APPENDIX — SUPPLEMENTARY MATERIAL

APPENDIX A – PROOF OF PROPOSITION 1

Before we prove this result, we need to state the following result from Richardson (2003). Given an ancestral set A, the *Markov blanket* of vertex X_v in A, $mb(X_v, A)$, is given by the district of X_v in $(\mathcal{G})_A$ (except X_v itself) along with all parents of elements of this district. Let a *total ordering* $\prec$ of the vertices of $\mathcal{G}$ be any ordering such that if $X_v \prec X_t$, then X_t is not an ancestor of X_v in $\mathcal{G}$. A probability measure is said to satisfy the *ordered local Markov condition* for $\mathcal{G}$ with respect to $\prec$ if, for any X_v and ancestral set A such that $X_t \in A\backslash\{X_v\} \Rightarrow X_t \prec X_v$, we have X_v is independent of $A\backslash(mb(X_v, A) \cup \{X_v\})$ given $mb(X_v, A)$. The main result from Richardson (2003) states:

Theorem 1. *The ordered local Markov condition is equivalent to the global Markov condition in ADMGs*[6].

Proof of Proposition 1: The proof is done by induction on $|X_V|$, with the case $|X_V| = 1$ being trivial. We will show that if $P(X_V)$ is a probability function that factorizes according to (5), as given by an ADMG $\mathcal{G}$, then $P(X_V)$ is Markov with respect to $\mathcal{G}$. To prove this, first notice there must be some X_v with no children in $\mathcal{G}$, since the graph is acyclic. Let X_{D_i} be the district of X_v. By assumption,

$$\begin{array}{rcl} P(X_V) & = & P_F(X_v \mid X_{D_i} \cup pa_{\mathcal{G}}(X_{D_i})) \\ & \times & P_F(X_{D_i}\backslash X_v \mid pa_{\mathcal{G}}(X_{D_i})\backslash X_{D_i}) \\ & \times & \prod_{j\neq i} P_j(X_{D_j} \mid pa_{\mathcal{G}}(X_{D_j})\backslash X_{D_j}) \end{array} \tag{12}$$

Since X_v is childless, it does not appear in any of the factors in the expression above, except for the first. Hence,

$$\begin{array}{rcl} P(X_V\backslash X_v) & = & P_F(X_{D_i}\backslash X_v \mid pa_{\mathcal{G}}(X_{D_i})\backslash X_{D_i}) \\ & \times & \prod_{j\neq i} P_j(X_{D_j} \mid pa_{\mathcal{G}}(X_{D_j})\backslash X_{D_j}) \end{array} \tag{13}$$

which by induction hypothesis is Markov with respect to the marginal graph $(\mathcal{G})_{X_V\backslash X_v}$.

One minor detail about the induction hypothesis: it is true that $(\mathcal{G})_{X_V\backslash X_v}$ might have more districts than $\mathcal{G}$ after removing X_v: this might happen if removing X_v results on having $X_{D_i}\backslash X_v$ becoming disconnected in $(\mathcal{G}_{X_V\backslash X_v})_{\leftrightarrow}$. However, the result still holds by further factorizing $P_F(X_{D_i}\backslash X_v \mid pa_{\mathcal{G}}(X_{D_i})\backslash X_{D_i})$ according to the newly formed districts of $X_{D_i}\backslash X_v$ – which is possible by the construction of $P_F(\cdot)$ and $\mathcal{G}_i$.

[6]Notice this reduces to the standard notion of local independence in DAGs, where a vertex is independent of its (non-parental) non-descendants given its parents, from which d-separation statements can be derived (Lauritzen, 1996, Pearl, 1988).

By the ordered local Markov property for ADMGs and any ordering $\prec$ where X_v is the last vertex, probability function $P(X_V)$ will be Markov with respect to $\mathcal{G}$ if, according to $P(X_V)$, the Markov blanket of X_v in $\mathcal{G}$ makes X_v independent of the remaining vertices. But this true by construction, since this Markov blanket is contained in $X_{D_i} \cup pa_{\mathcal{G}}(X_{D_i})$ according to Theorem 1. □

APPENDIX B – BINARY CASE: RELATION TO COMPLETE PARAMETERIZATION

A complete parameterization for binary ADMG models is described by Richardson (2009). As we will see, parameters are defined in the context of different marginals, analogous to the purely bi-directed case (Drton and Richardson, 2008).

As in the bi-directed case, the joint probability distribution is given by an inclusion-exclusion scheme:

$$P(X_V = \alpha(V)) = \sum_{C:\alpha^{-1}(0)\subseteq C\subseteq V} g(C) \tag{14}$$

where $g(C)$ is given by

$$(-1)^{|C\backslash\alpha^{-1}(0)|} \prod_{H\in[C]_{\mathcal{G}}} P(X_H = 0 \mid X_{tail(H)} = \alpha(tail(H)))$$

and $\alpha(V)$ is a binary vector in $\{0,1\}^{|X_V|}$, $\alpha^{-1}(0)$ being a function that indicates which elements in X_V were assigned to be zero.

Each C indicates which elements are set to zero in the respective term of the summation. Depending on C, the factorization changes. $[C]_{\mathcal{G}}$ is a set of subsets of X_V: one subset per district, each subset being barren in $\mathcal{G}$. The corresponding $tail(H)$ is the Markov blanket for the ancestral set that contains H as its set of childless vertices.

As in our discussion of standard CDNs, Equation (14) can be interpreted as the CDF-to-probability transformation (3). It can be rewritten as

$$P(X_V = \alpha(V)) = \sum_{C:\alpha^{-1}(0)\subseteq C\subseteq V} (-1)^{|C\backslash\alpha^{-1}(0)|}\times$$

$$\prod_{H\in D_i\cap[C]_{\mathcal{G}}} P(X_{D_i}\backslash tail(H) \leq \alpha(V) \mid X_{tail(H)} =$$

$$\alpha(tail(H)))$$

Hence, this parameterization can also be interpreted as a CDF parameterization. One important difference is that each term in the summation uses only a subset of each district, $X_{D_i}\backslash tail(H)$ instead of X_{D_i}. Notice that some elements of X_{D_i} appear in the conditioning set (i.e., $tail(H)$ contains some of the remaining elements of X_{D_i}, on top of the respective parents).

The need for using subsets comes from the necessity of enforcing independence constraints entailed by bi-directed paths. As in the CDN model, the MCDN criterion factorizes each CDF according to its cliques as an indirect way of accounting for such constraints. Hence, we do not construct factorizations for different marginals: each factor within a summation term in (14) includes all elements of each district. We enforce that they remain barren by the transformation in Section 3.3 – which is unnecessary in Richardson (2009) because only barren subsets are being considered.

To understand how the parameterizations coincide, or which constraints analogous to (4) emerge in our parameterization, consider first the following example. Using the results from Richardson (2009), the graph in Figure 2(a) needs the specification of the following marginals:

$$\begin{array}{rcl} P(X_1, X_4) & = & P(X_1)P(X_4) \\ P(X_1, X_3, X_4) & = & P(X_3, X_4 \mid X_1)P(X_1) \\ P(X_1, X_2, X_4) & = & P(X_1, X_2 \mid X_4)P(X_4) \\ P(X_1, X_2, X_3, X_4) & = & P(X_1, X_2 \mid X_4)P(X_3, X_4 \mid X_1) \\ P(X_1, X_3) & = & P(X_3 \mid X_1)P(X_1) \\ P(X_2, X_4) & = & P(X_2 \mid X_4)P(X_4) \end{array} \tag{15}$$

As an example, the probability $P(X_{14} = 0, X_{23} = 1) \equiv P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0)$ can be derived from the above factorizations and (14) as

$$\begin{array}{l} P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0) \\ = \; P(X_1 \le 0, X_2 \le 1, X_3 \le 1, X_4 \le 0) - \\ \quad P(X_1 \le 0, X_2 \le 1, X_3 \le 0, X_4 \le 0) - \\ \quad P(X_1 \le 0, X_2 \le 0, X_3 \le 1, X_4 \le 0) + \\ \quad P(X_1 \le 0, X_2 \le 0, X_3 \le 0, X_4 \le 0) \\ = \; P(X_1 = 0, X_4 = 0) - \\ \quad P(X_1 = 0, X_3 = 0, X_4 = 0) - \\ \quad P(X_1 = 0, X_2 = 0, X_4 = 0) + \\ \quad P(X_1 = 0, X_2 = 0, X_3 = 0, X_4 = 0) \\ = \; P(X_1 = 0)P(X_4 = 0) - \\ \quad P(X_{34} = 0 \mid X_1 = 0)P(X_1 = 0) - \\ \quad P(X_{12} = 0 \mid X_4 = 0)P(X_4 = 0) + \\ \quad P(X_{12} = 0 \mid X_4 = 0)P(X_{34} = 0 \mid X_1 = 0) \end{array}$$

where the last line comes from the pool of possible factorizations (15). The corresponding probability using the MCDN parameterization is

$$\begin{array}{l} = \; P(X_1 = 0, X_2 = 1 \mid X_4 = 0) \times \\ \quad P(X_3 = 1, X_4 = 0 \mid X_1 = 0) \\ = \; [P(X_1 \le 0, X_2 \le 1 \mid X_4 = 0) - \\ \quad P(X_1 \le 0, X_2 \le 0 \mid X_4 = 0)] \times \\ \quad [P(X_3 \le 1, X_4 \le 0 \mid X_1 = 0) - \\ \quad P(X_3 \le 0, X_4 \le 0 \mid X_1 = 0)] \\ = \; (P(X_1 = 0 | X_4 = 0) - \\ \quad P(X_1 = 0, X_2 = 0 \mid X_4 = 0)) \times \\ \quad (P(X_4 = 0 \mid X_1 = 0) - \\ \quad P(X_3 = 0, X_4 = 0 \mid X_1 = 0)) \\ = \; (P(X_1 = 0) - P(X_1 = 0, X_2 = 0 \mid X_4 = 0)) \times \\ \quad (P(X_4 = 0) - P(X_3 = 0, X_4 = 0 \mid X_1 = 0)) \\ = \; P(X_1 = 0)P(X_4 = 0) - \\ \quad P(X_{34} = 0 \mid X_1 = 0)P(X_1 = 0) - \\ \quad P(X_{12} = 0 \mid X_4 = 0)P(X_4 = 0) + \\ \quad P(X_{12} = 0 \mid X_4 = 0)P(X_{34} = 0 \mid X_1 = 0) \end{array}$$

where the first line comes from the factorization of $P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0)$ according to (5) and the fourth line comes from the Markov properties of each $\mathcal{G}_i$ factor. Although these parameterizations have the same high-level parameters, they still do not coincide, as shown in the next example.

For a more complicated case where an extra constraint appears in our parameterization, consider Figure 3(a). In Richardson (2009), it is shown that one of the parameters of the complete parameterization is $P(X_1 = 0, X_3 = 0 \mid X_2 = 0, X_4 = 0, X_5 = 0)$, which reflects the fact that X_1 and X_5 are dependent given all other variables. This also true in our case, except that according to Figure 3(c), our corresponding CDF is given by

$$\begin{array}{l} F(x_1 \mid X_2)F(x_1, x_3)F(x_2, x_3)F(x_3, x_4)F(x_4, x_5) \times \\ F(x_3 \mid X_5)F(x_2 \mid X_4) \end{array}$$

which, evaluated at $X_{12345} = 0$, gives

$$\begin{array}{l} P(X_1 = 0 \mid X_2 = 0)P(X_1 = 0, X_3 = 0) \times \\ P(X_2 = 0, X_3 = 0)P(X_3 = 0, X_4 = 0) \times \\ P(X_4 = 0, X_5 = 0)P(X_3 = 0 \mid X_5 = 0) \times \\ P(X_2 = 0 \mid X_4 = 0) \end{array}$$

implying that $P(X_{12345} = 0)$ factorizes as $f(X_1, X_2, X_3, X_4)g(X_2, X_3, X_4, X_5)$, the generalization to (4).