Appendix

A PROOFS FOR SECTION 4.3: ELEMENTARY EXACT TRANSFORMATIONS

Lemma 4. *The identity mapping and permuting the labels of variables are both exact transformations. That is, if $\mathcal{M}_X$ is an SEM and $\pi : \mathbb{I}_X \to \mathbb{I}_X$ is a bijection then the transformation*

$$\tau : \mathcal{X} \to \mathcal{Y}$$
$$(x_i : i \in \mathbb{I}_X) \mapsto (x_{\pi(i)} : i \in \mathbb{I}_X)$$

naturally gives rise to an SEM $\mathcal{M}_Y$ that is an exact τ-transformation of $\mathcal{M}_X$, corresponding to relabelling the variables.

Proof of Lemma 4. Consider the SEM $\mathcal{M}_Y$ obtained from $\mathcal{M}_X$ by replacing, for all $i \in \mathbb{I}_X$, any occurrence of X_i in the structural equations $\mathcal{S}_X$ and interventions $\mathcal{I}_X$ by $Y_{\pi(i)}$ and leaving the distribution over the exogenous variables unchanged. □

Proof of Lemma 5 (Transitivity of exact transformations). Let $\omega_{ZY} : \mathcal{I}_Y \to \mathcal{I}_Z$ and $\omega_{YX} : \mathcal{I}_X \to \mathcal{I}_Y$ be the mappings between interventions corresponding to the exact transformations τ_{ZY} and τ_{YX} respectively and define $\omega_{ZX} = \omega_{ZY} \circ \omega_{YX} : \mathcal{I}_X \to \mathcal{I}_Z$. Then ω_{ZX} is surjective and order-preserving since both ω_{ZY} and ω_{YX} are surjective and order-preserving. Since τ_{ZY} and τ_{YX} are exact it follows that for all $i \in \mathcal{I}_X$

$$\mathbb{P}^i_{\tau_{ZX}(X)} = \mathbb{P}^{\omega_{ZY}(\omega_{YX}(i))}_{\tau_{ZY}(\tau_{YX}(X))} = \mathbb{P}^{\mathrm{do}(\omega_{ZX}(i))}_Z$$

i. e. $\mathcal{M}_Z$ is an τ_{ZX}-exact transformation of $\mathcal{M}_X$. □

B PROOFS FOR SECTION 5.1: MARGINALISATION OF VARIABLES

Proof of Theorem 9 (Marginalisation of childless variables). By Lemma 5 it suffices to proof this for marginalisation of one childless variable. Without loss of generality, let X_1 be the childless variable to be marginalised out.

Let $\mathcal{M}_Y = (\mathcal{S}_Y, \mathcal{I}_Y, \mathbb{P}_F)$ be the SEM where

- the structural equations $\mathcal{S}_Y$ are obtained from $\mathcal{S}_X$ by removing the structural equation corresponding to the childless variable X_1;
- $\mathcal{I}_Y$ is the image of the map $\omega : \mathcal{I}_X \to \mathcal{I}_Y$ that drops any reference to the variable X_1 (e. g. $\mathrm{do}(X_1 = x_1, X_2 = x_2) \in \mathcal{I}_X$ would be mapped to $\mathrm{do}(X_2 = x_2) \in \mathcal{I}_Y$);
- $F = (E_i : i \in \mathbb{I}_X \setminus \{1\})$ are the remaining noise variables distributed according to their marginal distribution under $\mathbb{P}_E$.

By construction, ω is surjective and order-preserving. Let $i \in \mathcal{I}_X$ be any intervention. The variable X_1 being childless ensures that the law on the remaining variables $X_k, k \in \mathbb{I}_X \setminus \{1\}$ that we obtain by *marginalisation* of the childless variable, i. e. $\mathbb{P}^i_{\tau(X)}$, is equivalent to the law one obtains by simply *dropping* the childless variable, which is exactly what the law under $\mathcal{M}_Y$ amounts to, i. e. $\mathbb{P}^{\omega(\mathrm{do}(i))}_Y$. □

Proof of Theorem 10 (Marginalisation of non-intervened variables). By Lemma 5 it suffices to proof this for marginalisation of one never-intervened-upon variable. Without loss of generality, let X_1 be the never-intervened-upon variable to be marginalised out. By acyclicity of the SEM $\mathcal{M}_X$, the structural equation corresponding to variable X_1 is of the form $X_1 = f_1\left(\mathbf{X}_{\mathrm{pa}(1)}, E_1\right)$ and X_1 does not appear in the structural equation for any of its ancestors.

Now let $\mathcal{M}_Y = (\mathcal{S}_Y, \mathcal{I}_Y, \mathcal{P}_F)$ be the SEM where

- $\mathcal{I}_Y = \mathcal{I}_X$;

- $F_i = ((E_i, E_1) : i \in \mathbb{I}_X \setminus \{1\})$ are the noise variables distributed as implied by $\mathbb{P}_E$;
- the structural equations $\mathcal{S}_Y$ are obtained from $\mathcal{S}_X$ by removing the structural equation of X_1 and replacing any occurrence of X_1 in the right-hand side of the structural equations of children of X_1 by $f_1\left(\mathbf{X}_{\mathrm{pa}(1)}, E_1\right)$, yielding $X_i = f_i\left(f_1\left(\mathbf{X}_{\mathrm{pa}(1)}, E_1\right), \mathbf{X}_{\mathrm{pa}(i)}, E_i\right)$.

Note that the structural equations of the resulting SEM are still acyclic and are all of the form $X_i = h_i\left(\mathbf{X}_{\setminus i}, F_i\right)$.

Then $\mathcal{M}_Y$ is, by construction, an τ-exact transformation of $\mathcal{M}_X$ for $\omega = \mathrm{id}$. □

C PROOF FOR SECTION 5.2: MICRO- TO MACRO-LEVEL

Proof of Theorem 11. We begin by defining a mapping between interventions

$$\begin{aligned} \omega : \mathcal{I}_X &\to \mathcal{I}_Y \\ \varnothing &\mapsto \varnothing \\ \mathrm{do}(W = w) &\mapsto \mathrm{do}\left(\widehat{W} = \frac{1}{n}\sum_{i=1}^{n} w_i\right) \\ \mathrm{do}(Z = z) &\mapsto \mathrm{do}\left(\widehat{Z} = \frac{1}{m}\sum_{i=1}^{m} z_i\right) \\ \mathrm{do}(W = w, Z = z) &\mapsto \mathrm{do}\left(\widehat{W} = \frac{1}{n}\sum_{i=1}^{n} w_i,\ \widehat{Z} = \frac{1}{m}\sum_{i=1}^{m} z_i\right) \end{aligned}$$

Note that ω is surjective and order-preserving (in fact, it is an order embedding). Therefore, it only remains to show that the distributions implied by $\tau(X)$ under any intervention $i \in \mathcal{I}_X$ agree with the corresponding distributions implied by $\mathcal{M}_Y$. That is, we have to show that

$$\mathbb{P}^{i}_{\tau(X)} = \mathbb{P}^{\mathrm{do}(\omega(i))}_{Y} \quad \forall i \in \mathcal{I}_X$$

In the observational setting, the distribution over $\mathcal{Y}$ is implied by the following equations:

$$\begin{aligned} \widehat{W} &= \frac{1}{n}\sum_{i=1}^{n} W_i = \frac{1}{n}\sum_{i=1}^{n} E_i \\ \widehat{Z} &= \frac{1}{m}\sum_{i=1}^{m} Z_i = \frac{1}{m}\sum_{i=1}^{m}\left(\sum_{j=1}^{n} A_{ij}W_j + F_i\right) = \frac{a}{m}\widehat{W} + \frac{1}{m}\sum_{i=1}^{m} F_i \end{aligned}$$

Since the distributions of the exogenous variables in $\mathcal{M}_Y$ are given by $\widehat{E} \sim \frac{1}{n}\sum_{i=1}^{n} E_i$, $\widehat{F} \sim \frac{1}{m}\sum_{i=1}^{m} F_i$, it follows that $\mathbb{P}^{\mathrm{do}(\varnothing)}_{\tau(X)}$ and $\mathbb{P}^{\mathrm{do}(\varnothing)}_{Y}$ agree. Similarly, the push-forward measure on $\mathcal{Y}$ induced by the intervention $\mathrm{do}(W = w) \in \mathcal{I}_X$ is given by

$$\begin{aligned} \widehat{W} &= \frac{1}{n}\sum_{i=1}^{n} W_i = \frac{1}{n}\sum_{i=1}^{n} w_i \\ \widehat{Z} &= \frac{1}{m}\sum_{i=1}^{m} Z_i = \frac{1}{m}\sum_{i=1}^{m}\left(\sum_{j=1}^{n} A_{ij}W_j + F_i\right) = \frac{a}{m}\widehat{W} + \frac{1}{m}\sum_{i=1}^{m} F_i \end{aligned}$$

which is the same as the distribution induced by the ω-corresponding intervention $\mathrm{do}\left(\widehat{W} = \frac{1}{n}\sum_{i=1}^{n} w_i\right)$ in $\mathcal{M}_Y$.

Similar reasoning shows that this also holds for the interventions $\mathrm{do}(Z = z)$ and $\mathrm{do}(W = w, Z = z)$.

□

D PROOF FOR SECTION 5.3: STATIONARY BEHAVIOUR OF DYNAMICAL PROCESSES

Proof of Theorem 12. We begin by defining a mapping between interventions

$$\omega : \mathcal{I}_X \to \mathcal{I}_Y$$
$$\mathrm{do}(X_t^j = x_j \;\; \forall t \in \mathbb{Z}, \, \forall j \in J) \mapsto \mathrm{do}(Y^j = x_j \; \forall j \in J)$$

Note that ω is surjective and order-preserving (in fact, it is an order embedding). Therefore, it only remains to show that the distributions implied by $\tau(X)$ under any intervention $i \in \mathcal{I}_X$ agree with the corresponding distributions implied by $\mathcal{M}_Y$. That is, we have to show that

$$\mathbb{P}^i_{\tau(X)} = \mathbb{P}_Y^{\mathrm{do}(\omega(i))} \quad \forall i \in \mathcal{I}_X$$

For this we consider, without loss of generality, the distribution arising from performing the $\mathcal{M}_X$-level intervention

$$i = \mathrm{do}(X_t^j = x_j \; \forall t \in \mathbb{Z}, \forall j \leq m \leq n) \in \mathcal{I}_X$$

for $m \in [n]$ (for $m = 0$ this amounts to the null-intervention).

Since A is a contraction mapping, it follows from Lemma 15 that for any intervention in $\mathcal{I}_X$, the sequence of random variables X_t defined by $\mathcal{M}_X$ converges everywhere. That is, there exists a random variable X_* such that $X_t \xrightarrow[t\to\infty]{\text{everywhere}} X_*$. In the case of the intervention i above, the random variable X_* satisfies:

$$\begin{cases} X_*^k = x_k & \text{if } k \leq m \\ X_*^k = \sum_j A_{kj} X_*^j + E^k & \text{if } m < k \leq n \end{cases} \tag{1}$$

Since $\tau(X) = \lim_{t\to\infty} X_t$, it follows from the definition of X_* that $\tau(X) = X_*$, and hence $\tau(X)$ also satisfies the equations above. It follows (rewriting the second line in Equation 1 above) that under the push-forward measure $\mathbb{P}^i_{\tau(X)} = \tau\left(\mathbb{P}_X^{\mathrm{do}(i)}\right)$ the distribution of the random variable $\tau(X) = X_*$ is given by:

$$\begin{cases} X_*^k = x_k & \text{if } k \leq m \\ X_*^k = \frac{\sum_{j\neq k} A_{kj} X_*^j}{1-A_{kk}} + \frac{E^k}{1-A_{kk}} & \text{if } m < k \leq n \end{cases}$$

We need to compare this to the law of Y as implied by $\mathcal{M}_Y$ under the intervention $\omega(i)$, i. e. $\mathbb{P}_Y^{\mathrm{do}(\omega(i))}$. The $\mathcal{M}_Y$-level intervention $\omega(i)$ corresponding to i is

$$\omega(i) = \mathrm{do}(Y^j = x_j \; \forall j \leq m \leq n) \in \mathcal{I}_Y$$

and so the structural equations of $\mathcal{M}_Y$ under the intervention $\omega(\mathrm{do}(i))$ are

$$\begin{cases} Y^k = x_k & \text{if } k \leq m \\ Y^k = \frac{\sum_{j\neq k} A_{kj} Y^j}{1-A_{kk}} + \frac{F^k}{1-A_{kk}} & \text{if } m < k \leq n \end{cases}$$

Since $F \sim E$ it indeed follows that $\tau(X) \sim Y$, i. e. $\mathbb{P}^i_{\tau(X)} = \mathbb{P}_Y^{\mathrm{do}(\omega(i))}$.

Thus $\mathcal{M}_Y$ is an exact τ-transformation of $\mathcal{M}_X$. □

D.1 CONTRACTION MAPPING AND CONVERGENCE

The following Lemmata show that A being a contraction mapping ensures that the sequence $(X_t)_{t\in\mathbb{Z}}$ defined by $\mathcal{M}_X$ in Theorem 12 converges everywhere under any intervention $i \in \mathcal{I}_X$. That is, for any realisation $(x_t)_{t\in\mathbb{Z}}$ of this sequence, its limit $\lim_{t\to\infty} x_t$ as a sequence of elements of $\mathbb{R}^n$ exists.

Lemma 13. *Suppose that the function*

$$\begin{aligned} f : \mathbb{R}^n &\to \mathbb{R}^m \\ x &\mapsto f(x) \end{aligned}$$

is a contraction mapping. Then, for any $e \in \mathbb{R}^m$, so is the function

$$\begin{aligned} f^* : \mathbb{R}^n &\to \mathbb{R}^m \\ x &\mapsto f(x) + e \end{aligned}$$

Proof. By definition, there exists $c < 1$ such that for any $x, y \in \mathbb{R}^n$,

$$\|f^*(x) - f^*(y)\| = \|(f(x) + e) - (f(y) + e)\| = \|f(x) - f(y)\| \leq c\|x - y\|$$

and hence f^* is a contraction mapping. □

Lemma 14. *Suppose that the function*

$$\begin{aligned} f : \mathbb{R}^n &\to \mathbb{R}^n \\ x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} &\mapsto \begin{pmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{pmatrix} \end{aligned}$$

is a contraction mapping. Then for any $m \leq n$, and $x_i^* \in \mathbb{R},\ i \in [m]$, so is the function

$$\begin{aligned} f^* : \mathbb{R}^n &\to \mathbb{R}^n \\ x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} &\mapsto \begin{pmatrix} x_1^* \\ \vdots \\ x_m^* \\ f_{m+1}(x) \\ \vdots \\ f_n(x) \end{pmatrix} \end{aligned}$$

Proof. By definition, there exists $c < 1$ such that for any $x, y \in \mathbb{R}^n$,

$$\begin{aligned} \|f^*(x) - f^*(y)\| = \left\| \begin{pmatrix} x_1^* \\ \vdots \\ x_m^* \\ f_{m+1}(x) \\ \vdots \\ f_n(x) \end{pmatrix} - \begin{pmatrix} x_1^* \\ \vdots \\ x_m^* \\ f_{m+1}(y) \\ \vdots \\ f_n(y) \end{pmatrix} \right\| = \left\| \begin{pmatrix} 0 \\ \vdots \\ 0 \\ f_{m+1}(x) - f_{m+1}(y) \\ \vdots \\ f_n(x) - f_n(y) \end{pmatrix} \right\| &\leq \left\| \begin{pmatrix} f_1(x) - f_1(y) \\ \vdots \\ f_n(x) - f_n(y) \end{pmatrix} \right\| \\ &= \|f(x) - f(y)\| \\ &\leq c\|x - y\| \end{aligned}$$

and hence f^* is a contraction mapping. □

Lemma 15. *Consider the SEM $\mathcal{M}_X$ in Theorem 12, and suppose that the linear map $A : \mathbb{R}^n \to \mathbb{R}^n$ is a contraction mapping. Then, for any intervention $i \in \mathcal{I}_X$, the sequence of X_t converges everywhere.*

Proof. Consider, without loss of generality, the intervention

$$\operatorname{do}(X_t^j = x_j \ \forall t \in \mathbb{Z}, \forall j \leq m \leq n) \in \mathcal{I}_X$$

for $m \in [n]$ (for $m = 0$ this amounts to the null-intervention). The structural equations under this intervention are

$$\begin{cases} X_{t+1}^k = x_k & \text{if } k \leq m \\ X_{t+1}^k = \sum_j A_{kj} X_t^j + E^k & \text{if } m < k \leq n \end{cases}$$

and thus the sequence X_t can be seen to transition according to the function $f = g \circ h$, where

$$\begin{aligned} h : \mathbb{R}^n &\to \mathbb{R}^n \\ v &\mapsto w = Av + E \end{aligned}$$

$$\begin{aligned} g : \mathbb{R}^n &\to \mathbb{R}^n \\ w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} &\mapsto \begin{pmatrix} x_1 \\ \vdots \\ x_m \\ w_{m+1} \\ \vdots \\ w_n \end{pmatrix} \end{aligned}$$

By Lemma 13 and Lemma 14, f is a contraction mapping for any fixed E. Thus, by the contraction mapping theorem, the sequence of X_t converges everywhere to a unique fixed point. ☐