 0$) and negatively with accounting incentives ($cov(\mathbf{z},\mathbf{x_2}) < 0$), β_{12}^d might be negative even in the presence of a complementarity between delegation and accounting incentives, i.e., $\beta_{12} > 0$. Hence, the demand specification needs to control for environmen-

[5] The equivalence between the regression and the conditional correlation follows from the regression anatomy (Angrist & Pischke, 2008). That is, $\beta_{12}^d = \frac{cov(x_1,x_2|z)}{var(x_2|z)} = cor(x_1,x_2|z)\frac{stdev(x_1)}{stdev(x_2|z)}$. The semi-partial correlation is directly related to the partial or conditional correlation: $\sqrt{1-cor^2(x_1,z)}cor(x_1,x_2|z) = cor(x_1|z,x_2|z)$ Thus, β_{12}^d is proportional to the conditional correlation $cor(x_1|z,x_2|z)$.

tal uncertainty. Similarly, a researcher has to control for all other contingency factors that effect both delegation and incentives.

2.4. Summary

A researcher's decision to use a performance specification or a demand specification is often framed as a trade-off between the lack of power of the demand specification when the practices are far from the optimal levels and the lack of power of the performance specification when the practices are close to the optimal levels (Grabner & Moers, 2013; Aral et al., 2012; Johansson, 2018). While our analysis confirms the existence of this trade-off, one key question that remains, is *at what levels of optimality* one method dominates the other one. For example, how far from the optimal levels should firm choices be before the performance specification has more power than the demand specification and is thus preferred?

An additional decision rule by researchers seems to be the correlated omitted variable problem. The correlated omitted variable bias of the demand specification is a well known problem (Grabner & Moers, 2013; Arora, 1996; Carree et al., 2011), which would suggest a preference for the performance specification when such omitted variables are expected to be present. However, our analysis above highlights the often ignored omitted variable bias in the performance specification. We show that both specifications are vulnerable to the same omitted variable bias. That is, both specifications will be biased if they do not control appropriately for an environmental factor that is a contingency factor for both practices of interest and when firms (to some extent) take into account these contingencies when they design their accounting system. A second question is therefore: *how vulnerable* is each method to correlated omitted variables? To address these two questions, we perform a simulation study.

3. Simulation Study

In the simulation study that follows, we investigate the three performance specifications and the demand specification with respect to: (1) Type I errors, i.e., incorrectly rejecting a true null-hypothesis; and (2) power, i.e., the ability to reject a false null-hypothesis. Our main analysis focuses on the key question of how the level of optimality affects Type I errors and power of the theoretically appropriate *demand* and *performance 1* specifications. We further compare these specifications to the *performance 2* and *performance 3* specification that researchers have used to test for a complementarity between management accounting practices.

Following the main analysis, we test the robustness of the findings to variations in the parameters of the objective function (1). In addition, we investigate to what extent the demand specification is vulnerable to a violation of the second order optimality condition and whether the properties of the specifications change with binary practices. Lastly, we investigate to what extent the theoretically appropriate *demand* and *performance 1* specification are vulnerable to correlated omitted variables.

3.1. Simulation Algorithm

In this section, we describe the simulation algorithm and calibrate the parameters to the observed correlations in nine management accounting studies that test for an interdependency between management control practices and cite Grabner & Moers (2013). The simulation algorithm is based on the objective function (1) for a single firm from the formal analysis. For completeness, we reproduce it here:

$$y = \beta_0 + (\beta_1 + \gamma_1 z + \epsilon_1)x_1 + (\beta_2 + \gamma_2 z + \epsilon_2)x_2 + \beta_{12}x_1x_2 - \frac{1}{2}\delta_1 x_1^2 - \frac{1}{2}\delta_2 x_2^2 + \nu$$

The structural parameters, β, γ, and δ are the same for each simulated firm in a sample, which allows for a clean analysis of our questions. For the baseline analysis, we set $\beta_0, \beta_1, \beta_2$ equal to 0 because these parameters do not interact with the effect of the interdependency. Keeping $\beta_1 = \beta_2 = 0$ also ensures that the mean optimal level of the practices is in the middle of the distribution that generates the values of the practices, and minimises any ceiling or floor biases in the algorithm. The contingency effects are initially set at $\gamma_1 = 0.33$ and $\gamma_2 = 0$.

We will compare samples where the contingency effect on the second practise varies from being absent ($\gamma_2 = 0$) to being positive ($\gamma_2 = 0.33$) or negative ($\gamma_2 = -0.33$).

The values for z, ϵ_1, and ϵ_2 are different for each firm and are simulated from a normal distribution with mean 0 and standard deviations 1, σ_{ϵ_1}, and σ_{ϵ_2} respectively. The δ_i's and σ_{ϵ_i}'s determine the scale and unobserved heterogeneity of the effects, respectively, and interact with the interdependency. In the baseline analysis, we set these parameters equal to 1 and vary them in follow-up simulations to investigate the robustness of the baseline analysis. Finally, ν is normally distributed with standard deviation, σ_ν. The baseline objective function in the simulation is thus:

$$y = (0.33z + \epsilon_1)x_1 + \epsilon_2 x_2 + \beta_{12}x_1x_2 - \frac{1}{2}x_1^2 - \frac{1}{2}x_2^2 + \nu \tag{4}$$

Note that in this baseline objective function, there is no omitted variable bias in the four specifications with respect to environmental factors. Next, we show how the baseline scenario is a good reflection of a typical study in the management accounting survey literature. To investigate whether the different specifications have the power to detect a true effect while maintaining nominal Type I error rates, we will compare samples with a complementarity effect ($\beta_{12} = .25$) to samples without a complementarity effect ($\beta_{12} = 0$).

To examine how the level of optimality affects Type I errors and power, we vary the level of optimality, O. The simulation algorithm mimics the process where firms experiment with different combination of delegation, x_1, and accounting incentives, x_2, and keep the configuration that results in the highest business unit profit y. When O is larger, a firm has experimented with more configurations and therefore the probability that the firm selects the optimal configuration of accounting practices is higher.

The full procedure for one observation is described in mathematical form in algorithm (5). In order to generate a sample of 300 observations, the algorithm is run 300 times.

$$
\begin{aligned}
& z \sim \mathcal{N}(0,1); \epsilon_i \ \sim \mathcal{N}(0, \sigma_{\epsilon_i}) \\
\forall 0 < o \leq O : & x_i^o \sim U[-5, 5] \\
\forall 0 < o \leq O : & \hat{y^o} = \beta_0 + (\beta_1 + \gamma_1 z + \epsilon_1) x_1^o + (\beta_2 + \gamma_2 z + \epsilon_2) x_2^o + \beta_{12} x_1^o x_2^o - \frac{1}{2}\delta_1 {x^o}_1^2 - \frac{1}{2}\delta_2 {x^o}_2^2 \\
\forall 0 < o \leq O : & y^o \sim \mathcal{N}(\hat{y^o}, \sigma_\nu) \\
& y^{max} = max(y^1, y^2, ..., y^O) \\
& \text{if } y^o = y^{max} \text{then } y = y^o, x_i = x_i^o
\end{aligned} \quad (5)
$$

For each firm, the algorithm generates values for x_1^o, x_2^o from two independent uniform distributions between -5 and 5. The range allows the randomly generated values for the practices to be far from the optimal level in the baseline analysis. Next, the algorithm calculates the performance, y^o, according to the objective function (1). For each firm, we repeat this process O times and keep the values of the repetition for which y^o is the highest as the observation in the sample. As a result, the more tries a firm has with different accounting systems, i.e., the greater O, the more likely it is they will adopt the optimal combination of the two practices. The parameter O is the same for each observation in a sample and is the key parameter capturing the probability of how close the firms in the sample are to the optimal level of the practices.

In Figure 1 we plot a sample of 300 observations of $\mathbf{x_1}$ and $\mathbf{x_2}$ for 6 levels of optimality, $O = 2, 4, 8, 16, 32, 64$ for the baseline scenario with a complementarity ($\beta_{12} = 0.25$). The figure shows the within sample heterogeneity of the practices in a simulated sample. If we take the sample with $O = 64$ as close to the optimal distribution, we see that there are always a number of observations outside the optimal distribution with lower levels of optimality. Nevertheless, there are also observations that are closer to the optimal level for each level of optimality. That is, increases in the level of optimality increase the probability that an observation will be closer to optimal which also shows in the stronger positive relation between $\mathbf{x_1}$ and $\mathbf{x_2}$ for higher levels of optimality.

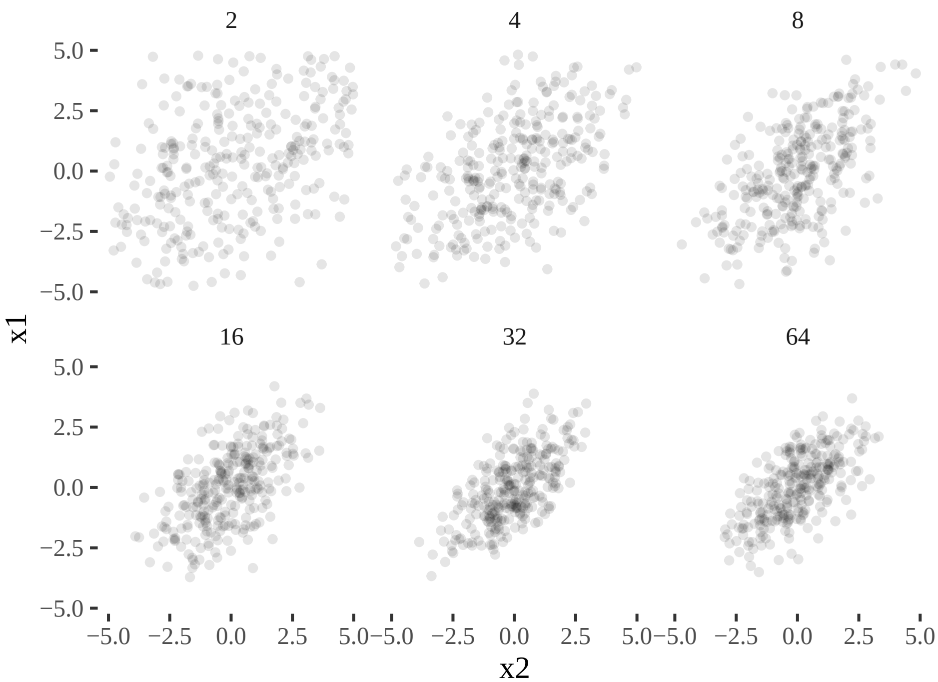

Figure 1: The figure shows a scatterplot of the distribution of 300 observations of x_1 and x_2 for different levels of optimality, $O = 2, 4, 8, 16, 32, 64$. The complementarity effect is present ($\beta_{12} = 0.25$). The decreasing marginal costs are set as $\delta_1 = \delta_2 = 1$. The effect of the environmental variable only affects one of the choices ($\gamma_1 = .33, \gamma_2 = 0$). The unobserved variation parameters are set at $\sigma_{\epsilon_1} = \sigma_{\epsilon_2} = \sigma_\nu = 1$.

3.2. Calibration to Empirical Studies

In order to calibrate the simulation, we collect the correlations of published survey studies that cite Grabner & Moers (2013) and have an interdependency hypothesis. The median number of observations in the studies is 250 and the average is 265. To give the methods the benefit of the doubt, the simulation will create samples of 300 firm observations where each observation follows the algorithm above. Next, we collect the correlation matrix from 9 studies that report full correlation matrices with correlations between practices, and between practices and environmental factors (Dekker et al., 2016; Grabner, 2014; Grabner & Speckbacher, 2016; Bedford & Malmi, 2015; Heinicke et al., 2016; Bedford et al., 2019; Abernethy et al., 2015; Sponem & Lambert, 2016; Samagaio et al., 2018). To get an estimate of observed correlations in empirical studies, we calculate the median and 90th percentile absolute correlation between practices and between a practice and a contingency factor. We consider the median value of those statistics to be typical for studies on interdependencies in management accounting. The median absolute correlation between two practices is 0.22 and the median 90th percentile correlation is 0.39. The median absolute correlation between a practice and an environmental factor is 0.16 and the median 90th percentile correlation is 0.30

To illustrate the effect of the optimality parameter and the strength of the associations generated by the simulation algorithm, we generate 100 samples with each 300 observations under the baseline scenario outlined above for 6 levels of optimality: $O \in \{2, 4, 8, 16, 32, 64\}$. For each sample, we calculate three statistics to illustrate that the simulated samples reflect a typical management accounting study. We plot the statistics for each simulated sample in Figure 2. The first statistic is the correlation between x_1 and x_2 (Panel A) and the second is the correlation between x_1 and z (Panel B). The absolute correlation between x_1 and x_2 varies between 0 and .5 and is smaller than the median 90th percentile (.39) in a typical study except at the higher levels of optimality. Similarly the correlation between x_1 and z varies between 0 and .4 and is smaller than the median 90th percentile (.30) in a typical study except at the higher levels of optimality. Hence, we conclude that the associations in our simulations are representative for the management accounting literature.

The third statistic is the non-optimality ratio and quantifies the extent to which deviations from the optimal level $x_1^*|z$ cannot be explained by the optimality conditions (2).[6] The non-optimality ratio measures how much of the variation in the practices is caused by the algorithmic search process and how far the practices in a sample are from the optimal levels. The resulting ratio is comparable to an R^2 statistic because the non-optimality ratio equals 1 when the difference between x_1 and x_1^* can be fully explained by the absence of optimality in a sample, and the ratio equals 0 when $x_1 = x_1^*$. Figure 2 Panel C shows the ratio for the 100 samples for different levels of the optimality parameter O. The ratio declines for higher values of the optimality parameter and for all values of O at least 20% of the variation cannot be explained by the optimality condition. In other words, the generated samples always have substantial deviations from optimal practices, which should give performance specifications a chance of detecting performance differences between optimal and sub-optimal accounting systems.

3.3. Power and Type I Error

In the next section, we will compare the power and Type I error rate of the four specifications when varying the optimality parameter, O. Because the accounting literature is concerned with testing the hypothesis that there is a (no) complementarity between two management control practices, a focus on power and error rates is appropriate. For completeness, we repeat the four

[6] The total deviations from the optimal level conditional on the contingency factor is calculated as the sum of $(x_1 - E(x_1^*|z))^2$ for each observation in the sample, where $E(x_1^*|z)$ is given by the optimality condition (2). The amount of variation explained by the unobserved factors, ϵ_1 and ϵ_2, can also be calculated from the optimality conditions (2), i.e., $Var(x_1^*|z) = \frac{\delta_2^2 \sigma_{\epsilon_1}^2 + \beta_{12}^2 \sigma_{\epsilon_1}^2}{(\delta_1 \delta_2 - \beta_{12}^2)^2}$ The non-optimality ratio is then calculated as the ratio between the deviations from optimality not explained by the unobserved factors and the total deviations from optimality.

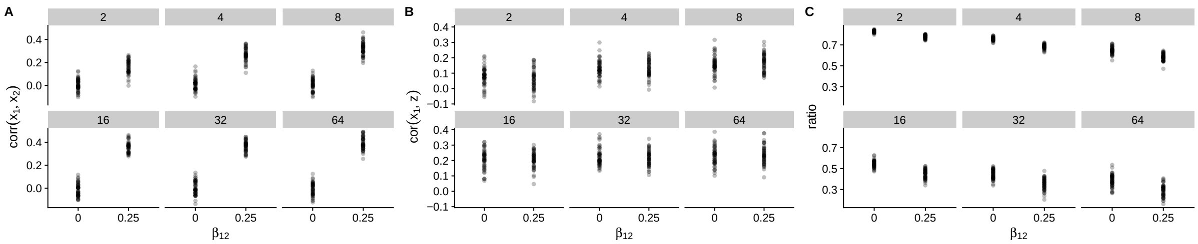

Figure 2: The figure shows the value of the correlation between $\mathbf{x_1}$ and $\mathbf{x_2}$ (Panel A), the correlation between $\mathbf{x_1}$ and $\mathbf{z}$ (Panel B), and the non-optimality ratio (Panel C), for 100 samples for 6 different levels (2, 4, 8, 16, 32, 64) of the optimality parameter, O. The complementarity effect is either present ($\beta_{12} = 0.25$) or absent ($\beta_{12} = 0$). The decreasing marginal costs are set as $\delta_1 = \delta_2 = 1$. The effect of the environmental variable only affects one of the choices ($\gamma_1 = .33, \gamma_2 = 0$). The unobserved variation parameters are set at $\sigma_{\epsilon_1} = \sigma_{\epsilon_2} = \sigma_\nu = 1$.

specifications here

$$x_1 = \beta_1^d + \beta_{12}^d x_2 + \gamma_1^d z + \epsilon^d$$

$$y = \beta_0^{p1} + (\beta_1^{p1} + \gamma_1^{p1} z)x_1 + (\beta_2^{p1} + \gamma_2 z)x_2 + \beta_{12}^{p1} x_1 x_2 + \delta_1^{p1} x_1^2 + \delta_2^{p1} x_2^2 + \alpha^{p1} z + \nu^{p1}$$

$$y = \beta_0^{p2} + (\beta_1^{p2} + \gamma_1^{p2} z)x_1 + (\beta_2^{p2} + \gamma_2 z)x_2 + \beta_{12}^{p2} x_1 x_2 + \alpha^{p2} z + \nu^{p2}$$

$$y = \beta_0^{p3} + \beta_1^{p3} x_1 + \beta_2^{p3} x_2 + \beta_{12}^{p3} x_1 x_2 + \alpha^{p3} z + \nu^{p3}$$

The β_{12} coefficients for each specification provide the test for the presence of an interdependency. The simulation generates 1000 samples for each combination of parameters. For each combination, we report the distribution of the t-statistic for the interdependency coefficient and compare it to the traditional cut-off value for the 5% level of significance ($|t| > 1.97$). We also calculate the power and Type I error rate of the specifications to investigate the performance of the four specifications in more detail. The power is the percentage of samples with a complementarity effect where the p-value is lower than 0.05 and the estimated coefficient is positive.[7] The Type I error is the percentage of samples without a

[7] An important caveat is that the power of a study will also be influenced by the size of the effect, measurement error, random variation, and the number of observations in the study. In the simulations, we assumed a fixed effect ($\beta_{12} = 0.25$), no measurement error, fixed the parameters that control random variation, σ_{ϵ_1}, σ_{ϵ_2}, and σ_ν, and the number of observations per sample. As a result, the absolute percentages in the results should be interpreted with

complementarity effect where the p-value is lower than 0.05 (irrespective of the sign of the estimated coefficient).

4. Results

4.1. Performance and Demand Specification

4.1.1. Power

Figure 3 Panel A shows a boxplot for the distribution of t-statistics for each type of test and each combination of parameters. The dot of the boxplot shows the median t-statistic of the 1000 samples, the gap between the whiskers shows the interquartile range, and the ends of the whiskers show the minimum and maximum t-statistic. Each boxplot can be compared to the zero line and the dotted lines representing a 95% confidence interval around a null effect. When $\beta_{12} = 0.25$, we expect the distribution of the t-statistic to be above the dotted lines because it shows that the test reliably reports a significant positive interdependency (power).

Figure 3 Panel A reveals the basic trade-off between the demand specification and the performance specification: with low levels of optimality the performance function is more likely to detect a true complementarity effect while the demand function is more likely to detect a true complementarity effect with high levels of optimality (Grabner & Moers, 2013; Aral et al., 2012; Johansson, 2018). The boxplots of the t-statistics are above the 95% confidence interval for the *performance 1* specification with lower levels of optimality and they are above the 95% confidence interval for the *demand* specification for higher levels of optimality.

Interestingly, even at relatively low levels of optimality, i.e., $O = 4$, the *demand* specification has similar power to the *performance 1* specification. Figure 2 shows that with $O = 4$ the optimality condition can at best explain 40% of the variation in the distance between the observed level of the practices and the optimal level of the practices. This implies that, as long as firms avoid the

caution. This study is mainly interested in the relative differences between specifications, focusing on a clean setting and a fair comparison.

worst possible combinations of management accounting practices, the demand specification is more likely to detect a true effect.

The *performance 2* specification without the quadratic terms fares worse than the *performance 1* specification at lower optimality levels. The boxplots fall almost entirely within the 95% confidence interval around 0 with low levels of optimality and a true effect. The omission of the quadratic terms decreases the ability of the *performance 2* specification to detect a real interdependency ($\beta_{12} = 0.25$) at lower levels of optimality. Surprisingly, the *performance 2* specification has more power to detect a real interdependency than the *performance 1* specification at higher levels of optimality. The counterintuitive reason for this is that the bias associated with the omission of the quadratic terms (see equation 3) inflates the estimate of the complementarity with a factor proportional to the correlation between x_1 and x_2. In effect, the bias in *performance 2* inadvertently picks up the same signal, i.e. the relation between the two practices, as the demand specification. Nevertheless, in terms of power, the *demand* specification dominates the *performance 2* specification. The *performance 3* specification performs as poorly as the *performance 2* specification, in terms of power.

To evaluate the demand and performance specifications in more detail, we report the power of the four specifications in Table 1. The results show that the "optimality trade-off" between the demand specification and the performance specification that is discussed in the literature (Grabner & Moers, 2013; Aral et al., 2012; Johansson, 2018) is really a second order problem. Both the *demand* and the *performance 1* specification are able to detect a true interdependency with more than 80% probability for low to medium levels of optimality. However, the *demand* specification is more likely to detect a true effect at higher levels of optimality. Strictly speaking, the *demand* specification is more likely to detect a true effect at all levels of optimality but the lowest level, i.e., at $O = 2$. From a power perspective, the *demand* specification dominates the *performance 1* specification. Finally, the *performance 2* and *performance 3* specifications are not able to detect the true interdependency for most of the parameter space

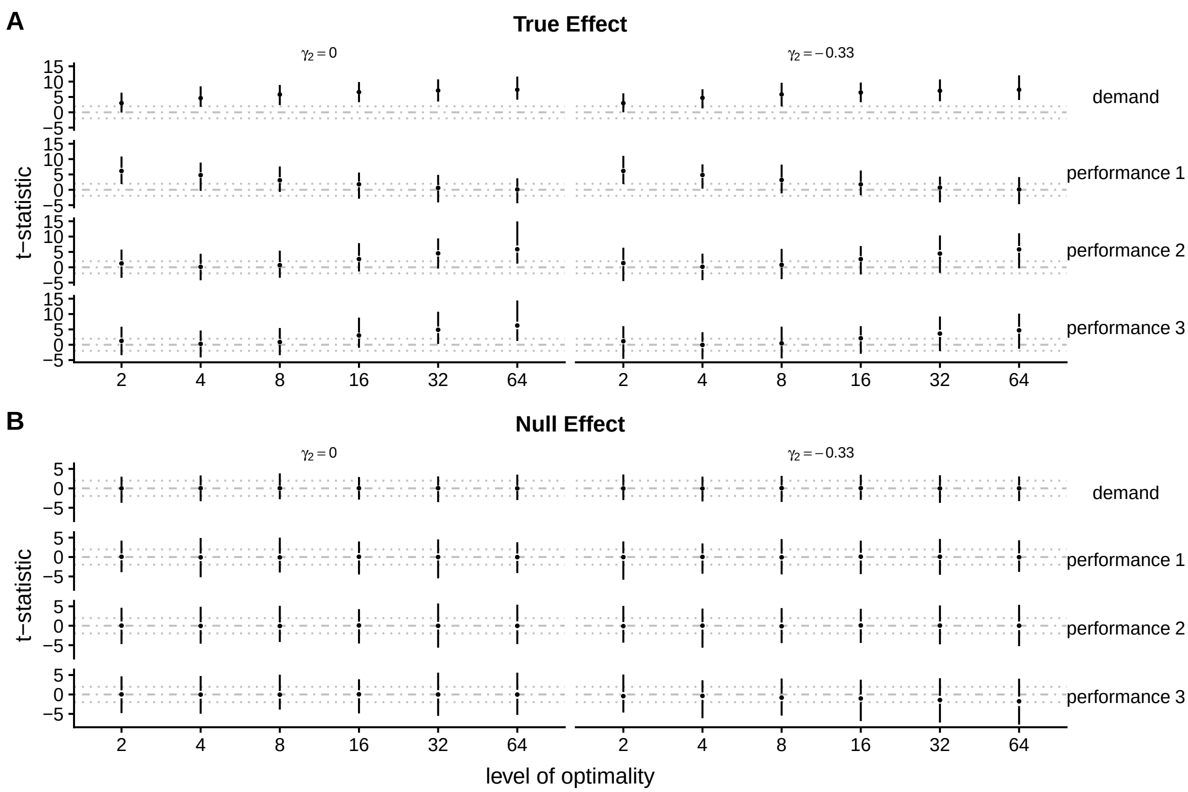

Figure 3: t-statistic of the performance and demand specification to test for complementarities when there is a complementarity effect ($\beta_{12} = .25$) in Panel A or a null effect ($\beta_{12} = 0$) in Panel B. The boxplots represent the median (the dot), the interquartile range (the gap), and the minimum and maximum (the whiskers). O is varied between 2, 4, 8, 16, 32, and 64. The effect of the environmental variable, $\mathbf{z}$, on the second choice is either absent ($\gamma_1 = .33, \gamma_2 = 0$), or negative ($\gamma_1 = .33$ and $\gamma_2 = -.33$).

Table 1: Power and Type I Error Rate for Baseline Simulation

specification	γ_2	Level of Optimality 2	4	8	16	32	64
Power							
demand	-0.33	0.86	1.00	1.00	1.00	1.00	1.00
demand	0.00	0.84	1.00	1.00	1.00	1.00	1.00
demand	0.33	0.84	1.00	1.00	1.00	1.00	1.00
performance 1	-0.33	1.00	0.98	0.82	0.44	0.16	0.05
performance 1	0.00	1.00	0.99	0.81	0.45	0.17	0.06
performance 1	0.33	1.00	0.98	0.83	0.44	0.16	0.06
performance 2	-0.33	0.34	0.11	0.19	0.69	0.95	0.99
performance 2	0.00	0.33	0.10	0.19	0.69	0.96	0.99
performance 2	0.33	0.37	0.12	0.20	0.70	0.96	1.00
performance 3	-0.33	0.28	0.07	0.14	0.55	0.86	0.95
performance 3	0.00	0.35	0.12	0.25	0.75	0.97	0.99
performance 3	0.33	0.45	0.25	0.47	0.94	1.00	1.00
Type I							
demand	-0.33	0.04	0.03	0.04	0.05	0.04	0.04
demand	0.00	0.04	0.06	0.03	0.04	0.04	0.05
demand	0.33	0.04	0.04	0.05	0.05	0.05	0.05
performance 1	-0.33	0.15	0.15	0.15	0.16	0.15	0.15
performance 1	0.00	0.15	0.17	0.16	0.17	0.14	0.14
performance 1	0.33	0.14	0.14	0.16	0.15	0.13	0.14
performance 2	-0.33	0.20	0.18	0.18	0.16	0.20	0.25
performance 2	0.00	0.20	0.20	0.17	0.17	0.19	0.22
performance 2	0.33	0.20	0.18	0.18	0.16	0.18	0.24
performance 3	-0.33	0.22	0.19	0.23	0.28	0.39	0.48
performance 3	0.00	0.18	0.20	0.16	0.17	0.22	0.25
performance 3	0.33	0.20	0.20	0.22	0.30	0.42	0.45

Note:
Type I error rates and power for the demand and performance specifications at different levels optimality: 2, 4, 8, 16, 32, 64. The parameters are the same as in Figure 3.

under investigation, except at high levels of optimality. As stated before, the power that these specifications have at higher levels of optimality is due to the bias from omitting the quadratic terms which depends on the correlation between x_1 and x_2. While this bias helps with increasing power, it comes at the cost of high Type I error rates, as we show next.

4.1.2. Type I error

Figure 3 Panel B shows a boxplot for the distribution of t-statistics for each type of test and each combination of parameters. When $\beta_{12} = 0$, we expect the distribution of the t-statistic to be centred on 0, as well as 95% of the distribution to be between the two dotted lines. When the distribution is not centred on 0, the test is biased. When the distribution is too wide, the test reports too many significant interdependencies in the absence of a true effect (Type I error).

Figure 3 Panel B shows that the demand specification, the theoretically appropriate *performance 1* and the *performance 2* specifications have an average t-statistic close to 0 in the absence of an interdependency. The boxplots are centred around the zero line for these three specifications. Recall that the omission of the quadratic terms does not bias the estimate of a complementarity in the absence of the complementarity ($\beta_{12} = 0$).

The *performance 3* specification fares worse than all other specifications. The omission of the interaction terms x_1z and x_2z biases the estimate of the interdependency when $\gamma_1\gamma_2 \neq 0$. The bias can be easily seen in Figure 3 Panel B. The distribution of t-statistics for samples without a complementarity effect no longer centres on 0 when the environmental factor has a contingency effect on the performance of both practices and the bias increases with higher levels of optimality. The bias is negative when γ_2 is negative and positive when γ_2 is positive. The latter case is not depicted in Figure 3.

To evaluate the Type I errors of the demand and performance specifications in more detail, we report the percentage of samples for which the estimate of β_{12} is significantly positive or negative in Table 1. Under the parameters in the simulation study, the *demand* specification has Type I error rates slightly below or equal to 0.05. This means that the number of false positives are consistent with the nominal p-value of 5%. The theoretically appropriate *performance 1* specification tends to have error rates elevated by a factor two to three relative to the *demand* specification. The most worrying results are for the *performance 2* and *3* specifications. Dropping the quadratic effects increases the error rates in the *performance 2* specification to around .20, four times the nominal error rate. The most commonly used specification in the literature, *performance 3*, has even higher Type I error rates that increase with higher levels of optimality and when the environmental factor has a contingency effect on both practices. The misspecification in *performance 3* leads to a bias in the estimated coefficient for the interdependency and thus the t-statistic. In conclusion, given the parameters in the simulation study only the *demand* specification rejects the null hypothesis at nominal 5% level of significance. The theoretically derived *performance 1* specification has elevated error rates, and the two other performance specifications are even more vulnerable to false positives.

4.1.3. Take-away of baseline simulation

The results of the simulation using the baseline model reveal the following: In contrast to arguments put forward in the literature, the *demand* specification has significant power at all levels of optimality. The assumption regarding the level of optimality to decide between specifications is a second-order consideration. Unless the researcher can argue that the sample has quasi-random assignment of practices, the demand specification should be preferred. In addition, the Type I error rates of the *demand* specification are consistent with the nominal p-value of 5%. The theoretically appropriate *performance 1* specification has significant power at low to medium levels of optimality, which is in line with arguments in the literature. However, the problem with this specification is the elevated Type I errors. These Type I errors only get worse once the performance specification is misspecified, as in *performance 2* and *performance 3*.

One way to interpret the elevated Type I errors of the performance specifications is to think about how a reader should react when observing a study with a significant positive interaction x_1x_2. To demonstrate the problems with the *performance 3* specification, we provide one dramatic example for $O = 4$ and $\gamma_2 = -0.33$. If we assume that a priori, we are indifferent between a null effect and a true interdependency of $\beta_{12} = 0.25$, and we observe one study that reports a significant positive interaction x_1x_2, the study is more likely to be from a sample where the null holds than from a sample where there is a true interdependency!

In sum, the results regarding power and Type I errors indicate that the *demand* specification performs better on both dimensions compared to all three performance specifications.

4.2. Parameter Variations

In this section, we explore the robustness of the above conclusions to variations in the parameters of the objective functions. Given the large number of possible variations, we restrict ourselves to theoretically driven comparisons. We do not consider the condition $O = 64$ in the following results because generating a sample of $O = 64$ is as computationally costly as generating a sample of each of the other levels of optimality. Furthermore, the results in Table 1 and Figure 3 show that the results do not qualitatively differ between $O = 32$ and $O = 64$.

4.2.1. Performance Variation

We first investigate whether an increase in the variance of performance, σ_ν, changes the above conclusions. This increase in variance has two possible effects. The first effect is a decrease in the importance of the management practices for performance, which decreases the power of the performance specification. The second effect follows from the first. When the management practices are less important, the level of optimality effects are less pronounced, and the correlations between x_1, x_2, and z are weaker, which in turn decreases the power of the demand specification and the omitted correlated variable bias in the *performance*

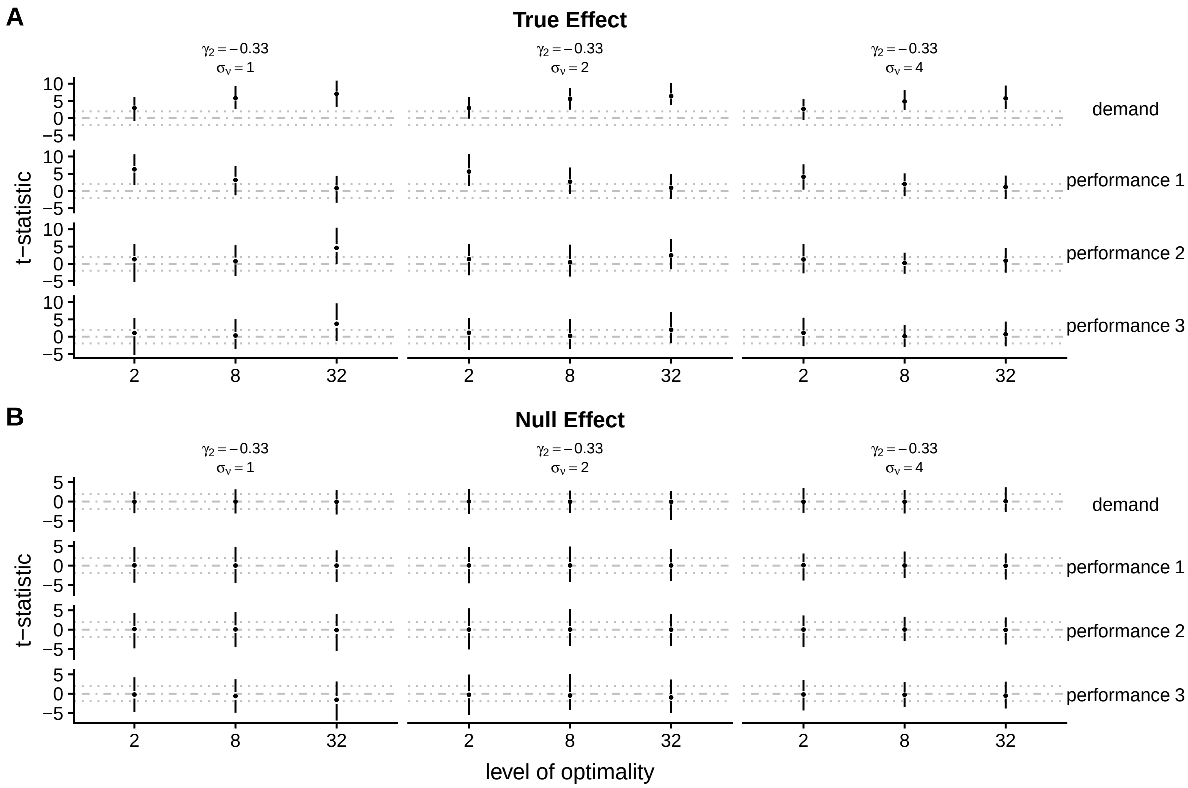

Figure 4: t-statistic of the performance and demand specification to test for complementarities when there is a complementarity effect ($\beta_{12} = .25$) in Panel A or a null effect ($\beta_{12} = 0$) in Panel B. The boxplots represent the median (the dot), the interquartile range (the gap), and the minimum and maximum (the whiskers). O is varied between 2, 8, 32. The effect of the environmental variable, $\mathbf{z}$, on the second choice is negative ($\gamma_1 = .0.33$ and $\gamma_2 = -.33$).

2 and *performance 3* specification.

To investigate the role of σ_ν, we vary the parameter between 1, 2, and 4 while keeping the other parameters the same as in Figure 3. For clarity of exposition, we limit the number of optimality variations ($O = 2, 8, 32$) and the number of variations of the contingency effect $x_2 z$ ($\gamma_2 = -.33$) in Figure 4.

The results in Figure 4 and Table 2 are qualitatively the same as the results in Figure 3 and Table 1. Surprisingly, in the presence of an interdependency, the increase in performance variance hardly affects the power of the *demand*

Table 2: Power and Type I Error Rate with Performance Variation

		Level of Optimality		
specification	γ_2	2	8	32
	Power			
demand	-0.33	0.75	1.00	1.00
demand	0.33	0.75	1.00	1.00
performance 1	-0.33	0.98	0.53	0.20
performance 1	0.33	0.97	0.49	0.17
performance 2	-0.33	0.30	0.06	0.16
performance 2	0.33	0.32	0.05	0.14
performance 3	-0.33	0.25	0.04	0.12
performance 3	0.33	0.39	0.11	0.37
	Type I			
demand	-0.33	0.04	0.05	0.04
demand	0.33	0.05	0.05	0.05
performance 1	-0.33	0.09	0.06	0.06
performance 1	0.33	0.08	0.06	0.07
performance 2	-0.33	0.12	0.07	0.07
performance 2	0.33	0.13	0.07	0.07
performance 3	-0.33	0.12	0.09	0.10
performance 3	0.33	0.14	0.08	0.11

Note:
Type I error rates and power for the demand and performance specifications at different levels optimality: 2, 8, 32. The parameters are the same as in Figure 4. Only the results for $\sigma_\nu = 4$ are reported.

specification. At all but the lowest level of optimality, the *demand* specification correctly identifies the interdependency for every simulated sample with a real interdependency, although the t-statistics decrease with higher performance variance. The drop-off in the t-statistic is steeper for the *performance 1* specification to the extent that power drops to around 20% when $O = 32$ and $\sigma_\nu = 4$. In summary, these results indicate that the impact of increasing the performance variance is a major decrease in the power of the performance specifications and only a minor decrease in the power of the demand specification.

Regarding Type I errors, we find that the *demand*, the *performance 1*, and the *performance 2* specification are unbiased. In addition, the *demand* specification has false positive rates below or close to the nominal rates, while the *performance 1* specification still has elevated Type I error rates. The *performance 2* and *performance 3* specification exhibit the same problem as before, namely elevated false positive rates as a result of misspecification. The increase in unobserved performance variation does lessen the impact of this bias. In summary, the conclusion that the demand specification outperforms the performance specifications is not changed when the noise in performance increases.

4.2.2. Marginal Costs

In this section, we vary the size of the increase in marginal costs, $\delta_1 = \delta_2$. We keep the parameters equal for both management control practices but they become smaller in size. There are two consequences of lowering the increase in marginal costs. First, decreasing $\delta_1 = \delta_2$ increases the importance of the complementarity between the management control practices. Second, the bias from omitting the quadratic terms is smaller. In the baseline scenario, we used $\delta_1 = \delta_2 = 1$. In this section, we compare the baseline scenario to two other values, i.e., 0.25 and 0. $\delta_1 = \delta_2 = .25$ is the largest value for which the parameters violate the second-order optimality condition $\beta_{12} < \sqrt{\delta_1 \delta_2}$. When the second-order optimality condition is violated, the optimal level for the practices will cluster towards 5 and -5 and away from 0. The extreme case is when $\delta_1 = \delta_2 = 0$ which may impede the inference by the *demand* specification even further.

Table 3: Power and Type I Error Rate and Marginal Costs

		Level of Optimality		
specification	δ_i	2	8	32
	Power			
demand	0.00	0.99	1.00	1.00
demand	0.25	1.00	1.00	1.00
demand	1.00	0.85	1.00	1.00
performance 1	0.00	1.00	0.72	0.01
performance 1	0.25	1.00	0.23	0.00
performance 1	1.00	1.00	0.83	0.16
	Type I			
demand	0.00	0.05	0.06	0.05
demand	0.25	0.06	0.05	0.05
demand	1.00	0.04	0.05	0.04
performance 1	0.00	0.11	0.09	0.07
performance 1	0.25	0.14	0.10	0.11
performance 1	1.00	0.15	0.16	0.14

Note:
Type I error rates and power for the *demand* and performance specifications at different levels optimality: 2, 8, 32. The parameters are the same as in Figure 5. The results are aggregated over the values for γ_2 (-0.33, 0.33).

Figure 5 and Table 3 report the power and Type I error rate of the different specifications for samples of firms with slowly increasing ($\delta_i = .25$) or fixed marginal costs ($\delta_i = 0$). Because the results showed no trends for different values of γ_2, we aggregated the results over all values of γ_2 in the table. The power of the *demand* specification is generally better for all but the lowest level of optimality, while the *performance 1* specification suffers from steep decreases in power with higher levels of optimality. The *demand* specification does suffer from slightly elevated false positives when the marginal costs of the choices are constant or only increasing slowly. However, the overall conclusions are unaffected.

4.2.3. Binary Practices

An alternative approach to deal with accounting practices with fixed or decreasing marginal costs is to treat them as binary decisions, because the optimal

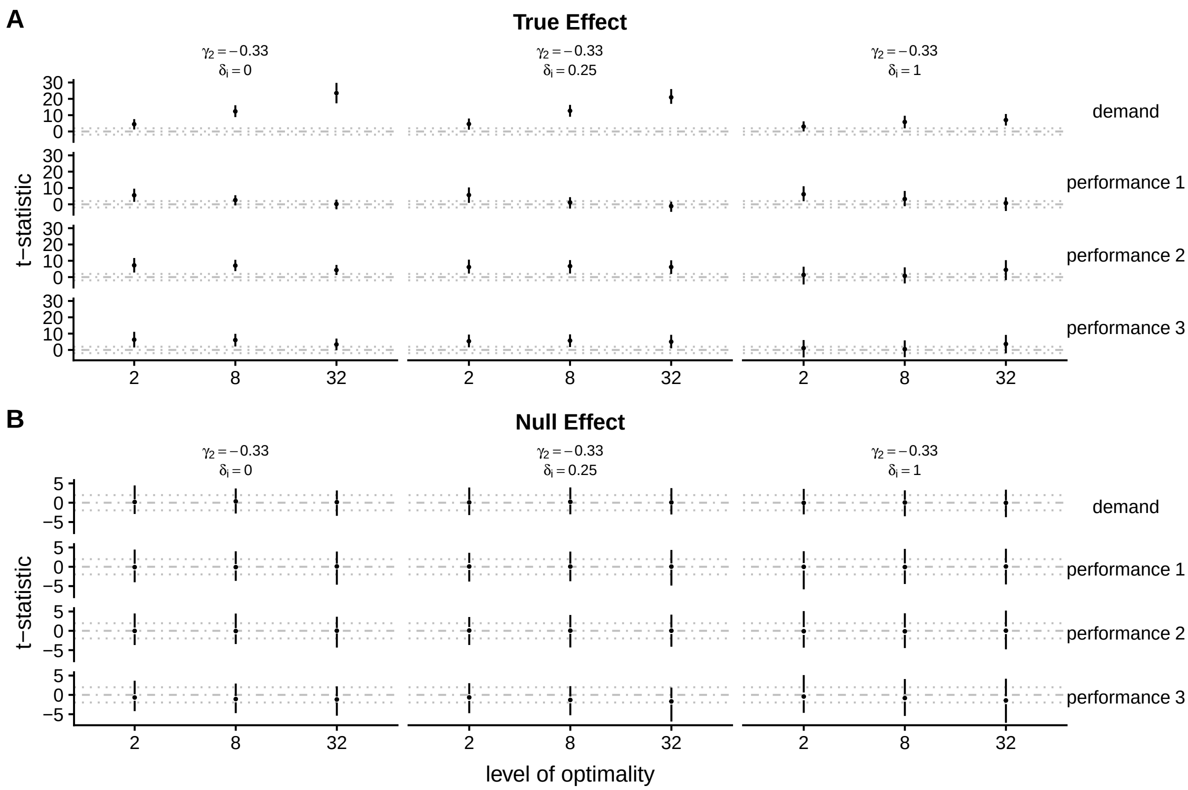

Figure 5: t-statistic of the performance and demand specification to test for complementarities when there is a complementarity effect ($\beta_{12} = .5$) in Panel A or a null effect ($\beta_{12} = 0$) in Panel B. The boxplots represent the median (the dot), the interquartile range (the gap), and the minimum and maximum (the whiskers). N is varied between 2, 8, 32. The effect of the environmental variable, $\mathbf{z}$, on the second choice is negative ($\gamma_1 = .33$ and $\gamma_2 = -.33$). The change in marginal costs varied between $\delta_1 = \delta_2 = 0, .25, 1$

decision will either be to use the practice to its full extent or not use it at all. When the accounting practices are binary, the quadratic terms in the *performance 1* specification are no longer identified. Hence, we will compare the *demand* with the *performance 2* specification. In the *demand* specification, we have to take into account that the dependent variable is now a binary outcome. Hence, we compare the linear probability model we have used so far, to the logit and probit estimates of the same specification. However, these alternative estimation methods should produce the same results in our setting.

We simulate new samples for the parameters in the main analysis with the following levels of optimality: 2, 4, 8, 16. The most important change in algorithm (5) is that we generate the accounting choices, x_i^o, no longer from a uniform distribution but let them be -1 or 1 with equal probability. We also set the parameters δ_1 and δ_2 equal to 0.

The results are shown in Table 4. Because the results again showed no differential trends for different values of γ_2, we aggregated the results over all values of γ_2. The power of the tests is lower for all tests compared to the same parameters in Table 1 for continuous accounting practices. As before, we find that at all but the lowest level of optimality, the *demand* specification has more power to detect a true effect, independent of the functional form used to estimate the complementarity. All specifications have similar and acceptable Type I error rates. In this specific case, the *performance 2* specification does not suffer from elevated Type I error rates. As before, with binary practices, we reach the same conclusion that the demand specification is superior to the performance specification as long as firms largely avoid accounting systems where the relation between the practices is the opposite of the optimal relation.

4.3. Correlated Omitted Variable

In this section, we investigate to what extent an omitted correlated environmental variable affects our conclusions for the *demand* and *performance 1* specification. The baseline simulation showed that the omission of a contingency factor that affects the performance of both practices, as in *performance*

Table 4: Power and Type I Error Rate with Discrete Practices

specification	Level of Optimality 2	4	8	16
Power				
demand	0.27	0.71	0.95	0.99
demand (logit)	0.27	0.71	0.95	0.99
demand (probit)	0.27	0.71	0.95	0.99
performance 2	0.54	0.43	0.23	0.10
Type I				
demand	0.05	0.06	0.05	0.05
demand (logit)	0.05	0.06	0.05	0.05
demand (probit)	0.05	0.06	0.05	0.05
performance 2	0.05	0.04	0.05	0.05

Note:
Type I error rates and power for the demand and performance specifications at different levels optimality: 2, 4, 8, 16. The practices can only take two values: 1 and -1. $\delta_1 = \delta_2 = 0$. The results are aggregated over the parameter values of γ_2 (-0.33, 0, 0.33).

3, leads to Type I errors. However, this correlated omitted variable problem was due to a misspecification, not due to not having access to the data about the variable. To test the vulnerability of the *demand* specification and the *performance 1* specification to spurious correlations in a well designed study, we run the following simulation: We introduce a new unobserved (to the researcher) environmental factor w that impacts the performance effect of x_1 with θ_1 and the performance effect of x_2 with θ_2.

We set the parameters assuming a well designed study that controls for most but not all of the environmental factors affecting the performance of both choices. Based on the nine studies in the calibration section, we set values for the contingency effect of the unobserved factor that reflect a large ($\theta_1 = 0.3, \theta_2 = -0.3$), medium ($\theta_1 = 0.3, \theta_2 = -0.2$), and small ($\theta_1 = 0.3, \theta_2 = -0.1$) spurious correlation. The previous methodology literature has long argued that an omitted environmental factor will bias the demand specification. The negative bias due to the omitted correlated variable increases the probability of

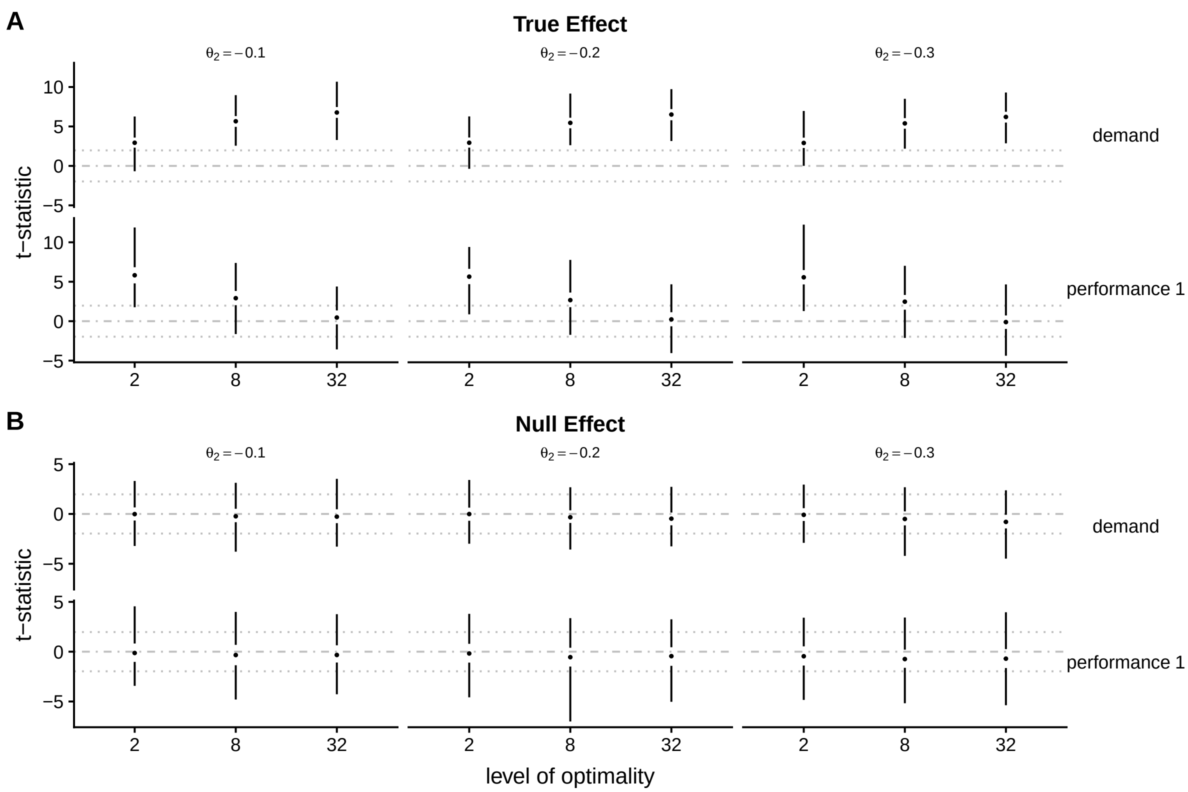

Figure 6: t-statistic of the performance and demand specification to test for complementarities when there is complementarity effect ($\beta_{12} = .25$), in Panel A or a null effect ($\beta_{12} = 0$) in Panel B. The boxplots represent the median (the dot), the interquartile range (the gap), and the minimum and maximum (the whiskers). O is varied between 2, 8, 32. The effect of the unobserved environmental variable, $\mathbf{w}$, on the choices varies from a strong correlation ($\theta_1 = .3$, $\theta_2 = -.3$), a medium correlation ($\theta_1 = .3$ and $\theta_2 = -.2$), or a weak correlation ($\theta_1 = .3$, $\theta_2 = -.1$).

Table 5: Power and Type I Error Rate with Correlated Omitted Variable

specification	θ_2	Level of Optimality 2	8	32
Power				
demand	-0.3	0.84	1.00	1.00
demand	-0.2	0.84	1.00	1.00
demand	-0.1	0.86	1.00	1.00
performance 1	-0.3	0.99	0.63	0.06
performance 1	-0.2	0.99	0.70	0.10
performance 1	-0.1	1.00	0.76	0.11
Type I				
demand	-0.3	0.06	0.07	0.12
demand	-0.2	0.05	0.05	0.08
demand	-0.1	0.04	0.05	0.06
performance 1	-0.3	0.18	0.21	0.21
performance 1	-0.2	0.17	0.20	0.18
performance 1	-0.1	0.14	0.18	0.16

Note:
Type I error rates and power for the *demand* and *performance 1* specifications at different levels optimality: 2, 8, 32. The parameters are the same as in Figure 6.

reporting a substitution effect in the absence of a true effect and decreases the power to reject the null hypothesis when there is a true complementarity.[8] We theoretically argued above that under the same condition, i.e., $\theta_1\theta_2 \neq 0$, the performance specification will be biased as well. The results of this simulation reveal to what extent both specifications are vulnerable to this bias.

The results are reported in Figure 6 and Table 5. These results are consistent with our previous findings. The *demand* specification has at all but the lowest levels of optimality more power to detect a true interdependency than the *performance 1* specification. One logical finding is that the *demand* specification becomes vulnerable to the omitted variable bias and thus higher Type

[8]A positive bias, i.e., $\theta_1\theta_2 > 0$, increases the probability of reporting a complementarity effect in the absence of a true effect. The Type I errors that might occur because of this correlated omitted variable problem are, by construction, identical to those for $\theta_1\theta_2 < 0$. Using $\theta_1\theta_2 < 0$ in our simulation allows us to capture two problems at once, i.e., potentially elevated Type I errors and reduced power.

I error rates at higher levels of optimality. This effect is most pronounced with the largest spurious correlation ($\theta_1 = .3$, $\theta_2 = -.3$) and $O = 32$, where we find that in 12% of the simulations with a null effect the *demand* specification reports a significant effect. This finding reinforces that the *demand* specification will only have appropriate Type I error rates as part of a well designed study that controls for environmental factors that affect both choices, especially when firms adopt the optimal management controls. However, consistent with our arguments, the *performance 1* specification exhibits the exact same bias and, as a result, the *demand* specification still has superior Type I error rates compared to the *performance 1* specification.

4.4. Robustness and Limitations of the Demand Specification

In this section we explore under which part of the parameter space the demand specification is less robust than the performance specification. In addition, we investigate whether our main conclusion that the demand specification generally dominates the performance specification, holds for objective functions with three practices and three complementarities, and for a combination of discrete and continuous practices. There are two main conclusions to this section. First, the principal weakness of the demand specification is elevated Type I error rates when it is optimal for more than half of the sample to choose the maximum level for both practices. This occurs when the practices have no increasing marginal costs (or alternatively, no decreasing marginal returns) and when contingency effects are small. Given how extreme such a worst case scenario is, we believe this result only marginally qualifies our main conclusion. In the majority of the management accounting studies we belief that variation in contingency factors and/or decreasing marginal returns will dominate over the direct benefits of the practices. If this is the case, the demand specification has more power to detect a true complementarity at all but the lowest level of optimality while keeping Type I error rates at the nominal 5% level.

The second conclusion of this section is that the demand specification generally deals well with discrete practices or an objective function with more than two practices and complementarities. Thus, our overall conclusion and recommendations generalise to this expanded objective function. Nevertheless, in this setting the demand specification suffers from the same inflated Type I error rate when it is optimal for more than half of the sample to choose the maximum level for any two practices. Because the decision to adopt a binary practice is inherently all or nothing, the issue is more pronounced for a test of complementarity between two discrete practices. Similarly as before, when the benefits of the practices vary enough in the sample due to contingency effects, the demand specification outperforms the performance specification at all but the lowest level of optimality.

The technical reason for the main weakness of the demand specification is that the corner solution of adopting both practices to their full extent violates the second order optimality condition as discussed in 2.1. The three subsections that follow explain the technical details of the simulations and explain how we establish quantitatively that the worst case scenario is unlikely to occur in a typical management accounting setting. The first section investigates an expanded set of parameter combinations to show the robustness of the baseline results in 4.1.3. Under this simulation, corner solutions are possible for all four combinations and they are all equally likely to occur in the sample.[9] The second section changes the first simulation so that one corner solution (i.e. both practices should be adopted to their full extent) is more likely. The third simulation introduces a third practice in the objective function and allows for discrete practices when one corner solution is more likely.

4.4.1. Robustness

In a first simulation, we vary the following parameters. β_{12} equals 0 or 0.25. $\delta_1 = \delta_2$ equal 0, .25 or 1. $\gamma_1 = 0.33$ and γ_2 equals 0, -0.33, or 0.33. $\sigma_{\epsilon_1} = \sigma_{\epsilon_2}$ varies between 0.5, 1, and 2 and σ_ν equals 1 or 2. Note that these

[9]The four possible corner solutions are: (1) Fully adopt practice 1 and 2, (2) Fully adopt practice 1 and do not adopt practice 2 at all, (3) fully adopt practice 2 and do no adopt practice 1 at all, and (4) do not adopt practice 1 and 2 at all.

parameter combinations no longer reflect the calibration to a typical management accounting study in section 3.2. For computational reasons, we restrict ourselves to comparing the two theoretically derived specifications: the *demand* specification and the *performance 1* specification.

The results in Table 6 largely confirm that only for lower levels of optimality the *performance 1* specification has more power to detect a true effect than the *demand* specification. The major issue for the *performance 1* specification are the inferior Type I error rates as reported in Panel B. The biggest issue for the *demand* specification is the slightly elevated Type I error rate when there are no increases in the marginal costs of the practices ($\delta_i = 0$) and the unobserved heterogeneity is relatively small ($\sigma_{\epsilon_i} = 0.5$). This is not surprising because these are the conditions where the optimal level for the practices are corner solutions and they violate the second-order optimality condition underlying the demand specification.

4.4.2. Corner Solutions

In the next simulation, we exaggerate the effect of the corner solutions by setting the main effects of the two practices, β_1 and β_2 equal to 0.5. This choice favours one of the four corners, namely where a firm would adopt both practices to their full extent. More specifically under the worst case scenario, i.e., $\sigma_{\epsilon_1} = \sigma_{\epsilon_2} = 0.5$ and $\delta_1 = \delta_2 = 0$, 80% of observations should adopt practice 1 and 80% should adopt practice 2.[10] This also implies that 64% of the sample optimally adopts both practices to their full extent even in the absence of a complementarity effect. In other words, for almost two thirds of the sample the question whether the practices are complements is irrelevant.

The results in Table 7 show that the introduction of the main effects does not affect the power of the demand specification relative to the performance speci-

[10]The distribution of the marginal effect of x_1 on performance, $\frac{\partial y}{\partial x_1}$, is given by a normal distribution with mean β_1 and standard deviation $\sqrt{\sigma_{\epsilon_1}^2 + \gamma_1^2}$. In the worst case scenario for the demand specification this implies a standard deviation of $\sqrt{.25 + 0.11} \approx 0.6$. The probability that the practice is beneficial for a firm at all levels of the practice is then given by $\Phi(\frac{\beta_1}{0.6}) = \Phi(\frac{0.5}{0.6}) \approx 0.8$ where Φ is the cumulative standard normal distribution.

Table 6: Power and Type I Error Rate without Main Effects

			demand specification					performance specification				
σ_{ϵ_i}	δ_i	σ_ν	2	4	8	16	32	2	4	8	16	32
						Power						
0.5	0.00	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.64	0.14
0.5	0.00	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.76	0.30
0.5	0.25	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.91	0.37	0.08
0.5	0.25	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.89	0.57	0.29
0.5	1.00	1	0.92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.88	0.56
0.5	1.00	2	0.91	0.99	1.00	1.00	1.00	1.00	1.00	0.91	0.60	0.35
1.0	0.00	1	0.99	1.00	1.00	1.00	1.00	1.00	0.98	0.72	0.13	0.01
1.0	0.00	2	0.97	1.00	1.00	1.00	1.00	1.00	0.97	0.79	0.25	0.04
1.0	0.25	1	0.99	1.00	1.00	1.00	1.00	1.00	0.92	0.33	0.01	0.00
1.0	0.25	2	0.98	1.00	1.00	1.00	1.00	1.00	0.92	0.54	0.10	0.01
1.0	1.00	1	0.85	1.00	1.00	1.00	1.00	1.00	0.98	0.82	0.43	0.17
1.0	1.00	2	0.84	0.99	1.00	1.00	1.00	1.00	0.95	0.70	0.39	0.21
2.0	0.00	1	0.66	0.98	1.00	1.00	1.00	0.73	0.63	0.24	0.02	0.00
2.0	0.00	2	0.64	0.98	1.00	1.00	1.00	0.74	0.64	0.28	0.04	0.00
2.0	0.25	1	0.70	0.99	1.00	1.00	1.00	0.74	0.51	0.08	0.00	0.00
2.0	0.25	2	0.67	0.99	1.00	1.00	1.00	0.74	0.55	0.15	0.01	0.00
2.0	1.00	1	0.58	0.97	1.00	1.00	1.00	0.81	0.59	0.18	0.02	0.00
2.0	1.00	2	0.56	0.96	1.00	1.00	1.00	0.79	0.58	0.20	0.04	0.01
						Type I						
0.5	0.00	1	0.07	0.10	0.09	0.07	0.06	0.10	0.09	0.07	0.08	0.08
0.5	0.00	2	0.06	0.08	0.07	0.06	0.07	0.09	0.09	0.07	0.06	0.07
0.5	0.25	1	0.06	0.05	0.05	0.05	0.05	0.11	0.08	0.10	0.10	0.11
0.5	0.25	2	0.05	0.07	0.05	0.05	0.05	0.09	0.08	0.06	0.07	0.08
0.5	1.00	1	0.03	0.04	0.04	0.04	0.03	0.14	0.13	0.13	0.10	0.09
0.5	1.00	2	0.04	0.04	0.04	0.04	0.04	0.09	0.09	0.08	0.06	0.05
1.0	0.00	1	0.05	0.05	0.04	0.05	0.05	0.10	0.10	0.08	0.07	0.07
1.0	0.00	2	0.05	0.05	0.05	0.05	0.05	0.10	0.10	0.09	0.09	0.07
1.0	0.25	1	0.06	0.05	0.06	0.05	0.05	0.11	0.12	0.11	0.11	0.12
1.0	0.25	2	0.05	0.05	0.05	0.05	0.05	0.10	0.10	0.11	0.09	0.10
1.0	1.00	1	0.05	0.04	0.04	0.05	0.04	0.14	0.16	0.16	0.16	0.15
1.0	1.00	2	0.05	0.04	0.04	0.04	0.04	0.14	0.12	0.13	0.11	0.10
2.0	0.00	1	0.05	0.06	0.04	0.05	0.04	0.11	0.11	0.09	0.08	0.07
2.0	0.00	2	0.05	0.05	0.04	0.04	0.04	0.11	0.11	0.10	0.08	0.08
2.0	0.25	1	0.05	0.05	0.04	0.04	0.04	0.12	0.12	0.12	0.10	0.10
2.0	0.25	2	0.05	0.05	0.04	0.04	0.05	0.12	0.11	0.10	0.10	0.09
2.0	1.00	1	0.04	0.05	0.05	0.06	0.05	0.14	0.14	0.17	0.17	0.20
2.0	1.00	2	0.06	0.05	0.05	0.05	0.05	0.13	0.15	0.14	0.16	0.14

Note:
Power and Type I error rate for the different levels of optimality (2, 4, 8, 16, 32) when $\beta_1 = \beta_2 = 0$ for the demand and performance 1 specification. The effect of the environmental variable, $\mathbf{z}$, on the second choice is either negative ($\gamma_1 = .33$ and $\gamma_2 = -.33$) or positive ($\gamma_1 = 0.33$ and $\gamma_2 = 0.33$). The results are averaged over the values of γ_2.

Table 7: Power and Type I Error Rate with Main Effects

			demand specification					performance specification				
σ_{ϵ_i}	δ_i	σ_ν	2	4	8	16	32	2	4	8	16	32
							Power					
0.5	0.00	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.78	0.17
0.5	0.00	2	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.78	0.28
0.5	0.25	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94	0.31	0.04
0.5	0.25	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94	0.54	0.18
0.5	1.00	1	0.89	0.99	1.00	1.00	1.00	1.00	1.00	0.99	0.85	0.51
0.5	1.00	2	0.87	0.99	1.00	1.00	1.00	1.00	1.00	0.91	0.57	0.33
1.0	0.00	1	0.96	1.00	1.00	1.00	1.00	1.00	0.99	0.80	0.15	0.01
1.0	0.00	2	0.92	1.00	1.00	1.00	1.00	1.00	0.98	0.83	0.26	0.04
1.0	0.25	1	0.97	1.00	1.00	1.00	1.00	1.00	0.95	0.29	0.01	0.00
1.0	0.25	2	0.94	1.00	1.00	1.00	1.00	1.00	0.94	0.54	0.08	0.01
1.0	1.00	1	0.80	0.99	1.00	1.00	1.00	1.00	0.98	0.79	0.40	0.15
1.0	1.00	2	0.78	0.99	1.00	1.00	1.00	1.00	0.95	0.68	0.35	0.18
2.0	0.00	1	0.60	0.97	1.00	1.00	1.00	0.76	0.68	0.24	0.02	0.00
2.0	0.00	2	0.57	0.96	1.00	1.00	1.00	0.75	0.67	0.30	0.04	0.01
2.0	0.25	1	0.62	0.99	1.00	1.00	1.00	0.75	0.50	0.08	0.00	0.00
2.0	0.25	2	0.61	0.98	1.00	1.00	1.00	0.74	0.55	0.13	0.01	0.00
2.0	1.00	1	0.54	0.96	1.00	1.00	1.00	0.81	0.60	0.17	0.01	0.00
2.0	1.00	2	0.53	0.94	1.00	1.00	1.00	0.79	0.58	0.19	0.03	0.01
							Type I					
0.5	0.00	1	0.17	0.37	0.36	0.23	0.14	0.12	0.08	0.05	0.03	0.04
0.5	0.00	2	0.11	0.24	0.23	0.19	0.13	0.09	0.07	0.05	0.05	0.04
0.5	0.25	1	0.07	0.11	0.08	0.06	0.05	0.14	0.10	0.07	0.07	0.07
0.5	0.25	2	0.07	0.08	0.07	0.06	0.05	0.08	0.07	0.06	0.05	0.05
0.5	1.00	1	0.04	0.04	0.05	0.04	0.04	0.14	0.13	0.12	0.10	0.08
0.5	1.00	2	0.04	0.03	0.04	0.04	0.04	0.10	0.08	0.08	0.06	0.06
1.0	0.00	1	0.08	0.10	0.09	0.08	0.07	0.12	0.09	0.06	0.05	0.05
1.0	0.00	2	0.07	0.09	0.09	0.08	0.06	0.10	0.09	0.08	0.06	0.05
1.0	0.25	1	0.06	0.08	0.08	0.06	0.05	0.12	0.11	0.09	0.09	0.08
1.0	0.25	2	0.06	0.08	0.06	0.06	0.05	0.11	0.10	0.09	0.08	0.08
1.0	1.00	1	0.04	0.04	0.04	0.05	0.04	0.15	0.15	0.15	0.14	0.13
1.0	1.00	2	0.04	0.04	0.04	0.05	0.04	0.13	0.12	0.12	0.10	0.09
2.0	0.00	1	0.05	0.05	0.05	0.04	0.05	0.12	0.11	0.08	0.07	0.06
2.0	0.00	2	0.05	0.06	0.05	0.05	0.04	0.12	0.10	0.09	0.08	0.07
2.0	0.25	1	0.05	0.06	0.05	0.05	0.04	0.12	0.11	0.10	0.08	0.08
2.0	0.25	2	0.05	0.05	0.04	0.05	0.04	0.12	0.11	0.11	0.10	0.08
2.0	1.00	1	0.05	0.04	0.06	0.05	0.05	0.13	0.16	0.17	0.16	0.18
2.0	1.00	2	0.05	0.06	0.05	0.04	0.05	0.12	0.13	0.15	0.14	0.15

Note:
Power and Type I error rate for the different levels of optimality (2, 4, 8, 16, 32) when $\beta_1 = \beta_2 = 0.5$ for the demand and performance 1 specification. The effect of the environmental variable, $\mathbf{z}$, on the second choice is either negative ($\gamma_1 = .33$ and $\gamma_2 = -.33$) or positive ($\gamma_1 = 0.33$ and $\gamma_2 = 0.33$). The results are averaged over the values of γ_2.

fication. Table 7 shows that the Type I error rate of the demand specification breaks down under the worst case scenario and one corner solution outlined above. Increases in unobserved variation in performances (σ_{ϵ_i}) and marginal costs (δ_i) largely mitigate the problem.

4.4.3. Two Discrete Practices and One Continuous Practice

Given that corner solutions violate the assumption of the demand specification, we run a last simulation to test the robustness of the demand specification. We extend the objective function to three practices where the first and the third practice are discrete practices. Table 8 reports the power of both specifications to detect a true substitution effect between the first and the second practice ($\beta_{12} = -0.25$) and a true complementarity between the first and the third practice ($\beta_{13} = 0.25$). Table 8 further reports the resulting Type I error rates in the absence of these effects. In both panels there is a complementarity between the second and the third practice ($\beta_{23} = 0.25$). β_{12} represents a complementarity between a continuous and a discrete practice. β_{13} represents a complementarity between two discrete practices. The full objective function with x_1 and x_3 as binary practices is reproduced below.

$$\begin{aligned} y =& (0.5 + 0.33z + \epsilon_1)x_1 + (0.5 - 0.33z + \epsilon_2)x_2 \\ &+ (0.5 + 0.33z + \epsilon_3)x_3 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + 0.25x_2x_3 - \frac{1}{2}\delta_2 x_2^2 + \nu \end{aligned}$$

The parameters are the same as in the previous simulation with two exceptions. First, we need to set the parameters for the direct effect ($\beta_3 = 0.5$) and the contingency effects for the third practice ($\sigma_{\epsilon_3} = \sigma_{\epsilon_1} = \sigma_{\epsilon_2}$, $\gamma_3 = 0.33$). Second, only the continuous, second practice can have increasing marginal costs.

The results in Table 8 are consistent with Table 7. The relative advantage of the demand specification to detect a true effect has not changed. The Type I error rates again show that the demand specification does not adequately control the false positive rate when there is low unobserved heterogeneity ($\sigma_{\epsilon_i} = 0.5$). When almost two thirds of the sample optimally adopts two practices (to their full extent) in the absence of an interdependency between these practices, the demand specification should not be trusted. Untabulated results show that the problem is less pronounced when there is more unobserved heterogeneity in performance ($\sigma_\nu = 2$ or $\sigma_\nu = 4$). For the interdependency with only one discrete practice, the problem for the demand specification dissipates when marginal costs of the continuous practice increase ($\delta_2 = 1$).

The results largely confirm the superiority of the demand specification for a typical management accounting study with one more caveat. When the two practices are discrete and should be adopted together in the absence of any interdependencies by the majority of the sample ($\sigma_{\epsilon_i} = 0.5$), the demand specification does not adequately control the Type I error rate. We argue that this is an uncommon occurrence for most management accounting practices.

4.5. Contingent Complementarity

In this section we extend the baseline model by allowing for a contingency effect on the complementarity. That is, we assume that not only the main effect of the practices depends on the environmental factor, z, but also the complementarity itself. The simplified objective function for the simulations is then as follows, where the parameter γ_{12} governs how the complementarity depends on the environmental factor:

$$y = (0.33z + \epsilon_1)x_1 + (\gamma_2 z + \epsilon_2)x_2 + (\beta_{12} + \gamma_{12} z)x_1 x_2 - \frac{1}{2}x_1^2 - \frac{1}{2}x_2^2 + \nu \quad (6)$$

The management accounting literature has shown great interest in investigating whether the strength of complementarity effects varies by environmental factors. For instance Grabner (2014) reports that subjective performance evaluation and performance based pay are complements for creativity dependent firms but not for other firms. Some researchers test for contingent complementarity effects by splitting the sample into observations with high and low values for the environmental factor and running separate regressions for both samples using either the performance specification or the demand specification. For

Table 8: Power and Type I Error Rate with Main Effects and Discrete Practices

			demand specification					performance specification				
σ_{ϵ_i}	δ_2	σ_ν	2	4	8	16	32	2	4	8	16	32
					1 Discrete Practice - Power							
0.5	0.00	1	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
0.5	0.25	1	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.97
0.5	1.00	1	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98
1.0	0.00	1	0.92	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.77
1.0	0.25	1	0.92	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.92	0.68
1.0	1.00	1	0.86	1.00	1.00	1.00	1.00	1.00	0.99	0.97	0.89	0.76
2.0	0.00	1	0.50	0.93	1.00	1.00	1.00	0.74	0.79	0.66	0.39	0.18
2.0	0.25	1	0.54	0.96	1.00	1.00	1.00	0.73	0.75	0.60	0.30	0.15
2.0	1.00	1	0.51	0.93	1.00	1.00	1.00	0.72	0.70	0.54	0.35	0.18
					1 Discrete Practice - Type I							
0.5	0.00	1	0.09	0.20	0.27	0.28	0.18	0.10	0.08	0.06	0.04	0.03
0.5	0.25	1	0.08	0.15	0.19	0.19	0.14	0.10	0.08	0.06	0.05	0.04
0.5	1.00	1	0.05	0.06	0.06	0.04	0.04	0.12	0.10	0.07	0.06	0.07
1.0	0.00	1	0.05	0.07	0.08	0.09	0.08	0.10	0.08	0.07	0.04	0.05
1.0	0.25	1	0.07	0.06	0.06	0.07	0.07	0.09	0.07	0.06	0.05	0.05
1.0	1.00	1	0.05	0.06	0.06	0.05	0.04	0.08	0.09	0.07	0.06	0.05
2.0	0.00	1	0.06	0.04	0.04	0.04	0.06	0.09	0.08	0.08	0.06	0.06
2.0	0.25	1	0.06	0.07	0.05	0.05	0.06	0.07	0.09	0.07	0.05	0.05
2.0	1.00	1	0.04	0.06	0.05	0.05	0.05	0.08	0.09	0.08	0.07	0.06
					2 Discrete Practices - Power							
0.5	0.00	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.5	0.25	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.5	1.00	1	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.0	0.00	1	0.97	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.91
1.0	0.25	1	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92
1.0	1.00	1	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
2.0	0.00	1	0.61	0.98	1.00	1.00	1.00	0.79	0.88	0.84	0.60	0.35
2.0	0.25	1	0.58	0.98	1.00	1.00	1.00	0.84	0.90	0.85	0.61	0.30
2.0	1.00	1	0.51	0.97	1.00	1.00	1.00	0.85	0.91	0.91	0.76	0.46
					2 Discrete Practices - Type I							
0.5	0.00	1	0.14	0.21	0.24	0.13	0.08	0.11	0.08	0.06	0.04	0.06
0.5	0.25	1	0.13	0.27	0.29	0.17	0.11	0.09	0.08	0.07	0.06	0.04
0.5	1.00	1	0.08	0.22	0.32	0.33	0.26	0.11	0.13	0.10	0.06	0.06
1.0	0.00	1	0.08	0.09	0.06	0.07	0.07	0.09	0.08	0.08	0.05	0.05
1.0	0.25	1	0.06	0.09	0.09	0.08	0.07	0.12	0.10	0.08	0.06	0.04
1.0	1.00	1	0.05	0.10	0.11	0.10	0.08	0.09	0.09	0.10	0.07	0.05
2.0	0.00	1	0.06	0.05	0.05	0.06	0.06	0.09	0.09	0.07	0.07	0.06
2.0	0.25	1	0.04	0.05	0.05	0.05	0.06	0.11	0.08	0.07	0.07	0.05
2.0	1.00	1	0.06	0.06	0.05	0.05	0.05	0.10	0.10	0.08	0.06	0.06

Note:
Power and Type I error rate for the different levels of optimality (2, 4, 8, 16, 32) when $\beta_1 = \beta_2 = 0.5$ for the demand and performance 1 specification. Power refers to the proportion of samples with a significantly negative estimate for the interdependency between a discrete and a continuous practice ($\beta_{12} = -0.25$) and a significantly positive estimate for the interdependency between two discrete practices ($\beta_{13} = 0.25$). Type I refers to the proportion of samples with a significant estimate for the complementarity when $\beta_{12} = \beta_{13} = 0$. β_1, β_2, and β_3 equal 0.5. Because two of the practices are discrete $\delta_1 = \delta_3 = 0$.

Table 9: Power and Type I Error Rate for Contingent Complementarity

		Level of Optimality				
specification	β_{12}	2	4	8	16	32
		Power				
demand	0.00	0.94	1.00	1.00	1.00	1.00
demand	0.25	0.92	1.00	1.00	1.00	1.00
performance 1	0.00	1.00	1.00	0.99	0.81	0.37
performance 1	0.25	1.00	1.00	0.99	0.82	0.48
performance 2	0.00	0.72	0.54	0.82	0.97	0.99
performance 2	0.25	0.80	0.67	0.87	0.94	0.94
		Type I				
demand	0.00	0.05	0.05	0.05	0.04	0.05
demand	0.25	0.05	0.03	0.05	0.04	0.04
performance 1	0.00	0.15	0.17	0.18	0.16	0.15
performance 1	0.25	0.15	0.17	0.17	0.18	0.17
performance 2	0.00	0.19	0.19	0.19	0.17	0.17
performance 2	0.25	0.19	0.18	0.17	0.17	0.19

Note:
Power is the proportion of samples reporting a significantly negative (when $\gamma_{12} = -0.33$) or a significantly positive (when $\gamma_{12} = 0.33$) contingent complementarity. Type I is the proportion of samples reporting a significant contingent complementarity when $\gamma_{12} = 0$. The remaining parameters are the same as in the baseline simulation in Figure 3.

instance, they expect a significant complementarity in one sample and no significant complementarity in the other sample. Our results and conclusions so far about the power and Type I error respectively directly apply to this approach, where these issues can now also affect the comparison between the subsamples. For example, the approach might miss a true effect in one subsample if the power of the specification is low, or alternatively it might erroneously report an effect when the Type I error rate of the specification is elevated.

Others try to estimate the parameter γ_{12} more directly. In this approach, the demand specification and performance specification include an additional term to capture the contingent complementarity: respectively $\gamma^d_{12}x_2z$ and $\gamma^p_{12}x_1x_2z$. The full *demand*, *performance 1*, and *performance 2* specification are given below. We do not report the *performance 3* specification because it is unlikely that a researcher would exclude x_1z or x_2z from the regression model when estimating the coefficient of $x_1 x_2 z$.

$$x_1 = \beta_0^d + \beta_{12}^d x_2 + \gamma_{12}^d x_2 z + \gamma_1^d z + \epsilon^d$$
$$y = \beta_0^{p1} + (\beta_1^{p1} + \gamma_1^{p1} z)x_1 + (\beta_2^{p1} + \gamma_2 z)x_2 + (\beta_{12}^{p1} + \gamma_{12}^{p1} z)x_1 x_2 + \delta_1^{p1} x_1^2 + \delta_2^{p1} x_2^2 + \alpha^{p1} z + \nu^{p1}$$
$$y = \beta_0^{p2} + (\beta_1^{p2} + \gamma_1^{p2} z)x_1 + (\beta_2^{p2} + \gamma_2 z)x_2 + (\beta_{12}^{p2} + \gamma_{12}^{p2} z)x_1 x_2 + \alpha^{p2} z + \nu^{p2}$$

Table 9 reports the results of a simulation to evaluate this approach similar to the baseline simulation. In the simulation, we vary the level of optimality O $(2, 4, 8, 16, 32)$, the complementarity, independent of the environmental factor, β_{12} (0 or 0.25), the direct contingency effects, γ_1 (0.33) and γ_2 (-0.33 or 0.33), and the contingency effect on the complementarity, γ_{12}. Table 9 shows the power to detect a contingent complementarity ($\gamma_{12} = 0.33$ or -0.33) and the the Type I error rate in the absence of a contingent complementarity ($\gamma_{12} = 0$).

The results show that demand specification also outperforms the performance specifications to test for a contingent complementarity at all but the lowest levels of optimality. As before, the demand specification has similar or more power to detect a true effect while maintaining better control over Type I error rates at all levels of optimality.

5. Going Forward

In this section, we use the above results to provide guidance for future studies on interdependencies between management control practices. First, we explain why it will almost always be better to rely on the demand specification in a research setting without a natural experiment, and why researchers should report the results of the demand specification in addition to the results of the performance specification. Second, we explain how researchers can appropriately control for contingency effects in the performance specifications. Third, we show how the bootstrap and corrections to standard error and degrees of freedom calculations can dramatically improve the Type I error rates of the performance specification. Fourth, we show how researchers can combine the results of the demand and performance specification if only one of them reports a significant interdependency.

5.1. Reporting the Demand Specification

Our recommendation applies to studies that test for complementarity (or substitution) with cross-sectional data where it is unlikely that firms have almost randomly chosen the accounting system. We recommend to always report the demand specification as it will have higher power to detect the complementarity while maintaining appropriate Type I error rates. This recommendation does not put any extra burden on the researcher with respect to data collection because the demand specification does not require additional data compared to the correctly specified performance specification.

A researcher might be interested in using the performance specification because they are interested in outcomes that are not the ultimate objective of the firm. For instance, researchers who are interested in which accounting system causes stress in employees (Shields et al., 2000) can use the performance specification to investigate this research question. The success of this approach hinges on whether the intermediate outcome (e.g. stress) is related to the final objective (e.g. job performance). If the intermediate outcome and the final objective are strongly correlated, the intermediate outcome will behave as a noisy measure of the final objective. In that case, the additional noise will be subsumed in the residual term. As we showed in Figure 4 the additional noise hampers the performance specification more than the demand specification. In other words, if the intermediate outcome is strongly related to the final outcome, all the problems with the performance specifications will emerge and the demand specification will be superior to detect an interdependency.

5.2. Controlling for Contingency Factors

One advantage of the performance specification is that it can estimate directly the size of the complementarity in terms of a performance increase. In order to report this estimate, a researcher should estimate the performance specification with the inclusion of all interactions between the accounting practices, such as delegation and incentives, as well as contingency factors that affect the performance of the practices, such as environmental uncertainty.

Bedford et al. (2016) provide an example of how to control for environmental dynamism in a complementarity test with the performance specification. For instance, their Table 6 reports the test for the complementarity effect of interactive control and firm structure on management control effectiveness, as well as controls for the interaction between environmental dynamism and interactive control and between environmental dynamism and firm structure.

When there are a large number of contingency factors that need to be included, the performance specification has to include a large number of interactions which can yield unstable estimates. We recommend that researchers use a dimension reduction technique such as principal component analysis on the contingency factors to reduce the number of variables and interactions in the regression. This recommendation assumes that researchers are mainly interested in estimating the complementarity and only include the contingency factors as control variables.

5.3. Correcting Type I Errors

Further, we address the Type I error rates of the performance specification. Even after adjusting for contingency factors, our results show that the performance specification consistently reports a higher proportion of Type I errors than the nominal false positive rate of 5%. We propose two solutions to this problem. The first solution is the bootstrap approach, which relies on repeated resampling of the data to empirically estimate the true distribution of the interdependency parameter (Efron & Hastie, 2017). The second solution relies on adjustments to the estimates of the standard error and degrees of freedoms that are used to calculate the t-statistic and p-value of the interdependency estimate (Young, 2016).

The bootstrap approach requires to run the performance specification repeatedly after resampling the data with replacement. The assumption of the bootstrap approach is that the data in the sample are representative of the broader population and that resampling from this sample approximates the variation in the population. To show the impact of bootstrapping, we run for each resampled dataset the performance specification, keep the estimate β_{12}^{p3} , and use the distribution of these estimates to decide whether the complementarity is statistically significant. If the 95% bias-corrected and accelerated confidence interval does not include 0, the complementarity is considered significant (Efron & Hastie, 2017). We use the recommended 2000 repetitions to get accurate estimates of the confidence interval (Efron & Hastie, 2017).

We test the Type I error rate and the power of the bootstrap approach for the *performance 1* specification on a subset of the parameters of the main analysis. The computational burden of the bootstrap approach forces us to limit the parameter space. Specifically, we limit the effect of the contingency factor z on x_2 to two values ($\gamma_2 = 0.33$ and $\gamma_2 = -0.33$) and we limit the values of the level of optimality to three values ($O = 2$, $O = 8$, and $O = 32$). The results in Table 10 show that the bootstrap approach reduces the Type I error rates of the performance specification considerably, close to the nominal 5%. Unfortunately, the improvement in error rates comes at the cost of reduced power to detect a true complementarity.

Young (2016) proposes a computationally efficient alternative adjustment. The adjustment uses the result that the standard error of a coefficient estimate is chi-squared distributed when the residuals of the regression are normally distributed. Young (2016) proposes an adjustment to the standard errors to allow deviations from normality and account for the possibility of influential outliers. In addition, he proposes a further adjustment to the degrees of freedom of the chi-squared distribution and thus of the t-statistic of the coefficient, to account for the fact that the data do not contain completely independent observations. Recall that one of the problems for the performance specification is that optimally designed accounting systems are completely determined by the contingency factors and thus firms in the same environment have the same

Table 10: Power and Type I Error Rate with Bootstrap

		Level of Optimality		
specification	γ_2	2	8	32
	Power			
performance 1	-0.33	0.99	0.68	0.09
performance 1	0.33	0.99	0.65	0.08
	Type I			
performance 1	-0.33	0.06	0.07	0.06
performance 1	0.33	0.05	0.07	0.06

Note:
Type I error rates and power for the *performance 1* specification at different levels of optimality: 2, 8, 32. γ_2 is set at either -0.33 and 0.33. The other parameters are the same as in Figure 3.

optimal accounting system. In other words, firms in the same environment are not independent from each other. As a result, the adjustments to the standard errors and degrees of freedom are appropriate to improve the performance specification.

We rerun the main analysis for the *demand* and *performance 1* specification. Table 11 reports the Type I error rate and power of both specifications with and without the corrections. For ease of exposition, we have aggregated the simulations over the different values of γ_2. The results show that the corrections almost bring the Type I error rates of the *performance 1* specification to the nominal level of 5%. As with the bootstrap results, this comes at the cost of a decrease in the power to detect a true complementarity.[11]

Because the power and Type I error rate of the bootstrap approach and the Young (2016) corrections are indistinguishable, researchers can report results

[11] The corrected standard errors are 2.8% ($O = 2$) to 1% ($O = 64$) smaller for the demand specification and 34% ($O = 2$) to 25% ($O = 64$) higher for the performance specification. The corrected degrees of freedom are 53% ($O = 2$) to 64% ($O = 64$) smaller for the demand specification and 77% ($O = 2$) to 85% ($O = 64$) for the performance specification. The correction of the standard errors have the largest impact on the improvement of the performance specification error rate. In practice, the degree of freedom correction could be as beneficial in smaller samples.

Table 11: Power and Type I Error Rate with Nearly Exact Correction

specification	Level of Optimality 2	4	8	16	32	64
Power						
demand	0.85	1.00	1.00	1.00	1.00	1.00
demand corrected	0.86	1.00	1.00	1.00	1.00	1.00
performance 1	1.00	0.98	0.83	0.45	0.16	0.07
performance 1 corrected	0.99	0.93	0.66	0.29	0.09	0.03
combined	1.00	1.00	1.00	1.00	1.00	1.00
combined corrected	1.00	1.00	1.00	1.00	1.00	1.00
Type I						
demand	0.04	0.04	0.05	0.05	0.05	0.05
demand corrected	0.05	0.05	0.05	0.05	0.05	0.05
performance 1	0.16	0.16	0.15	0.16	0.14	0.14
performance 1 corrected	0.06	0.06	0.06	0.06	0.06	0.06
combined	0.20	0.20	0.20	0.20	0.18	0.18
combined corrected	0.11	0.10	0.10	0.11	0.10	0.11

Note:
Type I error rates and power for the demand and performance function specification at different levels optimality: 2, 4, 8, 16, 32, 64. The parameters are the same as in the main analysis in the Figure refmain.

on the performance specification adjusting the OLS results with either of those methods. The bootstrap approach has the advantage that it is more flexible and applicable to other functional forms and multiple equation models. Most modern statistical packages have the functionality to run bootstrap tests. The corrections to the standard error and degrees of freedom are generally faster but limited to linear models. STATA users can use the script on Alwyn Young's website, while R users can use the functions on the Github page of this paper.[12]

5.4. Combining Performance and Demand Specification

Because the power of the performance specification decreases with the level of optimality while the power of the demand specification increases, a researcher might be tempted to run both regressions and decide that a complementarity is real if either of the regressions reports a significant effect. We assess this

[12] `http://personal.lse.ac.uk/YoungA/` and `https://github.com/stijnmasschelein/complementarity-simulation/blob/master/R/nearly_exact_correction.R`

combined strategy in Table 11. The results show that independent of the level of optimality, the combined approach with and without the correction leads to inflated Type I error rates. Further investigation of the results shows that for a given sample the probability of a false positive in the demand specification is independent of the probability of a false positive in the performance specification. Therefore, we recommend that researchers use a significance level of 2.5% for each test to keep the Type I error rate at 5% when combining the demand specification and the performance specification. Untabulated results show that this strategy contains the overall Type I error rate to the nominal 5%.

6. Summary and Discussion

This study builds on earlier studies on complementarity theory (Milgrom & Roberts, 1995; Grabner & Moers, 2013), to provide guidance on how to test for the presence of interdependencies in management control systems. The results of the simulation study show that in most common scenarios the assumptions of optimality should not be the main driver in deciding between the demand or the performance specification. In fact, unless researchers can make the case that a large number of firms have a management control system where the relation between practices is the opposite of the optimal relation or that contingency effects are small, the demand specification should be the preferred specification. A straight-forward check on the optimality assumption is to investigate the correlation between the practices and the environmental variables. Non-trivial levels of optimality in the sample will induce correlations between management accounting practices and environmental factors when there are contingency effects.[13]

When performance data is available, the performance specification can be

[13]The absence of any correlations does not imply the absence of high levels of optimality, as multiple contingency effects can cancel each other out. Another clarification about the superiority of the demand specification is that it does not imply that the demand specification provides evidence for high levels of optimality. A statistically significant interdependency effect in the demand specification only implies that firms on average avoid the worst possible management control systems, not that they have on average the optimal control system.

estimated as an additional test in combination with the demand specification. The performance specification can be expected to yield acceptable estimates when there is considerable variation between firms in the same environment. However, the results of this study show the importance of adequately controlling for contingency factors and adjusting the estimates of standard errors. As far as we know, the current accounting literature does not fully address these problems which lead to substantial increases in Type I error rates and a loss of power.

The most important weakness of the demand specification is that it assumes that the performance benefits of management control practices are decreasing with more extensive use. If this second-order condition does not hold, the demand specification might have elevated levels of false positives. We suggest that researchers verify the distribution of the management practices to check whether the second-order condition holds. If one of the management practices has more observations at the extremes of the measurement scale than at the centre, the second order condition might be violated and the results of both the demand and the performance specification should be interpreted with some caution. A conservative approach is to treat the practices as binary choices.

This study has several limitations. The recommended approach, the demand specification, does not allow to estimate the performance effect of the interdependency directly. More sophisticated models are needed to reliably estimate this performance effect. The economics literature has proposed and used a multiple equation model that combines both demand and performance functions (Athey & Stern, 1998; Gentzkow, 2007; Kretschmer et al., 2012; Miravete & Pernias, 2006). Further research can investigate the properties of these statistical models for the management control setting. An additional advantage of the models is that they can incorporate the effect of unobserved contingency factors and unobserved interdependent practices. A more detailed discussion of these issues goes beyond the scope of this study.

Another limitation of the current study is that the level of optimality is implemented with a naive algorithm that lacks external validity. Better theoretical models of how firms choose management accounting practices can improve upon our understanding of the distribution of management accounting practices and their interdependencies. The approach of Hemmer and Labro (2019) is one potential avenue to explore further. They model the firm's choices as decisions under incomplete information with Bayesian updating when more information becomes available. In these models, firms can end up with ex-post sub-optimal management control systems because they lack the appropriate information to choose the optimal system. The parameter governing the lack of information can replace our optimality parameter, O. Further innovations in these models can allow researchers to estimate directly the extent to which firms lack information and choose sub-optimal management control practices.

Appendix A. Omitted Variable Bias

The expected bias of omitting the term $x_1 z$ on the estimate of β_{12}^{p3} in the *performance 3* specification is given by $\gamma_1 \frac{cov(\widetilde{x_1 x_2}, \widetilde{x_1 z})}{var(\widetilde{x_1 x_2})}$ (Chenhall & Moers, 2007) where $\widetilde{x_1 x_2}$ and $\widetilde{x_1 z}$ are the residual vectors of regressing x_1, x_2, and z on $x_1 z$ and $x_1 z$ respectively (Angrist & Pischke, 2008; Cunningham, 2018). Assume that x_1, x_2, and z are multivariate standard normal with correlations r_{12}, r_{2z}, and r_{1z}. We can use the Isserlis (1918) theorem to derive the covariance between an interaction and one of the variables. Specifically, We make use of the fact that the expectation of the product of an odd number of standard normal variables equals 0. In the simplest case, $E(x_1) = E(x_2) = E(z) = 0$.

$$
\begin{aligned}
cov(x_1 x_2, x_1) &= E(x_1^2 x_2) - E(x_1 x_2)E(x_1) \\
&= 0 - 0 \\
cov(x_1 x_2, x_2) &= 0 \\
cov(x_1 x_2, z) &= 0 \\
cov(x_1 z, x_1) &= 0 \\
cov(x_1 z, x_2) &= 0 \\
cov(x_1 z, z) &= 0
\end{aligned}
$$

Given that the covariance between any product of two variables and a single variable is 0, the expectation of the residual terms $\widetilde{x_1 x_2}$ and $\widetilde{x_1 z}$ equal $x_1 x_2$ and $x_1 z$ respectively.

$$
\frac{cov(\widetilde{x_1 x_2}, \widetilde{x_1 z})}{var(\widetilde{x_1 x_2})} = \frac{cov(x_1 x_2, x_1 z)}{var(x_1 x_2)}
$$

where,

$$
\begin{aligned}
cov(x_1x_2, x_1z) &= E(x_1^2x_2z) - E(x_1x_2)E(x_1z) \\
&= E(x_1^2)E(x_2z) + E(x_1x_2)E(x_1z) + E(x_1z)E(x_1x_2) - E(x_1x_2)E(x_1z) \\
&= E(x_1^2)E(x_2z) + E(x_1x_2)E(x_1z) \\
&= r_{2z} + r_{12}r_{1z}
\end{aligned}
$$

In the second line, we use the Isserlis (1918) theorem for the product of an even number of multivariate standard normal variables. In the last line we make use of the fact that E(x_1^2) $= var(x_1) = 1$, and the fact that the correlation of two multivariate standard normal distributions equals the expected value of their dot product. Similarly, we can derive the variance of x_1x_2.

$$
\begin{aligned}
var(x_1x_2) &= E(x_1^2x_2^2) - E(x_1x_2)E(x_1x_2) \\
&= E(x_1^2)E(x_2^2) + E(x_1x_2)E(x_1x_2) + E(x_1x_2)E(x_1x_2) - E(x_1x_2)E(x_1x_2) \\
&= 1 + r_{12}^2
\end{aligned}
$$

Putting this all together, the bias of omitting the term x_1z on the estimate of β_{12}^{p3} is $\gamma_1 \frac{r_{2z}+r_{12}r_{1z}}{1+r_{12}^2}$

Similarly, the bias of omitting the term x_1z on the estimate of β_{12}^{p3} is $\gamma_2 \frac{r_{1z}+r_{12}r_{2z}}{1+r_{12}^2}$

The bias of omitting the terms x_1^2 and x_2^2 is $\frac{\delta_1 r_{12}}{1+r_{12}^2}$ and $\frac{\delta_2 r_{12}}{1+r_{12}^2}$ respectively.

References

Abernethy, M. A., Dekker, H. C., & Schulz, A. K.-D. (2015). Are employee selection and incentive contracts complements or substitutes? *Journal of Accounting Research*, *53*, 633–668. doi:10.1111/1475-679X.12090.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Aral, S., Brynjolfsson, E., & Wu, L. (2012). Three-way complementarities: Performance pay, human resource analytics, and information technology. *Management Science*, *58*, 913–931. doi:10.1287/mnsc.1110.1460.

Arora, A. (1996). Testing for complementarities in reduced-form regressions: A note. *Economics Letters*, *50*, 51–55. doi:10.1016/0165-1765(95)00707-5.

Athey, S., & Stern, S. (1998). An empirical framework for testing theories about complimentarity in organizational design. URL: https://www.nber.org/papers/w6600 - NBER Working Paper Series.

Bedford, D. S., Bisbe, J., & Sweeney, B. (2019). Performance measurement systems as generators of cognitive conflict in ambidextrous firms. *Accounting, Organizations and Society*, *72*, 21–37. doi:10.1016/j.aos.2018.05.010.

Bedford, D. S., & Malmi, T. (2015). Configurations of control: An exploratory analysis. *Management Accounting Research*, *27*, 2–26. doi:10.1016/j.mar.2015.04.002.

Bedford, D. S., Malmi, T., & Sandelin, M. (2016). Management control effectiveness and strategy: An empirical analysis of packages and systems. *Accounting, Organizations and Society*, *51*, 12–28. doi:10.1016/j.aos.2016.04.002.

Bouwens, J., & Van Lent, L. (2007). Assessing the performance of business unit managers. *Journal of Accounting Research*, *45*, 667–697. doi:10.1111/j.1475-679X.2007.00251.x.

Brynjolfsson, E., & Milgrom, P. (2013). Complementarity in organizations. In R. Gibbons, & J. Roberts (Eds.), *The Handbook of Organizational Economics*. Princeton: Princeton University Press. doi:10.1515/9781400845354-003.

Carree, M., Lokshin, B., & Belderbos, R. (2011). A note on testing for complementarity and substitutability in the case of multiple practices. *Journal of Productivity Analysis*, *35*, 263–269. doi:10.1007/s11123-010-0189-8.

Cassiman, B., & Veugelers, R. (2006). In search of complementarity in innovation strategy: internal R&D and external knowledge acquisition. *Management Science*, *52*, 68–82. doi:10.1287/mnsc.1050.0470.

Chenhall, R. H. (2003). Management control systems design within its organizational context: findings from contingency-based research and directions for the future. *Accounting, Organizations and Society*, *28*, 127–168. doi:10.1016/S0361-3682(01)00027-7.

Chenhall, R. H., & Moers, F. (2007). The issue of endogeneity within theory-based, quantitative management accounting research. *European Accounting Review*, *16*, 173–196. doi:10.1080/09638180701265937.

Cunningham, S. (2018). *Causal inference: The mixtape*. URL: http://scunning.com/mixtape.html.

Dekker, H. C., Ding, R., & Groot, T. (2016). Collaborative performance management in interfirm relationships. *Journal of Management Accounting Research*, *28*, 25–48. doi:10.2308/jmar-51492.

Efron, B., & Hastie, T. (2017). *Computer age statistical inference*. URL: https://web.stanford.edu/~hastie/CASI_files/PDF/casi.pdf.

Gentzkow, M. (2007). Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review*, *97*, 713–744. doi:10.1257/aer.97.3.713.

Grabner, I. (2014). Incentive system design in creativity-dependent firms. *The Accounting Review*, *89*, 1729–1750. doi:10.2308/accr-50756.

Grabner, I., & Moers, F. (2013). Management control as a system or a package? Conceptual and empirical issues. *Accounting, Organizations and Society*, *38*, 407–419. doi:10.1016/j.aos.2013.09.002.

Grabner, I., & Speckbacher, G. (2016). The cost of creativity: A control perspective. *Accounting, Organizations and Society*, *48*, 31–42. doi:10.1016/j.aos.2015.11.001.

Heinicke, A., Guenther, T. W., & Widener, S. K. (2016). An examination of the relationship between the extent of a flexible culture and the levers of control system: The key role of beliefs control. *Management Accounting Research*, *33*, 25–41. doi:10.1016/j.mar.2016.03.005.

Hemmer, T., & Labro, E. (2019). Management by the numbers: A formal approach to deriving Informational and distributional properties of unmanaged earnings. *Journal of Accounting Research*, *57*, 5–51. doi:10.1111/1475-679X.12249.

Hofmann, C., & van Lent, L. (2017). Organizational design and control choices. In M. A. Hitt (Ed.), *The Oxford Handbook of Strategy Implementation*.

Indjejikian, R. J., & Matejka, M. (2012). Accounting decentralization and performance evaluation of business unit managers. *Accounting Review*, *87*, 261–290. doi:10.2308/accr-10168.

Isserlis, L. (1918). On a formula of the product moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, *12*, 134–139. doi:10.1093/biomet/12.1-2.134.

Johansson, T. (2018). Testing for control system interdependence with structural equation modeling: Conceptual developments and evidence on the levers of control framework. *Journal of Accounting Literature*, *41*, 47–62. doi:10.1016/j.acclit.2018.02.002.

Kretschmer, T., Miravete, E. J., & Pernas, J. C. (2012). Competitive pressure and the adoption of complementary innovations. *American Economic Review*, *102*, 1540–1570. doi:10.1257/aer.102.4.1540.

Matejka, M., & Ray, K. (2017). Balancing difficulty of performance targets: theory and evidence. *Review of Accounting Studies*, *22*, 1666–1697. doi:10.1007/s11142-017-9420-4.

Milgrom, P., & Roberts, J. (1995). Complementarities and fit strategy, structure, and organizational change in manufacturing. *Journal of Accounting and Economics*, *19*, 179–208. doi:10.1016/0165-4101(94)00382-F.

Miravete, E. J., & Pernias, J. C. (2006). Innovation complementarity and scale of production. *The Journal of Industrial Economics*, *54*, 1–29. doi:10.1111/j.1467-6451.2006.00273.x.

Moers, F. (2006). Performance measure properties and delegation. *Accounting Review*, *81*, 897–924. doi:10.2308/accr.2006.81.4.897.

Otley, D. (2016). The contingency theory of management accounting and control: 19802014. *Management Accounting Research*, *31*, 45–62. doi:10.1016/j.mar.2016.02.001.

Samagaio, A., Crespo, N. F., & Rodrigues, R. (2018). Management control systems in high-tech start-ups: An empirical investigation. *Journal of Business Research*, *89*, 351–360. doi:10.1016/j.jbusres.2017.12.028.

Shields, M. D., Deng, F. J., & Kato, Y. (2000). The design and effects of control systems: tests of direct- and indirect-effects models. *Accounting, Organizations and Society*, *25*, 185–202. doi:10.1016/S0361-3682(99)00041-0.

Simons, R. (1994). *Levers of control: How managers use innovative control systems to drive strategic renewal*. Harvard Business Press.

Simons, R. (2000). *Performance measurement and control systems for implementing strategy*. Upper Saddle River, NJ: Prentice Hall.

Sponem, S., & Lambert, C. (2016). Exploring differences in budget characteristics, roles and satisfaction: A configurational approach. *Management Accounting Research*, *30*, 47–61. doi:10.1016/j.mar.2015.11.003.

Widener, S. K. (2007). An empirical analysis of the levers of control framework. *Accounting, Organizations and Society*, *32*, 757–788. doi:10.1016/j.aos.2007.01.001.

Young, A. (2016). Improved, nearly exact, statistical inference with robust and clustered covariance matrices using efffective degrees of freedom corrections. URL: http://personal.lse.ac.uk/YoungA/Improved.pdf - Working Paper.