This document contains errata from Jonas Peters: "Restricted Structural Equation Models for Causal Inference", Diss. ETH No. 20756.

1 Example 2.10.

This is the new corrected version of Example 2.10. The original version had slightly different coefficients which (by chance) generated new unwanted conditional independences.

Example 1 *In this example, we consider five variables, namely* $\mathbf{X} = (A, B_1, B_2, B_3, C)$. *Again we generate their distribution with a linear Gaussian SEM with structure and coefficients shown in Figure 1 and unit variances for the noise variables, i.e.* $\mathbf{var}_{\mathcal{G}_1}(N_X) = 1$ *for all* $X \in \mathbf{X}$. *In* $\mathcal{L}(\mathbf{X})$ *we find the independence constraints*

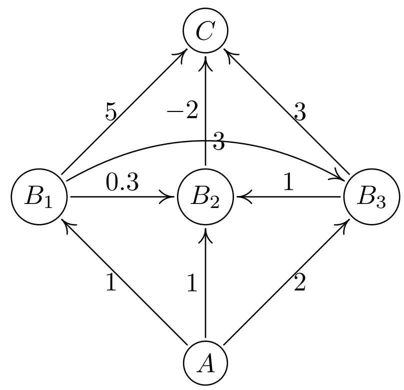

Figure 1: Graph $\mathcal{G}_1$ used to generate the joint distribution of Example **??**.

$$A \perp\!\!\!\perp C \,|\, \{B_1, B_2, B_3\} \quad (1)$$

$$A \perp\!\!\!\perp C \,|\, \{B_1\} \quad (2)$$

It turns out that the obtained distribution can also be generated by an SEM with structure shown in Figure 2. The coefficients and noise variances for the SEM with graph $\mathcal{G}_2$ *can be computed analytically from the coefficients in* $\mathcal{G}_1$ *using the covariance matrix of the distribution. In Figure 2 we show rounded values for the coefficients. For the variances use* $\mathbf{var}_{\mathcal{G}_2}(N_A) = 1$, $\mathbf{var}_{\mathcal{G}_2}(N_{B_1}) = 1$, $\mathbf{var}_{\mathcal{G}_2}(N_{B_2}) = 0.2$, $\mathbf{var}_{\mathcal{G}_2}(N_{B_3}) = 0.8333$ *and* $\mathbf{var}_{\mathcal{G}_2}(N_C) = 6$. *The distribution is not faithful to any of the graphs. The first independence constraint* (1) *is encoded in* $\mathcal{G}_1$ *(Figure 1), the second one* (2) *in* $\mathcal{G}_2$ *(Figure 2). We cannot leave out any of the edges since this would introduce new independences that are not in* $\mathcal{L}(\mathbf{X})$. *Thus,* $\mathcal{G}_1$ *and* $\mathcal{G}_2$ *have a minimal number of edges, but they are not Markov equivalent.*

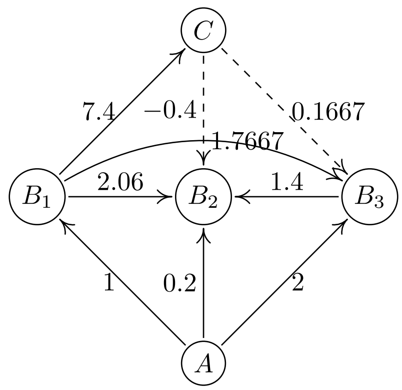

Figure 2: The distribution from Example **??** can also be generated by an SEM from this graph $\mathcal{G}_2$ (the dashed arrows are different from $\mathcal{G}_1$). Both graphs have the minimal number of edges, but are not Markov equivalent.

2 Theorem 3.3

The condition of a strictly positive density was missing in the original version of this thesis. This condition is necessary although this might not be apparent on first sight of the original paper [Shimizu et al., 2006]. The corrected version of Theorem 3.3 reads

Theorem 2 (Shimizu et al. [2006]) *Assume an SEM with graph* $\mathcal{G}_0$

$$X_j = \sum_{k \in \mathbf{PA}_j^{\mathcal{G}_0}} \beta_{jk} X_k + N_j \,, \qquad j = 1, \ldots, p \tag{3}$$

where all N_j *are jointly independent and non-Gaussian distributed with strictly positive density. Additionally, for each* $j \in \{1, \ldots, p\}$ *we require* $\beta_{jk} \neq 0$ *for all* $k \in \mathbf{PA}_j^{\mathcal{G}_0}$*. Then, the graph* $\mathcal{G}_0$ *is identifiable from the joint distribution.*

References

S. Shimizu, P.O. Hoyer, A. Hyvärinen, and A.J. Kerminen. A linear non-Gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7:2003–2030, 2006.