

8-2022

Full Disclosure: Model Uncertainty in Adjusting for Confounders

Dieudonne Dusenge
University of Arkansas, Fayetteville

Follow this and additional works at: <https://scholarworks.uark.edu/etd>



Part of the [Accounting Commons](#), [Finance Commons](#), and the [Risk Analysis Commons](#)

Citation

Dusenge, D. (2022). Full Disclosure: Model Uncertainty in Adjusting for Confounders. *Graduate Theses and Dissertations* Retrieved from <https://scholarworks.uark.edu/etd/4646>

This Dissertation is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact uarepos@uark.edu.

Full Disclosure: Model Uncertainty in Adjusting for Confounders

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Accounting

by

Dieudonne Dusenge
Philander Smith College
Bachelor of Arts in Mathematics, 2014
University of Arkansas
Master of Arts in Economics, 2016

August 2022
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

Cory Cassell, Ph.D.
Dissertation Chair

Jonathan Shipman, Ph.D.
Committee Member

Hyunseok Jung, Ph.D.
Committee Member

Abstract: This study examines the role of knowledge about underlying causal relationships in classifying controls in order to mitigate omitted- and included-variable biases. Using simulations and accounting examples, the study shows that the researcher may not distinguish good and bad controls because the underlying causal relationships are unobservable, and remedying strategies (such as the relative timing of measurement) may not remove the uncertainty in the classification. Because of the uncertainty about which controls to use, two or more models will be credible as will the distinct estimates derived from them. Next, the study shows that the current standard practice of singling out a preferred model (e.g., one about which the researcher is most confident) masks model uncertainty and is generally not supported by theories of rational decision making under uncertainty. Specifically, relying on the estimate from a preferred model will often involve higher risks in statements made from the study than relying on another estimate or combination of estimates. Accordingly, the study recommends that the researcher disclose all the relevant credible estimates without signalling which estimate is more important or choose an applicable theory that recommends which estimate to rely on in making conclusions and other important statements from their study.

Key words: Causal inferences, economic significance, confounder adjustment, model uncertainty, expected utility, risk, model averaging, credible estimates

Table of Contents

1. Introduction	6
2. Background	8
2.1 Sources Of Adjustment Uncertainty And Remedying Strategies	8
2.2 Accounting Examples	13
3. Case 1: Simulation of Model Uncertainty	21
3.1 Invisible Underlying Process and Credible Models	21
3.2 OLS Results of Credible Models for Case 1	22
3.3 Minimizing sampling error using average estimates over multiple simulations	24
3.4 Model Space as State Space and Potential Errors as Risks	25
3.5 Deterministic Action: No Model Uncertainty	26
3.6 Recommended Action Based on the Expected Utility	27
4. Case 2: Other Model Uncertainty	32
4.1 Invisible Underlying Process and Credible Models	32
4.2 OLS Results of Credible Models for Case 2	33
4.3 Summary of multiple simulations and risks for Case 2	35
4.4 Recommended Course of Action Based on Expected Utility	35
4.5 Recommended Action Based on Wald's Maximin Criterion	37
5. Discussion	38
5.1 Basis for Singling out the Preferred Model	38

5.2 Background Information as Input for Model Weights _____	40
5.3 On Quantifying Beliefs _____	41
6. Solution _____	43
6.1 Full Disclosure And Formal Analysis _____	43
6.2 A More Practical Solution _____	46
6.3 Future Considerations _____	47
7. Conclusion _____	48
References _____	49

Table of Tables

Table 1: Case 1 _____	50
Panel A: Correlation Matrix for Case 1 _____	52
Panel B: Auxiliary Regressions for Case 1: Statistical Equivalence of Controls _____	52
Panel C: OLS Results of Credible Models for Case 1 _____	53
Panel D: Summary of exposure estimates from multiple simulations for Case 1 _____	53
Panel E: Actions and Possible Risks for Case 1 _____	54
Panel F: Recommended Action Using Expected Utility and Assuming the Absolute Error Loss Function for Case 1 _____	54
Panel G: Recommended Action Using Expected Utility and Assuming the Squared Error Loss Function for Case 1 _____	57
Panel H: Recommended Action Using Expected Utility and Assuming the Zero-One Loss Function for Case 1 _____	57
Table 2: Case 2 _____	56
Panel A: Correlation Matrix for Case2 _____	56
Panel B: Auxiliary Regressions for Case 2: Statistical Equivalence of Controls _____	58
Panel C: OLS Results of Credible Models for Case 2 _____	59
Panel D: Summary of exposure estimates from multiple simulations for Case 2 _____	59
Panel E: Actions and Possible Risks for Case 2 _____	60
Panel F: Recommended Action Using Expected Utility and Assuming the Absolute Error Loss Function for Case 2 _____	60
Panel G: Recommended Action Using Expected Utility and Assuming the Squared Error Loss Function for Case 2 _____	61

Panel H: Recommended Action Using Expected Utility and Assuming the Zero-One Loss

Function for Case 2 _____ 62

Table of Figures

Figure 1: Reporting Exposure Estimates as Course of Action in a Decision Tree _____ 55

Figure 2: Course of Action and Risks with Probabilities of Occurrence _____ 56

1. Introduction

In establishing a causal relationship between a factor (x) and an outcome (y), a major concern in observational studies is the possibility that another factor, a confounder, may explain the observed relationship. To rule out alternative explanations, a lot of effort is spent to identify and adjust for the confounder(s). Traditionally, the issue of confounding variables has seemed harmless because one could in theory adjust models by including all potential confounders and eliminate any omitted variable bias (OVB), a strategy labelled by Bertomeu et al. (2016) as “kitchen-sink.” However, using causal diagrams formalized by Pearl (1995), studies have shown that this kitchen-sink approach can actually lead to biases (Bertomeu et al., 2016; Ding & Miratrix, 2014; Gow et al., 2016; Greenland et al., 1998; Shpitser et al., 2012; VanderWeele, 2019; Whited et al., 2021). Even though the inclusion of some variables may cause bias, accountings studies asymmetrically focus on OVB. For example, Gow et al. (2016) note:

“A typical paper in accounting research will include many variables to “control for” the potential confounding of causal effects. While many of these variables should be considered confounders, less attention is given to explaining why it is reasonable to assume that they are not mediators or colliders. Such a discussion is important because the inclusion of “controls” that are mediators or colliders will generally lead to bias. (p. 484).”

An implicit assumption for focusing on OVB at the expense of collider and other biases is that OVB is more serious, but the other biases can be just as severe (Gow et al., 2016). To alleviate the asymmetric consideration, Whited et al. (2021) and related studies suggest that researchers consider underlying causal relationships in order to select good controls (whose inclusion may reduce OVB) while avoiding bad controls (whose inclusion may lead to “included

variable bias” and other complications). If researchers can identify the underlying causal relationships, then the classification of good and bad controls is straightforward and there is no uncertainty about which model to use.

The goal of this study is to examine further the key element required to successfully classify controls: the knowledge about underlying causal relationships, including the relative timing of measurement of the variables. Generally, such knowledge is limited or imperfect because the underlying process is unobservable and typically not identified by data (Bertomeu et al., 2016; Gow et al., 2016; Mackinnon et al., 2000; VanderWeele, 2019; Whited et al., 2021). The limited knowledge ultimately causes uncertainty about which model to use.

Equipped with confounder adjustment strategies, the researcher typically selects a set of variables they believe meets the adjustment criteria the most. Baseline (or preferred) models are chosen in this manner in many studies. Paradoxically, the same (or a different) researcher may identify a distinct set of variables that seem equally appropriate. In other words, because of model uncertainty, two researchers may properly consider the underlying causal relationships and still arrive at two or more different credible models to adjust for confounds with two or more different credible estimates of the effect.

I illustrate model uncertainty using hypothetical research questions in accounting. The first question examines the effect of aggressive tax planning on firm value and the second examines the effect of information asymmetry on conservatism. For each question, I show that it is possible to have two or more credible models with different exposure estimates, even after relying on prior accounting research to properly select controls.

The current standard practice within accounting research generally consists of reporting the estimate from the model the researcher chooses. This practice does not acknowledge model

uncertainty because it dismisses estimates from credible models that are not chosen. Certainly, some studies show results of alternative specifications, but it is not clear the extent to which such specifications specifically address the uncertainty about whether a control should be included or excluded, or whether they relate to other econometric problems (such as the consideration of an alternative proxy for the exposure or the outcome). To the extent that researchers take into account model uncertainty in their overall analyses, the communication of results and accompanying inferences may, at other times, fail to indicate the uncertainty or the risk therein. Overall, there is a tendency to base conclusions on a preferred, baseline, model even when sensitivity analyses highlight some of the uncertainty.

In some cases, as long as other credible estimates are disclosed, the impact of model uncertainty on relevant causal inferences might be clear. In other cases, however, reliance on a single estimate to draw inferences will be misleading. For this reason, it might be helpful to distinguish causal inferences related to the existence of an effect from those related to the economic significance of the effect.

From an econometric viewpoint, the existence of an effect is directly linked to its size (and the standard error). As such, bias is problematic to causal inferences related to both. However, inferences related to existence of an effect might be more straightforward because they are binary. For example, two different credible estimates may lead the reader to conclude that the effect exists if the estimates are significant and of the same sign. Even if one estimate is not significant, it will be clear to the reader that there is a need to qualify the conclusion about such existence, as long as the researcher properly communicates the degree of uncertainty in the models from which the estimates were derived.

For economic significance, model uncertainty is more problematic because, even if the two credible estimates are significant and of the same sign, two distinct conclusions or implications would invariably be drawn from the estimates. Presumably, a manager who learns that \$1 saved in taxes paid increases firm value by \$1.3 is likely to have a different level of commitment to tax saving than another who learns the saving increases firm value by \$1.1. That is because the difference might imply a range of other alternatives to tax planning. To say that confidence intervals should address this problem ignores that confidence intervals also change from one model to the next and diverts attention from potential bias.

Failing to disclose estimates from other credible models or focusing on the “preferred” model (i.e., one about which the researcher is most confident) ultimately assumes that: 1) the preferred model is actually the true model; or 2) under uncertainty about the true model, the researcher should draw inferences based on the estimate from the preferred model. However, since the underlying causal process is unobservable and can hardly be identified with data, there is no guarantee the preferred model or any model is the true model. Without a measured and repeatable framework, researchers may understate the uncertainty in inferences of interest or mask the risk involved in drawing conclusions.

For a measured and repeatable framework, I consider decision theories which recommend actions that should be taken under uncertainty using formal reasoning. If reporting or relying on the estimate from the preferred model is appropriate, it should be justified or recommended by the applicable decision theory. Using two simulations, I show that under uncertainty the emphasis on the preferred model is justified in very limited situations. I show that it is justified primarily if the researcher uses expected utility theory and assumes a zero-one loss function. However, a zero-one loss function, which treats small and large errors equally, is inappropriate

for continuous (exposure) measures. Under uncertainty, the emphasis on the preferred model is also justified if the researcher uses expected utility theory, assumes the absolute error loss function, AND the preferred model is far more credible than other candidate models. I show the last condition is very limiting.

While other theories and other loss functions may give more weight to the preferred model, the action the theories recommend typically combines information from all credible models. More importantly, the theories do not suggest that singling out the preferred model should be the default choice. Therefore, by singling out the preferred model, researchers may mechanically assume that causal relationships are perfectly known, or under uncertainty, use limiting if not questionable decision criteria.

After showing that there is seldom a basis for singling out the preferred model, I propose two alternative solutions. The first alternative is to report all estimates from relevant credible models without signaling or implying which estimate is more important. This option is desirable because deriving a single estimate based on a particular decision theory requires more arithmetic. The second alternative I propose is to choose an applicable decision theory and report the estimate the theory recommends. In particular, I propose to use the standard (Savage's) expected utility theory, which recommends the action with the minimum expected risk (or highest expected utility). In practice, the theory recommends averaging exposure estimates using model weights, reflecting the researcher's beliefs about the goodness of the models. Using expected utility theory guarantees that, in the long run, the action involving the lowest risk is chosen.

Two major problems with the proposed solution are that the consideration of all credible models becomes impractical as the number of doubtful confounders increases and the assignment of weight to models can be arbitrary. To address the problems, I propose that researchers focus

on cases when the uncertainty about variable inclusion is severe. First, with this compromise, the number of credible estimates becomes tractable. Furthermore, as long as the researcher focuses on very uncertain cases, all models will have even probabilities of being adequate, resulting in a uniform distribution of model weights. Accordingly, the expected utility and the squared (absolute) error loss imply a simple averaging (median) of exposure estimates. This practical solution is a compromise because researchers will lose information from the less credible models if they only focus on severe cases of uncertainty.

I caveat that the solutions I recommend do not remove model uncertainty and will not reveal the “true” exposure estimate. In fact, they are predicated on the idea that sometimes model uncertainty cannot be removed. Instead, the solutions properly account for model uncertainty: the first entails disclosing relevant credible estimates, and the second recommends an estimate that involves the lowest risk in making statements (conclusions) when multiple estimates are credible. The solutions contrast with the current default of relying on the estimate from the preferred model, which may involve more risk.

This study contributes to a stream of accounting research that seeks to identify ways researchers can improve causal inferences drawn from observational studies. The relevant studies note that while many researchers understand assumptions required for drawing causal inferences, they rarely address the assumptions. The studies recommend using theory and other strategies to symmetrically consider controls (i.e., consider both OVB and included variable bias with similar effort) and select good controls that alleviate OVB and avoid bad controls that cause included variable bias (Bertomeu et al., 2016; Gow et al., 2016; Whited et al., 2021). However, they are silent on what to do when the strategies are limited, potentially leaving an impression that reliance on theory and the other strategies are a panacea for adjustment uncertainty. The key

innovation from my study is that it explicitly addresses the limitation of theory in selecting appropriate controls and the resulting model uncertainty, while offering a few suggestions on how to account for the limitations in drawing causal inferences. In doing so, it supplements the call by Ohlson (2021) to report more about the evidence (model uncertainty in this case) that might qualify or cast doubt on conclusions reached.

2. Background

2.1 Sources Of Adjustment Uncertainty And Remediating Strategies

2.1.1 *The Traditional Approach to Causal Inferences*

My study is related to a stream of research in accounting and other fields that aims to improve how researchers draw inferences from observational studies. More specifically, it builds on the subset of studies that address the difficulties involved in drawing causal inferences (Bertomeu et al., 2016; Ding & Miratrix, 2014; Gow et al., 2016; Greenland et al., 1998; Shpitser et al., 2012; VanderWeele, 2019; Whited et al., 2021). Predictive inferences are certainly important, but Gow et al. (2016) note that nearly 90% of studies published in top accounting journals seek to draw causal inferences.¹

Gow et al. (2016) also note that accounting researchers have improved how they draw causal inferences by using designs that better mitigate confounding bias, which arises when another factor may help explain the observed relationship between the exposure (x) and the outcome (y). However, they conclude that credible research designs are rare since they rely on strong assumptions. Relatedly, while it is understood that standard regressions (OLS) and quasi-natural experimental (QNE) methods rely on untestable assumptions, the reasonableness of the assumptions is rarely explicitly addressed (Bertomeu et al., 2016; Gow et al., 2016). For scope limitation, this study examines challenges in drawing causal inferences within regression methods (OLS), which are still the standard in accounting research (Gow et al., 2016).

A recurring theme in the highlighted studies is that, in seeking to draw causal inferences, accounting researchers tend to heavily focus on including variables to rule out confounding

¹ Their examination finds that, of the 106 original papers using observational data, 91 sought to draw causal inferences.

factors or omitted variable bias (OVB) (Bertomeu et al., 2016; Gow et al., 2016; Whited et al., 2021). An implicit assumption behind the emphasis is that, as long as confounding factors have been identified and controlled for, derived causal estimates will be valid, and that addition of unnecessary variables is harmless. In other words, there should be no uncertainty about which model to use since the safe strategy is to control for all candidate confounders.

On the contrary, inclusion of some variables will cause bias (Bertomeu et al., 2016; Gow et al., 2016; Greenland et al., 1998; Pearl, 1995; VanderWeele, 2019; Whited et al., 2021). As such, researchers need to distinguish confounders from these variables whose inclusion will cause “included variable bias.” Whited et al. (2021) refer to confounders as good controls and these other variables as bad controls. Among bad controls are variables that are mediators and colliders, and variables that represent the same constructs as x or y . Since bad controls and good controls both vary with x and y , the studies recommend that researchers consider the underlying causal relationships in order to distinguish the controls (Bertomeu et al., 2016; Gow et al., 2016; Greenland et al., 1998; Pearl, 1995; VanderWeele, 2019; Whited et al., 2021).

2.1.2 Using Causal Diagrams and Theory can Improve Causal Inferences

If the researcher has knowledge of the structure relating all variables to each other, they can use the backdoor path criterion to determine which covariates will sufficiently adjust for confounding without causing included variable bias (Bertomeu et al., 2016; Gow et al., 2016; Greenland et al., 1998; Pearl, 1995; VanderWeele, 2019; Whited et al., 2021). Thus, there should be no uncertainty about which model to use if researchers can identify the underlying causal relationships.

However, detailed knowledge about causal relationships is rarely attainable because the process is unobservable. Accordingly, the studies recommend using theory to help uncover the

(causal) relationships in selecting proper controls (Bertomeu et al., 2016; Whited et al., 2021). An implicit assumption in relying on theory is that it is sufficient to uncover underlying causal relationships necessary for control selection. For example, while the message from Bertomeu et al. (2016) is not that theory is panacea for correct variable inclusion, in stating that consideration of theory and the causal chain would provide clear predictions about which variables to exclude, there is a presumption that theory removes barriers in drawing causal inferences.² Overall, the studies do not explicitly address the limitation of reliance on theory in selecting which controls should be included or excluded in uncovering causal effects.

Ultimately, since theory does not “prove” the causal relationships, it may not always uncover them. On the one hand, issues examined in accounting research relate to a complex world where it is difficult to uncover detailed relationships among constructs. On the other hand, there are many factors that explain why existing theory may be misleading, including the prevalence of false positives (Ohlson, 2021). Thus, even after relying on theory, distinguishing good and bad controls may be difficult and sometimes impossible, resulting in uncertainty about which model to use. Gow et al. (2016, p. 479) state that “Statistical methods alone cannot solve the inference issues that arise in observational data.” In the same context, researchers should be aware that reliance on theory too cannot solve causal inferences that arise in observational data. This study complements the highlighted studies by examining how accounting researchers might address or integrate model uncertainty related to confounder adjustment in drawing causal inferences.

² Bertomeu et al. (2016) acknowledge the limitation of theory when stating that institutional knowledge might indicate the plausibility of certain assumptions, but it does not resolve untestable assumptions.

2.1.3 *Other Strategies For Proper Variable Inclusion*

Under the limitation of (detailed) theory, studies have recommended remedying strategies to use. One strategy is to control for covariates that are measured (determined) prior to the exposure (VanderWeele, 2019; Whited et al., 2021). Whited et al. (2021) refer to this strategy as the pre-treatment criterion. However, since relative timing is not always detectable, researchers may not always rely on the pre-treatment criterion in classifying controls.

Furthermore, even when the relative timing of measurement can be determined, the use of the pre-treatment criterion may cause other biases. M bias, so-called as it takes an M shape in causal diagrams, is a special case of collider bias. To illustrate M bias, suppose a variable (u_y) causes another variable (m) that is measured before the exposure x . Also suppose m causes y , but not x . Lastly, suppose there is another predetermined variable (u_x) that causes m and x , but not y . In the described structure, m shares a cause (u_x) with x and shares one (u_y) with y . As such, m is statistically associated with both x and y and may appear as a confounder. Reliance on the pre-treatment criterion would suggest that the researcher control for m . However, controlling for m will cause bias (Ding & Miratrix, 2014; Sjölander, 2009; VanderWeele, 2019). In short, while the pre-treatment criterion is helpful when the full causal diagram is unavailable, relying on it carries some risk.

A similar strategy to the pre-treatment criterion aims at controlling for only pre-measured covariates that are actual instead of apparent common causes of both the exposure and the outcome (Glymour et al., 2008). The strategy is less stringent than knowing the full causal diagram but still requires extensive knowledge to distinguish actual and apparent causes. It might be impractical to require predetermined variables that are known causes of the exposure and the outcome because those variables may be unobservable. In many cases, variables that are known

causes of just the exposure (not necessarily the outcome) may be sufficient and observable (VanderWeele & Shpitser, 2011). However, there are other cases where such causes of the exposure will amplify residual biases (Myers et al., 2011; Pearl, 2012).³

2.1.4 Residual Model Uncertainty Implies Multiple Credible Estimates

The above discussion suggests that reliance on theory and other strategies might not always reveal which controls to include or exclude in order to derive valid causal estimates. Two researchers may duly consider the causal relationships and still arrive at two or more distinct credible models to use, with two or more distinct credible exposure estimates. Yet, instead of considering estimates of all credible models, researchers generally report the estimate of a single model, largely ignoring the uncertainty. In the words of Ohlson (2021), studies do not provide the slightest hint that a paper's take-away could change due to misclassifying a control as good when it is bad or vice versa.⁴

A potential criticism of this categorization is that researchers are aware of model uncertainty and in doubt will pick a model that provides conservative estimates of the exposure effect; that is, they should lean towards variable inclusion. Gow et al. (2016) note the argument is typically used for focusing more on OVB and less on the included variable bias.⁵ As their study further notes, the claim is only correct if the indirect and direct effects are of the same sign.

³ Recall that instrumental variables are associated with the exposure and affect the outcome only through the exposure. When an instrument is valid, it replaces the exposure instead of being added to the model with it. Hence, the concern is generally that a variable may not be a valid instrument. Here the concern is the opposite. When controlling for variables that cause the exposure (i.e., confounders), the controls should not be valid instruments (i.e., affect the outcome only through the exposure). If they are and there is some residual confounding bias, the adjustment will amplify the bias (Myers et al., 2011; Pearl, 2012). However, while it's worth knowing that valid instruments are bad controls, determining whether a variable is an instrument or not is very challenging.

⁴ Similarly, Ohlson (2021) notes that abstracts do not provide the slightest hint that a paper's take-away could be qualified based on insights gained from preliminary data analyses.

⁵ Note the undue focus on OVB documented by the studies is due to researchers not considering underlying causal relationships using theory. It is not due to the ambiguity about whether a variable should be included. This ambiguity is assumed to be resolved after considering theory.

Otherwise, controlling for a mediator of a different sign may overstate the causal effect by suppressing the indirect effect. But more importantly, the claim about conservative estimates is only valid for mediators, not colliders, same constructs, or other cases presented above.

Furthermore, confounder bias is not necessarily more severe than the other biases (Gow et al., 2016). In the same vein, Whited et al. (2021) note that including bad controls may actually modify the intended research question and the interpretations that can be made from the study.

Lastly, one could argue that under model uncertainty it is better to report the estimate from the model the researcher believes most likely includes good controls and excludes bad controls, even if other models are credible. I refer to such a model as the preferred model. After illustrating model uncertainty in accounting research, I examine the reasonableness of singling out the preferred model by reviewing key theories of decision-making under uncertainty.

2.2 Accounting Examples

In this section, I discuss two hypothetical research questions to illustrate model uncertainty in accounting research. In particular, I show how researchers can use theory to help identify underlying causal relationships and properly classify controls. More importantly, I show that after relying on theory and timing of measurement, researchers may have difficulty in making the classification, resulting in model uncertainty. Accordingly, the illustrations will show what circumstances are likely to involve model uncertainty. It should be noted that while I attempt to find areas where two sides of a debate about a specific relationship have comparable support, a more careful examination might reveal a winning side. The point should be that ambiguity in theory about a potential confounder leads to ambiguity about whether a control for the confounder should be included or excluded, resulting in multiple credible models.

2.2.1 Does Aggressive Tax Planning Increase Firm Value?

In examining the effect of aggressive tax planning (x) on firm value (y), a researcher should control for factors (z) that may help explain the observed association. Aggressive tax planning may increase firm value because managers can use it to reduce taxes paid to the tax authority. The saved taxes may in turn be used for value-increasing opportunities.

A potential concern is that corporate transparency might help explain the relationship in question. In particular, Kerr (2019) argues and finds that corporate transparency reduces aggressive tax planning at the firm- and country-level.⁶ Since corporate transparency may affect firm value through other channels (through cash-flow or cost of equity components), it is a common cause of both firm value and aggressive tax planning. Thus, in order to mitigate OVB, the researcher should include corporate transparency in the model as a good control.

A parallel concern is possible reverse causality between corporate transparency and aggressive tax planning. Balakrishnan et al. (2019) argue and find that aggressive tax planning reduces corporate transparency, as tax planning may involve financial and organizational complexity. If the finding represents the true underlying process, controlling for corporate transparency suppresses the indirect negative effect (of lower transparency) on firm value. In other words, corporate transparency, a mediator, is a bad control that may cause included variable bias.

Based on the arguments and findings of both studies, there will be, on the one end, researchers on the side of Kerr (2019) who believe corporate transparency is a good control; on the other end, there will be those on the side of Balakrishnan et al. (2019) who believe it is a bad

⁶ Kerr (2019) defines corporate transparency as availability of firm-specific information to outside investors. I use the same definition.

control. In the middle, however, there will be researchers who recognize the merit of each study and acknowledge two possibilities.

To help the analysis, let's assume there is a firm characteristic (another factor) that is associated with aggressive tax planning and firm value. The researcher should consider the underlying causal relationships to determine whether the firm characteristic is a good or a bad control. Assuming the two potential confounders, there are four credible models:

$$Firm\ value = a_1 + b_1 ATP + e_1 \quad (1)$$

$$Firm\ value = a_2 + b_2 ATP + c_2 FirmChar + e_2 \quad (2)$$

$$Firm\ value = a_3 + b_3 ATP + d_3 CT + e_3 \quad (3)$$

$$Firm\ value = a_4 + b_4 ATP + c_4 FirmChar + d_4 CT + e_4 \quad (4)$$

In some cases, the underlying causal nature of a variable might be so widely accepted that it is impractical to entertain an alternative possibility. For example, if it is widely accepted that the firm characteristic in question is a good control, it might be impractical to consider models that exclude it as credible. In that case, only model (2) and model (4) would be the relevant credible models.

2.2.2 Does Information Asymmetry Affect Conservatism?

The second example is borrowed from Bertomeu et al. (2016) and based on LaFond and Watts (2008), who argue and find that information asymmetry affects conservatism. They also find that size significantly reduces the strength of the relationship. After discussing various theories related to the role of size, they conclude that size should not be controlled for in their model.

As Bertomeu et al. (2016) note, depending on competing conditions, size may be a good control or a bad control. Specifically, size is a good control (or a confounder) if 1) it is a cause of

information asymmetry and of conservatism or 2) the two are “mechanically” associated.^{7,8} In either of the cases, failing to control for size may result in information asymmetry taking some of the credit of size.

Alternatively, size is bad control if 1) it is affected by information asymmetry such that it constitutes an indirect effect on conservatism, 2) it is an effect of conservatism (collider bias), or 3) it represents the same construct as information asymmetry or conservatism.^{9,10} Whited et al. (2021) provide detailed explanations for why each of the three cases would lead to bias. LaFond and Watts (2008) essentially claim case 3): size represents the same construct as information asymmetry and is a bad control. Some might have a hard time accepting that a commonly used control such as size could be a bad control that leads to bias. However, the appropriateness of size as a control depends on these underlying causal relationships, any of which may have theoretical or empirical justification.

Regarding the impact of size in LaFond and Watts (2008), Bertomeu et al. (2016) conclude that strong causal inferences should not be drawn because the theory relating size to information asymmetry and conservatism is ambiguous. In other words, there is uncertainty about whether the right model should include or exclude size, and the exposure estimates that drive causal inferences directly depend on which model is used. Claiming that controlling for size leads to conservative estimates is valid only if the indirect effect through size is of the same sign as the direct effect. If the direct and indirect effects are opposite in sign, controlling for size

⁷ Firm size may cause information asymmetry, consistent with the notion that larger firms attract analyst following, which leads to lower information asymmetry. Size may cause conservatism through the political cost hypothesis explanation (Watts & Zimmerman, 1978).

⁸ If there are unknown risk factors (or other factors) that drive firm size and information asymmetry, the resulting association between the two constitutes confounding.

⁹ The notion that information asymmetry affects size is parallel to the notion that information asymmetry affects cost of capital (Easley & O’Hara, 2004), which affects firm value.

¹⁰ The notion that conservatism affects size is consistent with the argument in Khan and Watts (2009) that conservatism results in cumulative understatement of firm value.

would suppress the indirect effect and overstate the effect of information asymmetry. In other cases, controlling for size may lead to collider bias or other “included variable biases” that can be as severe (Gow et al., 2016). If like before, there is a firm characteristic (another factor) that is associated with information asymmetry and conservatism, the resulting model space will consist of four credible models as before:

$$\text{Conservatism} = \theta_1 + \beta_1 \text{InfoAsym} + \varepsilon_1 \quad (1)$$

$$\text{Conservatism} = \theta_2 + \beta_2 \text{InfoAsym} + \alpha_2 \text{FirmChar} + \varepsilon_2 \quad (2)$$

$$\text{Conservatism} = \theta_3 + \beta_3 \text{InfoAsym} + \delta_3 \text{Size} + \varepsilon_3 \quad (3)$$

$$\text{Conservatism} = \theta_4 + \beta_4 \text{InfoAsym} + \alpha_4 \text{FirmChar} + \delta_4 \text{Size} + \varepsilon_4 \quad (4)$$

2.3 Decision Theory and the Recommended Action

Because of model uncertainty and the possibility that two or more estimates are credible, the researcher faces a decision about which estimate to report or rely on in reaching conclusions. On the one hand, if one estimate is reported and the others dismissed, the researcher faces the risk that one of the dismissed estimates is actually the valid estimate of the causal effect (i.e., it was derived from a model that includes good controls and excludes bad controls). Thus, it seems credible estimates should be reported in order to qualify statements made from any single estimate. On the other hand, how should multiple credible estimates be relied upon for decision-making? Such decisions with imperfect information are classic cases of decision-making under uncertainty.

I argue that relying on a specific estimate is rational if such action is recommended by an applicable decision theory, a formal approach to decision-making under uncertainty. In fact, many econometric approaches are justified by formal analysis involving decision theories. The benefit of formal analysis is that, even if researchers (as decision-makers) disagree on the

appropriate action recommended by the theory, the assumptions made will be clearly stated and open for discussion. Accordingly, I examine the extent to which the standard approach of relying on the preferred model is justified by a formal approach rooted in decision theory.

The notion of model uncertainty has been extensively examined in Bayesian empirical methods, although the purpose is generally related to prediction.¹¹ In particular, studies have noted that it is riskier to settle on a particular model instead of drawing predictive strengths by combining (averaging) multiple models (Draper et al., 1995; Hoeting et al. 1999; Kaplan 2021). The studies note that averaging models is justified by the expected utility within decision theory – the benefit accrued from using one model versus a combination of models (Kaplan 2021). Since Bayesian model averaging (BMA) is the solution to model uncertainty in the context of predictive inferences, I later discuss how it handles the uncertainty in the context of causal inferences, the focus of this study. Particularly, I note the limitations of standard BMA, namely the reliance on model fit, in addressing the uncertainty due to confounder adjustment. However, the approach I adopt is rooted in this extensive and tested line of research.

Studies consider expected utility theory as arguably the standard approach to decision-making under uncertainty (Gilboa et al., 2008). Focusing on the standard theory allows me to use formal reasoning that has likely received substantial support.¹² However, I consider a common alternative to expected utility. In particular, I examine whether focusing on the preferred model is justified by Wald’s maximin criterion. Many textbooks cover Savage’s utility theory and Wald’s criterion as two prominent theories of decision-making under uncertainty (Luce & Raiffa, 1957; Resnik, 1987; Savage, 1954).

¹¹ The studies stress that the overarching goal of statistics is prediction (See Kaplan, 2021).

¹² The expected utility theory has many applications that vary with the chosen objective, also known as loss functions. To the extent that multiple objectives can be rational, the theory allows for a variety of rational behaviour.

The problem of the decision-maker may be stated as follows: there is a state-space comprising possible representations of the true state of the world, a risk space comprising risks of any conceivable situation, and a preference relation. The preference is about which action (also called gamble) the decision-maker would choose among alternatives. Each action involves possible payoffs or risks depending on which state obtains.

The expected utility theory states what a decision-maker should do in order to maximize their utility given their beliefs about the states. In particular, the theory suggests using subjective probabilities about the possible states and choosing an action with the highest expected utility or the lowest expected risk (Luce & Raiffa, 1957; Savage, 1954).

A major benefit of using expected utility is that it promotes means-end rationality, with an objective and clear means about achieving the objective. More importantly, it makes for good policy based on the law of large numbers. Specifically, in the long run, the decision-maker chooses the act with the lowest risk (Briggs, 2014).¹³ Wald proposed an alternative to Savage's theory that does not depend on the subjective beliefs of the decision-maker. According to Wald's criterion, the decision-maker chooses the act with the lowest worst-case risk (Luce & Raiffa, 1957; Resnik, 1987). Wald's criterion is also desirable because it is easy to apply.¹⁴

2.4 Introduction to Model Uncertainty Simulations

In the following sections, I use simulated data to illustrate the causes and implications of model uncertainty and the application of the decision theories described above. In Case 1 and Case 2, I simulate data that correspond to the underlying causal relationships described in Appendix A. In particular, one of the cases illustrates the ambiguity about whether a control

¹³ In the long run, average risk is overwhelmingly likely close to the expected value. Thus, in the long run, the best action will involve the lowest risk (Briggs, 2014).

¹⁴ While Wald's maximin criterion is perceived as pessimistic and unrealistic in some scenarios, Cohen and Jeffray (1980) argue that it is rational for decision-makers under extreme ignorance.

mitigates OVB or whether it causes M bias while the other case illustrates the ambiguity about whether the control will suppress indirect effects. Since the whole point is that the researcher does not observe the relationships, I only disclose the statistical associations at this point. I disclose the actual directions of causality and other details in appendix A. It might be useful to defer looking at the actual data generating process in the appendix until the discussion of the specific case is finished; this would help to mirror the experience of the researcher when they look at the data.

The simulations should be applied to accounting examples with caution. In generating the data, I picked particular directions magnitudes of causality among variables, but the choices are only for illustration. When applied to accounting examples they do not indicate which side of the argument I identify with or whether the signs or magnitudes of the coefficients are consistent with particular theories. Finally, by using simulations, I isolate other problems, such as noise in proxies, missing observations, etc., that are important in empirical studies, but beyond the scope of this study. For the discussion, I also assume the researcher is honest and exercises due care.

3. Case 1: Simulation of Model Uncertainty

3.1 Invisible Underlying Process and Credible Models

In Case 1, variables z_1 and z_2 are associated with the exposure x_1 and the outcome y_1 . In the traditional approach, such associations would make them good controls that mitigate OVB. Based on earlier discussions, however, four different scenarios are quite possible: 1) both z_1 and z_2 are bad controls, 2) z_1 is a good control and z_2 a bad control, 3) z_1 is a bad control and z_2 a good control, or 4) both are good controls.

$$y_1 = \beta_{11} x_1 + e_{11} \quad (1)$$

$$y_1 = \beta_{12} x_1 + \delta_{12} z_1 + e_{12} \quad (2)$$

$$y_1 = \beta_{13} x_1 + \gamma_{13} z_2 + e_{13} \quad (3)$$

$$y_1 = \beta_{14} x_1 + \delta_{14} z_1 + \gamma_{14} z_2 + e_{14} \quad (4)$$

Linking this setup to the first accounting example, x_1 may represent aggressive tax planning and y_1 firm value; z_1 may represent the firm characteristic and z_2 corporate transparency. Since model 1 and model 2 do not control for z_2 , they would rely on corporate transparency being a bad control. The two models would be supported by the arguments and findings in Balakrishnan et al. (2019) that aggressive tax planning reduces corporate transparency. That is, z_2 is a mediator. On the other hand, model 3 and model 4 would rely on corporate transparency being a good control. Such a position would be supported by Kerr (2019).

For the second example, x_1 may represent information asymmetry and y_1 conservatism; z_1 may represent the firm characteristic and z_2 firm size. Since model 1 and model 2 would not control for size (z_2), they would assume it is a bad control. Alternatively, by including size, model 3 and model 4 would assume it is a good control.

3.2 OLS Results of Credible Models for Case 1

Panel A of Table 1 shows the correlation matrix for the simulated data in Case 1. A glance at the panel suggests a possibility for confounding since factors z_1 and z_2 are correlated with exposure x_1 and outcome y_1 .

A mediator is an outcome of x that also causes y . Examining the possibility that z_1 and z_2 are mediators in a purely statistical manner reveals contradicting results. In the first and second columns of Panel B, the independent variable is x_1 and the dependent variables are z_1 and z_2 respectively. The significant coefficients for x_1 in both columns would suggest that z_1 and z_2 are outcomes of x_1 or mediators (if x_1 causes y_1). In column 3, the dependent variable is x_1 and the independent variables are z_1 and z_2 . The significant coefficients for z_1 and z_2 would, in contrast, suggest that they are causes of x or confounders (if they are also associated with y_1). Therefore, additional information related to underlying causal relationships is needed in order to classify z_1 and z_2 with certainty.

Panel C of Table 1 shows OLS results of the four credible models. An estimate is valid if the model from which it is derived correctly classifies controls. The estimate of 0.706 in column 1 is valid if z_1 and z_2 are actually (in the true underlying process) bad controls. The estimate of 0.452 in column 2 is valid if z_1 is actually a good control and z_2 a bad control.

Interestingly, the estimate of the exposure in column 3, which includes a control for confounder z_2 , is bigger than the one in column 1, which does not include any control. This observation contrasts the argument made for including controls for the sake of conservatism in estimates. The upward bias arises because the relationship between the exposure x_1 and the potential confounder z_2 is negative (see Panel B). The same behavior is observed in column 4 relative to column 2, where the exposure estimate is bigger (instead of smaller) with more

controls. Nonetheless, exposure estimates are model-dependent because they vary with inclusion or exclusion of certain controls. Since the actual states of the variables are not observable, it is difficult to know with certainty which estimate to rely on.

The results can be readily linked to the first example; the second simulation case will be linked to the second example. The different estimates in the columns would suggest different effects of aggressive tax planning on firm value depending on whether the firm characteristic (z1) and corporate transparency (z2) are good controls. For comparison, the estimate in column 1 (0.706) assumes the firm characteristic and corporate transparency are both bad controls while the estimate in column 3 (0.765) assumes the firm characteristic is a bad control and corporate transparency a good control.

With respect to model uncertainty, the results in the two columns (1&3) represent two possible and opposing channels in the examined relationship. In one channel, if aggressive tax planning (ATP) reduces corporate transparency (CT) (Balakrishnan et al., 2019), omitting CT (z2) results in an estimate of the full effect of ATP (0.706): a direct positive effect consisting of saved tax dollars, netted by an indirect negative effect of opaque information environment (caused by complex tax-saving structures). This estimate of the full effect (from model 1) is valid if the described channel represents the underlying process. On the other hand, controlling for CT isolates the direct positive effect of ATP and suppresses the indirect negative, resulting in a partial (unnetted) effect of ATP. Not only would this partial estimate (from model 3) be invalid (to the extent the researcher is concerned with “the value” of tax planning), but it would also be anything but conservative. Consistent with Gow et al. (2016), including more controls does not result in conservative estimates when the exposure and the potential confounders are negatively related.

In the alternative channel, corporate transparency (CT) reduces aggressive tax planning (ATP). Thus, failing to control for CT (z_2) results in a negative bias and a smaller exposure estimate (0.706) from model 1. Controlling for it results in an unbiased larger estimate (0.765) from model 3. While the estimate is not conservative, it is the valid one if CT curbs ATP (Kerr, 2019). In short, the researcher needs to know which channel represents the underlying process in deciding whether controlling for CT is appropriate and will result in a valid estimate.

Without such knowledge, a prudent manager would consider each of the estimated effects (benefits) of ATP to alternative value-increasing activities, as tax evasion may also entail other intangible consequences (e.g., a tainted reputation) (Graham et al., 2014). In addition, while confidence intervals acknowledge some uncertainty by indicating a range of exposure estimates, distinct confidence intervals will be derived from the distinct credible models. Like exposure estimates, confidence intervals are model-dependent. Ultimately, model uncertainty and the presence of multiple credible estimates invariably impact causal inferences about economic significance.

3.3 Minimizing sampling error using average estimates over multiple simulations

The results in Panel A-C come from one sample, which is what researchers typically have. Thus, the estimates contain sampling error, which would be weighed against other econometric issues. To minimize the sampling error, I use averages of exposure estimates across 1000 simulations. Each of the 1000 simulations consists of the same relationships among factors specified by the equations in Appendix A and the same sample size of 10,000 observations, but a different random seed. I run regressions for the four credible models for each of the 1000 simulated samples, obtain exposure estimates, and find the averages (and the standard deviation) of those estimates. The result is one average exposure estimate for each of the four credible

models. Obviously, using 1000 samples is not possible in practice, but the goal here is to minimize sampling error when calculating errors and model risk in next section. Generally, researchers do not have to worry about sampling error if they have a large sample, and they can otherwise rely on methods that account for it.

Panel D shows the summary of exposure estimates for the 1000 simulated samples for Case 1. The average exposure estimates are 0.690, 0.440, 0.750, and 0.500 for model 1 through model 4, respectively. As expected, these average exposure estimates are very close to exposure estimates in Panel C, and the small differences are attributable to sampling error in Panel C.

3.4 Model Space as State Space and Potential Errors as Risks

It is important to note that the framework presented here is only intended to theoretically show how actions result in possible risks and which action results in the lowest risk using a loss function of interest. Thus, observations made are borrowed from prior theoretical work (e.g., expected utility and Bayesian statistics), and interested readers should refer to such work for elaborate proofs and other details.

As I have proposed, when there is model uncertainty, relying on a specific model to make important statements or conclusions is a course of action that should be evaluated using formal analysis of decision-making under uncertainty. Using the framework laid out by decision theories, the four credible models are possible states of the world whose occurrence the researcher is uncertain about. Each action of the researcher entails possible risks whose occurrence depends on which state is actually the true state.

The models (as states) and possible risks of each reporting action are depicted in Figure 1. The numbers are also summarized in Panel E of Table 1. In the panel, a row indicates which of

the model estimates (average of estimates across 1000 simulations) the researcher relies on; a column indicates which of the states obtains. Risks are computed as the absolute value of the difference between the estimate from rows (from the model the researcher relies on) and the estimate from the columns (from actual states). For instance, suppose the researcher chooses to rely on the estimate from model 4 (row 4). If z_1 and z_2 are actually bad controls (model 1 is true or column 1), the risk of this action is an (absolute) error of about 0.190 – the difference between the estimate of 0.500 from model 4 and the estimate of 0.690 from model 1, both estimates being averages across 1000 simulations. If z_1 is good control and z_2 a bad control (model 2 is true), the risk is an error of 0.060 – the difference between the estimate of 0.500 from model 4 and the estimate of 0.440 from model 2. The risks of this action are depicted in the bottom-four branches of the tree in Figure 1. In sum, the researcher faces various possible risks when they report or rely on a specific model, and the magnitude of the risk depends both on which estimate they report and on which estimate would be derived from the true model.

3.5 Deterministic Action: No Model Uncertainty

If the underlying causal relationships are perfectly known, the model representing the true state is certain, and others are irrelevant. For Case 1, if one researcher believes z_1 and z_2 are good controls without a doubt, the states represented in other models except model 4 are irrelevant to them. Their only logical action is to rely on the estimate of 0.500 from model 4 since they know the risk of this action is zero. Any other action is unjustified because it would knowingly involve a bigger risk.

Although the decision seems obvious in this manner, certainty is a state of belief of the researcher. That is, nothing prevents a different researcher from “feeling” certain that the true underlying process is represented by model 2 (z_1 is good and z_2 is bad). Therefore, if the

researcher picks an estimate from a single preferred model, they may be said to act as if they are certain about the underlying process. In other words, singling out the preferred model may be justified if the underlying causal relationships are perfectly known, or controls are clearly distinguishable.

3.6 Recommended Action Based on the Expected Utility

3.6.1 Illustrative Model Probabilities

Recall that according to expected utility, the decision-maker chooses an action with the lowest expected risk or highest expected utility after applying their subjective probability (beliefs) about the states. Thus, in order to use the theory, the researcher will need to quantify their beliefs about the relative goodness of the controls. Beliefs about the controls translate into beliefs about the models.

In this sub-section, I show how the researcher can use their subjective beliefs to derive model weights and compute expected values of risks. While a rationale is provided in quantifying the beliefs, the exercise is mostly illustrative.¹⁵

Suppose the researcher believes that z_1 is a good control (i.e., a confounder) and assigns a subjective probability of 0.7 to this belief. That is, they admit the possibility that z_1 is a bad control (e.g., a mediator or same construct) with a probability of 0.3. Suppose on the other hand, they believe z_2 is likely a bad control with a probability of 0.6 but admit the possibility it is a good control with a probability of 0.4.

Model weights can be derived directly from their beliefs about the goodness of controls. Illustrative model weights are depicted in Figure 2. For instance, the weight of model 1 is the probability that z_1 and z_2 are bad controls. If the goodness of z_1 does not affect that of z_2 (i.e.,

¹⁵ This method of deriving model weights is borrowed from Bayesian model averaging techniques.

independence), the weight of model 1 is the product of two probabilities: the probability that z_1 is bad ($1 - 0.7$) times the probability that z_2 is bad ($1 - 0.4$), or 0.18. The weight for model 2 is the probability that z_1 is good (0.7) times the probability that z_2 is bad ($1 - 0.4$), or 0.42. And so on.

In the current standard practice, the researcher relies on the model they believe is most likely to include good controls and exclude bad controls. Since model 2 has the largest weight (0.42) among the credible models, researchers will typically use it in the main analyses as such model. I refer to this model as the preferred model.

3.6.2 Recommended Action Assuming the Absolute Error Loss Function

Risks and their probabilities of occurrence are depicted in Figure 2. The probabilities of occurrence are in the boxes next to the respective risks. For instance, suppose again the researcher chose to rely on the estimate from model 4 (bottom-four branches). If z_1 and z_2 happened to be bad controls (model 1 is true, top of the bottom-four branches), the risk of this action would be an error of 0.190. The probability of that happening is 0.18, corresponding to the researcher's level of confidence (belief) that z_1 and z_2 are bad controls. If z_1 happened to be a good control and z_2 a bad control (model 2 is true, second of the bottom-four branches), the risk of the action would be an error of 0.060. The probability of that happening is 0.42, corresponding to the researcher's level of confidence that z_1 is a good control and z_2 a bad control. The expected value of the risks for relying on the estimate from model 4 is the sum of the possible risks times their probabilities of occurrence.

Table 1 Panel F shows that relying on the estimate (average across 1000 simulations) of 0.500 from model 4 (row 4) involves the smallest expected risk, a 0.089 error. It can be shown that the (weighted) median estimate will involve the smallest expected value when risks are measured as their actual magnitudes. Such function for measuring risks is called absolute error

loss function. The median estimate is found by sorting all credible estimates (in ascending order, by estimate magnitude) and picking the estimate in the middle based on the weights of the models; that is the estimate from the first model that crosses 0.5 in cumulative weight.

3.6.3 Recommended Action Assuming Other Loss Functions

So far it was assumed that decision-makers value risks using the absolute error loss function. An alternative valuation function is the squared error loss function, which measures risks as their squared magnitudes. The OLS estimator exemplifies the wide use of the function. Under the squared error loss function, the expected value of risks (Mean Squared Error or MSE) is minimized by the weighted average estimate.¹⁶

Table 1 Panel G shows the expected value of risks if the researcher assumes the squared error loss function. The weighted average of the exposure estimates (0.539) achieves the smallest expected value of the risks (about 0.014). In contrast, the expected value of the risk of 0.024 for the preferred model (model 2) is higher. Therefore, singling out the preferred model is not necessarily justified if the researcher uses expected utility and assumes a standard loss function.

Another loss function relevant to reporting choices is the zero-one loss, generally used in classifying categorical outcomes. The function assigns 1-point risk (penalty) to the wrong classification and 0 to the right classification.

Before turning to the recommended action, it would be helpful to refer back to the actual magnitudes of risks from reporting a given estimate (Panel E of Table 1). If the researcher relies on the estimate from model 4, the risk of the action is 0.190 when z_1 and z_2 are actually bad

¹⁶ In many applications, using Savage's theory (expected utility) implies a weighted average because the squared error loss function is the default function in those applications. But as the prior subsection shows the theory recommends the median estimate when the absolute error loss function is assumed.

controls (column 1); the risk of the action is 0.060 when z1 is actually good control and z2 a bad control (column 2). Note the difference in magnitudes of the risks.

Panel H shows the recommended action assuming zero-one loss. Refer to the expected value of the risk for relying on the estimate from model 4 (row 4). The panel shows that the weights of model 1 (column 1) and model 2 (column 2) are both multiplied by 1 (the risk or penalty). This is because the function treats a 0.190 error and a 0.060 error equally, even though they are different in magnitudes (or severity). The panel shows that the recommended action is to rely on the estimate of 0.440 from model 2. Recall that model 2 has the largest weight among all the credible models (i.e., the preferred model). It can be shown that assuming the zero-one loss function, the recommended action is always the estimate from the model with the largest weight (the mode estimate).¹⁷ Therefore, it may be said that focusing on the preferred model under uncertainty is implied by expected utility assuming the zero-one loss function.

3.6.4 Recommended Course of Action Based on Wald's Maximin Criterion

According to Wald's maximin criterion, the researcher should be concerned with the worst-case risk, the largest possible error in deciding which estimate to rely on.

As Table 1 Panel E showed, when the researcher relies on the estimate from model 1 (row 1), their worst-case risk is an error of 0.250 (row 1, column 2); this is when z1 is a confounder but z2 is not in the underlying process. The worst-case risks are 0.310 for the estimate from model 2, 0.310 for the estimate from model 3, and 0.250 for the estimate from model 4. Since relying on the estimate from model 1 (0.690) and the estimate from model 4 (0.500) tie for the least worst-case risks (0.250), the researcher should rely on both or use another

¹⁷ The mode estimate is one derived from the model with the highest weight. The preferred model, as defined in this study, is generally such model.

criterion to break the tie. One important observation is that Wald's criterion does not recommend the action based on the researcher's belief about the relative goodness of the models.

4. Case 2: Other Model Uncertainty

4.1 Invisible Underlying Process and Credible Models

In the second simulation, I create data to highlight some of the uncertainty that remains when the researcher can determine the relative timing of measurement. The simulation also allows me to show other instances when reliance on the preferred model is likely to be inappropriate. As before, it might be useful to defer looking at the actual data generating process in the appendix until the discussion of this case is finished.

The relationship of interest is between exposure x_2 and outcome y_2 . There are two potential confounders z_3 and z_4 , both associated with x_2 and y_2 . Variables z_3 and z_4 are measured before x_2 and y_2 (other details of the process will be revealed in Appendix A). Assume the researcher knows this relative timing (without such knowledge, they would obviously be under more model uncertainty). However, even with this knowledge, z_3 or z_4 may be a variable that causes M bias or one that amplifies residual bias.

In the traditional approach, the statistical associations would suggest z_3 and z_4 are important controls that mitigate OVB. The pre-treatment criterion would also suggest they are good controls. However, to the extent the other biases may arise even when relative timing has been identified, four different scenarios are possible: 1) z_3 and z_4 are both bad controls, 2) z_3 is a good control and z_4 a bad control 3) z_3 is a bad control and z_4 a good control, or 4) both are good controls:

$$y_2 = \beta_{21} x_2 + e_{21} \quad (1)$$

$$y_2 = \beta_{22} x_2 + \delta_{22} z_3 + e_{22} \quad (2)$$

$$y_2 = \beta_{23} x_2 + \lambda_{23} z_4 + e_{23} \quad (3)$$

$$y_2 = \beta_{24} x_2 + \delta_{24} z_3 + \lambda_{24} z_4 + e_{24} \quad (4)$$

4.2 OLS Results of Credible Models for Case 2

Panel A of Table 2 shows the correlation matrix for y_2 , x_2 , z_3 , and z_4 . Note z_3 and z_4 are correlated with x_2 and y_2 . If the concern were only about OVB, both variables would be included as good controls, given the correlations.

Panel B shows what reliance on statistical associations alone would suggest (i.e., if they were sufficient for adjusting for confounders). In the first and second columns of Panel B, the independent variable is x_2 and the dependent variables are z_3 and z_4 respectively. The significant coefficients for x_2 in both columns would suggest that z_3 and z_4 are outcomes of x_2 or mediators (if x_2 causes y_2). In the third column, the independent variables are z_3 and z_4 and the dependent variable is x_2 . The significant coefficients for z_3 and z_4 would suggest that both factors are (unlike in the first two columns) causes of the exposure x_2 . Based on statistical associations alone, it is unclear whether z_3 or z_4 are good controls.

Panel C shows various estimates of the exposure x_2 after adjusting for z_3 , z_4 , neither, or both. As before, an estimate is valid if the model from which it is derived correctly classifies the controls. The estimate of 1.069 in column 1 is valid if z_3 and z_4 are actually bad controls. The estimate of 0.801 in column 2 is valid if z_3 is actually a good control and z_4 a bad control, and so on. Since the actual states of the variables are not observable, it is difficult to know with certainty which estimate to rely on.

If the researcher actually knows z_3 and z_4 are measured prior to x_2 (assumed known), they may rule out some possibilities. With respect to the second accounting example, they may conclude that since size leads (is measured before) information asymmetry, then size could not be an indirect effect of information asymmetry or a mediator. Note, however, that this is only one theory as LaFond and Watts (2008) maintain size and information asymmetry actually

represent the same construct. With mediator-bias ruled out, size and the firm characteristic would have a greater chance of being confounders. Then, controlling for both would suggest placing more weight on inferences based on model 4. In other words, the pre-treatment criterion may help in classifying controls, consistent with Whited et al. (2021). However, knowledge about the relative timing is not sufficient to rule out the other biases (e.g., M bias). Therefore, model 4 does not necessarily get all of the weight in drawing inferences.

In the context of the second example, columns in Panel C show different estimated effects of information asymmetry (x_2) on conservatism (y_2), depending on whether the firm characteristic and size are good controls.¹⁸ Although only two potential confounders are considered, in practice many more are considered. Therefore, on the margin, whether size is included will affect whether or the degree to which the researcher concludes there is a non-trivial effect of information asymmetry on conservatism.

The findings of the examination above have further implications. For example, in examining the effect of information asymmetry on cost of capital (Easley & O' Hara, 2004; Francis et al., 2004), one will need to consider both the direct effect of information asymmetry, which is likely positive, and the indirect effect of conservatism, which is likely negative. However, the prior analysis suggests the indirect effect of information asymmetry through conservatism materially depends on whether size was a good or bad control in the relationship between information asymmetry and conservatism. In other words, whether size is a good control impacts the extent to which the direct effect of information asymmetry on cost capital

¹⁸ It would be helpful to think of the control for firm size as being multiplied by -1 and y_2 , x_2 , z_3 , z_4 (size) being all positively correlated, since size is generally understood as being negatively related to information asymmetry and possibly conservatism.

outweigh the indirect effect through conservatism. Thus, both the significance and the economic significance of exposure estimates matter in inferences that can be drawn.

4.3 Summary of multiple simulations and risks for Case 2

Panel D shows the summary of exposure estimates for the 1000 simulated samples for Case 2. I follow the same procedure as in section 3.3. The average exposure estimates are 1.066, 0.799, 0.975, and 0.599 for model 1 through model 4, respectively. As the panel shows, the average exposure estimates are very close to exposure estimates in Panel C, where the small differences are attributable to sampling error in Panel C.

Panel E shows the risks of relying on a given credible estimate. As before, it is important to note that the framework presented here is only theoretical and has been shown in prior work. The risks are calculated based on average estimates from Panel D.

To illustrate the deterministic course of action, assume the timing of measurement leads the researcher to classify z_3 and z_4 as good controls with certainty (model 4 is true, column 4). In that case, it is clear that relying on the estimate of 0.599 from model 4 is the recommended course of action since it involves zero risk (row 4, column 4). Relying on the estimate of 1.069 from model 1, for example, would knowingly involve a risk of 0.467 (row 1, column 4).

Therefore, by singling out one model, the researcher may be said to assume the causal relationships are perfectly known. However, the certainty of the researcher does not necessarily reflect the actual underlying causal relationships.

4.4 Recommended Course of Action Based on Expected Utility

4.4.1 Model as States of the World with Illustrative Probabilities

Suppose that the researcher applies their knowledge about the variables besides variables' relative timing. For example, the researcher may have knowledge about earlier causes of z_3 (e.g.,

from prior studies) that increases their confidence in classifying it as a good control. In light of the information, they may assign a 0.9 probability to z3 being a good control. While this is a very high probability, the researcher admits some possibility that predetermined variables may actually be bad controls.

On the other hand, suppose they know less about z4. For example, on top of knowing z4 is measured before y2, they may only theoretically link it to x2 but not to y2. Based on this limited knowledge, they may assign a 0.6 probability to adjusting for z4 as a good control. Conversely, they assign a probability of 0.4 to not adjusting for z4.

The resulting model probabilities are 0.04 $[(1-0.9)*(1 - 0.4)]$ for the model adjusting for neither, 0.36 for the model adjusting for z3 but not z4, 0.06 for the model adjusting for z4 but not z3, and 0.54 for the model adjusting for both. Since model 4 (M4) has the largest weight, in the current standard practice it would be considered the preferred model. The tiny weights for models 1 (0.04) and 3 (0.06) illustrate how models that exclude near-certain confounders (i.e., z3) will naturally have trivial weights. This is reassuring because as background information becomes perfect, the true model will have all of the weight.

4.4.2 Recommended Action Assuming the Absolute Error Loss, Squared Error Loss, and Zero-one Loss Functions

Table 2 Panel F shows the risks of reporting an estimate if the researcher uses the expected utility and assumes the absolute error loss function. The expected value of the risks is the sum of risks (from Panel E) times their probabilities of occurrence. The panel shows that reporting or relying on the estimate of 0.599 from model 4 is the recommended action since it involves the smallest expected value of the risks (0.113). This is the median estimate given the model weights. Note also that the estimate is from the preferred model. This is because the

model has a weight of 0.54, which crosses 0.5 in cumulative weight (after sorting the estimates). Generally, if the preferred model has a weight of 0.5 or more, the estimate derived from it is automatically the median. Thus, it may be said that singling out the preferred model is justified by expected utility assuming the absolute error loss function, provided the preferred model has a far bigger weight.¹⁹

Panel G of Table 2 shows the risks of relying on an estimate if the researcher uses the expected utility and assumes the squared error loss function. Risks are squared before being multiplied by model weights. The last row indicates that reporting the weighted average estimate of 0.713 involves the lowest expected value of risks (0.019). Therefore, it is the recommended course of action. It is recommended over the estimate from the preferred model (model 4), which has a higher expected value of the risks (0.032).

Panel H shows respective results assuming the zero-one loss function. The mode estimate of 0.599 from the preferred model is the recommended action, as it involves the lowest expected value of the risks (0.460).

4.5 Recommended Action Based on Wald's Maximin Criterion

Table 2 Panel E showed that the worst-case risk if the researcher relies on the estimate from model 1 (row 1) is a 0.467 error (row 1, column 4); this is if z_1 and z_2 are actually confounders based on the underlying causal process. The worst-case risks are 0.267 for the estimate from model 2, 0.376 for the estimate from model 3, and 0.467 for the estimate from model 4. According to Wald's criterion, it is recommended to rely on the estimate of 0.799 from model 2 (row 2), which has the smallest worst-case risk (error of 0.267).

¹⁹ The expected utility approach and the absolute error loss function automatically recommend the median.

5. Discussion

5.1 Basis for Singling out the Preferred Model

In Case 1 and Case 2, it was shown that singling out one model is the recommended and obvious action if information about the underlying causal relationships is perfect. Specifically, relying on estimates from other models would knowingly involve (larger) unwanted risks. However, since many factors are interrelated within accounting research, having an accurate structure of the complex underlying causal relationships is implausible. This study also showed how remedying strategies may not remove the uncertainty. Therefore, singling out a model on the basis that controls are clearly distinguishable is generally unwarranted and the researcher may mask the risk of the statements they make. Without using formal analysis, they also forgo the chance of relying on estimates that may involve lower risks.

After admitting model uncertainty and using a formal analysis, it was shown in Case 1 and Case 2 that singling out the preferred model is justified in limited circumstances. First, it was shown that the expected utility theory, the standard approach, generally does not recommend singling out the preferred model. Specifically, Case 2 showed that assuming the absolute error loss function, singling out the preferred model is recommended if the model is far better than the other credible models. In that case, the estimate derived from it was guaranteed to be the median (which always involves the lowest risk for the function). Since expected utility is the standard approach to decision making under uncertainty and the absolute error loss function is commonly used for continuous measures, I discuss further whether this circumstance would justify treating the preferred model as the default under uncertainty.

If only two models are credible, the researcher will often be more confident about one over the other. In that case, relying on the estimate from the better model is clearly justified, as

its weight is automatically greater than 0.5, and the estimate derived from it automatically the median. However, as the number of potential confounders increase, so too does the number of credible models. When this happens, the preferred model is less likely to be far better (i.e., have a weight of 0.5 or more). For example, if there are 4 potential confounders and the researcher has a 0.8 belief (confidence) about each, the preferred model will have a weight of about 0.4 ($0.8 \times 0.8 \times 0.8 \times 0.8$).²⁰ Suppose now there are 12 potential confounders. If the researcher has perfect information about 8 of them and a 0.8 belief for each of the remaining 4, the preferred model will again have a weight of about 0.4. In short, even for solid information about causal relationships among the potential confounders, singling out the preferred model is unlikely to be supported by expected utility and the absolute error loss function.

Furthermore, it was shown in both cases that the recommended action is automatically to report the estimate from the model with the highest weight when the zero-one loss is assumed (even if the weight is less than 0.5). This is because the mode estimate (from the model with highest weight) minimizes risk for the zero-one loss. Therefore, the researcher only needs to have a model about which they are most confident when the zero-one loss function is assumed. However, assuming such function implies that an error of 5% the size of the true exposure effect and an error of 20% should be treated equally. This makes the function inappropriate for continuous exposures.

Finally, it was shown that the Wald's criterion does not recommend reliance on the preferred model unless doing so involves the smallest worst-case risk. However, the estimate

²⁰ This preferred model may consist of all of the four potential confounders ($0.8 \times 0.8 \times 0.8 \times 0.8$), where the probabilities are about the controls being good. It may also consist of two potential confounders ($0.8 \times 0.8 \times 0.8 \times 0.8$), where the first two probabilities are about the first two controls being good and the last two probabilities are about the last two controls being bad.

from the preferred model does not have a unique chance of achieving this objective, as the criterion does not take advantage of the relative goodness of the models (i.e., model weights).

5.2 Background Information as Input for Model Weights

In examining causal relationships, Rubin (2007) argues that in order to mirror randomized experiments researchers should design models using only background information. He insists that model design should be done prior to seeing any outcome data. It is a similar position that Whited et al. (2021) take.²¹ They note that model fit does not necessarily indicate model quality, as bad controls may contribute to the explanatory power of the model. Consistent with the two studies, Case 1 showed adding control z1 or z2 increases fit even though it's unclear which is a good control. Thus, including a control for corporate transparency in the model examining the effect of aggressive tax planning on firm value could increase the fit of the model, even if corporate transparency were a mediator (or suppressor) and a bad control. Similarly, in case 2, including a control for size may increase the fit of the model examining the relationship between information asymmetry and conservatism even if size were a bad control.

Perhaps the primary source of background information for the researcher is the related studies on the subject. Thus, if more studies support the position that corporate transparency curbs aggressive tax planning, the researcher should have more confidence in classifying corporate transparency as a good control. However, reliance on prior studies should depend on the quantity and quality of the studies, which includes how well the studies considered underlying causal relationships.

Background information and confidence in classifying controls should also depend on the difficulty of determining the relative timing of controls. When the researcher has determined the

²¹ Consistent with both studies, adding predictors of the outcome that are not associated with the exposure will automatically increase model fit even if they do not adjust for confounding.

relative timing, they should still consider the possibility that a predetermined variable is a bad control (e.g., one that causes M bias or amplifies some residual biases).

Another potential source of background information is the researcher's industry experience. It is known that industry experience gives rise to many research ideas, suggesting it is a big part of the researcher's toolbox. It may, in fact, complement and, in some respects, compete with their reliance on prior studies. Finally, workshop discussions, peers' comments, and other feedback can be used to form or modify the beliefs about the goodness of controls. To some degree, such discussions already contribute to the models of published studies.

5.3 On Quantifying Beliefs

In using the expected utility approach, the challenge is that quantifying beliefs allows researchers much wiggle room that can influence which weighted average estimate they obtain. Specifically, the theory does not place constraints on the subjective probabilities that represent beliefs, and as such, some beliefs will be inappropriate (Gilboa et al., 2008).

Savage (1954) admitted that personal probability is flawed since it is difficult to determine probabilities (e.g., a certain person will be the next president) with any degree of accuracy. However, he added that interesting and useful theories of modern science (e.g., geometry, relativity, quantum mechanics) are inexact, albeit to different degrees.

Similarly, Gelman and Hennig (2017) argue that instead of using objective and subjective qualifications, it might be more fruitful to consider attributes inherent in the two interpretations of probability. They also note that subjectivity is otherwise accepted in many contexts, such as the treatment of influential observations (e.g., winsorizing versus trimming at 1% or 5%), missing values, and even p values. In other cases, judgment is left at the hands of the softwares that set default parameters. They suggest that objectivity implies attributes of consensus,

transparency, and correspondence to observable reality, while subjectivity implies attributes of awareness of multiple perspectives. Since researchers may have different perspectives about the underlying causal relationships (even after relying on the same prior literature), their beliefs are naturally applied using subjective probability.

On the other hand, it is possible to reduce researchers' degrees of freedom in applying subjective probability by requiring a range of probabilities or multiple priors (Gilboa & Schmeidler, 1989).

Towards further making (model) weights objective, it is worth examining the extent to which data can be used to update weights using the Bayesian formula. In theory, the researcher may quantify their prior beliefs based on their judgment about the likely timing of variables or their industry experience, then use the relative number of studies supporting their position as the likelihood function to update their beliefs. However, based on the "no outcome data in sight" principle, it would be inappropriate to update the weights of the models based on how well the models fit the outcome data. Bayesian model averaging techniques invariably favor models with better fit regardless of whether the models include bad controls. In other words, they do not resolve model uncertainty. Thus, observations required for updating prior beliefs using the Bayesian formula should relate to background information, not model fit.

Furthermore, the Bayesian formula is not the only way beliefs are updated. In particular, beliefs may change after hearing new arguments. For example, it is possible that, upon discussion with peers, the researcher has a weakened belief that x_1 causes z_1 . Aragónes et al. (2005) argue that, due to complexity in some situations, "fact-free" learning (which does not rely on hard data) is sometimes warranted.

6. Solution

6.1 Full Disclosure And Formal Analysis

Using simulations, I show that singling out the preferred model is generally not supported by formal analysis based on decision theories. Therefore, in order to properly account for model uncertainty, it is necessary to consider the broader range of credible estimates. When the researcher would rely on a single estimate, they should consider one that is justified by formal analysis and is likely to involve less risk in statements made.

Although studies show the sensitivity of results to alternative model specifications, the practice does not necessarily address model uncertainty. In a hypothetical example, a main result may consist of bid-ask spreads to proxy for information asymmetry – the exposure where conservatism is the outcome – while an additional result consists of abnormal trading volume as the proxy for information asymmetry. While the combined results provide robustness to the appropriateness of proxies, they would not address whether controlling for factors, such as size, is appropriate or whether doing so might lead to bias. That is because model uncertainty is a result of limited knowledge about the underlying causal process relating the exposure and the outcome to other factors even when proxies are ideal.

Certainly, when proxies are not ideal, researchers will have a harder time identifying the underlying causal process. Consider, LaFond and Watts (2008), which states that the reason to exclude size in controls is the concern that size might represent the same construct as information asymmetry. If information asymmetry and size are well measured, the researcher will have a better understanding of how they relate to each other. However, better measurement will not always indicate whether information asymmetry affects size or vice versa, or whether an

entirely different factor affects both.²² Thus, by stating the rationale for excluding size, LaFond and Watts (2008) is one of the better examples that separate model uncertainty from other econometric issues and is able address it explicitly.

In other cases, studies may show results of extended models to imply that even conservative estimates support the hypotheses. However, appealing to conservatism of estimates is valid in relation to potential mediator bias only when direct and indirect effects are of the same sign and invalid to collider or other “included variable” biases (Gow et al., 2016). Thus, without specifying the sources and the seriousness of the uncertainty in a particular study, it is not clear the extent to which the sensitivity analyses in a given study actually address the uncertainty or whether they address a different problem.

The effectiveness of sensitivity analyses (and robustness tests) also depends on what inferences the study seeks to make. For example, if exposure estimates do not flip signs or become insignificant in different sets of models, one might reasonably make inferences about the existence and the direction of an effect. On the other hand, differences in estimated effects will invariably imply differences in economic significance and policy considerations. Failure to consider such differences may result in unbalanced budgets, misallocation of resources, and other real inefficiencies. In the current practice, after including sensitivity analyses (and laying out all the specifications they considered), researchers tend to pick one model from which to draw conclusions. For example, while LaFond and Watts (2008) discuss the resulting credible estimates depending on whether size is included, they rely on only the model that excludes size

²² For example, in audit research, even if restatements adequately proxied for misstatements, audit researchers would still need to understand the underlying relationship between misstatements, governance factors, and other firm characteristics when examining the relationship between misstatements and a factor of interest (e.g., Big 4 auditor). In medical terms, a doctor might clearly identify or measure a patient’s condition (i.e., exposure) and lifestyle (i.e., factor) without understanding how they relate to each other (e.g., whether the lifestyle caused the condition).

to draw their causal inferences.²³ Consistent with this notion, Ohlson (2021) notes that researchers sometimes fail to qualify their findings, even when the findings involve uncertainty.

To properly account for model uncertainty, researchers would need to disclose relevant credible estimates and the level of uncertainty involved in choosing the models from which the estimates were derived. At the same time, they should account for every relevant credible estimate in drawing causal inferences from the study.

An alternative solution to disclosing relevant credible estimates is to choose an applicable decision theory and rely on the estimate the theory recommends. There are many theories the researcher can choose from, with Savage's theory of expected utility being the standard in similar applications. Specifically, the researcher may use the expected utility approach and a standard loss function, like the squared error loss. In practice, they would integrate credible models by reporting or relying on the weighted average (median) of exposure estimates derived from the models. By doing so, the researcher takes into account their confidence about the relative goodness of each credible model. The approach is particularly desirable because it guarantees that in the long run the action with minimum risk is chosen.

With respect to Wald's criterion, the risks of relying on specific exposure estimates from one sample are confounded with sampling error. Accordingly, the calculated risks are also confounded with the error, which will influence what estimate (or credible model) the criterion recommends. However, when the sample size is large, the sampling error is negligible. Accounting for sampling error when applying Wald's criterion is left for future research.

As a caveat, full disclosure, the weighted average of exposure estimates, and the Wald's criterion do not remove model uncertainty. As such, neither solution purports to reveal the true

²³ The argument that conclusions in LaFond and Watts (2008) are too strong given the understanding of the underlying relationships is borrowed from Bertomeu et al. (2016).

exposure estimate. The solutions are, in fact, predicated on the notion that sometimes model uncertainty cannot be removed. However, the solutions would guide the researcher's actions in those cases.

6.2 A More Practical Solution

Unfortunately, consideration of all credible estimates may be unpractical. For example, if there are four doubtful confounders, the model space will consist of sixteen credible models. Moreover, for some of the controls, the level of uncertainty may not be significant enough to warrant such detailed consideration. Therefore, a tradeoff between integrating model uncertainty and practicality is needed.

As a practical solution, researchers should at the minimum identify cases when the uncertainty is severe. This means disclosing the resulting credible estimates and basing inferences on each without signaling a preference. Equal treatment of exposure estimates is necessary in this case because severe uncertainty (or ambiguity) about a control implies the model that includes the control and one that excludes it are equally good. On the other hand, if the collective research strongly suggests the control is good (bad), a compromise might lead to considering only the model that includes (excludes) the control.

An unexpected benefit of the compromise is that the choice of model weights becomes trivial. As the uncertainty increases, the weights of the models tend to a uniform distribution. Accordingly, the model-weighted averaging, implied by expected utility and the squared error loss function, approximates to a simple averaging of credible estimates. The model-weighted median, implied by expected utility and the absolute error loss function, approximates to a simple median. This helps address how to integrate model uncertainty in causal inferences without relying on substantial subjectivity from the researcher.

6.3 Future Considerations

A limitation of the proposed compromise is that consideration of only extreme levels of uncertainty leads to losing relevant information from the less credible models. While allowing researchers to use judgment in assigning model weights increases their degrees of freedom, it is possible to set standards in quantifying beliefs. As discussed earlier, subjective judgements (e.g., use of p values or whether to winsorize/trim and at which percentages) are commonly used in empirical methodology. One of the reasons such practices are acceptable is that there is consensus around cut-off points (e.g., 1% or 5%) even though it is understood that a different cut-off could change conclusions made (Gelman & Hennig, 2017). In the same vein, future studies could help set standards in how the researcher can quantify their judgment in applying the expected utility theory to address model uncertainty.

For example, if the researcher is ambiguous about whether a control is good or bad, the weight for including the control could be set at 0.5. If they are certain (and most if not all of their peers would agree) that a control is good, the weight for including the control could be set at 1. If they are confident (and the majority of their peers would agree that a control is good) the weight for including the control could be set at 0.75. In this way, the researcher may not arbitrarily assign a weight of, say, 0.8, and the consensus would limit the researcher's degrees of freedom.

7. Conclusion

Most investors would want to know all the relevant possible outcomes from an investment, not just the most likely. This is because their decisions could change if they learned a separate outcome is also possible. So too in research, readers might want to know the other credible estimates before relying on a study for decision-making. In investments, the expected return concept is an ideal solution when other returns are possible; it is the certainty equivalent justified by expected utility and the squared error loss function under uncertainty. Similarly, while the preferred model may carry more weight, the researcher should use formal analysis to recommend which estimate readers should rely on. Otherwise, when readers look at a study, they may assume controls were clearly distinguishable or that singling out the preferred model is automatically justified under model uncertainty.

References

- Aragones, E., I. Gilboa, A Postlewaite, and D. Schmeidler. 2005. Fact-free learning. *American Economic Review* 95 (5): 1355 – 1368.
- Balakrishnan, K., J. L. Blouin, and W. R. Guay. 2019. Tax aggressiveness and corporate transparency. *The Accounting Review* 94 (1): 45 – 69.
- Bertomeu, J., A. Beyer, and D. J. Taylor. 2016. From casual to causal inference in accounting research: The need for theoretical foundations. *Foundations and Trends in Accounting* 10 (2-4): 262 -313.
- Briggs, R. 2014. Normative theories of rational choice: Expected utility. *Stanford Encyclopedia of Philosophy Archive*.
- Cohen, M., and J. Jeffray. 1980. Rational behavior under complete ignorance. *Econometrica* 48 (5): 1281 – 1299.
- Ding, P, and L. Miratrix. 2014. To adjust or not to adjust? Sensitivity analysis of m-bias and butterfly-bias. *Journal of Causal Inference* 3 (1): 41 – 57.
- Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society* 57: 55–98.
- Easley, D. and M. O’Hara. 2004. Information and the cost of capital. *Journal of Finance* 59 (4): 1553 – 1583.
- Francis, J., R. Lafond, P. M. Olsson, and K. Schipper. 2004. Cost of equity and earnings attributes. *The Accounting Review* 79 (4): 967 – 1010.
- Gelman, A., and C. Henning. 2017. Beyond subjective and objective in statistics. *Journal of the Royal Statistical Society: Statistics in Society* 180: 967 – 1033.
- Gilboa, I. A. W. Postlewaite, and D. Schmeidler. 2008. Probability and uncertainty in economic modeling. *Journal of Economic Perspectives* 22 (3): 173 – 188.
- Gilboa, I., and D. Schmeidler. 1989. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18 (2): 141 – 153.
- Glymour, M. M., J. Weuve, and J. T. Chen. 2008. Methodological challenges in causal research on racial and ethnic patterns of cognitive trajectories: Measurement, selection, and bias. *Neuropsychology Review* 18: 194 – 213.
- Gow, I.D., Larcker, D.F. and Reiss, P.C. 2016. Causal inference in accounting research. *Journal of Accounting Research* 54: 477 – 523.

- Graham, J. R., M. Hanlon, T. Shevlin, and N. Schroff. 2014. Incentives for tax planning and avoidance: Evidence from the field. *The Accounting Review* 89 (3): 991 – 1023.
- Greenland, S., J. Pearl, and J. Robins. 1998. Causal diagrams for epidemiologic research. *Epidemiology* 10 (1): 37 – 48.
- Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. 1999. Bayesian model averaging: A tutorial. *Statistical Science* 14: 382 – 417.
- Kaplan, D. 2021. On the quantification of model uncertainty: a Bayesian perspective. *Psychometrika* 86: 215 – 238.
- Kerr, J. 2019. Transparency, information shocks, and tax avoidance. *Contemporary Accounting Research* 36 (2): 1146 – 1183.
- Khan, M., and Watts, R.L. 2009. Estimation and empirical properties of a firm-year measure of accounting Conservatism. *Journal of Accounting and Economics* 48 (2-3): 132 – 150.
- LaFond, R., and Watts, R. L. 2008. The information role of conservatism. *The Accounting Review* 83 (2), 447–478.
- Luce, R. D., and H. Raiffa. 1957. *Games and Decisions: Introduction and Critical Survey*. Dover Publications, Inc. New York, NY.
- Mackinnon, D. P., J. L. Krull, and C. M. Lockwood. 2000. Equivalence of the mediation, confounding and suppression effect. *Prevention Science* 1 (4): 173 – 181.
- Myers, J. A., J. A. Rassen, J. J. Gagne, K. F. Huybrechts, S. Schneeweiss, K. J. Rothman, M. M. Joffe, and R. J. Glynn. 2011. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *American Journal of Epidemiology* 174 (11): 1213 – 1222.
- Ohlson, J.A. 2021. Researchers’ data analysis choices: an excess of false positives? *Review of Accounting Studies* (2021).
- Pearl, J. 1995. Causal diagrams for empirical research. *Biometrika* 82 (4): 669 – 688.
- Pearl, J. 2012. On a class of bias amplifying variables that endanger effect estimates. Working paper.
- Resnik, M. 1987. *Choices: An Introduction to Decision Theory*. University of Minnesota Press. Minneapolis, MN.
- Rubin, D. B. 2007. The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Statistics in Medicine* 26 (1): 20 – 36.
- Savage, L. J. 1954. *Foundations of Statistics*. Dover Publications, Inc. New York, NY.

- Shpitser, I., T. VanderWeele, and J. M. Robbins. 2012. On the validity of covariate adjustment for estimating causal effects. Working paper.
- Sjölander, A. 2009. Letter to the editor: Propensity scores and m-structures. *Statistics in Medicine* 28: 1415 – 1424.
- VanderWeele, T. J. 2019. Principles of confounder selection. *European Journal of Epidemiology* 34 (3): 211 – 219.
- VanderWeele, T. J., and I. Shpitser. 2011. A new criterion for confounder selection. *Biometrics* 67 (4): 1406 – 1413.
- Watts, R. L., and Zimmerman, J. L. 1978. Towards a positive theory of the determination of accounting standards. *The Accounting Review* 53 (1): 112–134.
- Whited, R. L., Q. T. Swanquist, J. E. Shipman, and J. R. Moon. 2021. Out of control: The (over) use of controls in accounting research. *The Accounting Review*, forthcoming.

Table 1: Case

Panel A: Correlation Matrix for Case 1

	y1	x1	z1	z2
y1	1.000	0.487	0.468	0.009
x1	0.487	1.000	0.704	-0.270
z1	0.468	0.704	1.000	-0.191
z2	0.009	-0.270	-0.191	1.000

Panel A shows the correlation matrix for one sample with 10,000 observations. This is the random sample the researcher gets during data collection.

Panel B: Auxiliary Regressions for Case 1: Statistical Equivalence of Controls

	<i>Dependent variable</i>		
	z1 (1)	z2 (2)	x1 (3)
x1	0.498*** (0.005)	-0.302*** (0.011)	
z1			0.958*** (0.010)
z2			-0.126*** (0.006)
Constant	0.008 (0.007)	-0.004 (0.015)	-0.012 (0.010)
Observations	10,000	10,000	10,000
Adjusted R ²	0.496	0.073	0.515

This panel shows the equivalency between good and bad controls. Based on statistical associations alone, columns 1 and 2 test whether z1 and z2 are mediators (outcomes of x1) or bad controls; column 3 tests whether z1 and z2 are confounders (cause x1) or good controls. *, **, and *** indicate a statistically significant difference from zero at the 10%, 5%, and 1% level, respectively.

Panel C: OLS Results of Credible Models for Case 1

<i>Dependent variable:</i>				
y1				
	(1)	(2)	(3)	(4)
x1	0.706*** (0.013)	0.452*** (0.017)	0.765*** (0.013)	0.511*** (0.018)
z1		0.510*** (0.025)		0.511*** (0.024)
z2			0.197*** (0.012)	0.197*** (0.011)
Constant	0.706*** (0.013)	0.452*** (0.017)	0.765*** (0.013)	0.511*** (0.018)
Observations	10,000	10,000	10,000	10,000
Adjusted R ²	0.237	0.268	0.258	0.290

This panel shows results of the four credible models for Case 1. The result in a column (e.g., column 1) is the exposure estimate if it the model in the column reflects the underlying causal relationships or correctly classifies controls (e.g., z1 and z2 are bad controls). *, **, and *** indicate a statistically significant difference from zero at the 10%, 5%, and 1% level, respectively.

Panel D: Summary of exposure estimates from multiple simulations for Case 1

	Model 1	Model 2	Model 3	Model 4
Average of estimates	0.690	0.440	0.750	0.500
Std of estimates	0.0120	0.0174	0.0123	0.0173
Observations	1000	1000	1000	1000

This panel shows results of average exposure estimates across 1000 simulations for each of the four models for Case 1. Each of the 1000 observations consists of an exposure estimate from one simulation with 10,000 observations. The panel also shows the standard deviations of the estimates across the 1000 simulations. The purpose of the simulations is to minimize sampling error in showing risks and risks of choosing a specific model. Remaining results (tables) for this case are based on average estimates in this table instead of estimates in Panel C (which contain sampling error).

Panel E: Actions and Possible Risks for Case 1

		Actual State of the World			
		Model 1: 0.690	Model 2: 0.440	Model 3: 0.750	Model 4: 0.500
Course of Action	Model 1: 0.690	0.000	0.250	0.060	0.190
	Model 2: 0.440	0.250	0.000	0.310	0.060
	Model 3: 0.750	0.060	0.310	0.000	0.250
	Model 4: 0.500	0.190	0.060	0.250	0.000

This panel shows possible actions (i.e., relying on a model) and possible risks depending on which state obtains (i.e., represents the unobservable underlying causal relationships). A row indicates which of the models the researcher relies on (e.g., model 4 in row 4). A column indicates which of the states obtains (e.g., model 1 in column 1 or z1 and z2 are actually confounders). For each model, exposure effects are derived from averages of exposure estimates across 1000 simulations from Panel D. The risk is computed as the absolute value of the difference between the average estimate from the row (e.g., the average of 1000 estimates from model 4 or 0.500) and the average estimate from the column (e.g., the average of 1000 estimates from model 1 or 0.690).

Panel F: Recommended Action Using Expected Utility and Assuming the Absolute Error Loss Function for Case 1

		Actual State of the World				
		Model 1: 0.690	Model 2: 0.440	Model 3: 0.750	Model 4: 0.500	Risk
Course of Action	Model 1: 0.690	0.000x0.18 +	0.250x0.42 +	0.060x0.12 +	0.190x0.28=	0.165
	Model 2: 0.440	0.250x0.18 +	0.000x0.42 +	0.310x0.12 +	0.060x0.28=	0.099
	Model 3: 0.750	0.060x0.18 +	0.310x0.42 +	0.000x0.12 +	0.250x0.28=	0.211
	Model 4: 0.500	0.190x0.18 +	0.060x0.42 +	0.250x0.12 +	0.000x0.28=	0.089

This panel shows the expected value of the risks for relying on an exposure effect assuming the absolute error loss function. The expected value of the risks is computed as the sum of the possible risks (from prior panel) times their probabilities of occurrence. The probabilities of risks correspond to researcher’s beliefs about the underlying causal relationships. For example, the probability of 0.18 (model 1 column) corresponds to the researcher’s confidence that z1

and z_2 are bad controls (e.g., are mediators). The expected value of the risks (risk) for each action is in the last column.

Figure 1: Reporting Exposure Estimates as Course of Action in a Decision Tree

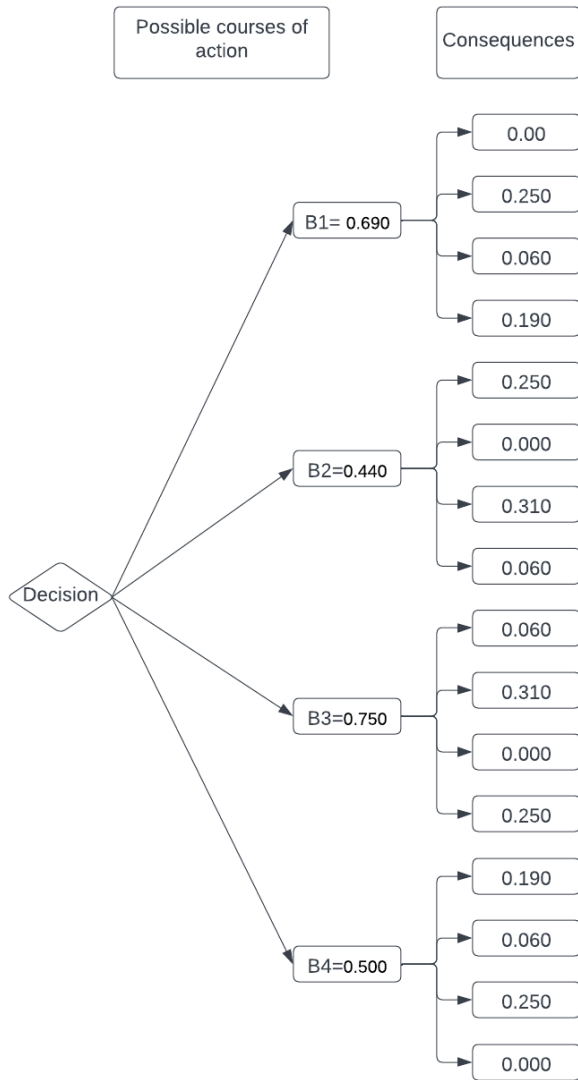


Figure 1 depicts possible risks of each reporting action for case 1, consisting of relying on one of the four exposure effects. Exposure estimates represent averages of exposure estimates across 1000 simulations from Panel D, and are free of sampling error (unlike OLS results). Actions of relying on specific exposure effects are in the boxes right next to the first (straight) set of arrows. The risks are in boxes right next to the second set of arrows. There are four possible risks for relying on each exposure effect depending on which model is the true model.

Figure 2: Course of Action and Risks with Probabilities of Occurrence

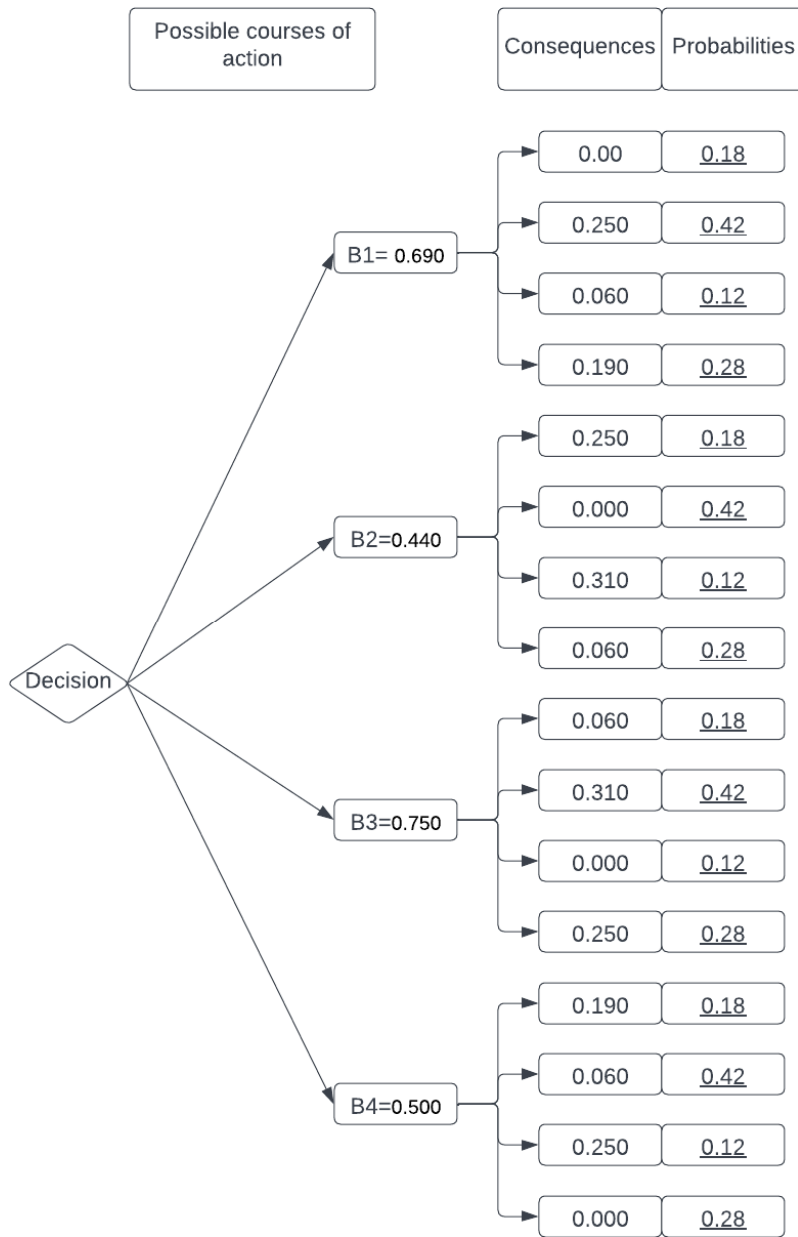


Figure 2 depicts possible risks of each reporting action for case 1 and the probabilities of those risks. The risks are in boxes next to the second set of arrows. There are four possible risks for relying on each exposure effect depending on which model is the true model. The probability of each risk is in a box right next to the respective risk.

Panel G: Recommended Action Using Expected Utility and Assuming the Squared Error Loss Function for Case 1

		Actual State of the World				
		Model 1: 0.690	Model 2: 0.440	Model 3: 0.750	Model 4: 0.500	Risk
Course of Action	Model 1: 0.690	$(0.000)^2 \times 0.18 +$	$(0.250)^2 \times 0.42 +$	$(0.060)^2 \times 0.12 +$	$(0.190)^2 \times 0.28 =$	0.0367
	Model 2: 0.440	$(0.250)^2 \times 0.18 +$	$(0.000)^2 \times 0.42 +$	$(0.310)^2 \times 0.12 +$	$(0.060)^2 \times 0.28 =$	0.0237
	Model 3: 0.750	$(0.060)^2 \times 0.18 +$	$(0.310)^2 \times 0.42 +$	$(0.000)^2 \times 0.12 +$	$(0.250)^2 \times 0.28 =$	0.0584
	Model 4: 0.500	$(0.190)^2 \times 0.18 +$	$(0.060)^2 \times 0.42 +$	$(0.250)^2 \times 0.12 +$	$(0.000)^2 \times 0.28 =$	0.0155
	Average: 0.539	$(0.151)^2 \times 0.18 +$	$(0.098)^2 \times 0.42 +$	$(0.211)^2 \times 0.12 +$	$(0.038)^2 \times 0.28 =$	0.0139

This panel shows the expected value of the risks for relying on an exposure effect assuming the squared error loss function. The expected value of the risks is computed as the sum of the possible risks squared times their probabilities of occurrence. The probabilities of risks correspond to researcher's beliefs about the underlying causal relationships. For example, the probability of 0.18 (model 1 in column 1) corresponds to the researcher's confidence that z1 and z2 are bad controls (e.g., are mediators). The expected value of the risks (risk) for each action is in the last column. Decimals in the risk column were extended to clearly show the weighted average estimate involves the smallest expected risk.

Panel H: Recommended Action Using Expected Utility and Assuming the Zero-One Loss Function for Case 1

		Actual State of the World				
		Model 1: 0.690	Model 2: 0.440	Model 3: 0.750	Model 4: 0.500	Risk
Course of Action	Model 1: 0.690	$0.000 \times 0.18 +$	$1.000 \times 0.42 +$	$1.000 \times 0.12 +$	$1.000 \times 0.28 =$	0.82
	Model 2: 0.440	$1.000 \times 0.18 +$	$0.000 \times 0.42 +$	$1.000 \times 0.12 +$	$1.000 \times 0.28 =$	0.58
	Model 3: 0.750	$1.000 \times 0.18 +$	$1.000 \times 0.42 +$	$0.000 \times 0.12 +$	$1.000 \times 0.28 =$	0.88
	Model 4: 0.500	$1.000 \times 0.18 +$	$1.000 \times 0.42 +$	$1.000 \times 0.12 +$	$0.000 \times 0.28 =$	0.72

This panel shows the expected value of the risks for relying on an exposure effect assuming the zero-one loss function. The expected value of the risks is computed as the sum of the possible risks (from prior panel) times their probabilities of occurrence. Risks are not

measured as the magnitude of the error. The zero-one loss function assigns 1 point (penalty) to a positive error and 0 point (penalty) to a zero error.

Table 2: Case 2

Panel A: Correlation Matrix for Case 2

	y2	x2	z3	z4
y2	1.000	0.763	0.661	0.430
x2	0.763	1.000	0.575	0.336
z3	0.661	0.575	1.000	-0.003
z4	0.430	0.336	-0.003	1.000

Panel A shows the correlation matrix for one sample with 10,000 observations.

Panel B: Auxiliary Regressions for Case 2: Statistical Equivalence of Controls

<i>Dependent variable:</i>			
	z3	z4	x2
	(1)	(2)	(3)
x2	0.336*** (0.005)	0.338*** (0.009)	
z3			0.985*** (0.013)
z4			0.336*** (0.007)
Constant	-0.008 (0.008)	-0.016 (0.016)	0.020 (0.013)
Observations	10,000	10,000	10,000
Adjusted R ²	0.331	0.113	0.445

This panel shows the equivalency between good and bad controls. Based on statistical associations alone, columns 1 and 2 test whether z3 and z4 are mediators (outcomes of x2) or bad controls; column 3 tests whether z3 and z4 are confounders (cause x2) or good controls. *, **, and *** indicate a statistically significant difference from zero at the 10%, 5%, and 1% level, respectively.

Panel C: OLS Results of Credible Models for Case 2

<i>Dependent variable:</i>				
	<i>y2</i>			
	(1)	(2)	(3)	(4)
x2	1.069*** (0.009)	0.801*** (0.010)	0.977*** (0.009)	0.597*** (0.010)
z3		0.798*** (0.017)		1.000*** (0.016)
z4			0.273*** (0.009)	0.401*** (0.008)
Constant	1.069*** (0.009)	0.801*** (0.010)	0.977*** (0.009)	0.597*** (0.010)
Observations	10,000	10,000	10,000	10,000
Adjusted R ²	0.582	0.656	0.616	0.725

This panel shows results of the four credible models for Case 2. The result in a column (e.g., column 1) is the exposure estimate if the model in the column reflects underlying causal relationships or correctly classifies controls (e.g., z3 and z4 are bad controls). *, **, and *** indicate a statistically significant difference from zero at the 10%, 5%, and 1% level, respectively.

Panel D: Summary of exposure estimates from multiple simulations for Case 2

	Model 1	Model 2	Model 3	Model 4
Average of estimates	1.066	0.799	0.975	0.599
Std of estimates	0.0093	0.0103	0.0093	0.0101
Observations	1000	1000	1000	1000

This panel shows results of average exposure estimates across 1000 simulations for each of the four models for Case 2. Each of the 1000 observations consists of an exposure estimate from one simulation with 10,000 observations. The panel also shows the standard deviations of the estimates across the 1000 simulations. The purpose of the simulations is to minimize sampling error in showing risks and risks of choosing a specific model. Remaining results (tables) for this case are based on average estimates in this table instead of estimates in Panel C (which contain sampling error).

Panel E: Actions and Possible Risks for Case 2

		Actual State of the World			
		Model 1: 1.066	Model 2: 0.799	Model 3: 0.975	Model 4: 0.599
Course of Action	Model 1: 1.066	0.000	0.267	0.091	0.467
	Model 2: 0.799	0.267	0.000	0.176	0.200
	Model 3: 0.975	0.091	0.176	0.000	0.376
	Model 4: 0.599	0.467	0.200	0.376	0.000

This panel shows possible actions (i.e., relying on a model) and possible risks depending on which state obtains (i.e., represents the unobservable underlying causal relationships). A row indicates which of the models the researcher relies on (e.g., model 4 in row 4). A column indicates which of the states obtains (e.g., model 1 in column 1 or z3 and z4 are actually confounders). For each model, exposure effects are derived from averages of exposure estimates across 1000 simulations from Panel D. The risk is computed as the absolute value of the difference between the average estimate from the row and the average estimate from the column.

Panel F: Recommended Action Using Expected Utility and Assuming the Absolute Error Loss Function for Case 2

		Actual State of the World				Risk
		Model 1: 1.066	Model 2: 0.799	Model 3: 0.975	Model 4: 0.599	
Course of Action	Model 1: 1.066	0.000x0.04 +	0.267x0.36 +	0.091x0.06 +	0.467x0.54 =	0.354
	Model 2: 0.799	0.267x0.04 +	0.000x0.36 +	0.176x0.06 +	0.200x0.54 =	0.129
	Model 3: 0.975	0.091x0.04 +	0.176x0.36 +	0.000x0.06 +	0.376x0.54 =	0.270
	Model 4: 0.599	0.467x0.04 +	0.200x0.36 +	0.376x0.06 +	0.000x0.54 =	0.113

This panel shows the expected value of the risks for relying on an exposure effect assuming the absolute error loss function. The expected value of the risks is computed as the sum of the possible risks (from prior panel) times their probabilities of occurrence. The probabilities of risks correspond to researcher's beliefs about the underlying causal relationships. For example, the probability of 0.04 (model 1 in column 1) corresponds to the researcher's

confidence that z3 and z4 are bad controls. The expected value of the risks (risk) for each action is in the last column.

Panel G: Recommended Action Using Expected Utility and Assuming the Squared Error Loss Function for Case 2

		Actual State of the World				
		Model 1: 1.066	Model 2: 0.799	Model 3: 0.975	Model 4: 0.599	Risk
Course of Action	Model 1: 1.066	$(0.000)^2x$ 0.04+	$(0.267)^2x$ 0.36 +	$(0.091)^2x$ 0.06+	$(0.467)^2x$ 0.54=	0.144
	Model 2: 0.799	$(0.267)^2x$ 0.04 +	$(0.000)^2x$ 0.36+	$(0.176)^2x$ 0.06+	$(0.200)^2x$ 0.54=	0.026
	Model 3: 0.975	$(0.091)^2x$ 0.04 +	$(0.176)^2x$ 0.36+	$(0.000)^2x$ 0.06+	$(0.376)^2x$ 0.54=	0.088
	Model 4: 0.599	$(0.467)^2x$ 0.04 +	$(0.200)^2x$ 0.36+	$(0.376)^2x$ 0.06+	$(0.000)^2x$ 0.54=	0.032
	Average: 0.713	$(0.354)^2x$ 0.04 +	$(0.087)^2x$ 0.36+	$(0.262)^2x$ 0.06+	$(0.113)^2x$ 0.54=	0.019

This panel shows the expected value of the risks for relying on an exposure effect assuming the squared error loss function. The expected value of the risks is computed as the sum of the possible risks squared times their probabilities of occurrence. The probabilities of risks correspond to researcher's beliefs about the underlying causal relationships. For example, the probability of 0.04 (model 1 in column 1) corresponds to the researcher's confidence that z3 and z4 are bad controls. The expected value of the risks (risk) for each action is in the last column.

Panel H: Recommended Action Using Expected Utility and Assuming the Zero-One Loss Function for Case 2

		Actual State of the World				Risk
		Model 1: 1.066	Model 2: 0.799	Model 3: 0.975	Model 4: 0.599	
Course of Action	Model 1: 1.066	0.000x0.04 +	1.000x0.36 +	1.000x0.06 +	1.000x0.54=	0.960
	Model 2: 0.799	1.000x0.04 +	0.000x0.36 +	1.000x0.06 +	1.000x0.54=	0.640
	Model 3: 0.975	1.000x0.04 +	1.000x0.36 +	0.000x0.06 +	1.000x0.54=	0.940
	Model 4: 0.599	1.000x0.04 +	1.000x0.36 +	1.000x0.06 +	0.000x0.54=	0.460

This panel shows the expected value of the risks for relying on an exposure effect assuming the zero-one loss function. The expected value of the risks is computed as the sum of the possible risks (from prior panel) times their probabilities of occurrence. Risks are not measured as the magnitude of the error. The zero-one loss function assigns 1 point (penalty) to a positive error and 0 point (penalty) to a zero error.

Appendix A: Details of the Underlying Causal Relationships

Case 1: Confounder-Mediator Uncertainty

In case 1, factor z_1 causes x_1 and y_1 : z_1 is a confounder. Factor x_1 causes factor z_2 , which in turn causes y_1 : z_2 is a mediator. This means that z_2 is in the process that generates y_1 . x_1 has a direct effect on y_1 and an indirect effect through z_2 . Without z_2 in this long model, there would be no indirect effect of x_1 . In other words, z_2 is a bad control because it affects y_1 while being an outcome of x_1 . Model 2 is the correct model since it controls for z_1 (good control) but not z_2 (bad control). The data generating process is summarized below. For replication of OLS results, the random seed is set at 100. To get 1000 simulations, I generate the data using the same process 1000 times with the random seed changing each time. I also set the sample size to 10000 for each of the simulations.

$$z_1 \sim N(0,1)$$

$$x_1 = z_1 + e_{x_1}, \quad e_{x_1} \sim N(0,1)$$

$$z_2 = -0.3x_1 + e_{z_2}, \quad e_{z_2} \sim N(0,2.25)$$

$$y_1 = 0.5x_1 + 0.5z_1 + 0.2z_2 + e_1, \quad e_1 \sim N(0,3)$$

Specifically, 0.5 unit of y_1 is the direct effect of 1 unit of x_1 and $0.2 \cdot z_2$ (or $0.2 \cdot -0.3 x_1$) unit of y_1 is the indirect effect. The true model for the effect of x_1 is hence:

$$y_1 = 0.44 x_1 + 0.5 z_1 + e'_1$$

Case 2: Confounder-M-bias Uncertainty

Variable u_x causes x_2 and z_4 , but not y_2 (i.e. not a confounder). Variable u_y causes y_2 and z_4 , but not x_2 . By sharing causes with x_2 and y_2 , z_4 is statistically correlated with both and will appear as a confounder. More importantly, controlling for z_4 will cause bias (M bias). Thus, model 2 is the correct model since it controls for z_3 (good control) but not z_4 (bad control). For

replication of OLS results, the random seed is set at 100. I obtain 1000 simulations as I do in Case 1.

$$ux \sim N(0,1)$$

$$uy \sim N(0,1)$$

$$z3 \sim N(0,1)$$

$$z4 = ux + uy + e_m, \quad e_m \sim N(0,1)$$

$$x2 = z3 + ux + e_{x2}, \quad e_{x2} \sim N(0,1)$$

$$y2 = 0.8 x2 + 0.8 z3 + uy + e_{y2}, \quad e_{y2} \sim N(0,1)$$

Note that $z4$ is not in the data generating process of $y2$ because it has no effect on $y2$, not directly, not as cause of $x2$, and not as an indirect effect of $x2$. It is a bad control because within its cluster, $y2$ and $x2$ are mechanically correlated.

With the second example in mind, I am asserting that size ($z4$) is not a cause of information asymmetry ($x2$) and does not affect conservatism ($y2$). It is possible that some management traits (ux) affect size ($z4$) and information asymmetry ($x2$) but size has no effect on conservatism otherwise. Similarly, it is possible that a firm characteristic or a piece of regulation (uy) affects size ($z4$) and conservatism ($y2$) but size itself ($z4$) has no effect on conservatism. It is not hard to imagine that size is an amalgam of endogenous factors, ux and uy , that include firm structure, its management, or outside forces.