Testable Implications of Causal Graphs in the Presence of Measurement Error

Karthika Mohan KARTHIKA.MOHAN@OREGONSTATE.EDU
Causal Intelligence and Reasoning Lab,
School of Electrical Engineering and Computer Science (EECS)
Oregon State University, Corvallis, USA

Samuel Zink ZINKSAM@OREGONSTATE.EDU
Causal Intelligence and Reasoning Lab
School of Electrical Engineering and Computer Science (EECS)
Oregon State University, Corvallis, USA

Abstract

Measurement errors from sensors, instruments and human factors pose significant challenges for causal inference. Classical testable implications such as d-separations require fully observed variables and cannot be readily evaluated when variables are corrupted by measurement error. This paper shows that causal models nevertheless admit testable implications in such settings. We derive novel testable implications for causal graphs and present a general theorem stating sufficient conditions for their existence. Our results enable model falsification when traditional methods are inapplicable, enabling assessment of model compatibility with observed data when measurement error is present.

1. Introduction

A variable is said to be corrupted by measurement error when its recorded value deviates from its true value. Such errors can arise from a variety of sources, including malfunctioning sensors, improperly calibrated instruments, environmental conditions (e.g., high humidity or extreme temperatures) that affect readings, human error and data transmission errors in digital systems.

Measurement error, missing data, selection bias and non-IID data (or interference) are common factors that can significantly degrade data quality. Compared with problems such as missing data and selection bias, relatively fewer techniques are available that guarantee consistent estimation in the presence of measurement error. An important contribution to causal inference under measurement error is Pearl (2010). Building on earlier work by Greenland and Lash (2008), the method proposed in Pearl (2010) computes $P(Z)$, the distribution over variable Z corrupted by measurement error, by leveraging a proxy variable W for Z. This approach relies on external information in the form of the conditional distribution $P(W \mid Z)$. Subsequent work by Kuroki and Pearl (2014) generalizes these results and extends them to settings where such external information is not required. While these results primarily focus on obtaining unbiased estimates of causal effects in the presence of unmeasured confounders or measurement error, to the best of our knowledge there has been little work that systematically characterizes the testable implications of models involving measurement error. We illustrate this issue with the following example.

Example 1 *Consider the causal graph in Figure 1, where U is a variable corrupted by measurement error. In this graph, the causal effect of $X = x$ on Y can be identified using the front-door criterion (Pearl, 2009). A natural question that arises is: What are the testable implications of this graph? The graph encodes d-separation statements such as $U \perp\!\!\!\perp Z \mid X$ and $Y \perp\!\!\!\perp X \mid U, Z$, which constitute testable implications of causal models (Tian and Pearl, 2002). However, because U is corrupted by measurement error, these conditional independence relations cannot be readily tested from observed data. This raises the following question: can we derive testable implications for causal graphs involving variables corrupted by measurement error? If so, how?*

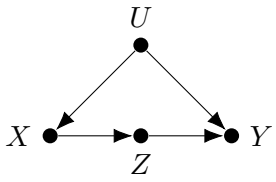

Figure 1: Front-door Causal Graph.

Motivated by this challenge, this work takes a step toward deriving testable implications for causal graphs involving measurement error. The importance of testable implications cannot be overstated. Causal graphs are often constructed, at least in part, based on expert domain knowledge. Identification algorithms rely heavily on the assumed graph structure and estimation is performed only when the causal query is identifiable. Consequently, misspecification of the underlying model can lead to inaccurate estimates.

Yet many existing testable implications become difficult or impossible to evaluate when key variables are corrupted by measurement error. This limitation creates an opportunity to extend the scope of testable implications beyond traditional constraints such as d-separation and Verma constraints (Shpitser and Pearl, 2008; Tian and Pearl, 2002; Pearl, 2009). In this work, we show that additional testable implications can be derived in causal models involving measurement error.

Before proceeding, it is useful to distinguish between testable implications and the statistical tests used to evaluate them. A testable implication is a population-level constraint implied by a causal model; if such a constraint fails, the model is falsified. In contrast, tests are the statistical procedures used to evaluate these implications in finite samples and must contend with sampling variability, noise, and methodological choices. For instance, d-separations are testable implications of a causal graph (Pearl, 2009), whereas assessing them empirically requires conditional independence tests such as chi-square or kernel-based methods. This paper is concerned with deriving testable implications, following the tradition of prior work on causal graphical models such as Tian and Pearl (2002), Shpitser and Pearl (2008) and Mohan and Pearl (2014). If a model possesses no testable implications, then it cannot, even in principle, be empirically tested.

2. Testable Implications of the Front-door Graph under Measurement Error

As noted earlier, d-separations constitute testable implications of causal models (Tian and Pearl, 2002). However, in the graph shown in Figure 1(a), relations such as $Z \perp\!\!\!\perp U \mid X$ are not directly testable when U is latent or corrupted by measurement error. In this section, we show that causal models may nevertheless admit testable implications even when such

standard criteria are difficult to evaluate. In this section, we assume that variables are binary. When referring to a variable X, we use the shorthand x and $\bar{x}$ to denote $X = 1$ and $X = 0$, respectively. This section focuses specifically on deriving testable implications for the graph in Figure 1.

2.1. External Information Assumption

As in Pearl (2010), we assume the availability of external information. Specifically, we assume that the conditional distribution of the outcome Y given the treatment X and the variable corrupted by measurement error U, namely $P(Y \mid X, U)$, is available. This assumption is reasonable in many practical settings.

For example, in a medical diagnosis scenario, Y may represent a particular medical condition (e.g., lung cancer), X a known risk factor (e.g., smoking) and U an unobserved genetic predisposition (Pearl, 2009). Although accurately measuring U or determining its population distribution can be difficult due to limited genetic testing and data collection, a domain expert, a medical professional in this case, would nevertheless provide reliable estimates of $P(Y \mid X, U)$ based on medical studies.

Similarly, in an educational setting, Y may represent student performance, X the quality of teaching and U factors such as parental involvement. While accurately measuring U or estimating $P(U)$ may be challenging due to the complexity of family dynamics, the conditional distribution $P(Y \mid X, U)$ may still be available from prior studies or expert domain knowledge.

2.2. Deriving Testable Implications in the Front-Door Graph

Examine the graph in Figure 1. The causal effect of X on Y is identifiable in the graph using the front-door criteria (Pearl, 2009). Let $P_{y|\hat{x}}$ denote the estimand, $\sum_z P(z|x) \sum_{x'} P(y|z, x')P(x')$. Applying the backdoor criterion on the same graph yields, $P(Y|do(x)) = \sum_u P(Y|x, u)P(u)$. Since both criteria identify the same causal effect, the resulting expressions must be equal. Thus,

$$P_{y|\hat{x}} = P(Y|x, u)P(u) + P(Y|x, \bar{u})(1 - P(u)) \tag{1}$$

In the above equation, the only unknown quantity is $P(u)$, since $P(Y \mid x, u)$ is assumed to be available from an external source. Solving for $P(u)$ yields,

$$P(u) = \frac{P_{y|\hat{x}} - P(Y|x, \bar{u})}{P(Y|x, u) - P(Y|x, \bar{u})} \tag{2}$$

Since X is binary, we have two causal effects, $P(Y|\hat{x})$ and $P(Y|\hat{\bar{x}})$. Each interventional distribution, $P(Y \mid \hat{x})$ and $P(Y \mid \hat{\bar{x}})$, induces its $P(U)$. These must match; otherwise, the assumed causal structure cannot generate the observed quantities. Requiring consistency leads directly to the constraint:

$$\boxed{\frac{P_{y|\hat{x}} - P(Y|x, \bar{u})}{P(Y|x, u) - P(Y|x, \bar{u})} = \frac{P_{y|\hat{\bar{x}}} - P(Y|\bar{x}, \bar{u})}{P(Y|\bar{x}, u) - P(Y|\bar{x}, \bar{u})}} \tag{3}$$

Remark 1

Although we assume that U is affected by measurement error, equation 2 allows the marginal distribution $P(U)$ to be identified without relying on direct measurements of U. As a result, the testable implication in equation 3 remains valid even when U is treated as a latent variable, provided its cardinality is known.

3. Towards More General Results

In this section, we discuss how the testability criterion derived above can be extended to more general cases where variables are not all binary and the graph structure is not limited to the front door.

3.1. Matrix Formulation

Continuing with Figure 1, our approach involves solving a system of linear equations; it is natural to adopt a matrix-based formulation. When referring to either vectors or matrices, we denote said matrix or vector in bold, i.e., $\mathbf{U}$. We define three objects: $\mathbf{I}_x$, $\mathbf{E}_x$, and $\mathbf{U}$, as follows, where $X = x$ denotes the value of X set by intervention.

The matrix containing the interventional distribution is $\mathbf{I}_x$. The number of rows in $\mathbf{I_x}$ depends on the cardinality of Y, denoted as $|Y|$ where the possible values are $y_0, y_1, \ldots, y_{|Y|-1}$.

Thus, we have, $\mathbf{I}_x = \begin{pmatrix} P(y_0 \mid \hat{x}) \\ P(y_1 \mid \hat{x}) \\ \vdots \\ P(y_{|Y|-1} \mid \hat{x}) \end{pmatrix}$.

Similarly, the matrix with the external information is $\mathbf{E}_x$. The number of rows will correspond to the cardinality of Y and the number of columns to the cardinality of U. Thus we have,

$$\mathbf{E}_x = \begin{pmatrix} P(y_0 \mid x, u_0) & P(y_0 \mid x, u_1) & \cdots & P(y_0 \mid x, u_{|U|-1}) \\ P(y_1 \mid x, u_0) & P(y_1 \mid x, u_1) & \cdots & P(y_1 \mid x, u_{|U|-1}) \\ \vdots & \vdots & & \vdots \\ P(y_{|Y|-1} \mid x, u_0) & P(y_{|Y|-1} \mid x, u_1) & \cdots & P(y_{|Y|-1} \mid x, u_{|U|-1}) \end{pmatrix}.$$

Finally, the column matrix with the unknown quantities is $\mathbf{U} = \begin{pmatrix} P(u_0) \\ P(u_1) \\ \vdots \\ P(u_{|U|-1}) \end{pmatrix}$.

Assembling these components, we obtain the linear system $\mathbf{I}_x = \mathbf{E}_x \mathbf{U}$. Solving this system yields a candidate distribution over U for the intervention $X = x$. This computation is carried out separately for each value of X. The causal model implies a set of numerical constraints that require the solutions obtained for different values of X to agree. These equalities form the testable implications of the model. Any disagreement among the resulting vectors $\mathbf{U}$ indicates that the assumed causal model is incompatible with the observed data. Theorem 2 below formalizes this requirement. Algorithm 1 operationalizes the procedure described.

Theorem 2 (Testable Implications under Measurement Error) *Let G be a causal graph with treatment set X, outcome set Y and let M denote the set of variables corrupted by measurement error. Suppose that, for each $x \in X$, the graph implies a linear system $\mathbf{I}_x = \mathbf{E}_x\mathbf{U}$, where $\mathbf{I}_x$ is the matrix comprising estimand for $P(Y \mid \hat{x})$, $\mathbf{E}_x$ is formed from available conditional distributions and $\mathbf{U}$ is an unknown probability vector over a set of variables U such that $U \cap M \neq \emptyset$. Δ_U denotes the probability simplex over U. If*

$$\bigcap_{x \in X} \{\mathbf{U} \in \Delta_U : \mathbf{I}_x = \mathbf{E}_x\mathbf{U}\} = \emptyset,$$

then G is falsified.

All proofs are presented after the main text (Section 7).

Algorithm 1 Consistency Check under Measurement Error

Input: For all x, matrices $\mathbf{I}_x$ and $\mathbf{E}_x$
Output: TRUE if no inconsistency is detected; FAIL otherwise
Initialize $\mathcal{S} \leftarrow \Delta_U$ ▷ *probability simplex over U*
for *each instantiation x of X* **do**
  Construct the system $\mathbf{I}_x = \mathbf{E}_x\mathbf{U}$
  Solve the system and obtain the solution set $\mathcal{U}_x$
  $\mathcal{S} \leftarrow \mathcal{S} \cap \mathcal{U}_x$
  if $\mathcal{S} = \varnothing$ **then**
    return FAIL
  end
end
return TRUE

Corollary 3 *Algorithm 1 is sound.*

We note that the algorithm operates in an idealized setting in which the quantities used to construct the linear systems are assumed to be known exactly. In practice, however, these quantities may need to be estimated from finite samples and are subject to sampling variability and noise. As such, the exact system-solving step above may need to be replaced by an appropriate numerical procedure, such as Total Least Squares or other noise-robust methods.

3.2. Extension to Causal Graphs beyond Front-door

We take inspiration from the notion of confounding equivalence (Pearl and Paz, 2014) and demonstrate the existence of testable implications in a backdoor graph when a confounder is corrupted by measurement error.

Examine the graph in Figure 2. The causal effect $P(Y \mid \hat{x})$ can be identified using the backdoor criterion as $P(Y \mid \hat{x}) = \sum_z P(Y \mid x, z)\, P(z)$. Let $P^z_{y|x}$ denote the estimand obtained by adjusting for the observed confounder Z. The measurement error corrupted

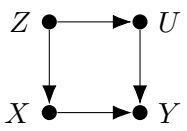

Figure 2: Backdoor Causal Graph

variable U also satisfies the graphical condition for backdoor adjustment and therefore induces an analogous estimand. Thus we have, $P^z_{y|x} = \sum_U P(Y|x,u)P(u)$.

We now construct the matrices $\mathbf{I}_x$, $\mathbf{E}_x$ and $\mathbf{U}$ as in Section 3.1, with one modification; here, $\mathbf{I_x}$ is a column vector whose entries correspond to the interventional distribution $P^z_{y|x}$. As in the earlier case, this yields the linear system $\mathbf{I}_x = \mathbf{E}_x\,\mathbf{U}$. Once these matrices are assembled, Theorem 2 can be applied directly.

3.3. Graph with Altered Backdoor

To illustrate how Theorem 2 generalizes to larger and more complex topologies, we increase the node and edge cardinalities and consider different configurations. Consider the graph in Figure 3, and suppose that U and R are variables corrupted by measurement error. In this setting, the causal effect $P(Y \mid \hat{x})$ is identifiable from the observed distribution by the front-door criterion; let $P_{y|\hat{x}}$ denote the resulting estimand.

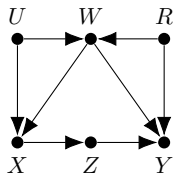

Figure 3: Causal Graph with Altered Backdoor Path

The same causal effect can also be identified by adjusting for the variables U and W, i.e., $P(Y \mid \hat{x}) = \sum_{u,w} P(Y \mid x,u,w)P(u,w)$. Equating this expression with the front-door estimand yields a linear system in the unknown distribution over (U,W).

We therefore construct the matrices $\mathbf{I}_x$ and $\mathbf{E}_x$, together with the unknown probability vector $\mathbf{U}$, analogously to Section 3.1 with $\mathbf{E}_x \in \mathbb{R}^{|Y|\times|U||W|}$ and $\mathbf{U} \in \mathbb{R}^{|U||W|}$. Although R is also corrupted by measurement error in this graph, it is not used in this particular representation of $P(Y \mid \hat{x})$ and therefore does not appear in the corresponding linear system.

$$\mathbf{E}_x = \begin{pmatrix} P(y_0 \mid u_0, x, w_0) & \cdots & P(y_0 \mid u_{|U|-1}, x, w_{|W|-1}) \\ P(y_1 \mid u_0, x, w_0) & \cdots & P(y_1 \mid u_{|U|-1}, x, w_{|W|-1}) \\ \vdots & & \vdots \\ P(y_{|Y|-1} \mid u_0, x, w_0) & \cdots & P(y_{|Y|-1} \mid u_{|U|-1}, x, w_{|W|-1}) \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} P(u_0, w_0) \\ \vdots \\ P(u_{|U|-1}, w_{|W|-1}) \end{pmatrix}$$

This yields the linear system $\mathbf{I}_x = \mathbf{E}_x\mathbf{U}$, allowing us to apply Algorithm 1.

3.4. Graph with Altered Frontdoor Path

Now consider the graph in Figure 4. Let U be a variable corrupted by measurement error. By repeated applications of the rules of do-calculus (Pearl, 2009), the causal effect $P(Y \mid \hat{x})$ can be calculated in two ways. Using the backdoor criterion, we obtain $P(Y \mid \hat{x}) = \sum_u P(Y \mid u,x)P(u)$ and by using a modified form of the frontdoor formula, we obtain, $P(Y \mid \hat{x}) = \sum_{z_1,z_2} P(Y \mid z_1,z_2,x)P(z_1 \mid x)\sum_{x'} P(z_2 \mid x',z_1)P(x')$ (derivation shown in section 7.4). Letting $P_{y|\hat{x}}$ denote the latter, we construct the matrices $\mathbf{I}_x$, $\mathbf{E}_x$ and U as in 3.1 and apply Algorithm 1.

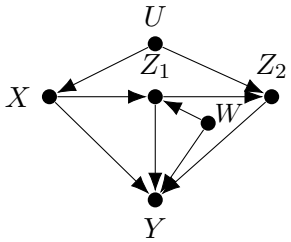

Figure 4: Causal Graph with Altered Frontdoor Path

4. Locating Model Misspecifications

When Algorithm 1 returns FAIL, we know that the observed data are incompatible with the assumed causal model. A natural question that arises is whether we can determine *where* the misspecification originates and thereby diagnose or correct the model. The following result characterizes plausible sources of such misspecification in terms of violated graphical assumptions that warrant further investigation.

Theorem 4 *Let G be a causal graph implying a system of equations of the form $\mathbf{I}_x = \mathbf{E}_x\mathbf{U}$ such that any assumptions only involving observed variables are correct. If Algorithm 1 returns* FAIL *on the system corresponding to G, then all candidate sources of misspecification are those that involve the measurement-error-corrupted variables used in deriving the system.*

Application to the Front-door Graph For the graph in Figure 1, the equality $\mathbf{I}_x = \mathbf{E}_x\mathbf{U}$ is obtained under the d-separation assumptions $Z \perp\!\!\!\perp U \mid X$ and $Y \perp\!\!\!\perp X \mid Z, U$. To see this, assume that these d-separation statements hold. Then,

$$\begin{aligned}\sum_z P(z \mid x) \sum_{x'} P(y \mid x', z)P(x') &= \sum_{z,u,x'} P(z \mid x, u)P(y \mid u, x, z)P(u \mid x')P(x') \\ &\quad \text{(using } Z \perp\!\!\!\perp U \mid X \text{ and } Y \perp\!\!\!\perp X \mid Z, U\text{)} \\ &= \sum_{z,u,x'} P(y, z \mid u, x)P(u, x') \\ &= \sum_u P(y \mid u, x)P(u).\end{aligned}$$

Thus, $\sum_z P(z \mid x) \sum_{x'} P(y \mid x', z)P(x') = \sum_u P(y \mid u, x)P(u)$, which establishes the equality $\mathbf{I}_x = \mathbf{E}_x\mathbf{U}$.

Consequently, there exists at least one distribution $P(U) \in \Delta_U$ satisfying the system for all $x \in X$. Therefore, under the d-separation assumptions $Z \perp\!\!\!\perp U \mid X$ and $Y \perp\!\!\!\perp X \mid Z, U$, the system supplied to Algorithm 1 admits a consistent solution. Consequently, if Algorithm 1 returns FAIL, then violations of these d-separation assumptions become plausible candidate sources of misspecification.

Theorem 4 becomes particularly useful in larger graphs containing many d-separation statements involving measurement-error-corrupted variables. Consider the larger graph in Figure 5, of which Figure 1 is a subgraph.

This graph contains multiple d-separation statements involving the measurement-error-corrupted variable U. However, not all such d-separations become candidate sources of misspecification when Algorithm 1 returns FAIL. In particular, the equality $\mathbf{I}_x = \mathbf{E}_x\mathbf{U}$ derived for the front-door graph relies only on the assumptions $Z \perp\!\!\!\perp U \mid X$ and $Y \perp\!\!\!\perp X \mid Z, U$. Consequently, only violations of these assumptions become plausible candidate sources of misspecification in this setting.

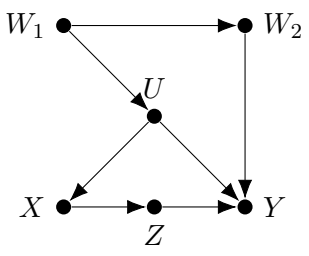

Figure 5: Extended Front-door Causal Graph

Other d-separation statements involving U that are not used in deriving the system need not be implicated. Thus, the derivation of $\mathbf{I}_x = \mathbf{E}_x\mathbf{U}$ helps localize the relevant assumptions underlying the consistency check performed by Algorithm 1.

Application to the Backdoor Causal Graph Consider the causal graph in Figure 2. The system $\mathbf{I}_x = \mathbf{E}_x\mathbf{U}$ is derived under the assumptions that $X \perp\!\!\!\perp U \mid Z$ and $Y \perp\!\!\!\perp Z \mid X, U$. As proof, consider the sequence of transformations.

$$\begin{aligned}\sum_z P(Y \mid z, x)P(z) &= \sum_{z,u} P(Y, u \mid z, x)P(z) \\ &= \sum_{z,u} P(Y \mid u, x)P(u \mid z)P(z) \\ &\quad \text{(using } U \perp\!\!\!\perp X \mid Z \text{ and } Y \perp\!\!\!\perp Z \mid U, X\text{)} \\ &= \sum_u P(Y \mid u, x)P(u)\end{aligned}$$

From here, the left-hand side implies $\mathbf{I}_x$ and the right-hand side implies $\mathbf{E}_x\mathbf{U}$. We can test this system using Algorithm 1 and if it returns FAIL, our sources of model misspecifications can be traced back to these d-separations.

Applications to Broader Graphs Next consider Figures 3 and 4. Figure 3, gives rise to the system $\mathbf{I}_x = \mathbf{E}_x\mathbf{U}$ under the d-separation assumptions $Z \perp\!\!\!\perp U \mid X$, $Z \perp\!\!\!\perp W \mid U, X$ and $Y \perp\!\!\!\perp X \mid W, Z, U$ as we can construct the following sequence under the d-separation assumptions

$$\begin{aligned}\sum_z P(z \mid x) \sum_{x'} P(y \mid x', z)P(x') &= \sum_{z,u,w,x'} P(z \mid x)P(y, u, w \mid x', z)P(x') \\ &= \sum_{z,u,w,x'} P(z \mid x)P(y \mid u, w, x', z)P(w \mid u, x', z)P(u \mid x', z)P(x') \\ &= \sum_{z,u,w,x'} P(z \mid x, u, w)P(y \mid u, w, x, z)P(w \mid u, x')P(u \mid x')P(x') \\ &\quad \text{(using } Z \perp\!\!\!\perp U \mid X \quad \text{and} \quad Z \perp\!\!\!\perp W \mid U, X \quad \text{and} \quad Y \perp\!\!\!\perp X \mid W, Z, U\text{)} \\ &= \sum_{u,w} P(y \mid u, w, x)P(u, w)\end{aligned}$$

Figure 4's system is obtained from the d-separations $U \perp\!\!\!\perp Z_1 \mid X$, $Z_2 \perp\!\!\!\perp X \mid U, Z_1$, and $Y \perp\!\!\!\perp U \mid Z_1, Z_2, X$

$$\sum_{z_1,z_2} P(Y \mid z_1, z_2, x)P(z_1 \mid x) \sum_{x'} P(z_2 \mid x', z_1)P(x')$$

$$= \sum_{z_1,z_2} P(Y \mid z_1, z_2, x) P(z_1 \mid x) \sum_{x',u} P(z_2 \mid u, x, z_1) P(u \mid x') P(x')$$

(using $X \perp\!\!\!\perp U \mid Z_1$ and $Z_2 \perp\!\!\!\perp X \mid U, Z_1$)

$$= \sum_{z_1,z_2,u} P(Y \mid z_1, z_2, x, u) P(z_1 \mid x) P(z_2 \mid u, x, z_1) P(u)$$

(using $Y \perp\!\!\!\perp U \mid Z_1, Z_2, X$)

$$= \sum_{z_1,z_2,u} P(Y, z_2 \mid z_1, x, u) P(z_1 \mid x) P(u)$$

$$= \sum_{u} P(Y \mid x, u) P(u)$$

In both cases, we expect that if our model is correct, there exists a consistent solution for every $x \in X$. However, if Algorithm 1 returns FAIL, we can infer from Theorem 4 that our model misspecification arose from the d-separations containing the measurement corrupted variable U.

5. Related Works

Measurement error has been studied in causal inference, primarily from the perspective of identification and bias correction. Early work by Pearl (2010) and Kuroki and Pearl (2014) shows that causal effects can be recovered in the presence of measurement error using proxy variables and additional assumptions. Pearl (2010) leverages external information about the measurement error mechanism, while Kuroki and Pearl (2014) presents conditions under which causal effects can be restored using proxy variables, including settings with two conditionally independent proxies. Miao et al. (2018) generalizes one such strategy of Kuroki and Pearl (2014), showing that with at least two independent proxy variables satisfying a rank condition, causal effects can be non parametrically identified without requiring identification of the measurement error mechanism. Related lines of work in epidemiology and statistics focus on sensitivity analysis and bias correction under mismeasured variables (Greenland and Lash, 2008; Carroll et al., 2006).

A separate line of work studies causal discovery under measurement error, where the objective is to discover the underlying causal graph from noisy observations. Silva (2006) considers linear latent variable models and develops methods to recover structure by exploiting constraints induced by latent variables and measurement error, typically returning an equivalence class of graphs rather than a unique DAG. Zhang et al. (2018) similarly considers linear latent variable models and studies under what conditions the measurement-error free model can be recovered from the observed measurement corrupted variables. Dai et al. (2022) considers models where data from random variables are generated by a linear, non-Gaussian acyclic model corrupted by measurement error (LiNGAM (Shimizu et al., 2006)) and provides a new condition called Transformed Independent Noise, which checks independence between specific linear combinations of some variables. Whereas in our paper, we make no assumptions on the parameterization of the SCM. Uhler et al. (2013) analyzes the role of the faithfulness assumption and characterizes conditions under which conditional

independence-based methods can recover the underlying structure, clarifying when unique identification is possible and when only equivalence classes can be obtained. Using the tensor-based methods of Kruskal (1976), Allman et al. (2009) demonstrated how latent variables can be recovered given at least 3 conditionally independent proxy variables. Extending the results of both Kuroki and Pearl (2014) and Allman et al. (2009), **?** present algebraic arguments that establish identifiability in small DAGs and for fixed state spaces parameter identifiability depends on the Markov equivalence class, exemplifying this with a DAG containing a hidden parent node to 4 observable child nodes. Anandkumar et al. (2014) develops additional tensor-based methods and combines the existing methods under a unified framework for learning latent variable models from higher-order moments, recovering latent structure under rank and separability conditions. These approaches provide guarantees under their respective assumptions and focus on identifying graph structure or latent representations from data. In contrast, our focus is on assessing whether a given causal model is compatible with observed data by deriving testable implications, which can be used to rule out candidate models, including those obtained from discovery procedures.

Rather than focusing on the identification of causal effects or recovery of graph structure, we study the testability of causal models in the presence of measurement error. Classical results on testable implications of causal graphs, such as d-separation and Verma constraints (Tian and Pearl, 2002; Shpitser and Pearl, 2008), rely on fully observed variables and are often not applicable when key variables are corrupted or latent. Scheines and Ramsey (2017), explores how said classical testable implications are affected in the presence of corrupted or latent variables. When there is too much measurement error to perform constraint-based tests Blom et al. (2018) develops techniques to circumvent the measurement by constructing an upper bound on the variance of random measurement error, which can be used to correct constraint-based causal discovery. The work in this paper doesn't seek to analyze the performance of typical tests or correct them for measurement error, but rather develops new tests for latent or corrupted variables. Other recent work has explored testability in related settings such as missing data (Mohan and Pearl, 2014; Nabi and Bhattacharya, 2023), but there has been little systematic study of testable implications in causal models with measurement error. This paper takes a step in that direction by deriving novel observable constraints that enable falsification of such models.

6. Conclusions and Future Directions

In this paper, we have demonstrated the existence of testable implications that diverge from the traditional ones, such as d-separation, within models that involve variables affected by measurement error. More specifically, we have identified testable implications in a variety of graphical structures, including front-door and back-door graphs. Additionally, we have introduced a theorem that establishes a sufficient criterion for the presence of these constraints. Our future research efforts will focus on generalizing these findings to arbitrary graphical structures.

7. Proofs

7.1. Proof of Theorem 2

Proof Suppose $\bigcap_{x \in X} \{U \in \Delta_U : \mathbf{I}_x = \mathbf{E}_x \mathbf{U}\} = \emptyset$.

This means there exists no probability distribution over U that simultaneously satisfies all the linear constraints $\mathbf{I}_x = \mathbf{E}_x \mathbf{U}$ for every $x \in X$. However, if G were the true causal model, then by the assumption that G implies $\mathbf{I_x} = \mathbf{E}_x \mathbf{U}$, the true marginal $P(U)$ would necessarily satisfy these constraints. Since no such distribution exists, G cannot be the true model. Therefore, the model is falsified. ■

7.2. Proof of Corollary 3

Proof Assume Algorithm 1 returns FAIL, then $\mathcal{S} = \varnothing$. Initially, $\mathcal{S} = \Delta_U$ where Δ_U is a probability simplex, i.e. $\{(p_0, \ldots, p_{|U|-1}) \in \mathbb{R}^{|U|} \mid \sum_{i=0}^{|U|-1} p_i = 1 \text{ and } \forall i,\, p_i \geq 0\}$. For every instantiation of $x \in X$, we only ever decrease the cardinality of $\mathcal{S}$ since we intersect with $\mathcal{U}_x$. Thus, if $\mathcal{S} = \varnothing$, then there is no solution to $\mathbf{I}_x = \mathbf{E}_x \mathbf{U}$, such that $\mathbf{U} \in \Delta_U$, which is impossible if our model is correct.

Thus, when we return Fail, we correctly identify that our model is incorrect.

■

7.3. Proof of Theorem 4

Proof Assume that G implies the system $\mathbf{I}_x = \mathbf{E}_x \mathbf{U}$ and any and all assumptions only involving the observed variables are correct. If Algorithm 1 returns FAIL. By Corollary 3, this means there exists no consistent solution for all instantiations of x for the system $\mathbf{I_x} = \mathbf{E}_x \mathbf{U}$. Let S be the set of assumptions that were used to construct the system. Since Algorithm 1 returned FAIL, it follows that there exists an $s \in S$, such that s is false. By hypothesis, every assumption in S that involves only observed variables is true. Since s is false, s cannot be one of these. Thus, s must contain a variable in U. ■

7.4. Derivation of causal effect using Figure 4

We can factorize the joint distribution for 4 as follows $P(y, x, w, u, z_1, z_2) = P(u)P(w)P(x \mid u)P(z_1 \mid x, w)P(z_2 \mid u, z_1)P(y \mid x, w, z_1, z_2)$ Upon the intervention of x our distribution can be transformed as follows

$$
\begin{aligned}
P(y \mid \hat{x}) &= \sum_{u, z_1, z_2, w} P(u)P(w)P(z_1 \mid \hat{x}, w)P(z_2 \mid u, z_1)P(y \mid \hat{x}, w, z_1, z_2) \\
&= \sum_{u, z_1, z_2, w} P(u)P(w)P(z_1 \mid x, w)P(z_2 \mid u, z_1)P(y \mid x, w, z_1, z_2) \\
&\quad \text{Using Rule 2 of Do-Calculus (Pearl, 2009)} \\
&= \sum_{u, z_1, z_2, w} P(w)P(z_1 \mid x, w)P(y \mid x, w, z_1, z_2) \sum_{x'} P(z_2 \mid u, z_1)p(u, x')
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{u,z_1,z_2,w} P(w)P(z_1 \mid x,w)P(y \mid x,w,z_1,z_2) \sum_{x'} P(z_2 \mid u,z_1)p(u \mid x')P(x') \\
&= \sum_{u,z_1,z_2,w} P(w)P(z_1 \mid x,w)P(y \mid x,w,z_1,z_2) \sum_{x'} P(z_2 \mid x',u,z_1)p(u \mid x',z_1)P(x') \\
&= \sum_{u,z_1,z_2,w} P(w)P(z_1 \mid x,w)P(y \mid x,w,z_1,z_2) \sum_{x'} P(z_2,u \mid x',z_1)P(x') \\
&= \sum_{z_1,z_2,w} P(w)P(z_1 \mid x,w)P(y \mid x,w,z_1,z_2) \sum_{x'} P(z_2 \mid x',z_1)P(x') \\
&= \sum_{z_1,z_2,w} P(w \mid x)P(z_1 \mid x,w)P(y \mid x,w,z_1,z_2) \sum_{x'} P(z_2 \mid x',z_1)P(x') \\
&= \sum_{z_1,z_2,w} P(z_1,w \mid x)P(y \mid x,w,z_1,z_2) \sum_{x'} P(z_2 \mid x',z_1)P(x') \\
&= \sum_{z_1,z_2,w} P(w \mid x,z_1)P(z_1 \mid x)P(y \mid x,w,z_1,z_2) \sum_{x'} P(z_2 \mid x',z_1)P(x') \\
&= \sum_{z_1,z_2,w} P(w \mid x,z_1,z_2)P(z_1 \mid x)P(y \mid x,w,z_1,z_2) \sum_{x'} P(z_2 \mid x',z_1)P(x') \\
&= \sum_{z_1,z_2,w} P(z_1 \mid x)P(y,w \mid x,z_1,z_2) \sum_{x'} P(z_2 \mid x',z_1)P(x') \\
&= \sum_{z_1,z_2} P(y \mid x,z_1,z_2)P(z_1 \mid x) \sum_{x'} P(z_2 \mid x',z_1)P(x')
\end{aligned}
$$

References

Elizabeth S Allman, Catherine Matias, and John A Rhodes. Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, (6A): 3099–3132, 2009.

Animashree Anandkumar, Rong Ge, Daniel Hsu, Sham M Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *The Journal of Machine Learning Research*, 15(1):2773–2832, 2014.

Tineke Blom, Anna Klimovskaia, Sara Magliacane, and Joris M. Mooij. An upper bound for random measurement error in causal discovery. In *UAI*, 2018.

Raymond J Carroll, David Ruppert, Leonard A Stefanski, and Ciprian M Crainiceanu. *Measurement error in nonlinear models: a modern perspective*. Chapman and Hall/CRC, 2006.

Haoyue Dai, Peter Spirtes, and Kun Zhang. Independence testing-based approach to causal discovery under measurement error and linear non-gaussian models. In *NeurIPS*, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/b05bffeb1ef937677ef0e32f027b4c80-Paper-Conference.pdf.

Sander Greenland and Timothy Lash. *Bias analysis. In Modern Epidemiology by Rothman KJ, Greenland S, Lash TL*. Lippincott Williams and Wilkins, Philadelphia PA, 2008.

Joseph B. Kruskal. More factors than subjects, tests and treatments: An indeterminacy theorem for canonical decomposition and individual differences scaling. *Psychometrika*, 42(3):281–293, 1976. doi: 10.1007/BF02293554.

Manabu Kuroki and Judea Pearl. Measurement bias and effect restoration in causal inference. *Biometrika*, 101(2):423–437, 2014.

Wang Miao, Zhi Geng, and Eric J Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993, 2018.

Karthika Mohan and Judea Pearl. On the Testability of Models with Missing Data. In *AISTATS*, pages 643–650, 2014. URL https://proceedings.mlr.press/v33/mohan14.html.

Razieh Nabi and Rohit Bhattacharya. On testability and goodness of fit tests in missing data models. In *UAI*, 2023.

Judea Pearl. *Causality*. Cambridge University Press, 2 edition, 2009.

Judea Pearl. On measurement bias in causal inference. In *UAI*, pages 425–432, 2010.

Judea Pearl and Azaria Paz. Confounding equivalence in causal inference. *Journal of Causal Inference*, 2(1):75–93, 2014. doi: doi:10.1515/jci-2013-0020. URL https://doi.org/10.1515/jci-2013-0020.

Richard Scheines and Joseph Ramsey. Measurement error and causal discovery. In *CEUR Workshop Proceedings*, volume 1792, pages 1–7, 2017.

Shohei Shimizu, Patrik O. Hoyer, Aapo Hyvärinen, and Antti Kerminen. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7 (72):2003–2030, 2006. URL http://jmlr.org/papers/v7/shimizu06a.html.

Ilya Shpitser and Judea Pearl. Dormant independence. In *AAAI Press*, 2008.

Ricardo Silva. Learning the structure of linear latent variable models. *Journal of Machine Learning Research*, 2006.

Jin Tian and Judea Pearl. On the testable implications of causal models with hidden variables. In *UAI*, 2002.

Caroline Uhler, Garvesh Raskutti, Peter Bühlmann, and Bin Yu. Geometry of the faithfulness assumption in causal inference. *The Annals of Statistics*, pages 436–463, 2013.

Kun Zhang, Mingming Gong, Joseph D Ramsey, Kayhan Batmanghelich, Peter Spirtes, and Clark Glymour. Causal discovery with linear non-gaussian models under measurement error: Structural identifiability results. In *UAI*, pages 1063–1072, 2018.