

# Real Rate

## **Kausale Analyse von Gleichungssystemen mit strukturellen neuronalen Netzen**

Holger Bartel

### **Abstract**

Diese Arbeit umfasst die mathematischen Grundlagen für die kausale Analyse von Gleichungssystemen, wie sie im RealRate-Bewertungsmodell verwendet werden.

Wir stellen die totalen Ableitungen eines Gleichungssystems als geschlossene Lösung in Matrixform dar. Dies erfolgt für die exogenen Variablen und die endogenen Variablen. Dies ermöglicht die simultane Berechnung sämtlicher totaler Ableitungen des Systems.

In dem so linearisierten strukturellen Modell entspricht der graphentheoretische Effekt eines Knoten auf den anderen der totalen Ableitung. Dies gilt auch für den zyklischen Fall. Dies zeigt den direkten Zusammenhang zwischen totalen Ableitungen und der im Rahmen der in der Graphentheorie üblichen Sensitivitäts- und Kausalitätsanalyse mittels do-Calculus. Anders als üblich, definieren wir den Eigeneffekt einer endogenen Variablen auch in zyklischen Systemen als eins. Zur Mediationsanalyse zerlegen wir schließlich den Effekt eines Knotens auf seine ausgehenden Kanten und geben hierfür eine geschlossene Formel an.

Für die Schätzung des durch den Graphen definierten strukturellen kausalen Modells nutzen wir ein strukturelles neuronales Netz. Dabei entsprechen die latenten Modellvariablen den hidden Units und die Nullrestriktionen sind nicht vorhandene Kanten zwischen den Neuronen. Wir geben die Zielfunktion, den Gradienten und die Hessesche Matrix für den nichtlinearen Optimierungsalgorithmus an. Die Hessesche wird auch für die empirische Prüfung der Identifikation verwendet.

**Keywords:** total derivative, graphical effect, graph theory, do-Calculus, structural neural network, linear Simultaneous Equations Model (SEM), Structural Causal Model (SCM)

## Inhaltsverzeichnis

1) Grundlagen .....	4
1.1 Gleichungssystem .....	4
1.2 Linearisierung .....	4
1.3 Normierung .....	6
1.4 DAG .....	6
1.5 Strukturelle und reduzierte Form .....	7
1.6 Spezifikation .....	7
1.7 Identifikation .....	8
1.8 Kausalität .....	9
1.9 Effekt .....	10
1.10 Notation .....	11
2) Totale Ableitungen .....	14
2.1 Exogene totale Ableitungen .....	14
2.2 Endogene totale Ableitungen .....	15
3) Graphentheoretische Effekte .....	20
3.1 Exogene Effekte .....	20
3.2 Endogene Effekte .....	21
3.3 Exogene finale Effekte .....	25
3.4 Endogene finale Effekte .....	26
4) Graphentheoretische Interpretation .....	29
4.1 Adjazenzmatrix .....	29
4.2 Pfadeffekte .....	29
4.3 Partieller, totaler und finaler Graph .....	31
4.4 Input-Output-Analyse .....	33
4.5 Relative Gain Array .....	34
5) Schätzung .....	37
5.1 Strukturierte simultane Regression (ökonometrische SEM) .....	38
5.2 Strukturelle Neuronale Netzte .....	43
5.3 Kovarianzstrukturmodelle (psychometrische SEM) .....	59
6) Zusammenfassung .....	69
7) Beispiele .....	71
Beispiel 1 Nichtlineares System .....	71
Beispiel 2 Unabhängige Teilsysteme .....	73

Beispiel 3 Fehlende Bilanzgleichung.....	75
Beispiel 4 Asymmetrische Bilanzgleichung .....	76
Beispiel 5 Symmetrische Bilanzänderungen.....	78
Beispiel 6 Zyklisches System .....	80
Beispiel 7 Unabhängigkeit von der Darstellung .....	83
Beispiel 8 Stark nichtlineares System .....	85
Beispiel 9 Kausalität .....	87
Anhang.....	95
A.1 Elemente von Potenzen triangulärer Matrizen.....	95
A.2 Elemente der Lösung linearer Gleichungssysteme .....	95
A.3 Inversion von Matrixsummen .....	96
A.4 Charakteristisches Polynom .....	97
A.5 Determinanten von Subsystemen .....	97
A.6 Geometrische Reihen von Matrizen.....	99
A.7 Spezielle Zielfunktion .....	100
Literatur.....	104

# 1) Grundlagen

## 1.1 Gleichungssystem

Wir betrachten die n homogenen Modellgleichungen:

$$y_j - M^j(\mathbf{y}, \mathbf{x}) = 0, \quad j = 1 \dots, n$$

bzw.

$$y_j = M^j(\mathbf{y}, \mathbf{x}), \quad j = 1 \dots, n,$$

die zusammen das Gleichungssystem bilden. Die Gleichungen müssen nicht notwendigerweise in Form von Funktionen gegeben sein, also kann  $y_j$  in seiner Gleichung auch auf der rechten Seite auftauchen oder die Gleichung ist in homogener Form gegeben. Es handelt sich um ein deterministisches, nichtlineares simultanes Gleichungssystem. Die Modellgleichungen seien linear unabhängig und differenzierbar. Die Gleichungen hängen von den n endogenen Variablen  $\mathbf{y}$  und den m exogenen Variablen  $\mathbf{x}$  ab.

Das System der Modellgleichungen lautet in Matrixnotation:

$$\mathbf{y} = M(\mathbf{y}, \mathbf{x}) = \begin{bmatrix} M^1(\mathbf{y}, \mathbf{x}) \\ \vdots \\ M^n(\mathbf{y}, \mathbf{x}) \end{bmatrix},$$

$$\underset{(n \times 1)}{\mathbf{y}} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \underset{(m \times 1)}{\mathbf{x}} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}.$$

Da die Modellgleichungen im Allgemeinen nichtlinear sind, können individuelle Effekte für die einzelnen Beobachtungen an der jeweiligen Stelle der exogenen Variablen bestimmt werden. Es ist also eine individuelle Analyse jeder einzelnen Beobachtung möglich.

## 1.2 Linearisierung

Die Gleichungen des Systems können durch Ableitung approximiert und linearisiert werden. Hierzu verwenden wir das totale Differential. Dies ist die Summe der partiellen Ableitungen nach allen Argumenten, jeweils multipliziert mit dem totalen Differential des Arguments:

$$dy_j = \sum_{h=1}^n \frac{\partial M^j}{\partial y_h} dy_h + \sum_{l=1}^m \frac{\partial M^j}{\partial x_l} dx_l, \quad j = 1, \dots, n.$$

Das totale Differential ist die Basis für die Bestimmung der totalen Ableitungen in Kapitel 2. Es ist zugleich die in Abschnitt 5.1. betrachtete linearisierte strukturierte Form. Das durch Ableitung linearisierte System lautet als Matrixgleichung:

$$(I_n - M_y)dy - M_x dx = 0$$

bzw.

$$dy = M_y dy + M_x dx.$$

mit den Differentialen

$$\underset{(n \times 1)}{dy} = \begin{bmatrix} dy_1 \\ \vdots \\ dy_n \end{bmatrix}, \quad \underset{(m \times 1)}{dx} = \begin{bmatrix} dx_1 \\ \vdots \\ dx_m \end{bmatrix}.$$

Wir benötigen die Invertierbarkeit von  $(I_n - M_y)$  und nehmen daher an, dass auch die linearisierten Modellgleichungen linear unabhängig sind.

Die Differentiale ergeben zusammen mit dem Mittelwert  $\mu$  die absoluten endogenen Variablen:

$$y = \mu + dy.$$

Die Differentiale  $dy$  sind Änderungen der Variablen. Sie entsprechen sie den um ihren Mittelwert  $\mu$  bereinigten Daten  $y$ . Die totalen Ableitungen werden also an der Stelle des Mittelwerts bestimmt.  $M_y = M_y|_{x,y} = [m_{yij}]_{i,j=1,\dots,n}$  ist die Jacobi Matrix der *partiellen* Ableitungen erster Ordnung der Gleichungen nach den endogenen Variablen an der Stelle  $(x, y)$  und  $M_x = M_x|_{x,y} = [m_{xij}]_{i,j=1,\dots,n}$  ist die entsprechende Jacobi-Matrix nach den exogenen Variablen.

Die „aktuelle“ Stelle  $(x, y)$  wird bestimmt, indem zusätzlich zu den exogenen Variablen  $x$  die endogenen Variablen  $y$  eingesetzt werden, die sich aus der Lösung des nichtlinearen Gleichungssystems bei gegebenem  $x$  ergeben. Aufgrund des Implizite-Funktionen-Theorems, siehe Abschnitt 2.1, gelten die an dieser Stelle bestimmten partiellen und totalen Ableitungen des linearisierten Modells auch für das nichtlineare System der Modellgleichungen. Dies gilt zumindest in der Nähe der Stelle der Ableitungen, sofern die Gleichungen linear unabhängig und kontinuierlich sind.

Um die gesamte Wirkung zwischen den Variablen zu bestimmen, reicht es nicht aus, die partiellen Ableitungen zu betrachten. Es sind alle Modellzusammenhänge über die totale Ableitung zu berücksichtigen. Das linearisierte System ist linear in den Differentialen, aber die Modellparameter  $M_x$  und  $M_y$  fließen nichtlinear in die totale Ableitung ein, wie in Abschnitt 1.5 gezeigt wird. So wird die Summe aus allen direkten und indirekten Wirkungen bestimmt. Die entsprechenden Matrizen der *totalen* Ableitungen werden mit  $dy/dx^T$  und  $dy/dx^T$  bezeichnet.

Das linearisierte System ist ein Metamodell des Ausgangsmodells. Nur im Falle der Linearität ist das Metamodell exakt. Die partiellen Ableitungen müssen konstant sein, also unabhängig von der

Stelle der Beobachtung. Im Falle der Schätzung, siehe Kapitel 5, kommt noch die Annahme der Homogenität hinzu: Die partiellen Ableitungen müssen also für alle Beobachtungen gleich sein. Dies erleichtert die Identifikation.

Die Linearisierung hat entscheidende Vorteile: Sie ermöglicht die einfache und kompakte Darstellung der partiellen Abteilungen und Effekte in Matrixform. Dies erhöht die Interpretierbarkeit und die Rechengeschwindigkeit. Wir verwenden diese approximativen Formeln daher für die Sensitivitätsanalyse, obwohl das Durchrechnen des Ausgangsmodells – gerade bei nicht infinitesimal kleinen Änderungen – die exakte Lösung liefert. Der nachfolgend dargestellt do-Calculus, also das Ersetzen einer endogenen Variable durch einen konstanten Wert zur Sensitivitätsanalyse, wäre nämlich auch im nichtlinearen Ausgangsmodell anwendbar.

### 1.3 Normierung

Die *linearisierte* Form kann durch Auflösen stets äquivalent so formuliert werden, dass jeweils  $dy_j$  nur auf der linken Seite seiner Gleichung vorkommt. Dies kann o. B. d. A. stets angenommen werden, siehe auch Beispiel 7. Wir nennen dies *Normierung*. Dies bedeutet, dass die Matrix  $M_y$  Nullen auf der Hauptdiagonale hat. Es handelt sich um eine Restriktion, die diese  $n$  redundanten Parameter eliminiert. Auch Bollen (1989), Seite 15, nimmt stets an, dass eine Variable keinen direkten Einfluss auf sich selbst hat. Die Normierungsannahme dient der vereinfachten Darstellung.

Die Normierungsannahme wird lediglich in Kapitel 5 im Rahmen der Identifikation der zu schätzenden strukturellen Parameter benötigt.

### 1.4 DAG

DAG steht für Directed Acyclic Graph, also einen gerichteten azyklischen Graphen. Der DAG ist ein häufig betrachteter und einfacher zu berechnender und zu interpretierender Spezialfall. In unserem Fall ergeben sich im DAG-Fall einfachere Formeln für die endogenen totalen Ableitungen bzw. Effekte, siehe Abschnitte 2.2, 3.2 und 3.4. Im Spezialfall eines DAG können die Variablen, bzw. die Gleichungen, des Gleichungssystems in eine topologische Reihenfolge gebracht werden, siehe [https://en.wikipedia.org/wiki/Topological\\_sorting](https://en.wikipedia.org/wiki/Topological_sorting). Diese bestimmt, in welcher Reihenfolge die Variablen zu berechnen sind, um schließlich die Lösung für das gesamte System zu erhalten. Diese Reihenfolge sei ohne Beschränkung der Allgemeinheit aufsteigend, so dass eine Variable nur von vorangehenden Variablen mit niedrigerem Zeilenindex beeinflusst wird.

Im DAG gilt: Die Matrix  $M_y$  der ersten partiellen Ableitungen der Modellgleichungen nach den endogenen Variablen ist eine strikte untere Dreiecksmatrix. Dies ist äquivalent zu einem gerichteten azyklischen Graphen DAG des Gleichungssystems:

$$\frac{\partial y_j}{\partial y_t} = 0 \text{ für } t \geq j.$$

Insbesondere ist im DAG die im vorangehenden Abschnitt dargestellte Normierung bereits automatisch erfüllt,  $\partial y_i / \partial y_j = 0$ ,  $i = 1, \dots, n$ . Die möglicherweise erwartete Identität  $\partial y_j / \partial y_j = 1$  gilt *nicht*, da die Ableitung einer Funktion nur in Bezug auf ihre Argumente definiert ist.

Der DAG ist ein einfacher Graph, also ein Graph ohne Schleifen und Multikanten. Als azyklischer Graph besitzt er die Markov-Eigenschaft, das heißt, in einem Knoten sind bereits alle Vorgängerinformationen enthalten. Man spricht von einem rekursiven Gleichungssystem.

Die Lösung des linearisierten Gleichungssystems ist im Fall des DAG besonders einfach. Die endogenen Variablen können schrittweise nacheinander bestimmt werden. Im zyklischen Fall hingegen muss das simultane Gleichungssystem mit einem startwertabhängigen Optimierungsverfahren bestimmt werden.

## 1.5 Strukturelle und reduzierte Form

Die linearisierte Form

$$dy = M_y dy + M_x dx.$$

des Modells wird auch *strukturelle* Form genannt und  $M_x$  und  $M_y$  sind die strukturellen Modellparameter. Die strukturelle Form zeichnet sich dadurch aus, dass die strukturellen Modellparameter und somit die Beziehungen zwischen den Variablen verwendet werden. Diese bestimmen die Kausalität, wie im nächsten Abschnitt erläutert wird.

Die *reduzierte* Form zeigt den gesamten Einfluss der exogenen auf die endogenen Variablen an.

$$dy = \frac{dy}{dx^T} dx.$$

Dabei ist  $dy/dx^T$  die *totale* Ableitung gemäß Abschnitt 2.2. Sie gibt im Gegensatz zur partiellen Ableitung die gesamthafte Wirkung zwischen den Variablen wider, unter Beachtung sämtlicher Abhängigkeiten.

## 1.6 Spezifikation

Für die korrekte Spezifikation der Effekte gibt es das graphische Back-Door-Kriterium, siehe Pearl (2009) Definition 4. Danach ist eine notwendige bzw. zulässige Menge an beobachtbaren Variablen in das Regressionsmodell aufzunehmen, um Confounding und somit die verzerrte Schätzung der Effekte zu vermeiden. Ein Confounder ist eine Variable, die einen direkten kausalen Pfad auf die Ausgangsvariable hat, sowie einen Pfad auf die Zielvariable, der nicht über die Ausgangsvariable geht. Confounder sind gemeinsame Vorgänger der beiden Effektvariablen. Confounder sind in das Regressionsmodell aufzunehmen. Umgekehrt dürfen keine Collider in das Regressionsmodell aufgenommen werden. Ein Collider ist eine Variable, in die zwei Pfade eingehen, die aus der Ausgangs- bzw. der Zielvariablen ausgehen. Collider sind gemeinsame Nachfolger der beiden Effektvariablen. Die Aufnahme eines Colliders würde über das Bedingen auf diesen eine künstliche Asso-

ziation erzeugen. Entsprechend ist die notwendige Menge an Regressionsvariablen ist wie folgt definiert:

Erstens müssen über die notwendige Menge alle Back-Door-Pfade, also Pfade, deren eines Ende in die Ausgangsvariable zeigt, blockiert sein. Zweitens darf in dieser Menge kein Nachfolger der Ausgangsvariablen enthalten sein.

Ersteres ist in unserem Fall sicher gestellt, da alle Eltern der Ausgangsvariablen in die Regression aufgenommen werden. Die zweite Anforderung ist ebenfalls erfüllt, da außer den Eltern der Ausgangsvariablen keine weiteren Variablen in die jeweilige Regressionsgleichung aufgenommen werden.

## 1.7 Identifikation

Es ist stets möglich, von der strukturellen Form zur reduzierten Form zu gelangen. Es ist jedoch nicht ohne weiteres möglich, von der reduzierten Form eindeutig auf die strukturelle Form zu schließen. Hierfür ist die Identifikation der strukturellen Parameter erforderlich. Die Identifikation des Modells hängt von seiner Formulierung ab. So können – ohne inhaltliche Änderung des Modells – endogene Variablen substituiert werden. Durch die reduzierte Anzahl von Variablen und Effekten zwischen diesen, kann möglicherweise eine Identifikation der verbleibenden Effekte erreicht werden.

Die Identifikation hängt davon ab, ob Variablen beobachtbar sind oder latent. Die in die Regression aufgenommenen Variablen müssen im Sinne der Graphentheorie beobachtbar sein. Da wir die kompletten Zusammenhänge über die Modellgleichungen vorgeben, sind unsere latenten Variablen stets berechenbar und in diesem speziellen Sinn auch beobachtbar. Sie fließen aber nicht in die zu minimierende Zielfunktion ein. Diese Struktur nutzen wir in Abschnitt 5.2 für die strukturellen neuronalen Netze. Da in diesem Sinne alle Variablen unseres Modells beobachtbar sind, ist auch eine hinreichende Bedingung für die Identifikation der Effekte erfüllt, siehe Pearl (2009) Theorem 2: Alle Pfade von der Ausgangsvariablen über deren Kinder, die zur Zielvariablen führen, müssen mindestens eine Kante enthalten, die eine beobachtbare Variable berührt.

Bollen und Bauldry (2010) schlagen vor, die Identifikation komplexerer Systeme über ein algebraisches Lösen mittels Computer-Algebra-Systemen sicherzustellen. In unserem Fall wäre dies gegeben, wenn man von den in Abschnitt 3.1 dargestellten exogenen Effekten  $E_x$  unter Berücksichtigung der Identifikationsrestriktionen auf die partiellen Ableitungen  $M_x$  und  $M_y$  schließen kann.

Im Fall der exakten Identifikation, also gleich vieler Parameter und Gleichungen, sind drei Fälle zu unterscheiden: Liegt eine eindeutige Lösung vor, so ist das System global identifiziert, siehe auch Beispiel 1. Liegen mehrere Lösungen für die Parameter vor, wie es zum Beispiel bei Quadratwurzeln geschehen kann, so ist das System lokal identifiziert. Liegt keine Lösung vor, so ist das System nicht identifiziert.

Typisch ist der Fall der Überidentifikation. Es gibt dann mehr Gleichungen als Parameter, in unserem Fall also mehr freie Effekte als freie partielle Ableitungen. Dann gibt es keine eindeutige Lösung. Die Überidentifikation kann jedoch auf Fälle der exakten Identifikation zurückgeführt werden, indem Sub-Gleichungssysteme betrachtet werden, in denen alle Parameter vorkommen. Es

werden also überschüssige Informationen weggeworfen. Dies führt zwar zu unterschiedlichen Parametern und ist somit nicht als Lösung verwendbar, jedoch ist eine eindeutige Lösung eines Sub-Systems hinreichend für die Sicherstellung der globalen Identifikation des gesamten Systems, siehe Bollen, K. A., Bauldry, S. (2010b).

Eine Rückrechnung aus den totalen Ableitungen der reduzierten Form auf die partiellen Ableitungen der strukturierten Form ist also im üblichen überbestimmten Fall nicht möglich. Dies liegt auch daran, dass die Identifikationsrestriktionen der strukturierten Form nicht wiedergewonnen werden können. Nur in einem Spezialfall sind partielle Ableitungen aus den Effekten ablesbar: Dies ist dann der Fall, wenn es nur genau einen Pfad zwischen den Variablen gibt und dieser die Länge eins hat. Dies sind diejenigen Elemente von  $E_x \circ ID_x$  und  $E_y \circ ID_y$ , die sonst keine weiteren Pfade aufweisen, siehe auch Beispiel 1. Das identifizierte strukturierte System muss also z. B. mittels kleinster Quadrate direkt geschätzt werden. Der Ansatz von Computer-Algebra-Systemen ist jedoch nicht einfach, da sie das Lösen von Matrix-Gleichungen mit nicht-kommutativen Variablen oft nicht unterstützen, siehe zum Beispiel SymPy <https://github.com/sympy/sympy/pull/12264>. Ferner sind die Identifikationsrestriktionen aufzuerlegen.

In der Praxis ist die empirische Schätzung der reduzierten Form einfach. Da wir jedoch an der strukturellen Form interessiert sind, werden wir in Kapitel 5 Methoden zur Schätzung der strukturierten Form behandeln. Dort werden auch Identifikationsmöglichkeiten für spezielle Fälle behandelt. Wir verwenden in der Praxis die empirische, lokale Identifikation über den Rang der Hesseschen Matrix, siehe Abschnitt 5.2. Die strukturelle Form ermöglicht die Schätzung der Stärke der Kausalität zwischen den Variablen. Die Struktur und Richtung der Kausalität ist jedoch durch den Modellierer vorzugeben, wie im nachfolgenden Abschnitt dargestellt.

## 1.8 Kausalität

Grundsätzlich ist jede endogene Modellvariable einer Gleichung zuzuordnen. Die Reihenfolge der Gleichungen spielt dabei keine Rolle, eine ggf. vorhandene topologische Reihenfolge kann automatisch erkannt werden. Zur Vereinfachung der Notation wird die Variable  $y_j$  durch die  $j$ -te Gleichung bestimmt. Die  $j$ -te Gleichung wird dann als implizite *Funktion* der  $j$ -te Variablen aufgefasst:  $y_j$  ist dann eine Funktion aller anderen in der  $j$ -ten Gleichung vorkommenden Variablen. Dies ist in Abschnitt 1.2 bei der Anwendung des totalen Differentials bereits stillschweigend verwendet worden. Das totale Differential bzw. die totale Ableitung können nur auf Funktionen angewendet werden.

Neben der rein mathematischen Lösung der Gleichung wird die zugeordnete endogene Variable zudem als *kausale* Zielvariable interpretiert: das totale Differential der endogenen Variable führt zu einer linearen Gleichung, die nach der endogenen Variable *aufgelöst* wird. Die kausale Zielvariable steht dann auf der linken Seite der Gleichung und die sie beeinflussenden Ausgangsvariablen stehen auf der rechten Seite der Gleichung. Das Gleichheitszeichen ( $=$ ) ist dann als kausale Richtung ( $\leftarrow$ ) zu lesen. Die Kausalität ist dann aus den Koeffizientenmatrizen  $M_x$  und  $M_y$  der linearen strukturellen Form ablesbar. Die Kausalität wird durch die Pfeilrichtung in den gerichteten Graphen dargestellt. Siehe Beispiel 9 für den Einfluss unterschiedlicher kausaler Formulierungen.

Die Schätzung eines identifizierten strukturellen Modells mit gegebener Kausalitätsstruktur anhand realer Daten zeigt schließlich, welche der vorhandenen (nicht auf null gesetzten) Koeffizienten an-

gepasst werden sollten. Dabei ist zu beachten, dass die die Kausalität bestimmenden Identifikationsrestriktionen *nicht* empirisch validiert werden können. Die Kausalitätsstruktur ist vom Modellierer vorzugeben. Bei gegebener Kausalität kann jedoch deren Stärke empirisch geschätzt werden.

## 1.9 Effekt

Das Gleichungssystem kann in einen gerichteten Graphen überführt werden. Die Knoten des Graphen sind die Variablen des Systems und die Richtungen der Graphen ergeben sich dadurch, welche Variable in welche Modellgleichung einfließt. In der Darstellung des sogenannten partiellen Graphen, siehe auch Abschnitt 4.3, wird die partielle Ableitung zwischen zwei Variablen auf der entsprechenden Kante eingezeichnet. Man erhält so einen gewichteten Graphen und die Gewichte entsprechen den Einträgen von  $M_x$  und  $M_y$ . Entsprechend werden im sogenannten totalen Graphen, siehe ebenfalls Abschnitt 4.3, die totalen Ableitungen, also die Effekte auf den Kanten eingezeichnet.

Die graphentheoretischen Effekte werden üblicherweise für den DAG und voneinander unabhängige Fehlern behandelt, da in diesem Fall deren Identifikation sichergestellt ist, siehe Pearl (2009) Abschnitt 3.2.3. Im zyklischen Fall sind weitere Annahmen erforderlich. So zeigen Spirtes (1994) und Koster (1995), dass d-Separation in zyklischen Modellen funktioniert, wenn das Modell linear ist und die Fehler normalverteilt sind. Dies ist der für uns relevante Fall. Alternativ zeigen Pearl and Dechter (2013), dass d-Separation in zyklischen Modellen funktioniert, wenn die Variablen diskret sind.

Wir verwenden den Effekt für linearisierte strukturelle Modelle gemäß Abschnitt 1.2 und zeigen in Kapitel 3, dass diese graphentheoretischen Effekte den totalen Ableitungen entsprechen. Dies gilt sowohl für den DAG-Fall als auch für zyklische Modelle, womit wir die Anwendbarkeit unserer Formeln auch auf den zyklischen Fall nachweisen. Der graphentheoretische Effekt, im Sinne von Pearl (2009) Definition 4, ist definiert als:

Der Effekte eines Knotens auf anderen Knoten ist die totale Ableitung nach diesem Knoten in demjenigen *Subsystem*, das durch Fixierung dieses Knotens entsteht. Dabei werden alle in diesen Knoten eingehenden Kanten ausgenullt, während die ausgehenden Kanten im System verbleiben.

Die entsprechende Zeile von  $M_x$  und  $M_y$  wird also ausgenullt. Die Dimension des Subsystems entspricht jedoch der des Ausgangssystems.

Es ist also zuerst die Variable zu löschen, nach der abgeleitet werden soll und *danach* ist die Ableitung im Subsystem nach dieser nun exogenen Variable zu bestimmen. Dieses Löschen einer Gleichung wird als do-Calculus bezeichnet, siehe Pearl (2009), da eine Variable kontrafaktisch, also entgegen der sie eigentlich bestimmenden Modellgleichungen auf einen bestimmten Wert gesetzt wird (englischer Imperativ „do“!). Daher wird die Gleichung dieser Variable aus dem System gelöscht und ihr Wert wie bei einer exogenen Variablen als fixer Inputparameter vorgegeben. Inkonsistenzen werden so vermieden und der Effekt bzw. die Kausalität dieser Variablen kann bestimmt werden. Ein solcher Ansatz im Rahmen von simultanen Gleichungssystemen SEM wird auch Structural Causal Model, SCM, genannt. Dieser Ansatz grenzt sich somit klar von dem weit verbreiteten Ansatz der bedingten Erwartungswerte ab, der nie kontrafaktisch wirkt. Algebraisch

entspricht der do-Calculus dem Ersetzen der Funktionsgleichung der betrachteten Variablen durch eine Konstante, an deren Stelle die Ableitung bestimmt werden soll.

In linearen Systemen ergeben sich so lineare Effekte, die einfach mit der unterstellten Änderung multipliziert werden. Im nichtlinearen Fall sind die Effekte hingegen im allgemein nicht identifizierbar, siehe Pearl (2009) Seite 121.

Ein anderer Ansatz wird von Bollen (1989), Seite 383 ff., verfolgt. Seine sogenannten *spezifischen* Effekte ergeben sich aus der Differenz der ursprünglichen Effekte und derjenigen Effekten, die resultieren, wenn man die *Zeilen und Spalten* von  $M_x$  und  $M_y$  auf null setzt, die der Ausgangsvariablen entsprechen. Es werden also ein- und ausgehende Kanten eliminiert. So erhält man nur die Wirkungen der Ausgangsvariablen.

Bollen (1989), Seite 376 ff., verwendet im Übrigen ähnliche Bezeichnungen wie wir: Die partiellen Ableitungen bezeichnet er als direkte Effekte und unsere Effekte bezeichnet er als totale Effekte. Die Differenz zwischen beiden bezeichnet er ferner als indirekte Effekte.

### 1.10 Notation

Bei den Matrizen der partiellen und totalen Ableitungen werden die Variablen, nach denen abgeleitet wird, in den Spalten dargestellt:

$$\begin{aligned}
 M_y &= \begin{matrix} (n \times n) \\ \left[ \begin{array}{ccc} \frac{\partial M^1}{\partial y_1} & \dots & \frac{\partial M^1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial M^n}{\partial y_1} & \dots & \frac{\partial M^n}{\partial y_n} \end{array} \right] \end{matrix}, & M_x &= \begin{matrix} (n \times m) \\ \left[ \begin{array}{ccc} \frac{\partial M^1}{\partial x_1} & \dots & \frac{\partial M^1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial M^n}{\partial x_1} & \dots & \frac{\partial M^n}{\partial x_m} \end{array} \right] \end{matrix}, \\
 \frac{dy}{dy^T} &= \begin{matrix} (n \times n) \\ \left[ \begin{array}{ccc} \frac{dM^1}{dy_1} & \dots & \frac{dM^1}{dy_n} \\ \vdots & \ddots & \vdots \\ \frac{dM^n}{dy_1} & \dots & \frac{dM^n}{dy_n} \end{array} \right] \end{matrix}, & \frac{dy}{dx^T} &= \begin{matrix} (n \times m) \\ \left[ \begin{array}{ccc} \frac{dM^1}{dx_1} & \dots & \frac{dM^1}{dx_m} \\ \vdots & \ddots & \vdots \\ \frac{dM^n}{dx_1} & \dots & \frac{dM^n}{dx_m} \end{array} \right] \end{matrix}.
 \end{aligned}$$

Die Ableitungen werden äquivalent notiert als:

$$\frac{\partial y_j}{\partial y_t} = \frac{\partial M^j}{\partial y_t}, \quad \frac{\partial y_j}{\partial x_i} = \frac{\partial M^j}{\partial x_i}, \quad \text{bzw.} \quad \frac{dy_j}{dy_t} = \frac{dM^j}{dy_t}, \quad \frac{dy_j}{dx_i} = \frac{dM^j}{dx_i},$$

wobei auf der rechten Seite der Gleichung explizit dargestellt wird, welche Funktion abgeleitet wurde.

Wir benötigen weitere Konventionen für die Notation, da wir neben dem ursprünglichen System der Modellgleichungen auch Subsysteme behandeln: Wir betrachten  $\iota = 1, \dots, n$  Subsysteme, bei denen jeweils die  $\iota$ -te Gleichung gelöscht wurde. Das Löschen einer Gleichung kann äquivalent abgebildet werden, indem man die entsprechende Zeile der Matrix  $M_y$  ausnullt. Wir vereinfachen die Notation,

indem wir die Nummerierung der verbliebenen Gleichungen des Subsystems wie die des Ausgangssystems bezeichnen.

Wir verwenden folgende Notationen: Matrizen werden in Großbuchstaben geschrieben, Vektoren werden fett geschrieben und wir verwenden den gleichen Buchstaben wie für die Matrix, jedoch als Kleinbuchstaben. Standardmäßig sind Vektoren Spaltenvektoren. Ein Zeilenvektor wird mit Zeilenindex in Klammern geschrieben. Ein Element der Matrix, also ein Skalar, wird mit entsprechendem Kleinbuchstaben geschrieben. Die Notation ist beispielhaft für die Matrix  $M_y$  dargestellt:

$$M_y = [m_{yj}]_{j,l=1,\dots,n} = \begin{bmatrix} \mathbf{m}_{y(1)} \\ \vdots \\ \mathbf{m}_{y(n)} \end{bmatrix} = [\mathbf{m}_{y1}, \dots, \mathbf{m}_{yn}].$$

$M_y$  ist die Ausgangsmatrix,

$M_{yt}$  ist die Matrix  $M_y$ , bei der die  $t$ -te Spalte ausgenullt wurde,

$M_{y(t)}$  ist die Matrix  $M_y$ , bei der die  $t$ -te Zeile ausgenullt wurde,

$M_{y\iota\iota} = M_{y(t)\iota}$  ist die Matrix  $M_y$ , bei der die  $t$ -te Zeile und Spalte ausgenullt wurden,

$M_y^k$  ist die mit  $k$  potenzierte die Matrix  $M_y$ ,

$M_{y\iota\iota}^k$  ist die potenzierte Matrix  $M_y$ , bei der die  $t$ -te Zeile, Spalte ausgenullt wurden,

$m_{yij}$  ist das  $i, j$ -te Element von  $M_y$  und von  $M_{y\iota\iota}$ ,

$m_{yij,k}$  ist das  $i, j$ -te Element von  $M_y^k$ ,

$\mathbf{m}_{yj}$  ist die  $j$ -te Spalte von  $M_y$ ,

$\mathbf{m}_{y(j)}$  ist die  $j$ -te Zeile von  $M_y$ ,

$\mathbf{m}_{yj,k}$  ist die  $j$ -te Spalte von  $M_y^k$ ,

$\mathbf{m}_{y(j),k}$  ist die  $j$ -te Zeile von  $M_y^k$ ,

$\mathbf{m}_{yj\iota}$  ist die  $j$ -te Spalte von  $M_y$ , bei der das  $t$ -te Element ausgenullt wurde,

$\mathbf{m}_{y(j)\iota}$  ist die  $j$ -te Zeile von  $M_y$ , bei der das  $t$ -te Element ausgenullt wurde,

$(\cdot)_{ij}$  ist das  $i, j$ -te Element des Ausdrucks in runden Klammern,

$(\cdot)_{(j)}$  ist die  $j$ -te Zeile des Ausdrucks in runden Klammern,

$(\cdot)_{\iota}$  ist die  $t$ -te Spalte des Ausdrucks in runden Klammern.

Im Fall der Partitionierung einer Matrix in Blockmatrizen ist es hingegen erforderlich, dass die Zeilen und Spalten tatsächlich herausgeschnitten und nicht nur ausgenullt wurden. Die Dimension ist entsprechend reduziert. Dies wird durch in eckige Klammern gesetzten Indizes deutlich gemacht, beispielsweise bei:

$M_{y[u]}$  ist die Matrix  $M_y$ , bei der die  $t$ -te Zeile und Spalte herausgeschnitten wurde.

Die zu  $M_y$  gehörende Identifikationsmatrix  $ID_y$  enthält an den Stellen eine Eins, an denen der Eintrag von  $M_y$  ungleich null ist; ansonsten sind die Einträge null. Die entsprechende Identifikations-

matrix der Effekte  $E_y$  bezeichnen wir mit  $ED_y$  und die Identifikationsmatrix der finalen Effekte  $E_{y,y}$  bezeichnen wir mit  $FD_y$ .

Für Zwecke der Überprüfung der Identifikation benötigen wir noch eine besondere Notation:  $M_{[j,-j]}$  ist die Matrix  $M$ , bei der die  $j$ -te Zeile herausgeschnitten wird und nur die Spalten behalten werden, bei denen in der  $j$ -ten Zeile eine null stand.

Die elementweise Multiplikation von Matrizen, also das Hadamard-Produkt, wird mit  $\circ$  bezeichnet. Für die elementweise Division von Matrizen wird  $/$  verwendet. Das Kronecker-Produkt wird mit  $\otimes$  bezeichnet.  $I_n$  bezeichnet die  $n$ -dimensionale Einheitsmatrix und  $\mathbf{i}_t$  ist der  $t$ -te Einheitsvektor mit passender Dimension.  $\mathbf{1}_{nm}$  bezeichnet die  $(n \times m)$ -Matrix aus Einsen und  $\mathbf{1}_t$  ist der Eins-Spaltenvektor mit passender Dimension, bei dem das  $t$ -te Element null ist.  $J^{ij}$  ist die Elementarmatrix mit einer Eins an der Stelle  $ij$  und Nullen sonst.  $\delta_{jt}$  ist das Kronecker-Delta, welches eins annimmt, wenn  $j = t$  und null sonst. Der Vektorisierungsoperator  $\text{vec}(\cdot)$  ordnet die Spalten einer Matrix untereinander an und  $\text{vec}^{-1}(\cdot)$  ist sein inverser Operator, der die ursprüngliche Matrix mit ihrer Ausgangsdimension wieder herstellt. Der Operator  $\text{diag}(\cdot)$  erstellt eine quadratische Matrix, deren Hauptdiagonale dem übergebenen Vektor entspricht und die sonst nur Nullen aufweist.  $\text{tr}(\cdot)$  bezeichnet die Spur bzw. Trace, also die Summe der Hauptdiagonalelemente einer quadratischen Matrix.

Bei der Darstellung des Matrix-Calculus wird abhängig von der Dimension des Zählers und Nenners von  $\partial \mathbf{y} / \partial \mathbf{x}$  entweder das Numerator-Layout oder das Denominator-Layout verwendet. Diese sind definiert als:

- Numerator-Layout: entsprechend  $\mathbf{y}$  und  $\mathbf{x}^T$
- Denominator-Layout: entsprechend  $\mathbf{y}^T$  und  $\mathbf{x}$ ,

siehe [https://en.wikipedia.org/wiki/Matrix\\_calculus#Layout\\_conventions](https://en.wikipedia.org/wiki/Matrix_calculus#Layout_conventions). Wir verwenden folgende Layouts und modifizieren die Notation mit etwaigem Transponieren des Nenners, um die Dimensionen deutlicher zu machen:

- Vektor zu Vektor: Numerator-Layout, Nenner transponiert notiert  

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \rightarrow \frac{\partial \mathbf{y}}{\partial \mathbf{x}^T}$$
- Matrix zu Skalar: Numerator-Layout  

$$\frac{\partial \mathbf{M}}{\partial v}$$
- Skalar zu Matrix: Denominator-Layout  

$$\frac{\partial v}{\partial \mathbf{M}}$$

Die beiden letzten Fälle werden zusammen auch als Mixed Layout bezeichnet.

## 2) Totale Ableitungen

### 2.1 Exogene totale Ableitungen

Die totalen Ableitungen einer Funktion sind in Bezug auf exogene Variablen definiert. Wir nennen sie auch kurz exogene totale Ableitungen um sie von den nachfolgend eingeführten endogenen Ableitungen zu unterscheiden. Die aufgrund der Kettenregel im Produkt vorkommenden totalen Ableitungen einer exogenen Variablen nach sich selbst sind eins. Alle anderen Ableitungen von exogenen Variablen nach anderen Variablen sind null. Durch Division des totalen Differentials mit  $dx_i$  erhält man die exogenen totalen Ableitungen:

$$\begin{aligned}\frac{dy_j}{dx_i} &= \frac{dM^j}{dx_i}, \quad j = 1, \dots, n, \quad i = 1, \dots, m \\ &= \sum_{h=1}^n \frac{\partial M^j}{\partial y_h} \frac{dy_h}{dx_i} + \sum_{l=1}^m \frac{\partial M^j}{\partial x_l} \frac{dx_l}{dx_i} \\ &= \sum_{h=1}^n \frac{\partial M^j}{\partial y_h} \frac{dy_h}{dx_i} + \frac{\partial M^j}{\partial x_i}\end{aligned}$$

mit

$$\frac{dx_i}{dx_i} = 1, \quad i = 1, \dots, m,$$

$$\frac{dx_l}{dx_i} = 0, \quad i, l = 1, \dots, m, \quad l \neq i.$$

Wir schreiben die exogenen totalen Ableitungen in Matrixform und lösen nach diesen auf:

$$\frac{dy}{dx^T} = M_y \frac{dy}{dx^T} + M_x$$

(n×m)

$$\Rightarrow \frac{dy}{dx^T} = (I_n - M_y)^{-1} M_x.$$

An dieser Stelle sei auf das Implizite-Funktionen-Theorem verwiesen, siehe [https://en.wikipedia.org/wiki/Implicit\\_function\\_theorem](https://en.wikipedia.org/wiki/Implicit_function_theorem) oder Chiang (1984), Part 3, Formel (8.23). Das Theorem liefert neben Aussagen zur Existenz und Differenzierbarkeit der impliziten Definitionen auch eine Formel für deren Ableitung. Diese entspricht genau den obigen exogenen Effekten, nämlich dem Produkt einer invertierten und einer negativen Jacobi-Matrix: Der Term  $(I_n - M_y)$  ist die Jacobi-Matrix der homogenen Funktionen nach den endogenen Variablen und  $-M_x$  ist die Jacobi-Matrix der homogenen Funktionen nach den exogenen Variablen. Durch die

Invertierung der endogenen Jacobi-Matrix kann man das System nach den partiellen Ableitungen der impliziten Funktionen lösen und erhält so die Effekte. Die impliziten Funktionen sind die Darstellung der endogenen Variablen ausschließlich als Funktionen der exogenen Variablen. Die Invertierbarkeit von  $(I_n - M_y)$  wird für das Implizite-Funktionen-Theorem *angenommen*. Auch Bollen (1989), Tabelle 2.1 nimmt die Invertierbarkeit an und weist auf Seite 379 darauf hin, dass dies nötig dafür ist, dass die (totalen) Effekte definiert sind.

Selbst wenn die Modellgleichungen nicht bekannt sind, sondern nur ihre Ableitungen, so spiegeln die hier bestimmten Effekte in der Nähe der Ableitungen das Verhalten des nichtlinearen Systems der Modellgleichungen wieder. Das Theorem vermeidet das explizite Substituieren der Modellgleichungen. In der Ökonomie macht die Methodik der komparativen Statik Gebrauch vom Implizite-Funktionen-Theorem. Die Lösung des Gleichungssystems wird als Gleichgewicht interpretiert und die totalen Ableitungen, ausgedrückt ausschließlich in Form der partiellen Ableitungen, zeigen, wie sich das Gleichgewicht ändert, wenn sich die exogenen Variablen ändern.

Eine andere Interpretation ist, dass die Auflösung des Systems die strukturelle Darstellung in seine reduzierte Form überführt, die direkt abgeleitet werden kann. Siehe auch Abschnitt 5.1 für weitere Erläuterungen zur reduzierten Form.

## 2.2 Endogene totale Ableitungen

Wir erweitern die totalen Ableitungen in Bezug auf die endogenen Variablen. Dies erfolgt durch Division des totalen Differentials mit  $dy_t$ . Damit die endogenen totalen Ableitungen wohl definiert sind, *definieren* wir die Eigenableitungen als eins. Ansonsten ergäbe sich die Bestimmungsgleichung  $dy/dy^T = M_y(dy/dy^T)$  mit der trivialen Lösung  $dy/dy^T = \mathbf{0}$ . Die Definition in Höhe von eins ist naheliegend, da sich die latente Variable sonst anders ändern würde als beabsichtigt. Wir verwenden eine entsprechende Annahme in Abschnitt 3.2, definieren die Eigeneffekte als eins und zeigen, dass die endogenen totalen Ableitungen den endogenen Effekten entsprechen. Wie wir in Abschnitt 3.2 erläutern, definiert Bollen (1989) die gesamten endogenen Effekte direkt und nicht über die Nebenbedingung der Eigeneffekte und kommt zu anderen Größen.

In der nachfolgenden Darstellung sind die als Eins definierten Eigeneffekte explizit dargestellt, um die Struktur zu verdeutlichen. Der nachfolgende letzte Schritt wird hier nicht benötigt, die Fallunterscheidung  $h \neq t$  verdeutlicht jedoch die Aufteilung in direkte und indirekte Wirkung: wie üblich, entsprechen die partiellen Ableitungen den direkten Wirkungen und die Produkte den indirekten Wirkungen. Das Bestimmungssystem der endogenen totalen Ableitungen ist also simultan definiert:

$$\frac{dy_j}{dy_t} = \frac{dM^j}{dy_t}, \quad j, t = 1, \dots, n$$

$$\equiv (1 - \delta_{jt}) \sum_{h=1}^n \frac{\partial M^j}{\partial y_h} \frac{dy_h}{dy_t} + 1 \cdot \delta_{jt}$$

$$= (1 - \delta_{jt}) \left( \sum_{\substack{h=1 \\ h \neq t}}^n \left( \frac{\partial M^j}{\partial y_h} \frac{dy_h}{dy_t} \right) + \frac{\partial M^j}{\partial y_t} \cdot 1 \right) + 1 \cdot \delta_{jt}$$

mit

$$\frac{dy_t}{dy_t} \equiv 1, \quad t = 1, \dots, n,$$

$$\frac{dx_i}{dy_j} = 0, \quad j = 1, \dots, n, \quad i = 1, \dots, m.$$

Die Bestimmungsgleichung der totalen Ableitungen lässt sich unter Verwendung des elementweisen Hadamard-Produkts wie folgt in Matrixform vereinfachen:

$$\frac{dy}{dy^T} = (1_{nn} - I_n) \circ \left( M_y \frac{dy}{dy^T} \right) + I_n$$

(n×n)

$$= M_y \frac{dy}{dy^T} - I_n \circ \left( M_y \frac{dy}{dy^T} \right) + I_n$$

$$\Leftrightarrow \frac{dy}{dy^T} = (I_n - M_y)^{-1} \left( I_n \circ \left( M_y \frac{dy}{dy^T} \right) + I_n \right).$$

Also hat die totale Ableitung die Gestalt:

$$\frac{dy}{dy^T} = (I_n - M_y)^{-1} D,$$

wobei D eine Diagonalmatrix dergestalt ist, so dass die Matrix der totalen Ableitungen definitionsgemäß Einsen auf der Hauptdiagonale hat,  $I_n \circ (dy/dy^T) = I_n$ . Also muss D eine spaltenweise Normierung des Faktors  $(I_n - M_y)^{-1}$  bewirken und daher muss gelten:

$$D = \left( I_n \circ (I_n - M_y)^{-1} \right)^{-1}$$

und damit folgt insgesamt:

$$\frac{dy}{dy^T} = (I_n - M_y)^{-1} \left( I_n \circ (I_n - M_y)^{-1} \right)^{-1}.$$

Eine alternative Darstellung ergibt sich unter Verwendung des diag- und vec-Operators und des Kronecker-Produkts:

$$\frac{dy}{dy^T} = (1_{nn} - I_n) \circ \left( M_y \frac{dy}{dy^T} \right) + I_n,$$

$$\Leftrightarrow \text{vec} \left( \frac{dy}{dy^T} \right) = \text{vec}(1_{nn} - I_n) \circ \text{vec} \left( M_y \frac{dy}{dy^T} \right) + \text{vec}(I_n)$$

$$= \text{vec}(1_{nn} - I_n) \circ \left( (I_n \otimes M_y) \text{vec} \left( \frac{dy}{dy^T} \right) \right) + \text{vec}(I_n)$$

$$= \text{diag}(\text{vec}(1_{nn} - I_n)) (I_n \otimes M_y) \text{vec} \left( \frac{dy}{dy^T} \right) + \text{vec}(I_n)$$

$$\Leftrightarrow \frac{dy}{dy^T} = \text{vec}^{-1} \left( \left( I_{n^2} - \text{diag}(\text{vec}(1_{nn} - I_n)) (I_n \otimes M_y) \right)^{-1} \text{vec}(I_n) \right).$$

(n×n)

Hier wurde der Vektorisierungsoperator  $\text{vec}$  verwendet, der die Spalten einer Matrix zu einem langen Vektor zusammenstellt. Sein inverser Operator erstellt aus dem Vektor wieder eine Matrix mit der ursprünglichen Dimension.

Die Inverse  $(I_n - M_y)^{-1}$ , deren Existenz wir stets annehmen, kann als unendliches Polynom geschrieben werden. Hierfür ist die Konvergenz der Neumann-Reihe erforderlich, siehe [https://en.wikipedia.org/wiki/Neumann\\_series](https://en.wikipedia.org/wiki/Neumann_series):

$$(I_n - M_y)^{-1} = I_n + \sum_{k=1}^{\infty} M_y^k.$$

Für den DAG-Fall ergibt sich eine einfachere Formel, da  $(I_n - M_y)$  und somit auch seine Inverse eine unitrianguläre Matrix, also mit Einsen auf der Hauptdiagonale, ist. Zudem ist  $M_y$  nilpotent vom Grad höchstens  $n$ , so dass die Konvergenz gewährleistet ist und die Reihe nach spätestens  $n$  Schritten abbricht und. Auf eine Matrixinversion kann im DAG also verzichtet werden:

$$\begin{aligned} \frac{dy}{dy^T} &= (I_n - M_y)^{-1} \\ &= I_n + \sum_{k=1}^{n-1} M_y^k. \end{aligned}$$

Der endogene Effekt ist im DAG die Summe der endogenen Pfadefekte, siehe Abschnitt 4.1, über alle Wege und Längen für die Pfade von  $y_i$  nach  $y_j$ . Setzt man anstatt  $M_y$  die ungewichtete Identifikationsmatrix  $ID_y^T$  ein, die nur aus Einsen und Nullen besteht, so erhält man die Anzahlen der  $k$ -Schritt-Pfade im gerichteten Graphen, siehe auch Abschnitt 4.1.

Ferner gilt im DAG die folgende Beziehung, die wir auch in Abschnitt 3.4 zur Vereinfachung der endogenen finalen Effekte im DAG verwenden:

$$\begin{aligned}
 \frac{dy}{dy^T} &= I_n + \sum_{k=1}^{n-1} M_y^k \\
 &= M_y + \sum_{k=2}^{n-1} M_y^k + I_n \\
 &= M_y + \sum_{k=1}^{n-1} M_y^{k+1} + I_n \\
 &= \left( I_n + \sum_{k=1}^{n-1} M_y^k \right) M_y + I_n \\
 &= (I_n - M_y)^{-1} M_y + I_n \\
 &= \frac{dy}{dy^T} M_y + I_n.
 \end{aligned}$$

Wenn man im DAG-Fall den Term  $(I_n - M_y)$  von rechts an die erste Gleichung heranmultipliziert und dann mit dem Ergebnis der Linksmultiplikation vergleicht, sieht man ferner, dass die Matrizen kommutieren:

$$M_y \frac{dy}{dy^T} = \frac{dy}{dy^T} M_y.$$

Tim Vieira (2017) zeigt in seinem Blog <https://timvieira.github.io/blog/post/2017/08/18/backprop-is-not-just-the-chain-rule/>, wie man die endogenen totalen Ableitungen mittels Lagrange-Multiplikator-Ansatz erhält. Dabei werden die endogenen Gleichungen als Nebenbedingung eines Optimierungsproblems verwendet. Die zugehörigen Lagrange-Multiplikatoren sind dann die totalen endogenen Ableitungen nach diesen Variablen. Diese sind, wie oben dargestellt, simultan definiert. Die Lösung entspricht der obigen,  $dy/dy^T = (I_n - M_y)^{-1}$ , abgesehen vom Normierungsfaktor, der bei uns im zyklischen Fall hinzukommt. Die Backpropagation-Lösung entspricht damit der in Abschnitt 3.2 behandelten Formel von Bollen (1989), Formel (8.81). Im für Backpropagation üblichen DAG-Fall erhält man diese Lösung auch ohne Matrix-Inversion mit geringem Rechenaufwand. Im zyklischen Fall steigt der Rechenaufwand hingegen auf  $O(n^3)$ .

Die totalen Ableitungen, sowohl die exogenen als auch die endogenen, lassen sich auch mittels automatischer Differenzierung bestimmen, siehe [https://en.wikipedia.org/wiki/Automatic\\_differentiation](https://en.wikipedia.org/wiki/Automatic_differentiation). Eine mögliche Implementation bietet PyTorch, siehe <https://pytorch.org>. Dabei erfolgt die Ableitung auf Basis des Programmcodes und liefert numerisch exakte Ergebnisse. In unserem Fall werden also die totalen Ableitungen der endo-

genen Variablen auf die exogenen und endogenen Variablen and der Stelle der aktuellen Daten direkt algorithmisch bestimmt. Hierfür ist als Input lediglich das Modell in absoluten Leveln erforderlich. Eine Einschränkung liegt darin, dass die totalen Ableitungen mittels automatischer Differenzierung in PyTorch zurzeit nur für den DAG-Fall bestimmt werden können. (ToDo:) Sollten die totalen Ableitungen in PyTorch künftig auch für den zyklischen Fall möglich sein, so ist dann noch zu prüfen, ob diese endogenen Eigeneffekte auf eins normiert sind.

### 3) Graphentheoretische Effekte

Wir unterscheiden in partielle Ableitungen und totale Ableitungen. Wie nachfolgend dargestellt, entsprechen die totalen Ableitungen den graphentheoretischen Effekten, wie sie in Abschnitt 1.9 definiert sind. Pearl (2001), Pearl (2009) Abschnitt 5.1 oder Pearl (2018) Abschnitt 4.1 nennt die partiellen Ableitungen Natural Direct Effect, NDE, und die totale Ableitungen Total Effect, TE. Ferner bezeichnet er die Differenz  $NIE = TE - NDE$  als Natural Indirect Effect.

Für das Verständnis der Zusammenhänge ist es hilfreich zu sehen, über welche Variablen – die sogenannten Mediatoren – die Ausgangsvariable auf die Zielvariable wirkt. Dies ist Bestandteil der graphischen Analyse des Systems. Die entsprechenden finalen Effekte werden in den Abschnitten 3.3 und 3.4 dargestellt.

#### 3.1 Exogene Effekte

In Abschnitt 1.9 wurden Effekte definiert als Ableitungen nach einem Knoten, dessen eingehende Kanten ausgenullt wurden. Bei den Ableitungen nach den exogenen Variablen liegt der einfache Spezialfall vor, dass sie nicht über eingehenden Kanten verfügen und somit im Rahmen des do-Calculus auch keine Modellgleichung gestrichen wird. Die Modellgleichungen werden mittels totalen Differentials approximativ linearisiert:

$$dy = M_y dy + M_x dx,$$

$$\Rightarrow dy = (I_n - M_y)^{-1} M_x dx.$$

Die Effekte der endogenen Variablen ergeben nach erfolgter Ableitung ein Gleichungssystem, das zu lösen ist:

$$\begin{aligned} E_x &= \frac{dy}{dx^T} \\ (n \times m) & \\ &= (I_n - M_y)^{-1} M_x. \end{aligned}$$

Die exogenen Effekte sind identisch zu den exogenen totalen Ableitungen.

Diese Formel findet sich auch in Bollen (1989), Formel (8.84). Dort bezieht sie sich auf die exogenen latenten Effekte und wurde über unendlich wiederholtes Einsetzen der Bestimmungsgleichung für die endogenen latenten Variablen in sich selbst abgeleitet.

Die absoluten Änderungen der endogenen Variablen lauten:

$$dy = E_x dx.$$

### 3.2 Endogene Effekte

Aus dem *heuristischen* Ansatz für die graphentheoretischen endogene Effekte  $E_y$  in Form von  $E_x = E_y M_x$  würde direkt  $E_y = (I_n - M_y)^{-1}$  folgen. Nachfolgend wird deutlich, dass diese Heuristik jedoch nur im DAG gilt. Wir betrachten jedoch den allgemeinen zyklischen Fall.

Die endogenen Effekte erhalten wir gemäß Definition in Abschnitt 1.9, indem die jeweilige endogene Variable  $y_t$  – also die  $t$ -te Zeile – aus dem linearisierten System ausgenullt wird, siehe die erste nachfolgende Zeile. Die entsprechende Zeile von  $M_x$  bzw.  $M_y$  bzw. das entsprechende Element von  $dy$  werden also ausgenullt, die Dimension des Subsystems jedoch beibehalten. Dann wird die totale Ableitung in diesem Subsystem nach dieser Variablen gebildet. Dafür wird die entsprechende Spalte separat geschrieben, so dass man nach den verbleibenden Variablen auflösen kann, siehe die zweite nachfolgende Zeile:

$$\begin{aligned} dy_t &= M_{y(t)} dy + M_{x(t)} dx, \quad t = 1 \dots, n \\ &= M_{y_{tu}} dy_t + \mathbf{m}_{y_{tu}} dy_t + M_{x(t)} dx \\ \Rightarrow dy_t &= (I_n - M_{y_{tu}})^{-1} \mathbf{m}_{y_{tu}} dy_t + (I_n - M_{y_{tu}})^{-1} M_{x(t)} dx. \end{aligned}$$

Die  $n$  Gleichungssysteme mit je  $n - 1$  Effekten ergeben zusammen mit den  $n$  definierten Eigeneffekten die  $n^2$  Effekte in  $E_y$ :

$$\begin{aligned} \mathbf{e}_{y_{tu}} &= \frac{dy_t}{dy_t}, \quad t = 1 \dots, n \\ (n \times 1) & \\ &= (I_n - M_{y_{tu}})^{-1} \mathbf{m}_{y_{tu}} \end{aligned}$$

$$e_{y_{tu}} \equiv 1, \quad t = 1, \dots, n,$$

$$E_y = [\mathbf{e}_{y_1}, \dots, \mathbf{e}_{y_n}] = [\mathbf{e}_{y_{11}}, \dots, \mathbf{e}_{y_{nn}}] + I_n.$$

Analog zur Definition der totalen Ableitung einer endogenen Variablen nach sich selbst als eins, definieren wir die Eigeneffekte als eins.

Wir zeigen nun, dass diese endogenen Effekte im allgemeinen zyklischen Fall den endogenen totalen Ableitungen entsprechen. Wir verwenden die beiden oben hergeleiteten Ausdrücke für die endogenen totalen Ableitungen:

$$\frac{dy}{dy^T} = (I_n - M_y)^{-1} (I_n \circ (I_n - M_y)^{-1})^{-1}$$

$$\frac{dy}{dy^T} = (1_{nn} - I_n) \circ \left( M_y \frac{dy}{dy^T} \right) + I_n.$$

Die in eckige Klammern gesetzten Indizes, beispielsweise nachfolgend bei  $M_{y[u]}$ , zeigen an, dass die entsprechenden Zeilen und Spalten tatsächlich herausgeschnitten wurden und nicht nur ausge-nullt wurden. Dies ist für die Partitionierung des Systems in Blockmatrizen erforderlich. Dann wird die Inverse der Blockmatrix gebildet, siehe [https://en.wikipedia.org/wiki/Block\\_matrix-#Block\\_matrix\\_inversion](https://en.wikipedia.org/wiki/Block_matrix-#Block_matrix_inversion). Dabei wird, ohne inhaltliche Änderung des Systems,  $\iota$  in die letzte Zeile und Spalte permutiert, das heißt, wir betrachten den Fall  $\iota = n$ :

$$I_n - M_y = \begin{bmatrix} I_{n-1} - M_{y[u]} & -\mathbf{m}_{y[\iota][\iota]} \\ -\mathbf{m}_{y(\iota)[\iota]} & 1 - m_{y\iota\iota} \end{bmatrix}$$

$$\Leftrightarrow (I_n - M_y)^{-1} = \begin{bmatrix} Z & Z \frac{\mathbf{m}_{y[\iota][\iota]}}{1 - m_{y\iota\iota}} \\ \frac{\mathbf{m}_{y(\iota)[\iota]}}{1 - m_{y\iota\iota}} Z & z \end{bmatrix}$$

mit

$$Z = \left( I_{n-1} - M_{y[u]} - \frac{\mathbf{m}_{y[\iota][\iota]} \mathbf{m}_{y(\iota)[\iota]}}{1 - m_{y\iota\iota}} \right)^{-1}$$

$$z = \frac{1}{1 - m_{y\iota\iota}} + \frac{\mathbf{m}_{y(\iota)[\iota]}}{1 - m_{y\iota\iota}} Z \frac{\mathbf{m}_{y[\iota][\iota]}}{1 - m_{y\iota\iota}}.$$

Da  $z$  das untere rechte Element von  $(I_n - M_y)^{-1}$  ist, gilt auch folgende Normierungseigenschaft für die  $\iota$ -te Spalte einer Matrix  $A$ :

$$A \left( I_n \circ (I_n - M_y)^{-1} \right)_\iota^{-1} = A_\iota / z.$$

Wir nutzen noch die Struktur der Inversen:

$$(I_n - M_y)(I_n - M_y)^{-1} = I_n$$

$$\Leftrightarrow \begin{bmatrix} I_{n-1} - M_{y[u]} & -\mathbf{m}_{y[\iota][\iota]} \\ -\mathbf{m}_{y(\iota)[\iota]} & 1 - m_{y\iota\iota} \end{bmatrix} \begin{bmatrix} Z & Z \frac{\mathbf{m}_{y[\iota][\iota]}}{1 - m_{y\iota\iota}} \\ \frac{\mathbf{m}_{y(\iota)[\iota]}}{1 - m_{y\iota\iota}} Z & z \end{bmatrix} = \begin{bmatrix} I_{n-1} & 0 \\ 0 & 1 \end{bmatrix}$$

und verwenden die Block-Gleichung oben rechts:

$$(I_{n-1} - M_{y[u]}) Z \frac{\mathbf{m}_{y[\iota][\iota]}}{1 - m_{y\iota\iota}} - \mathbf{m}_{y[\iota][\iota]} z = 0,$$

$$\Leftrightarrow \left( \frac{M_{y[u]}Z}{(1 - m_{yu})Z} + I_{n-1} \right) \mathbf{m}_{y[u]} = \frac{Z}{(1 - m_{yu})Z} \mathbf{m}_{y[u]},$$

wobei durch  $z \neq 0$  geteilt wurde, also muss notwendigerweise  $m_{yu} \neq 1$ ,  $u = 1, \dots, n$  sein.

Es ist zu zeigen:

$$\mathbf{e}_{y[u]} = \left( \frac{dy}{dy^T} \right)_u, \quad u = 1, \dots, n.$$

Wir setzen die Definition von  $\mathbf{e}_{y[u]}$  und den ersten Ausdruck für  $dy/dy^T$  ein. Im nächsten Schritt wird die Eigenschaft des zweiten Ausdrucks für  $dy/dy^T$  verwendet. Dann kann der Einheitsvektor  $\mathbf{i}_u$  auf beiden Seiten entfallen und die Normierungseigenschaft von  $z$  wird verwendet. Dann wird das System in Blockmatrizen geschrieben und die Blockgleichung zweimal angewendet.

$$\Leftrightarrow \mathbf{i}_u + (I_n - M_{yu})^{-1} \mathbf{m}_{yu} = \left( (I_n - M_y)^{-1} (I_n \circ (I_n - M_y)^{-1})^{-1} \right)_u$$

$$\Leftrightarrow \mathbf{i}_u + (I_n - M_{yu})^{-1} \mathbf{m}_{yu} = \left( (1_{nn} - I_n) \circ (M_y(I_n - M_y)^{-1} (I_n \circ (I_n - M_y)^{-1})^{-1}) + I_n \right)_u$$

$$\Leftrightarrow (I_n - M_{yu})^{-1} \mathbf{m}_{yu} = (1_{nn} - I_n)_u \circ (M_y(I_n - M_y)^{-1} (I_n \circ (I_n - M_y)^{-1})^{-1})_u$$

$$\Leftrightarrow (I_n - M_{yu})^{-1} \mathbf{m}_{yu} = \mathbf{1}_u \circ (M_y(I_n - M_y)^{-1} / z)$$

$$\Leftrightarrow \begin{bmatrix} I_{n-1} - M_{y[u]} & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{m}_{y[u]} \\ 0 \end{bmatrix} = \mathbf{1}_u \circ \left( \begin{bmatrix} M_{y[u]} & \mathbf{m}_{y[u]} \\ \mathbf{m}_{y^{(u)[u]}} & m_{yu} \end{bmatrix} \begin{bmatrix} I_{n-1} - M_{y[u]} & -\mathbf{m}_{y[u]} \\ -\mathbf{m}_{y^{(u)[u]}} & 1 - m_{yu} \end{bmatrix}^{-1} / z \right)$$

$$\Leftrightarrow \begin{bmatrix} (I_{n-1} - M_{y[u]})^{-1} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{m}_{y[u]} \\ 0 \end{bmatrix} = \mathbf{1}_u \circ \left( \begin{bmatrix} M_{y[u]} & \mathbf{m}_{y[u]} \\ \mathbf{m}_{y^{(u)[u]}} & m_{yu} \end{bmatrix} \begin{bmatrix} Z \frac{\mathbf{m}_{y[u]}}{1 - m_{yu}} \\ z \end{bmatrix} / z \right)$$

$$\Leftrightarrow \begin{bmatrix} (I_{n-1} - M_{y[u]})^{-1} \mathbf{m}_{y[u]} \\ 0 \end{bmatrix} = \begin{bmatrix} M_{y[u]}Z \frac{\mathbf{m}_{y[u]}}{(1 - m_{yu})Z} + \mathbf{m}_{y[u]} \\ 0 \end{bmatrix}$$

$$\Leftrightarrow (I_{n-1} - M_{y[u]})^{-1} \mathbf{m}_{y[u]} = \left( \frac{M_{y[u]}Z}{(1 - m_{yu})Z} + I_{n-1} \right) \mathbf{m}_{y[u]}$$

$$\Leftrightarrow \mathbf{m}_{y[u]} = (I_{n-1} - M_{y[u]}) \frac{Z}{(1 - m_{yu})Z} \mathbf{m}_{y[u]}$$

$$\Leftrightarrow \mathbf{m}_{y_t[t]} = \frac{Z}{(1 - m_{y_{tt}})_Z} \mathbf{m}_{y_t[t]} - \frac{M_{y_{[tt]}} Z}{(1 - m_{y_{tt}})_Z} \mathbf{m}_{y_t[t]} - \mathbf{m}_{y_t[t]} + \mathbf{m}_{y_t[t]}$$

$$\Leftrightarrow \mathbf{m}_{y_t[t]} = \frac{Z}{(1 - m_{y_{tt}})_Z} \mathbf{m}_{y_t[t]} - \frac{Z}{(1 - m_{y_{tt}})_Z} \mathbf{m}_{y_t[t]} + \mathbf{m}_{y_t[t]}$$

$$\Leftrightarrow \mathbf{m}_{y_t[t]} = \mathbf{m}_{y_t[t]}.$$

Dabei ist  $\mathbf{i}_t$  der  $t$ -te Einheitsvektor,  $\mathbf{1}_t$  ist der Eins-Vektor, bei dem das  $t$ -te Element null ist und  $(\cdot)_t$  ist die  $t$ -te Spalte des Ausdrucks in runden Klammern.

Damit stimmen die endogenen Effekte mit den totalen endogenen Ableitungen überein:

$$\begin{aligned} E_y &= [\mathbf{e}_{y_{11}}, \dots, \mathbf{e}_{y_{nn}}] + I_n \\ &= \left[ (I_n - M_{y_{11}})^{-1} \mathbf{m}_{y_{11}}, \dots, (I_n - M_{y_{nn}})^{-1} \mathbf{m}_{y_{nn}} \right] + I_n \\ &= \frac{dy}{dy^T} \\ &= (I_n - M_y)^{-1} (I_n \circ (I_n - M_y)^{-1})^{-1}. \end{aligned}$$

Interessanterweise erhält man also für die betrachtete Variable  $y_t$  aus dem Subsystem die gleichen Effekte wie aus dem gesamten System. Im Subsystem wurde die  $t$ -te Zeile und Spalte gestrichen, die  $t$ -te Spalte jedoch wieder heranzumultipliziert, so dass insgesamt die  $t$ -te Zeile fehlt. Diese spielt für die Effekte der  $t$ -ten Variablen keine Rolle! Das Löschen nur der  $t$ -ten Zeile entspricht dem Löschen der in  $y_t$  eingehenden Kanten, während die ausgehenden Kanten erhalten bleiben, da die entsprechende Spalte erhalten bleibt. Des Weiteren wird deutlich, dass die Multiplikation mit den gelöschten Spalten zugleich die Aufgabe der spaltenweisen Normierung übernimmt.

Der Vergleich mit der Literatur zeigt die Vorteile und Unterschiede unseres allgemeinen Ansatzes für die endogenen Effekte:

Die graphentheoretischen Effekte werden üblicherweise für den DAG und voneinander unabhängige Fehler behandelt, da in diesem Fall deren Identifikation sichergestellt ist, siehe Pearl (2009) Abschnitt 3.2.3. Im zyklischen Fall sind weitere Annahmen erforderlich. So zeigen Spirtes (1994) und Koster (1995), dass  $d$ -Separation in zyklischen Modellen funktioniert, wenn das Modell linear ist und die Fehler normalverteilt sind. Wir haben die Definition des Effekts gemäß Abschnitt 1.9 auch auf zyklische linearisierte strukturelle Modell angewandt. Als Bestätigung dieses Ansatzes dient die Identität der endogenen Effekte und der totalen Ableitungen. Dabei ist zu beachten, dass wir die Eigenableitungen in Abschnitt 2.2 und die Eigeneffekte in diesem Abschnitt 3.2 als eins *definieren*.

Bollen (1989), Formel (8.81), verwendet für die endogenen latenten Effekte hingegen anderen Größen! Er *definiert* die endogenen Effekte nicht über die Nebenbedingung der Eigeneffekte, sondern direkt als  $(I_n - M_y)^{-1} - I_n$ . Im DAG-Fall besteht der Unterschied lediglich darin, dass seine Ei-

geneffekte null sind. Im zyklischen Fall sind seine Eigeneffekte hingegen unbeschränkt, also im Allgemeinen ungleich null und auch ungleich eins. Es fehlt auch die entsprechende spaltenweise Normierung der Effektmatrix  $E_y$ .

Da unsere Ergebnisse auf der RAM-Darstellung beruhen, sind sie allgemeiner als in Bollen (1989) und umfassen auch Effekte von endogenen manifesten Variablen aufeinander, sowie von exogenen manifesten Variablen auf endogene manifeste Variablen. Diese sind in der LISREL-Notation von Bollen (1989) stets null. Bollen (1989), Seite 15, nimmt jedoch auch die Normierung gemäß Abschnitt 1.3 an, dass  $M_y$  Nullen auf der Hauptdiagonalen hat.

### 3.3 Exogene finale Effekte

Wir zerlegen die Effekte einer Variablen auf eine betrachtete „finale“ Variable auf ihre ausgehenden Kanten, also die direkt nachfolgenden Mediatoren. Die finale Variable kann jede endogene Variable des Systems sein. Sie ist inhaltlich von besonderem Interesse. Beispielsweise kann es das ökonomische Eigenkapital sein, das sich letztendlich aus dem Modell ergibt. Siehe auch Abschnitt 4.3 für die Darstellung als finaler Graph. Dies ermöglicht eine weitergehende Sensitivitäts- und Kausalitätsanalyse des Systems.

Jeder exogene Effekt, also jedes Element von

$$E_x = (I_n - M_y)^{-1} M_x$$

ist die Summe

$$\begin{aligned} \frac{dy_j}{dx_i} &= \left( (I_n - M_y)^{-1} \right)_{(j)} \mathbf{m}_{xi} \\ &= \sum_{h=1}^n e_{jhi,x} \end{aligned}$$

seiner finalen Effekte  $e_{jhi,x}$  von  $x_i$  über  $y_h$  nach  $y_j$ :

$$e_{jhi,x} = \left( (I_n - M_y)^{-1} \right)_{jh} \frac{\partial M^h}{\partial x_i},$$

bzw. im DAG:

$$e_{jhi,x} = \frac{dy_j}{dy_h} \frac{\partial M^h}{\partial x_i}.$$

Um die finalen Effekte in Matrixschreibweise zu formulieren ist eine der drei Dimensionen festzuhalten. Die exogenen finalen Effekte für eine gegebene Ausgangsvariable  $x_i$  lauten:

$$E_{x_i} = [e_{jhi,x}]_{j,h=1,\dots,n}, \quad i = 1, \dots, m$$

$$(n \times n)$$

$$= (I_n - M_y)^{-1} \circ (\mathbf{1}_n \mathbf{m}_{xi}^T),$$

bzw. für die übliche Betrachtung bei gegebener finaler Zielvariable  $y_j$ :

$$E_{y_jx} = [e_{jhi,x}]_{\substack{h=1,\dots,n \\ i=1,\dots,m}}, \quad j = 1, \dots, n$$

$$(n \times m)$$

$$= \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(m)} \right) \circ M_x$$

und gemäß Konstruktion gilt die Summationseigenschaft für die exogenen finalen Effekte:

$$E_x = [E_{x_1} \mathbf{1}_n, \dots, E_{x_m} \mathbf{1}_n] = \begin{bmatrix} \mathbf{1}_{(n)} E_{y_1x} \\ \vdots \\ \mathbf{1}_{(n)} E_{y_nx} \end{bmatrix}.$$

$$(n \times m)$$

### 3.4 Endogene finale Effekte

Wir zerlegen jeden endogenen Effekt in die Summe seiner ausgehenden finalen Effekte auf die finale Variable. Wir verwenden hierbei die graphentheoretische Darstellung, bei der der gelöschte Vektor an das Subsystem multipliziert wird. Die finalen Eigeneffekte müssen jedoch separat behandelt werden: Im Fall des DAG ist es gar nicht möglich, den als eins definierten Eigeneffekt zu zerlegen, da entweder der erste oder der zweite Faktor des Produkts der nachfolgend verwendeten Zerlegung null ist. Im zyklischen Fall erhält man ebenfalls Nullen, bis auf denjenigen finalen Eigeneffekt, der auch noch über sich selber läuft; er entspricht der partiellen Ableitung nach sich selbst und wir *definieren* ihn ebenfalls als null. Durch diese Definition wird für die endogenen finalen Effekte auch die Unabhängigkeit von der Darstellung des Systems erreicht, siehe Beispiel 7.

Jeder endogene Effekt, also jedes Element von

$$\mathbf{e}_{y_{\iota\iota}} = (I_n - M_{y_{\iota\iota}})^{-1} \mathbf{m}_{y_{\iota\iota}}, \quad \iota = 1, \dots, n,$$

$$e_{y_{\iota\iota}} \equiv 1$$

kann geschrieben werden als:

$$\frac{dy_j}{dy_{\iota}} = (1 - \delta_{j\iota}) \left( (I_n - M_{y_{\iota\iota}})^{-1} \right)_{(j)} \mathbf{m}_{y_{\iota\iota}} + \delta_{j\iota}$$

$$\begin{aligned}
&= (1 - \delta_{j\iota}) \sum_{h=1}^n \left( \left( (I_n - M_{y\iota})^{-1} \right)_{jh} (1 - \delta_{h\iota}) m_{yh\iota} \right) + \delta_{j\iota} \\
&= \sum_{h=1}^n \left( (1 - \delta_{j\iota}) \left( (I_n - M_{y\iota})^{-1} \right)_{jh} \frac{\partial M^h}{\partial y_\iota} \right) + \delta_{j\iota}
\end{aligned}$$

wobei der Faktor  $(1 - \delta_{h\iota})$  entfallen kann, da das  $jh$ -te Element von  $(I_n - M_{y\iota})^{-1}$  für  $h = \iota$  gleich null ist. Somit kann jeder endogene Effekt für  $j \neq \iota$  in die Summe seiner endogenen finalen Effekte zerlegt werden:

$$\frac{dy_j}{dy_\iota} = \sum_{h=1}^n e_{jh\iota,y}, \quad j \neq \iota$$

mit den finalen Effekten  $e_{jh\iota,y}$  von  $y_\iota$  über  $y_h$  auf die finale Variable  $y_j$ :

$$e_{jh\iota,y} = \left( (I_n - M_{y\iota})^{-1} \right)_{jh} \frac{\partial M^h}{\partial y_\iota}, \quad j, h, \iota = 1, \dots, n, \quad \text{ohne } j = h = \iota,$$

$$e_{\iota\iota,y} \equiv 0, \quad \iota = 1, \dots, n.$$

Im DAG kann der finale Effekt auch als Produkt von totaler und partieller Ableitung geschrieben werden. Die von  $y_\iota$  nach  $y_h$  ausgehende Kante wird mit der partiellen Ableitung gewichtet und der restliche Pfad von  $y_h$  zur Zielvariablen  $y_j$  wird mit dem entsprechenden Effekt gewichtet:

$$e_{jh\iota,y} = \frac{dy_j}{dy_h} \frac{\partial M^h}{\partial y_\iota}.$$

In der Matrixschreibweise wird die obige Definition durch den ersten Hadamard-Faktor  $(\mathbf{1}_{nn} - I_n)$  umgesetzt. Die endogenen finalen Effekte für eine gegebene Ausgangsvariable  $y_\iota$  lauten:

$$\begin{aligned}
E_{y_\iota} &= [e_{jh\iota,y}]_{j,h=1,\dots,n}, \quad i = 1, \dots, m \\
(n \times n) & \\
&= (\mathbf{1}_{nn} - I_n) \circ (I_n - M_{y\iota})^{-1} \circ (\mathbf{1}_n \mathbf{m}_{y_\iota}^T)
\end{aligned}$$

und dies ist im DAG eine strikte untere Dreiecksmatrix, deren  $\iota$  erste Spalten null sind.

In der üblichen Betrachtung bei gegebener finaler Zielvariable  $y_j$  hängen der zweite und dritte Hadamard-Faktor von  $\iota$  ab, so dass  $E_{y_j y}$  spaltenweise erstellt werden muss:

$$E_{y_j y} = [e_{jh\iota,y}]_{h,\iota=1,\dots,n}, \quad j = 1, \dots, n$$

$(n \times n)$

$$\begin{aligned}
&= \left[ (E_{y_j y})_1, \dots, (E_{y_j y})_n \right] \\
&= (\mathbf{1}_{nn} - I_n) \circ \left[ \left( (I_n - M_{y_{11}})^{-1} \right)_{(j)}^T, \dots, \left( (I_n - M_{y_{nn}})^{-1} \right)_{(j)}^T \right] \circ M_y
\end{aligned}$$

und dies ist im DAG eine strikte untere Dreiecksmatrix und alle Spalten größer als  $j$  sind gleich null, da es keinen Einfluss von nachfolgenden Outputvariablen auf  $y_j$  gibt. Im DAG gilt  $E_y = E_y M_y + I_n$ , siehe Abschnitt 2.2, und daher können die obigen Formeln für die endogenen finalen Effekte in diesem Fall auch mit  $M_y$  anstatt mit  $M_{y_u}$  geschrieben werden. Dies erleichtert die Berechnung von  $E_{y_j y}$ , da die Fallunterscheidung für die Spalte  $\iota$  entfällt. Also gilt im DAG:

$$E_{y_\iota} = (I_n - M_y)^{-1} \circ (\mathbf{1}_n \mathbf{m}_{y_\iota}^T),$$

$$E_{y_j y} = \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(n)} \right) \circ M_y.$$

Gemäß Konstruktion gilt die Summationseigenschaft für die endogenen finalen Effekte:

$$E_y = [E_{y_1} \mathbf{1}_n, \dots, E_{y_n} \mathbf{1}_n] = \begin{bmatrix} \mathbf{1}_{(n)} E_{y_1 y} \\ \vdots \\ \mathbf{1}_{(n)} E_{y_n y} \end{bmatrix}.$$

$(n \times n)$

## 4) Graphentheoretische Interpretation

Wir fassen vorab zentrale graphentheoretische Definitionen zusammen: Ein Weg im Graphen ist eine Sequenz von direkt verbundenen Knoten. Ein Pfad ist dadurch gekennzeichnet, dass jeder Knoten höchstens einmal besucht werden. Im DAG sind alle Wege Pfade. Sind Start- und Endknoten gleich, sonst aber alle Knoten verschieden, so liegt ein Zyklus vor. Hat eine Kante den gleichen Ausgangs- und Endpunkt, so ist dies eine Schleife. Ein Multigraph hat mehrere Kanten zwischen zwei Knoten. Siehe Brualdi and Cvetkovic (2008) Seiten 2–3.

Wir gehen stets von einfachen gerichteten Graphen aus, also ohne Schleifen und Multikanten.

### 4.1 Adjazenzmatrix

Die Transponierte von  $M_y$  ist die gewichtete Adjazenzmatrix  $A$  des gerichteten Graphen des Gleichungssystems  $M^j$ ,  $j = 1 \dots, n$ , wobei die Gewichte den partiellen Ableitungen, also den Einträgen von  $M_y$ , entsprechen:

$$A = M_y^T.$$

Die gewichtete Adjazenzmatrix ist die transponierte Jacobian Matrix und wird auch als Gradient bezeichnet, siehe Magnus and Neudecker (2007). Die Matrix  $M_y$  zeigt an, welche Modellgleichung durch welche anderen Modellgleichungen beeinflusst wird. Umgekehrt gibt die Matrix  $A$  an, welche Modellgleichungen andere Modellgleichungen beeinflussen. Dies führt zur Transponierten. Im DAG mit topologischer Reihenfolge ist die Adjazenzmatrix eine obere Dreiecksmatrix.

Ferner ist die ungewichtete Adjazenzmatrix  $\bar{A}$  von Relevanz. Sie enthält in ihrem  $i, j$ -ten Element die Anzahl der Kanten, die inzident sind zu den Knoten  $i$  und  $j$ , siehe Stanley (2013) Seite 10. Da wir von einfachen gerichteten Graphen ausgehen, also ohne Schleifen und Multikanten, besteht unsere Adjazenzmatrix nur aus Nullen und Einsen. Die ungewichtete und die gewichtete Adjazenzmatrix haben also die gleichen Nullelemente und die ungewichtete Adjazenzmatrix hat Einsen an den Stellen, wo die gewichtete Adjazenzmatrix keine Nullen hat. Dies lässt sich direkt über die Identifikationsmatrix formulieren:

$$\bar{A} = ID_y^T.$$

### 4.2 Pfadefekte

Mit Brualdi and Cvetkovic (2008) Theorem 3.1.2 gilt:

Die potenzierte  $(n \times n)$  Adjazenzmatrix  $A^k$  enthält im Element  $a_{ij,k}$  die Summe aller Pfadefekte  $p_{ij,ks}$  der Wege der Länge  $k$  von  $i$  nach  $j$ .

Brualdi and Cvetkovic (2008) Seite 51 nennen den einzelnen Pfadefekt Weggewichte. Der Pfadefekt ist definiert als das Produkt der Gewichte aller Kanten des Wegs. Dieser Begriff ist zu unterscheiden von den pfadspezifischen Effekten, siehe Pearl (2001) Definition 8, welches Effekte sind,

bei denen die nicht interessierenden Kanten festgehalten werden. Entsprechend gilt mit Stanley (2013) Theorem 1.1:

Die potenzierte ungewichtete Adjazenzmatrix  $\bar{A}^k$  enthält im Element  $a_{ij,k}$  die Anzahl der Wege der Länge  $k$  von  $i$  nach  $j$ .

Die Theoreme lassen sich direkt mit der Eigenschaft der Matrixmultiplikation beweisen, siehe Stanley (2013) Beweis zu Theorem 1.1:

$$a_{ij,k} = \sum_s a_{ii_1} a_{i_1 i_2} \cdots a_{i_{k-1} j} = \sum_s p_{ij,ks} = \sum_s m_{y_j i_{k-1}} \cdots m_{y_2 i_1} m_{y_1 i}$$

wobei über alle Sequenzen  $s = (i_1, \dots, i_{k-1})$  mit  $1 \leq i_l \leq n$  der Länge  $k$  summiert wird, die in  $i$  starten und in  $j$  enden.

Im DAG ist  $A^n = 0$ , siehe Brualdi and Cvetkovic (2008) Theorem 3.1.4. Es reicht im DAG aus, bis zur Potenz  $n - 1$  zu summieren, siehe Israel (2011). Die Interpretation der Effekte über die Summe von Pfadeffekten ist im DAG einfacher, da die Summe endlich ist.

Die mit  $k$  potenzierte Adjazenzmatrix ist also die Produktschrittsumme der Elemente der Adjazenzmatrix, für die Pfade der Länge  $k$  zwischen zwei Variablen. Diese Darstellung lässt sich wie folgt nachvollziehen: Eine Zeile der Adjazenzmatrix gibt an, welche Variablen durch diese Variable beeinflusst werden. Bei einer dieser Variablen angekommen, muss man in deren Zeile springen, um ablesen zu können, welche weiteren Variablen von dort aus beeinflusst werden können. Entsprechend wird der alte Spaltenindex zum neuen Zeilenindex. Eine Spalte der Adjazenzmatrix gibt entsprechend an, durch welche Variablen diese Variable beeinflusst wird. Die Ableitungsmatrix der Modellgleichungen, als Transponierte der gewichteten Adjazenzmatrix, ist entsprechend mit Vertauschung von Zeilen und Spalten zu interpretieren. Somit umfassen  $M_y^k = A^{kT}$  sämtliche  $k$ -Schritt-Effekte. Um die Pfade über alle erreichbaren Knoten einzubeziehen, muss noch über alle Längen  $k$  addiert werden, sowie die Einheitsmatrix als nullte Potenz addiert werden, siehe Neumann-Reihe in Abschnitt 2.2. Damit können die Effekte als Summe der Pfadeffekte bzw. als Summenprodukt der partiellen Ableitungen dargestellt werden:

$$\begin{aligned} (I_n - M_y)^{-1} &= I_n + \sum_{k=1}^{\infty} M_y^k \\ &= I_n + \sum_{k=1}^{\infty} A^{kT} \\ &= I_n + \sum_{k=1}^{\infty} [a_{ji,k}]_{j,i=1,\dots,n} \end{aligned}$$

$$\begin{aligned}
&= I_n + \left[ \sum_{k=1}^{\infty} \sum_s p_{ij,ks} \right]_{j,i=1,\dots,n} \\
&= I_n + \left[ \sum_{k=1}^{\infty} \sum_s m_{y_j i_{k-1}} \cdots m_{y_{i_2} i_1} m_{y_{i_1} i} \right]_{j,i=1,\dots,n} .
\end{aligned}$$

Bei Bayesianischen Netzwerken liegt eine ähnliche Struktur vor: Die Summation resultiert aus der Integration der gemeinsamen Dichtefunktion über alle anderen Variablen und das Produkt resultiert aus der Faktorisierung der gemeinsamen Dichtefunktion in bedingte unabhängige Verteilungen. Eine Nachricht ist dann die Summe des Produkts der bedingten unabhängigen Verteilungen mit den entsprechenden eingehenden Nachrichten über alle Zustände der Variablen. Die Struktur ist jedoch rekursiv, da jede eingehende Nachricht selbst wieder eine Produktschritte ihrer eingehenden Nachrichten ist. Die Lösung wird dann mittels des Sum-Product-Algorithmus, auch Belief Propagation genannt, bestimmt. Man erhält so abschließend die Randverteilung einer Variablen als Produkt aller anderen Nachrichten. In unserem Fall wurde die Rekursivität des totalen Differentials bereits durch die geschlossene Formel in Matrixform gelöst, die dem iterativen Einsetzen der totalen Differentiale der vorangehenden Variablen entspricht. Dies ist am Ende des nächsten Abschnitts dargestellt.

Auf diese Weise können alle endogenen und exogenen Effekte bestimmt werden, ohne dass das gesamte Gleichungssystem schrittweise gelöst werden muss. Man benötigt nur die endogenen und exogenen partiellen Ableitungen der Modellgleichungen. Die Bildung der Summe über alle Pfadefekte ist nichts anderes, als das Berechnen der totalen Ableitungen über der direkten und alle indirekten Effekte, welches in der Praxis häufig durch graphische Diagramme unterstützt wird.

### 4.3 Partieller, totaler und finaler Graph

Die Effekte können geeignet in Graphen dargestellt werden, um die Interpretation des Systems zu ermöglichen. Die endogenen manifesten Variablen haben Fehlerterme, die untereinander korreliert sind; die zugehörigen bidirektionalen Kanten lassen wir im Graphen zur Übersichtlichkeit weg.

Wir führen drei verschiedene graphische Darstellungen ein, die unterschiedliche Blicke auf das System ermöglichen:

- Im *partiellen Graphen* sind die partiellen Ableitungen auf den Kanten eingezeichnet. Die Variablen sind, wie auch in den anderen Graphen, als Knoten dargestellt. Es werden die Koeffizienten von  $M_x$  und  $M_y$  dargestellt.
- Im *totalen Graphen* sind die totalen Ableitungen, also die Effekte, auf den Kanten eingezeichnet. Die endogenen Eigeneffekte sind stets eins und werden daher nicht als Schleifen eingezeichnet. Es werden die Koeffizienten von  $E_x$  und  $E$  dargestellt.
- Im *finalen Graphen* werden die finalen Effekte auf die jeweils betrachtete „finale“ endogene Variable auf den ausgehenden Kanten eingezeichnet. Die finalen Effekte ergeben in Summe den Effekt (siehe totaler Graph) des ausgehenden Knotens. Der Effekt einer Variablen auf

die finale Variable wird also auf seine ausgehenden Kanten zerlegt. Dies ermöglicht eine Analyse, über welche Pfade die wesentlichen Wirkungen auf die finale Variable erfolgen. Die finalen Effekte sind sämtlich untereinander vergleichbar, da sie alle in der Einheit der finalen Variable gemessen werden. Für jede endogene Variable  $y_j$  gibt es einen eigenen finalen Graphen. Es werden die Koeffizienten von  $E_{y_jx}$  und  $E_{y_jy}$  dargestellt.

In Kapitel 7 sind diese Graphentypen anhand von Beispielen dargestellt.

Bei der Interpretation des finalen Graphen ist zu beachten, dass die Summationseigenschaft nur für *ausgehende*, nicht aber für *eingehende* Katen gilt. Für eine endogene Variable  $y_l$  ist die Summe der ausgehenden finalen Effekte auf  $y_j$  gleich ihrem Knoteneffekt:

$$\begin{aligned} \frac{dy_j}{dy_l} &= \sum_{h=1}^n e_{jh_l,y}, \quad j \neq l \\ &= \sum_{h=1}^n \left( (I_n - M_{y_{lu}})^{-1} \right)_{jh} \frac{\partial M^h}{\partial y_l}. \end{aligned}$$

Nur in dem im Allgemeinen *nicht* erfüllten Spezialfall:

$$\sum_{i=1}^n \frac{\partial M^l}{\partial y_i} = \frac{1}{\left( (I_n - M_y)^{-1} \right)_{lu}}$$

ist auch die Summe der in  $y_l$  *eingehenden* finalen Effekte auf  $y_j$  gleich ihrem Knoteneffekt:

$$\begin{aligned} \sum_{i=1}^n e_{ji_l,y} &= \sum_{i=1}^n \left( (I_n - M_{y_{lu}})^{-1} \right)_{ji} \frac{\partial M^l}{\partial y_i}, \quad j \neq i \\ &= \left( (I_n - M_{y_{lu}})^{-1} \right)_{ji} \sum_{i=1}^n \frac{\partial M^l}{\partial y_i} \\ &= \frac{\left( (I_n - M_{y_{lu}})^{-1} \right)_{ji}}{\left( (I_n - M_y)^{-1} \right)_{lu}} \\ &= \left( (I_n - M_y)^{-1} \left( I_n \circ (I_n - M_y)^{-1} \right)^{-1} \right)_{ji} \\ &= \frac{dy_j}{dy_l}, \quad j \neq l, i. \end{aligned}$$

Im Allgemeinen gilt jedoch:

$$\frac{dy_j}{dy_i} = \sum_{h=1}^n e_{jh,y} \neq \sum_{i=1}^n e_{ji,y}$$

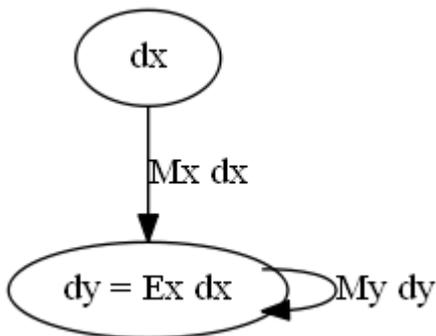
Es handelt sich also nicht um einen sogenannten Conserving Flow.

#### 4.4 Input-Output-Analyse

Das totale Differential

$$d\mathbf{y} = M_y d\mathbf{y} + M_x d\mathbf{x}$$

besteht aus dem Netto-Effekt  $d\mathbf{y} = E_x d\mathbf{x}$ , der sich zusammen setzt aus dem externen Effekt  $M_x d\mathbf{x}$  und dem internen Effekt  $M_y d\mathbf{y}$ , der wiederum den Netto-Effekt als Ausgangspunkt hat. Dies entspricht der Logik der Leontief Input-Output-Analyse, siehe [https://en.wikipedia.org/wiki/Input-output\\_model](https://en.wikipedia.org/wiki/Input-output_model). Als Graph der totalen Differentiale:



Das einzelne totale Differential kann geschrieben werden als:

$$\begin{aligned} dy_j &= \sum_{k=1}^n \frac{\partial M^j}{\partial y_k} dy_k + \sum_{l=1}^m \frac{\partial M^j}{\partial x_l} dx_l, \quad j = 1, \dots, n \\ &= \sum_{i=1}^m \frac{dy_j}{dx_i} dx_i \\ &= \sum_{i=1}^m \left( \sum_{h=1}^n \sum_{k=1}^n \sum_{v(h,j,k)} \prod_{l=1}^k \frac{\partial M^{v_l}}{\partial y_{v_{l-1}}} \frac{\partial M^h}{\partial x_i} + \frac{\partial M^j}{\partial x_i} \right) dx_i. \end{aligned}$$

Diese Darstellung erhält man auch, wenn man den folgenden Algorithmus anwendet: Das totale Differential wird  $n - 1$  mal ineinander eingesetzt, dabei sind die Änderungen  $dy_k$  selbst wieder

totale Differentiale. Durch Ausmultiplizieren und Zusammenfassen nach den  $dx_i$  erhält man die obige explizite Darstellung des totalen Differentials in Bezug auf die exogenen Variablen:

$$\begin{aligned}
 dy_j &= \sum_{k=1}^n \frac{\partial M^j}{\partial y_k} dy_k + \sum_{l=1}^m \frac{\partial M^j}{\partial x_l} dx_l \\
 &= \sum_{k=1}^n \frac{\partial M^j}{\partial y_k} \left( \sum_{k_1=1}^n \frac{\partial M^k}{\partial y_{k_1}} dy_{k_1} + \sum_{l_1=1}^m \frac{\partial M^k}{\partial x_{l_1}} dx_{l_1} \right) + \sum_{l=1}^m \frac{\partial M^j}{\partial x_l} dx_l \\
 &= \dots \\
 &= \sum_{i=1}^m \left( \sum_{k=1}^{n-1} \sum_{h=1}^{j-k} \sum_{v(h,j,k)} \prod_{l=1}^k \frac{\partial M^{v_l}}{\partial y_{v_{l-1}}} \frac{\partial M^h}{\partial x_i} + \frac{\partial M^j}{\partial x_i} \right) dx_i.
 \end{aligned}$$

Dieses iterative Vorgehen ist analog zu dem iterativen Einsetzen der bedingten Wahrscheinlichkeiten in die gemeinsame Wahrscheinlichkeitsdichte bei Bayesianischen Netzwerken.

#### 4.5 Relative Gain Array

Die folgenden Interaktionsmaße entstammen der Control System Theorie zu Multiple-Input-Multiple-Output (MIMO) Systemen.

Der Relative Gain Array, RGA, einer Matrix  $G$  ist definiert als:

$$RGA(G) = G \circ G^{-1T}.$$

Der RGA hat im Falle quadratischer Matrizen folgende Eigenschaften, siehe Skogestad and Postlethwaite (2007), Anhang A.4: Er ist skalierungs-unabhängig, die Zeilen- und Spaltensummen ergeben jeweils eins, er entspricht der Einheitsmatrix genau dann wenn  $G$  eine Dreiecksmatrix ist, subtrahiert man von einem Element von  $G$  den Kehrwert des Elements des RGA, so wird  $G$  nicht invertierbar und das RGA-Element groß, jede Permutation von  $G$  führt zur entsprechenden Permutation des RGA und es gilt:

$$RGA(G) = RGA(G^{-1T}) = RGA(G^{-1})^T.$$

Bei der Interpretation sind zwei unterschiedliche Wirkungen zu beachten. Zum einen gibt es die *totale* (open loop) Wirkung von  $y_i$  auf  $y_j$  und zum anderen gibt es die *partielle* (closed loop) Wirkung zurück von  $y_j$  auf  $y_i$ , so dass sich insgesamt ein Kreis ergibt. Dabei wird die Rückkopplung jedoch nur partiell betrachtet. Dies wird besonders deutlich bei der nachfolgenden RGA-Number. Dieses Festhalten der anderen Variablen zur Bestimmung der partiellen Ableitung wird als Tight Control bezeichnet. Die Rückkopplung wird in der Steuerungstheorie von einem Controller übernommen. Dieser soll möglichst einfach sein und nur eine Variable steuern. Um MIMO-

Systeme möglichst gut dezentral steuern zu können, werden Paare von Variablen, sogenannte Pairings, gesucht, die hohe Interaktionen innerhalb einer Gruppe und geringe Interaktionen zwischen den Gruppen haben. Die Spalten der Gain-Matrix  $G$  sind also so zu permutieren, dass der permutierte RGA möglichst der Einheitsmatrix entspricht, denn dann entsprechen sich die totale und die partielle Wirkung und es gibt keine weiteren Einflüsse. Diese perfekte Diagonaldominanz sichert jedoch nicht die Stabilität des Systems. Diese ist nur bei triangulären Systemen gewährleistet. Eine RGA-Number von null bedeutet aber nicht, dass das System triangulär ist, dies gilt nur für Systeme der Größe  $n \leq 2$ , siehe Skogestad and Postlethwaite (2007). Abweichungen von eins führen zu einer Über- oder Unterkompensation. Bei einem Wert von null ist entweder die totale Wirkung oder die partielle Wirkung null und es kann nicht gesteuert werden. Bei Werten kleiner null würde in die falsche Richtung gesteuert werden. Siehe auch Khaki-Sedigh and Moaveni (2009).

Die RGA-Number,  $\text{rga}(G)$ , ist definiert als die Summe der absoluten Werte aller Elemente des RGA:

$$\begin{aligned} \text{rga}(G) &= \|\text{RGA}(G) - I_n\|_1 \\ &\geq n. \end{aligned}$$

Die RGA-Number ist nur gleich  $n$ , wenn alle Elemente des RGA nichtnegativ sind.

Der RGA misst nur beidseitige Interaktionen, der Performance Relative Gain Array, PRGA, misst auch einseitige Interaktionen. Siehe Skogestad and Postlethwaite (2007):

$$\text{PRGA}(G) = (I_n \circ G)G^{-1}.$$

Die Normalized Interactions erhält man durch Subtraktion der Hauptdiagonalelemente und Division der Zeilen und Spalten durch diese Hauptdiagonalelemente:

$$\text{NORM}(G) = (G - (I_n \circ G))(I_n \circ G)^{-1}.$$

Der Niederlinski Index, NI, ist definiert als:

$$\text{NI}(G) = \frac{|G|}{|(I_n \circ G)|}.$$

Wir wenden diese Metriken nun auf die invertierbare Matrix  $(I_n - M_y)$  im DAG an, also wenn  $M_y$  eine strikte untere Dreiecksmatrix ist:

$$\begin{aligned} \text{RGA}(I_n - M_y) &= (I_n - M_y) \circ (I_n - M_y)^{-1T} \\ &= I_n, \end{aligned}$$

$$\begin{aligned} \text{rga}(I_n - M_y) &= \|\text{RGA}(I_n - M_y) - I_n\|_1 \\ &= 0, \end{aligned}$$

denn  $(I_n - M_y)$  ist eine untere unitrianguläre Matrix und  $(I_n - M_y)$  ist eine obere unitrianguläre Matrix. Trianguläre Systeme haben eine RGA-Number voll null. Umgekehrt bedeutet eine RGA-Number von null aber nicht, dass das System triangulär ist, siehe oben.

Im DAG gilt ferner:

$$\begin{aligned} \text{PRGA}(I_n - M_y) &= \left( I_n \circ (I_n - M_y) \right) (I_n - M_y)^{-1} \\ &= E_y \end{aligned}$$

und da die Hauptdiagonalmatrix von  $E_y$  die Einheitsmatrix ist und  $M_y$  nilpotent vom Grad  $n$  ist, gilt im DAG:

$$\begin{aligned} \text{NORM}(I_n - M_y) &= \left( I_n - M_y - \left( I_n \circ (I_n - M_y) \right) \right) \left( I_n \circ (I_n - M_y) \right)^{-1} \\ &= M_y \end{aligned}$$

und

$$\begin{aligned} \text{NI}(I_n - M_y) &= \frac{|I_n - M_y|}{\left| \left( I_n \circ (I_n - M_y) \right) \right|} \\ &= 1. \end{aligned}$$

Man sieht, dass diese Interaktionsmaße in azyklischen Graphen mit Adjazenzmatrix in strikter Dreiecksform kaum weitere Erkenntnisse bringen. Lediglich der Performance Relative Gain Array zeigt im DAG mit der Beziehung  $\text{PRGA}(I_n - M_y) = E_y$  die Bedeutung der endogenen Effektmatrix.

## 5) Schätzung

Das zu schätzende linearisierte Modell weicht vom Ausgangsmodell ab. Es ist somit nur ein Metamodell. Zum einen wird Linearität unterstellt, die partiellen Ableitungen seien also konstant und somit unabhängig von der Stelle der Beobachtung. Des Weiteren wird für die Schätzung Homogenität unterstellt, die partiellen Ableitungen müssen also für alle Beobachtungen gleich sein; die individuellen Effekte entsprechen also durchschnittlichen Effekten, siehe Pearl (2001), Pearl (2017). Weder Linearität, noch Homogenität sind aber bei nichtlinearen Modellen gegeben. Dies wird schon in dem einfachen Beispiel 1 in Kapitel 7 deutlich. Das linearisierte Modell ist also im Allgemeinen fehlspezifiziert. Die Schätzung kann in diesem Fall also nur eine grobe Indikation für die Validität des Ausgangsmodells sein. Wir verwenden die Linearisierung und Homogenität dennoch, da sie das Problem der Identifikation wesentlich vereinfacht.

Die linear approximierten absoluten Werte  $\mathbf{y} = \boldsymbol{\mu} + E_x d\mathbf{x}$  basieren auf dem theoretischen Modell und sind nicht an die beobachteten Werte angepasst. Wenn ein Fit an die beobachteten Werte gewünscht ist, so hat die Schätzung  $\hat{E}_x$  den Nachteil, dass diese für alle Beobachtungen gleich ist und die individuellen Sensitivitäten ignoriert werden. Dieser Nachteil könnte vermieden werden, wenn die individuellen Effekte um ihren Mittelwert bereinigt und dann wieder auf die konstante Schätzung addiert werden. Es bleibt jedoch das Problem, dass der Mittelwert nur für einige manifeste Variable beobachtbar ist. Daher sind stattdessen die Gleichungen des Ausgangsmodells anzupassen. Dies kann beispielsweise über gefittete Modifikationsindikatoren geschehen, wie am Ende von Abschnitt 5.2 dargestellt wird. Die Modellgleichungen sind mit den beobachteten Daten kompatibel, wenn geschätzten Effekte nicht signifikant von den mittleren theoretischen Effekten abweichen. Die Schätzung ist lediglich eine Nebenrechnung zur Überprüfung des verwendeten Modells. Diese Schätzung der Effekte wird nachfolgend dargestellt.

Die Modellgleichungen können fest gegeben sein, beispielsweise als Expertensystem von einem Experten postuliert. Es ist jedoch hilfreich, das Modell an realen Daten zu testen oder ausgehend von diesem Startzustand zu optimieren. Dies kann zum Beispiel mittels linearer Regression, neuronalen Netzen oder mit Kovarianzstrukturmodellen geschehen. Diese Verfahren werden in den nächsten Abschnitten behandelt.

Eine Schätzung könnte auf Basis des Originalmodells erfolgen. Jedoch wäre in diesem Fall das Problem der Identifikation des Modells schwer zu lösen. Hingegen gibt es für linearisierte Modelle geeignete Identifikationsverfahren. Zudem ist die Schätz- und Testtheorie für lineare Modelle besser entwickelt. Auch ist es einfacher, den Gradienten algebraisch zu bestimmen, was für einige iterative Schätzverfahren von Vorteil ist. Schließlich ist auch die Maximum-Likelihood-Theorie für lineare Modelle entwickelt.

Die durchschnittlichen Effekte  $E_x$  und  $E_y$ , auch Average Causal Effects, ACE genannt, sind im identifizierten reduzierten linearen Regressionsmodell einfach gleich dem Regressionskoeffizienten von  $y$  auf  $x$  und entsprechen der Summe aller Pfadeffekte gemäß Abschnitt 4.2 zwischen den betrachteten Variablen, siehe Pearl (2017) Theorem 1. Die partiellen Ableitungen  $M_x$  und  $M_y$ , auch Direct Effects genannt, sind die Koeffizienten des identifizierten strukturellen Modells, siehe Pearl (2017) Theorem 2 für azyklische Graphen. In linearen normalverteilten Modellen sind kontrafaktische Größen immer identifiziert, wenn das strukturelle Modell identifiziert ist.

Die Linearisierung erfolgt durch Ableitung, also durch Taylor-Approximation erster Ordnung. Das abgeleitete Modell ist für die Interpretation der Zusammenhänge zwischen den Variablen geeignet, denn die uns interessierenden Effekte basieren auf den Änderungen der Variablen. Wir schätzen aus diesen Gründen das durch Ableitung linearisierte Modell. Da die Änderungen um den Mittelwert betrachtet werden, gibt es im linearisierten Modell keine Konstanten, die die Mittelwerte modellieren.

Durch die Linearisierung wird auch direkt ein geeigneter Parameterraum vorgegeben, nämlich genau die partiellen Ableitungen zwischen den Variablen. In all den Fällen, in denen die partielle Ableitung gleich null ist, wird kein freier Parameter geschätzt. Diese Nullrestriktionen werden für die Identifikation des Modells verwendet. Die Ableitung gibt also eine geeignete Struktur vor, in deren Rahmen die Schätzung erfolgt. Zugleich wird jedoch auf etwas Struktur verzichtet, denn im Allgemeinen gibt es Beziehungen zwischen den partiellen Ableitungen, da die Modellvariablen auch noch in den Ableitungen vorkommen. Diese werden zur Vereinfachung ignoriert und alle Parameter werden frei geschätzt. Die Schätzung erfolgt mit Startwerten an der Stelle des Mittelwertes der Daten. Nach erfolgter Schätzung des linearisierten Modells ist das theoretische Ausgangsmodell geeignet anzupassen, so dass es besser mit den beobachteten Daten übereinstimmt. Dabei können dann sämtliche ursprünglichen Restriktionen beachtet werden.

Die Schätzung ermöglicht die empirische Überprüfung des Kausalitätsmodells: die sind geschätzten partiellen Ableitungen oder Effekte nicht signifikant von null verschieden, so ist dies ein Hinweis darauf, dass diese kausalen Beziehungen eliminiert werden können. In den Fußnoten zum Blog "ML beyond Curve Fitting: An Intro to Causal Inference and do-Calculus", siehe <https://www.inference.vc/untitled>, ist Literatur genannt, die sich mit der Überprüfung oder Ableitung von Kausalitäten aus Daten mittels Methoden des Machine Learnings befasst. Dabei geht es oft um das Finden von bedingten Unabhängigkeiten. Mit unserem Ansatz können wir zwar keine Kausalitäten finden, die außerhalb unseres Modells sind, aber wir können empirisch überprüfen, ob die angenommenen Kausalitäten signifikant von null verschieden sind.

## 5.1 Strukturierte simultane Regression (ökonometrische SEM)

Die lineare Regression misst die linearen Beziehungen zwischen beobachtbaren endogenen und exogenen Variablen. Das ökonomische Modell  $y_j = M^j(\mathbf{y}, \mathbf{x})$ ,  $j = 1 \dots, n$  ist jedoch im Allgemeinen nichtlinear. Wir betrachten daher die linearisierte Approximation in Form des totalen Differentials. Wir wollen die spezielle Struktur dieses ökonomischen Modells anhand von Daten überprüfen. Das linearisierte ökonomische Modell  $d\mathbf{y} = M_y d\mathbf{y} + M_x d\mathbf{x}$  wird als *strukturiertes* Modell bezeichnet, da Elemente der Koeffizientenmatrizen  $M_y$  und  $M_x$  aus der ökonomischen Theorie her restringiert sind, hier beispielsweise null betragen. Die Schätzung basiert hingegen auf dem *reduzierten* Modell, das sämtliche Effekte im Sinne der totalen Ableitungen von exogenen auf endogene Variable misst.

Das totale Differential als Regressionsgleichung lautet unter Ergänzung des Fehlerterms  $\mathbf{u}$  für die nicht erklärten Zusammenhänge in strukturierter Form:

$$d\mathbf{y} = M_y d\mathbf{y} + M_x d\mathbf{x} + \mathbf{u},$$

$$(I_n - M_y)dy - M_x dx = \mathbf{u},$$

bzw. in der Notation von Magnus and Neudecker (2007):

$$dy^T \Gamma + dx^T B = \mathbf{u},$$

$$\Gamma = I_n - M_y^T,$$

$$B = -M_x^T.$$

und wir treffen folgende übliche Annahmen:

$$E[\mathbf{u}|dx] = \mathbf{0},$$

$$E[\mathbf{u}\mathbf{u}^T|dx] = \Sigma_{\mathbf{u}},$$

wegen der Existenz endogener Variablen gilt jedoch  $E[\mathbf{u}|dy] \neq \mathbf{0}$ , und somit ist die OLS-Schätzung inkonsistent.

Die reduzierte Form ergibt sich durch Rechtsmultiplikation mit  $\Gamma^{-1}$  und  $\mathbf{v}$  ist der Fehlerterm der reduzierten Form:

$$dy = E_x dx + \mathbf{v},$$

bzw.

$$dy^T = dx^T \Pi + \mathbf{v},$$

$$\Pi = -B\Gamma^{-1}$$

$$= \left( (I_n - M_y)^{-1} M_x \right)^T$$

$$= E_x^T.$$

Sofern das System identifiziert ist, kann man von den Parametern  $\Pi^T = E_x = (I_n - M_y)^{-1} M_x$  der reduzierten Form auf die Parameter  $B, \Gamma$  bzw.  $M_y, M_x$  der strukturierten Form schließen. Zur Identifikation siehe Magnus and Neudecker (2007), Hendry, unser Abschnitt 1.5, sowie Shpitser und Pearl (2006), die die Identifikation mittels grafischer Kriterien überprüfen. So gelangt man von den totalen Ableitungen zu den partiellen Ableitungen. Bei Identifikation, sowohl für exakte als auch Überidentifikation, ist der nachfolgend erläuterte zweistufige Schätzer geeignet. Bei einer Unteridentifikation können die Parameter der betroffenen Gleichung nicht geschätzt werden. Wenn alle Gleichungen exakt identifiziert sind, ist die erste Stufe der zweistufigen Schätzung ausreichend. Man kann dann die totale Ableitung nach den partiellen Ableitungen eindeutig auflösen. Dies wird auch Indirect Least Squares, ILS, bezeichnet.

Die geschätzte  $(n \times m)$  Matrix  $E_x$  ist voll besetzt und hat  $nm$  verschiedene Elemente. Diese sind durch die Schätzung gegeben. Die gesuchte  $(n \times m)$  Matrix  $M_x$  hat  $nm$  Parameter. Zusätzlich ist die  $(n \times n)$  Matrix  $M_y$  mit Hauptdiagonale null, siehe Normierungsannahme in Abschnitt 1.3, und  $n(n - 1)$  freien Parametern gesucht. Man benötigt für die Identifikation der Parameter der strukturierten Form also noch  $n(n - 1)$  Restriktionen. Für die nachfolgend dargestellte Schätzung der strukturellen Parameter ist eine Spezifikation der  $n(n - 1)/2$  Parameter der Varianz- Kovarianzmatrix nicht erforderlich, anders als bei den später dargestellten Kovarianzstrukturmodellen. Die Restriktionen können beispielsweise auferlegt werden, indem man partielle Ableitungen als null festlegt, die gemäß der ökonomischen Struktur gleich null sein sollten, für die es also keine funktionalen Beziehungen zwischen den Modellgleichungen gibt.

Für die Identifikation der strukturellen Parameter gibt es die notwendige *Ordnungsbedingung*: Die Anzahl der ausgeschlossenen Variablen in einer Gleichung muss mindestens so groß sein, wie die um eins reduzierte Anzahl der eingeschlossenen endogenen Variablen dieser Gleichung.

Die notwendige und hinreichende *Rangbedingung* ist etwas aufwändiger zu prüfen. Die Rangbedingung untersucht den Rang der zusammengesetzten Matrix  $M$ . Sie gilt unter der hier gültigen Annahme, dass alle Fehler des Systems ohne Restriktionen der Kovarianzmatrix korreliert sind, siehe Bollen and Bauldry (2010). Dies ist äquivalent zur Betrachtung des Rangs der Matrix

$$M = [I_n + M_y, M_x].$$

Bezeichne  $M_{[j,-j]}$  die Matrix  $M$ , bei der die  $j$ -te Zeile herausgeschnitten wird und nur die Spalten behalten werden, bei denen in der  $j$ -ten Zeile eine null stand. Diese Matrix umfasst nur die Koeffizienten derjenigen Variablen, die zwar im System auftauchen, nicht aber in der betrachteten Gleichung. Die Identifikation wird also durch nicht auftauchende Variablen sichergestellt. Das Rangkriterium ist für die  $j$ -te Gleichung erfüllt, wenn voller Zeilenrang vorliegt:

$$\text{rk}(M_{[j,-j]}) = n - 1.$$

Die Rangbedingung sichert folgende Eigenschaft, siehe Susmel:

For each equation: Each of the variables excluded from the equation must appear in at least one of the other equations (no zero columns). Also, at least one of the variables excluded from the equation must appear in each of the other equations (no zero rows).

Die notwendige, aber nicht hinreichende *Ordnungsbedingung* lautet:

The number of excluded exogenous variables must be at least as great as the number of right-hand-side included endogenous variables.

Bei gegebener Identifikation kann nun die Schätzung erfolgen. Dabei sind die Differentiale als mittelwertbereinigte Daten zu behandeln, siehe Abschnitt 1.2:

$$dy = y - \bar{y},$$

$$dx = x - \bar{x}.$$

Eine Konstante ist im Regressionsmodell daher nicht erforderlich. Die Regression misst die Effekte zwischen den Variablen an der Stelle der Mittelwerte. Wir nehmen an, dass die Modellgleichungen für jedes Objekt und für jeden Zeitpunkt gelten. Daher kann man sämtliche Beobachtungen über den Querschnitt und die Zeitdimension zu insgesamt  $\tau$  Beobachtungen zusammenfassen.

Bei gegebenen Daten  $d\mathbf{y}_t$  und  $d\mathbf{x}_t$  für die Beobachtungen  $t = 1, \dots, \tau$ :

$$dY_{(\tau \times n)} = \begin{bmatrix} d\mathbf{y}_1^T \\ \vdots \\ d\mathbf{y}_\tau^T \end{bmatrix},$$

$$dX_{(\tau \times m)} = \begin{bmatrix} d\mathbf{x}_1^T \\ \vdots \\ d\mathbf{x}_\tau^T \end{bmatrix}$$

lautet die strukturierte Form mit der Fehlermatrix  $U$ :

$$dY\Gamma + dXB = U$$

und die reduzierte Form mit Fehlerterm  $V$  lautet:

$$dY = dX\Pi + V.$$

Die Schätzung erfolgt nun in zwei Schritten, mit dem Two-Stage Least Squares Schätzer, 2SLS. Beide Schritte umfassen eine gewöhnliche Kleinste-Quadrate-Schätzung, Ordinary Least Squares, OLS:

- 1) Schätzung der reduzierten Form  $dY = dX\Pi + V$  mittels OLS, mit Regressionskoeffizienten  $\widehat{\Pi} = (dX^T dX)^{-1} dX^T dY$  und Prognosen  $d\widehat{Y} = dX\widehat{\Pi}$ .
- 2) Schätzung der strukturierten Form  $d\mathbf{y}_j = [d\widehat{Y}_{[j]}, dX_j] \begin{bmatrix} -\bar{\mathbf{m}}_{yj} \\ -\bar{\mathbf{m}}_{xj} \end{bmatrix} + \mathbf{v}_j = dZ_j \bar{\mathbf{m}}_j$ ,  $j = 1, \dots, n$  mittels OLS, mit Regressionskoeffizienten  $\widehat{\bar{\mathbf{m}}}_j = (dZ_j^T dZ_j)^{-1} dZ_j^T d\mathbf{y}_j$ .

Im zweiten Schritt werden die strukturierten Gleichungen einzeln geschätzt. Dies geschieht unter Verzicht auf simultane Information. Eine Alternative wäre ein Full Information Maximum Likelihood-Schätzer, FIML, siehe Magnus and Neudecker (2007).  $dY_{[j]}$  ist die Matrix aller endogenen Variablen, mit Ausnahme von  $d\mathbf{y}_j$  und  $dX_j$  ist die Matrix der in der  $j$ -ten Gleichung vorkommenden exogenen Variablen.  $\bar{\mathbf{m}}_{yj}$  und  $\bar{\mathbf{m}}_{xj}$  sind die Vektoren der freien Parameter in der  $j$ -ten Gleichung, also der nicht auf null oder eins restringierten Koeffizienten. Es ist zu beachten, dass in der zweiten Schätzung die Prognosen  $d\widehat{Y}$  der endogenen Variablen  $dY$  verwendet werden. Dadurch wird das sogenannte Endogenitätsproblem gelöst, welches sonst zu einer verzerrten und inkonsistenten OLS-Schätzung geführt hätte. Die Prognosen sind eine orthogonale Projektion auf den Fehlerterm und somit – im Gegensatz zu den Originaldaten – nicht mehr mit den Fehlern korreliert.

$\hat{\Pi}$  ist die Matrix der Regressionskoeffizienten in einer linearen Regression und somit ein Schätzer für  $\Pi = E_x^T$ , also für die totalen Ableitungen. Die erste Regression schätzt also sämtliche exogenen totalen Ableitungen. In der Literatur der *nicht* strukturierten Modelle wird werden die Regressionskoeffizienten hingegen üblicherweise als *partielle* Ableitungen dargestellt, siehe zum Beispiel [https://en.wikipedia.org/wiki/Linear\\_regression](https://en.wikipedia.org/wiki/Linear_regression). Dies liegt daran, dass man mathematisch die partielle Ableitung nach einer einzelnen exogenen Variablen bildet und die anderen exogenen Variablen fest hält. Da diese Ableitung jedoch in der *reduzierten* Form stattfindet, ist sie als totale Ableitung zu interpretieren: Der Regressionskoeffizient gibt an, wie stark sich die endogene Variable ändert, wenn sich die exogene Variable ändert – unter Berücksichtigung des direkten und aller indirekten Effekte im System. Auch die Aussage, dass eine univariate Regression, also nur mit einer exogenen Variablen, die totale Ableitung darstellt, ist irreführend: Zwar findet die Schätzung in der reduzierten Form statt, jedoch handelt es sich offensichtlich um ein misspezifiziertes Modell, da die anderen erklärenden Variablen ignoriert wurden.

Die Effekte  $E_x = (I_n - M_y)^{-1} M_x$  der exogenen Variablen entsprechen, wie dargestellt, den Regressionskoeffizienten  $\hat{\Pi} = (dX^T dX)^{-1} dX^T dY$ . Intuitiv übernimmt  $(I_n - M_y)^{-1}$  die Rolle von  $(dX^T dX)^{-1}$  und  $M_x$  übernimmt die Rolle von  $dX^T dY$ . Wenn das Regressionsmodell eine Konstante enthält, ist  $(dX^T dX)^{-1}$  die Inverse der Kovarianzmatrix der exogenen Variablen und wird auch Precision Matrix genannt. Ihre Elemente sind die partiellen Korrelationen, siehe [https://en.wikipedia.org/wiki/Partial\\_correlation#Using\\_matrix\\_inversion](https://en.wikipedia.org/wiki/Partial_correlation#Using_matrix_inversion) und für einen Beweis siehe Huber (2015) <https://stats.stackexchange.com/a/156812/173285>. Bei den partiellen Korrelationen werden, im Gegensatz zum allgemeinen Korrelationskoeffizienten, die Korrelationen zwischen zwei Variablen berechnet, wobei der Effekt aller anderen Variablen gelöscht wird. Diese weiteren Regressionsvariablen des gesamten Systems eliminieren also gerade diese Effekte. Für eine Betrachtung der partiellen Korrelation im Kontext der Graphentheorie siehe Pearl (2017) Abschnitt 2.2. Insofern ist es plausibel, dass das gesamte System in  $(I_n - M_y)^{-1}$  die paarweisen totalen Ableitungen enthält, ohne dass man explizit Variable löschen muss. Diese Beobachtung korrespondiert mit unserer Erkenntnis, dass man über die Subsysteme zu den totalen Ableitungen gelangt, siehe Abschnitt 3.2.

Eine Einschränkung der strukturellen ökonometrischen Regressionsmodelle liegt jedoch darin, dass sämtliche endogenen Variablen als beobachtbare Daten vorliegen müssen, um die Schätzung durchführen zu können. Sind nur einige Variablen manifest, so stellt sich die Schätzung des Teilmodells wie folgt dar. Die manifesten endogenen Variablen  $dy_m$  werden mit der Filtermatrix aus den endogenen Variablen selektiert:

$$F_{y_m} = [0_{p,l}, I_p],$$

(p×n)

wobei ohne Beschränkung der Allgemeinheit angenommen wird, dass die p letzten endogenen Variablen manifest sind, so dass gilt:

$$dy_m = F_{y_m} dy,$$

(p×1)

$$dY_m = YF_{y_m}^T.$$

( $\tau \times p$ )

Für die beobachteten Variablen können die entsprechenden Zeilen der exogenen Effekte über die reduzierte Form des linearisierten Modells geschätzt werden:

$$E_x^T F_{y_m}^T = (dX^T dX)^{-1} dX^T dY_m.$$

( $m \times p$ )

Anmerkung: da alle Regressoren gleich sind, würde eine GLS-Schätzung wieder zur OLS-Schätzung führen, siehe Judge et al. (1988), Formel (14.4.5).  $M_x$  und  $M_y$  sind also so zu wählen, dass:

$$\left( (I_n - M_y)^{-1} M_x \right)^T F_{y_m}^T = (dX^T dX)^{-1} dX^T dY_m.$$

Die Schätzung des Teilmodells der beobachtbaren Variablen ist jedoch nicht von Interesse, da die Kausalität der latenten Variablen unberücksichtigt bleibt. Eine Alternative sind Modelle mit latenten Variablen. Wir verfolgen daher den Ansatz der im nächsten Abschnitt dargestellten neuronalen Netze. Im übernächsten Abschnitt folgen die Kovarianzstrukturmodelle. Beide Modelle sind speziell für latente Variablen geeignet.

## 5.2 Strukturelle Neuronale Netze

Die Verwendung neuronaler Netze ist schon deshalb naheliegend, da sich dessen Struktur direkt aus dem Graphen des strukturellen linearisierten Modells ergibt. Wie im vorangehenden Abschnitt erläutert, ist es notwendig, dass die strukturellen Modellparameter identifiziert sind, wenn man sie interpretieren möchte und das Modell nicht einfach nur zur Prognose verwenden möchte. Aufgrund dieser Identifikationsbedingungen sprechen wir von einem strukturellen neuronalen Netzwerk. Die Identifikationsrestriktionen werden übernommen, indem die entsprechenden Kanten aus dem neuronalen Netz entfernt werden. Anders als bei üblichen neuronalen Netzen sind daher auch die hidden Units als kausale Effekte identifiziert und interpretierbar.

Der zentrale Vorteil neuronaler Netze im Vergleich zu ökonomischen Strukturmodellen liegt darin, dass sie die Schätzung von Modellen mit latenten Variablen ermöglichen. So sind die Input- und die Output-Variablen beobachtbar, während die dazwischen liegenden hidden Units zwar berechenbar, aber nicht beobachtbar sind. Siehe hierzu auch unser Hinweis in Abschnitt 1.7 zur Berechenbarkeit latenter Variablen bei gegebenen Modellgleichungen. Die Input-Variablen  $dx$  bilden die sogenannte Input-Schicht. Eine Schicht ist dadurch gekennzeichnet, dass sich die Variablen untereinander nicht beeinflussen und sie nur Variablen der vorangehenden Schicht aufnehmen können. Anders als in neuronalen Netzen üblich, verwenden wir keine eigene Schicht für die Output-Variablen. Wir wollen eine Darstellung, in der diejenigen endogenen Variablen, die beobachtbar sind, in die Zielfunktion aufgenommen werden können, auch wenn sie noch andere endogene Variablen beeinflussen. Dabei soll die strukturelle Form erhalten bleiben. Die im Rahmen des supervised Learnings zur Schätzung genutzte Zielfunktion verwendet nur die beobachtbaren Daten.

Neuronale Netze verwenden als Input für ein Neuron, also einen Knoten, die gewichtete Summe der eingehenden Variablen und wenden darauf üblicherweise eine s-förmige, also sigmoidale, Funktion an und geben den resultierenden Wert an nachfolgende Knoten weiter. Unsere Modellgleichungen sind jedoch von allgemeinerer Art, sie verknüpfen die Inputs nichtlinear anstatt per gewichtete Summe. Durch die Linearisierung des strukturellen Modells, jeweils an der aktuellen Stelle, erreicht man die handhabbare lineare Form, wobei dann die sigmoidale Funktion gar nicht mehr erforderlich ist. In unserem Fall ist diese Aktivierungsfunktion dann einfach die Identität, was den Gradienten vereinfacht. Es liegt also ein lineares strukturelles neuronales Netz vor. Das Gesamtmodell wirkt trotz dieser linearen Darstellung nichtlinear, da die Ableitungen in jedem Iterationsschritt neu bestimmt werden. Die Ableitung hängt also nicht nur von den beobachteten Daten  $dx$ ,  $dy_m$ , sondern auch von den aktuellen Parametern  $M_x$  und  $M_y$  selbst ab.

Bei Verwendung eines neuronalen Netzes werden die Gewichte in jedem Batch-Lauf, siehe unten, an der Stelle des Mittelwertes der Daten bestimmt. In unserem Fall sind die Gewichte die partiellen Ableitungen und die Daten müssen mittelwertbereinigt werden, damit sie die Änderungen von diesem Mittelwert repräsentieren und die Gleichung des totalen Differentials geschätzt werden kann. Wir verwenden also die mittelwertbereinigten exogenen bzw. endogenen Variablen  $dx$  und  $dy_m$ . Das nichtlineare Ausgangsmodell operiert auf nicht mittelwertbereinigten Daten  $x$  und  $y_m$ .

Wir stellen also die linearisierte strukturelle Form  $dy = M_y dy + M_x dx$  als neuronales Netz dar. Damit sind  $M_y$  und  $M_x$  zugleich die Matrizen der Gewichte des neuronalen Netzes, separiert in endogene Variablen und Input-Schicht. Das strukturelle neuronale Netz ist ein lineares Netzwerk, da die Aktivierungsfunktionen stets die Identitäten sind. Es hat maximal  $n + m$  Schichten und jede Schicht hat Input- und Output-Dimension von höchstens  $\max(n; m)$ . Die Gewichte sind zudem durch die Identifikationsmatrizen  $ID_x$  und  $ID_y$  restringiert.

Als weitere Abweichung von neuronalen Netzen schätzen wir die Modellgleichungen explizit und weisen – wie in Regressionsmodellen üblich – den endogenen manifesten Variablen daher Fehlerterme zu, die untereinander korreliert sein können.

Unsere im Allgemeinen zyklischen Modelle entsprechen recurrent Networks. (ToDo mittels PyTorch:) Um diese mittels Backpropagation through time schätzen zu können, werden diese üblicherweise „ausgerollt“, so dass sie einem Feedforward Netzwerk entsprechen. Bei der Berechnung wird dann schrittweise vorgegangen, das heißt, die Knoten werden schrittweise nacheinander berechnet. Wir gehen anders vor. Anstatt das Modell auszurollen und Backpropagation zu verwenden, bestimmen wir die Zielfunktion und eine geschlossene Formel für den gesamten Gradienten in Form der totalen Ableitung an der aktuellen Stelle. Mittels einer nichtlinearen Optimierungsverfahrens erfolgt dann die Iteration der Modellparameter.

Die Schätzung anhand realer Daten zeigt schließlich, welche der vorhandenen (nicht auf null gesetzten) Gewichte, also hier die partiellen Ableitungen, angepasst werden sollten. So nutzen wir spezielle neuronale Netze, also Methoden der künstlichen Intelligenz, um das als Expertensystem vorgegebene Modell anzupassen. Während im Bereich der künstlichen Intelligenz und des maschinellen Lernens üblicherweise große Datenmengen vorliegen und große neuronale Netze trainiert werden, beschränkt sich unser lineares strukturelles neuronales Netz auf genau die direkten Kausalitäten und bereits mit wenigen Beobachtungen geschätzt werden.

Das neuronale Netz der reduzierten Form wird nicht geschätzt, da man daraus ohne Identifikationsrestriktionen nicht auf die strukturellen Parameter schließen kann. Es könnte allerdings für die reine Prognose der endogenen Variablen verwendet werden. Die linearisierte reduzierte Form lautet  $dy = E_x dx$ . Dies entspricht einem linearen neuronalen Netz mit einer Schicht und diese hat die Dimension  $(n, m)$  und ist durch  $ED_x$  restringiert.

Für die Schätzung ist die Identifikation der Gewichte erforderlich. Im Fall vorhandener latenter Variablen gibt es keine für alle Fälle gültige notwendige und hinreichende algebraische Methode zur Identifikation, siehe Bollen (1989) Seite 331. Am Ende des Abschnitts werden ergänzend Kriterien dargestellt, die zumindest in speziellen Fällen Aussagen zur Identifikation ermöglichen.

Wir verwenden daher den allgemeinen den *empirischen Ansatz für die lokale Identifikation*: Die Hessesche Matrix muss vollen Rang haben. Siehe Rothenberg (1971), Bollen (1989) Seite 248 oder Bollen and Bauldry (2010), die diesen Ansatz im Rahmen der im nächsten Abschnitt dargestellten psychometrischen Strukturmodelle behandeln. Wir verwenden die algebraisch abgeleitete Hessesche an der Stelle der Gewichte  $M_x, M_y$  und deren Kovarianzmatrix  $\Sigma_\theta$  bei gegebenen Beobachtungen  $dx, dy_m$ . In einem ersten Schritt wird die Identifikation an der Stelle der theoretischen Gewichte bei identitärer Kovarianzmatrix geprüft, in einem zweiten Schritt dann an der Stelle der geschätzten Gewichte und deren geschätzter Kovarianzmatrix.

Die Verwendung der empirischen Identifikation ist in der Praxis Standard und funktioniert gut, siehe Bollen (1989) Seite 249. Allerdings kann es aufgrund numerischer Effekte sein, dass der empirische Rang voll ist, obwohl keine Identifikation vorliegt und umgekehrt kann Identifikation vorliegen, aber der empirische Rang nicht voll sein. Probleme werden im ersten Fall dadurch sichtbar, dass die geschätzten Konfidenzintervalle der Parameter sehr groß werden, siehe Bollen (1989) Seite 250. Dies liegt, wie nachfolgend dargestellt, daran, dass Kovarianzmatrix der Gewichte direkt aus der inversen Hesseschen folgt, noch multipliziert mit der zweifachen Fehlervarianz. Daher empfiehlt es sich, mittels verschiedener Startwerte zu überprüfen, ob eine Konvergenz zum gleichen Schätzer erfolgt.

In einem linearen neuronalen Netzwerk, wie es hier vorliegt, sind alle Minima globale Minima, siehe Kawaguchi (2016). Das heißt, die Werte der Zielfunktion sind an diesen Stellen gleich. Daher ist das Problem nicht wohl formuliert, denn die Lösung ist nicht eindeutig. Die nachfolgend verwendete Tikhonov-Regularisierung sorgt für ein wohl formuliertes Problem, bei dem es nur noch ein globales Optimum gibt. Es geht jedoch die Konsistenz des NLS verloren, da ein Bias eingeführt wird. Im Gegenzug wird der Mean Squared Error minimiert,  $SSE/\tau = MSE = BIAS^2 + VAR$ . Dies ist das typische Vorgehen bei solchen Shrinkage-Schätzern. Durch die Regularisierung wird das aufgrund seiner vielen gleich guten globalen Minima nicht eindeutige Problem zu einem eindeutigen Problem. Zudem wird das Problem der numerischen Stabilität aufgrund der in der Regel stark korrelierten endogenen beobachtbaren Variablen vermieden. Ein zu kleines  $\alpha$  zu nahe bei null führt dazu, dass der Rang der Hesseschen Matrix der geschätzten Gewichte nicht voll ist und somit das numerische Kriterium der Identifikation nicht erfüllt ist. Ein ausreichend hohes  $\alpha$  führt zu einem vollen Rang der Hesseschen Matrix, wodurch das empirische Kriterium der lokalen Identifikation erfüllt ist.

Durch das Shrinkage wird verdeckt, dass es Gewichte geben kann, die zwar aufgrund des Shrinkage formal identifiziert sind, jedoch nicht mehr durch Daten bestimmt werden, sondern direkt auf das

Shrinkage-Ziel, also den theoretischen Wert gesetzt werden. Es handelt sich um reine Shrinkage-Identifikation. Dies erkennt man an dem t-Wert, mit dem getestet wird, ob der geschätzte Koeffizient dem theoretischen Wert entspricht: er ist stets null, da die Differenz im Zähler null ist, bei einer von null verschiedenen Standardabweichung. Auf Grund der fehlenden Daten-Information sind die Standardabweichungen jedoch üblicherweise recht groß. Daher ist es – für theoretische Werte ungleich Null – sinnvoll, die Power der Schätzung zu überprüfen: diese ist gegeben, wenn der theoretische Wert innerhalb und gleichzeitig die Null außerhalb des Konfidenzintervalls liegt.

Die Identifikation erreichen wir durch Nullsetzen bestimmter Koeffizienten. Dies bedeutet für die iterative Schätzung, dass Gewichte, die in der vorangegangenen Iteration auf null restringiert waren, nicht durch den Gradientenabstieg verändert werden dürfen. Wir halten diese Koeffizienten daher auch im Laufe des Verfahrens bei null. Weitere Koeffizienten dürften nur aufgenommen werden, wenn sie die Identifikationsanforderungen erfüllen. Die Nullrestriktionen erhalten wir über zwei Wege: Zum einen folgen sie direkt aus den Nullelementen der partiellen Ableitungen  $M_y$  und  $M_x$ , also aus der ökonomischen Theorie. Zum anderen werden ganze Gleichungen substituiert und somit formal neue Modelle mit weniger Variablen geschaffen, die jedoch inhaltlich unverändert sind. Um den Suchbaum zu vereinfachen, werden über mehrere Top-Down-Suchen zufällig Variablen substituiert, bis die lokale Identifikation sichergestellt ist.

Die verschiedenen lokal identifizierten Modelle werden geschätzt. Die Ergebnisse der Schätzung können zusammengeführt werden, jedoch nicht bezüglich der geschätzten Koeffizienten in  $M_y$  und  $M_x$ , sondern nur bezüglich der Effekte  $E_y$  und  $E_x$ . Dies liegt gerade daran, dass die Effekte die totale Wirkung im System zeigen, unabhängig davon, ob Variablen substituiert wurden. Die Mischung der Effekte erfolgt proportional zu ihren geschätzten Varianzen. Diese wiederum basieren auf den Varianzen der Koeffizienten, die aus der Hauptdiagonale der Kovarianzmatrix zu entnehmen sind. Diese werden dann mittels Delta-Methode in die Varianzen der Effekte überführt, siehe [https://en.wikipedia.org/wiki/Delta\\_method](https://en.wikipedia.org/wiki/Delta_method):

Sei  $\theta$  der  $(q \times 1)$  Vektor der unbekanntenen freien (nicht-null) Parameter bzw. Gewichte in  $M_y$  und  $M_x$ . Sei entsprechend  $\phi$  der  $(r \times 1)$ -Vektor der nicht-null-Effekte in  $E_y$  und  $E_x$ . Ferner sei  $\Sigma_{SSE,\theta}$  die  $(q \times q)$  Kovarianzmatrix der Gewichte und  $\Sigma_{SSE,\phi}$  die  $(r \times r)$  Kovarianzmatrix der Effekte. Dann gilt:

$$\Sigma_{SSE,\phi} = \frac{\partial \phi}{\partial \theta^T} \Sigma_{SSE,\theta} \left( \frac{\partial \phi}{\partial \theta^T} \right)^T,$$

wobei die  $(r \times q)$  Jacobi-Matrix  $\partial \phi / \partial \theta^T$  entweder numerisch oder algebraisch bestimmt werden kann. Bollen (1989), gibt in den Formeln (8A.5) und (8A.6) die algebraische Jacobi-Matrix für die latenten indirekten Effekte an. In unserem allgemeinen Fall für die Effekte von latenten und manifesten Variablen lautet die  $(n^2 + nm \times q)$  Jacobi-Matrix für sämtliche Elemente von  $E_y$  und  $E_x$ :

$$\frac{\partial \phi}{\partial \theta^T} = \begin{bmatrix} \partial \phi_y \\ \partial \theta^T \\ \partial \phi_x \\ \partial \theta^T \end{bmatrix}$$

$(n^2 + nm \times q)$

mit

$$\frac{\partial \Phi_y}{\partial \theta^T} = \left( \left( (I_n - M_y)^{-1} \right)^T \otimes (I_n - M_y)^{-1} - I_{n^2} \right) V_{M_y} + V_{M_y}$$

(n<sup>2</sup>×q)

$$\frac{\partial \Phi_x}{\partial \theta^T} = \left( \left( (I_n - M_y)^{-1} M_x \right)^T \otimes (I_n - M_y)^{-1} \right) V_{M_y} + \left( I_m \otimes \left( (I_n - M_y)^{-1} - I_n \right) \right) V_{M_x} + V_{M_x}$$

(nm×q)

mit

$$V_{M_y} = \text{vec} \left[ \frac{\partial M_y}{\partial \theta_1}, \dots, \frac{\partial M_y}{\partial \theta_q} \right]$$

(n<sup>2</sup>×q)

$$V_{M_x} = \text{vec} \left[ \frac{\partial M_x}{\partial \theta_1}, \dots, \frac{\partial M_x}{\partial \theta_q} \right].$$

(nm×q)

Diese Matrizen werden dann noch auf die Zeilendimensionen  $r$ ,  $r_y$  und  $r_x$  reduziert, indem nur die Zeilen behalten werden, die zu nicht-Null-Elementen von  $\text{vec}(M_y)$  und  $\text{vec}(M_x)$  gehören.

Wir teilen die  $n$  endogenen Variablen  $\mathbf{dy}$  auf in  $p$  beobachtbare Output-Variablen  $\mathbf{dy}_m$  und die  $l$  latenten Variablen  $\mathbf{dy}_l$ , siehe auch nächster Abschnitt:

$$\mathbf{dy}_m = F_{y_m} \mathbf{dy},$$

$$E(\mathbf{dy}_m \mathbf{dy}_m^T) = F_{y_m} E(\mathbf{dy} \mathbf{dy}^T) F_{y_m}^T = F_{y_m} \Sigma_{dy} F_{y_m}^T = \Sigma_{\mathbf{dy}_m}.$$

Da  $\Sigma_{dy}$  nicht beobachtbar ist, wird die beobachtbare empirische Kovarianzmatrix  $\hat{\Sigma}_{\mathbf{dy}_m}$  der manifesten Variablen verwendet. Zudem verwenden wir die  $(p \times p)$  Skalierungsmatrix  $L^T$ , mit der in der Zielfunktion nur über die gewichteten beobachtbaren Variablen summiert wird. Die Gewichtung erfolgt über ein Whitening mittels Cholesky-Zerlegung. Die Cholesky-Zerlegung der Precision-Matrix, also der inversen Kovarianzmatrix der endogenen beobachtbaren Variablen, in Faktoren mit unterer  $(p \times p)$  Dreiecksmatrix  $L$  lautet:

$$\Sigma_{\mathbf{dy}_m}^{-1} = LL^T.$$

(p×p)

bzw.  $L^T \Sigma_{\mathbf{dy}_m} L = I_p$ . Für das Whitening der nichtlinearen Kleinste-Quadrate Schätzung wird also die Invertierbarkeit von  $\Sigma_{\mathbf{dy}_m}$  benötigt. Diese ist bei Verwendung der empirischen Kovarianzmatrix praktisch gewährleistet. Damit erhält man die gewhiteten, mittelwertbereinigten manifesten Variablen:

$$\mathbf{z} = L^T \mathbf{dy}_m.$$

(p×1)

Diese sind unkorreliert:  $E(\mathbf{z}\mathbf{z}^T) = E(L^T d\mathbf{y}_m d\mathbf{y}_m^T L) = L^T \Sigma_{d\mathbf{y}_m} L = I_p$ . Alternativ kann auf das Whitening verzichtet werden, dann sind die transformierten Variablen weiterhin korreliert, aber ihre Varianz wird auf eins normiert:  $\Sigma_{d\hat{\mathbf{y}}_m}^{-1}$  ist dann eine Diagonalmatrix und die Hauptdiagonalelemente sind die inversen Varianzen der beobachtbaren endogenen Variablen.

Bei der Iteration müssen die Identifikationsbedingungen eingehalten werden. Diese werden ausschließlich über Null-Restriktionen abgebildet, durch Hadamard-Multiplikation von  $M_y$  und  $M_x$  mit den Identifikationsmatrizen  $ID_y$  bzw.  $ID_x$ . Die entsprechenden Einträge sind 1 bzw. 0, je nachdem, ob aus der Theorie, also den Modellgleichungen, eine partielle Ableitung ungleich null vorliegt oder nicht. So wird festgelegt, welche Elemente freie Parameter sind und welche auf null restringiert sind.

Die  $l$  latenten Variablen werden mit dem neuronalen Netzwerk bestimmt, womit sich auch die  $p$  manifesten Output-Variablen ergeben. Die Prognosen werden mit einem Dach als Schätzer gekennzeichnet. Die Variablen liegen für die Beobachtungen  $t = 1, \dots, \tau$  vor. Wir fassen die Daten in Matrizen zusammen (im Vergleich zum vorigen Abschnitt transponiert):

$$\begin{matrix} dY & = & [dy_{jt}]_{j=1,\dots,n} & = & [dy_1, \dots, dy_\tau], \\ (n \times \tau) & & t=1,\dots,\tau & & \end{matrix}$$

$$\begin{matrix} dY_m & = & [dy_{m,jt}]_{j=1,\dots,p} & = & [dy_{m,1}, \dots, dy_{m,\tau}], \\ (p \times \tau) & & t=1,\dots,\tau & & \end{matrix}$$

$$\begin{matrix} d\hat{Y} & = & [d\hat{y}_{jt}]_{j=1,\dots,n} & = & [d\hat{y}_1, \dots, d\hat{y}_\tau] = (I_n - M_y)^{-1} M_x dX = E_x dX, \\ (n \times \tau) & & t=1,\dots,\tau & & \end{matrix}$$

$$\begin{matrix} d\hat{Y}_m & = & [d\hat{y}_{m,jt}]_{j=1,\dots,p} & = & F_{y_m} d\hat{Y} = [d\hat{y}_{m,1}, \dots, d\hat{y}_{m,\tau}] = \begin{bmatrix} d\hat{y}_{m(1)} \\ \vdots \\ d\hat{y}_{m(p)} \end{bmatrix}, \\ (p \times \tau) & & t=1,\dots,\tau & & \end{matrix}$$

$$\begin{matrix} dX & = & [dx_{it}]_{i=1,\dots,m} & = & [dx_1, \dots, dx_\tau] \\ (m \times \tau) & & t=1,\dots,\tau & & \end{matrix}$$

und die gewhiteten endogenen Daten und Prognosen lauten entsprechend:

$$\begin{matrix} Z & = & [z_{jt}]_{j=1,\dots,p} & = & [z_1, \dots, z_\tau] = L^T dY_m, \\ (p \times \tau) & & t=1,\dots,\tau & & \end{matrix}$$

$$\begin{matrix} \hat{Z} & = & [\hat{z}_{jt}]_{j=1,\dots,p} & = & [\hat{z}_1, \dots, \hat{z}_\tau] = L^T d\hat{Y}_m. \\ (p \times \tau) & & t=1,\dots,\tau & & \end{matrix}$$

Das Verhalten des Gradientenabstiegsverfahrens kann dadurch beeinflusst werden, nach wie viel Beobachtungen der nächste Iterationsschritt stattfindet. Gibt man die Daten sequentiell mit  $\tau = 1$  vor, so spricht man von einem on-line oder stochastischen Gradientenabstiegsverfahren. Gibt man Pakete von mehreren Beobachtungen vor, so spricht man von einem Batch-Verfahren und gibt man

alle Trainingsdaten auf einmal vor, so ist das der klassische Fall des Gradientenabstiegsverfahrens. Aufgrund der relativ geringen Anzahl der Beobachtungen wählen wir den letzten Ansatz.

Die Daten sind gegeben und fest. Zur Minimierung der Zielfunktion werden die Gewichte angepasst. Dies geschieht iterativ mit einem Gradientenabstiegsverfahren, im Kontext der neuronalen Netze als Backpropagation bezeichnet, siehe zum Beispiel Rojas (1996). Dafür muss die Zielfunktion nach den Gewichten abgeleitet werden, also der Gradient bestimmt werden. Wir leiten eine algebraische Formel für den Gradienten her. Sie ersetzt das iterative Backpropagation-Verfahren, das in den üblichen Machine-Learning Tools mittels algorithmischer Differenzierung umgesetzt ist.

Die Schätzung der Gewichte erfolgt mit dem Verfahren der nichtlinearen Kleinsten Quadrate, NLS (Nonlinear Least Squares), siehe [https://en.wikipedia.org/wiki/Non-linear\\_least\\_squares](https://en.wikipedia.org/wiki/Non-linear_least_squares). NLS ist konsistent, und seine Schätzer sind asymptotisch normalverteilt. Damit sind Hypothesentests der Koeffizienten möglich. Hierbei ist zu beachten, dass aufgrund der Identifikationsrestriktionen einzelne Elemente der Gradienten auf null festgelegt sind. Damit ist es auch möglich, die Gewichte der latenten Variablen zu schätzen, sofern diese identifiziert sind.

Die Zielfunktion, auch Loss-Funktion genannt, besteht aus dem Kleinste-Quadrate-Term und der additiven Tikhonov-Regularisierung. Der Kleinste-Quadrate-Term der Zielfunktion ist die Fehlerquadratsumme über alle Beobachtungen und alle mittels Whitening transformierte manifeste Variable:

$$\begin{aligned}
 \text{SSE} &= \sum_{i=1}^p \sum_{t=1}^{\tau} (\hat{Z}_{it} - Z_{it})^2 \\
 (1 \times 1) & \\
 &= \sum_{i=1}^p \sum_{t=1}^{\tau} \left( (L^T d\hat{Y}_m)_{it} - (L^T dY_m)_{it} \right)^2 \\
 &= \text{vec}(L^T d\hat{Y}_m - L^T dY_m) \left( \text{vec}(L^T d\hat{Y}_m - L^T dY_m) \right)^T \\
 &= \text{tr} \left( (L^T d\hat{Y}_m - L^T dY_m) (L^T d\hat{Y}_m - L^T dY_m)^T \right) \\
 &= \text{tr} \left( L^T (d\hat{Y}_m - dY_m) (d\hat{Y}_m - dY_m)^T L \right) \\
 &= \text{tr} \left( (d\hat{Y}_m - dY_m)^T L L^T (d\hat{Y}_m - dY_m) \right) \\
 &= \text{tr} \left( (d\hat{Y}_m - dY_m)^T \Sigma_{dY_m}^{-1} (d\hat{Y}_m - dY_m) \right).
 \end{aligned}$$

mit den Prognosen für die manifesten Variablen:

$$d\hat{Y}_m = F_{y_m} (I_n - M_y)^{-1} M_x dX = F_{y_m} E_x dX.$$

Die Zielfunktion wird ergänzt um die Tikhonov-Regularisierung, auch Ridge Regression oder Weight-Decay genannt. Diese schrumpft die Gewichte, in diesem Fall in Richtung der theoretischen Werte. Der additive Tikhonov-Term lautet:

$$\text{Tikh} = \alpha(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \Sigma_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0). \\ (1 \times 1)$$

Dabei ist  $\boldsymbol{\theta}_0$  der  $(q \times 1)$  Vektor sich aus der Theorie ergebende bekannte Parameter. Die Kovarianzmatrix der Gewichte schätzt man wie folgt aus der Hesseschen Matrix  $H_{\text{SSE}}$  der SSE-Zielfunktion in Bezug auf die Gewichte, siehe Judge et al. (1988), Formel (12.2.79):

$$\hat{\Sigma}_{\text{SSE},\boldsymbol{\theta}} = 2\hat{\sigma}^2 H_{\text{SSE}}^{-1}$$

mit der Fehlervarianz  $\hat{\sigma}^2$ , wobei anstatt  $q$  die geringeren nachfolgend dargestellten effektiven Freiheitsgrade  $df_{\text{eff}}$  verwendet werden:

$$\hat{\sigma}^2 = \text{SSE}(\hat{\boldsymbol{\theta}}) / (\tau - df_{\text{eff}}).$$

Diese Kovarianzmatrix wird sowohl im Rahmen der Schätzung für die Regularisierung als auch für die Bestimmung der Konfidenzintervalle der Gewichte sowie für das Ziehen der zufälligen Startwerte im Rahmen der Optimierung verwendet. Für die Schätzung wird  $\hat{\Sigma}_{\text{SSE},\boldsymbol{\theta}}$  wird mittels einer Vorab-Schätzung bestimmt, in welcher wiederum als Proxy für die Kovarianzmatrix eine Einheitsmatrix angesetzt wird. Dies stellt eine niedrige Konditionszahl der Hesseschen Matrix sicher. Man erhält so einen Prior  $N(\boldsymbol{\theta}_0, \hat{\Sigma}_{\text{SSE},\boldsymbol{\theta}})$ , der im Sinne der Bayesianischen Schätzung als normalverteilt angenommenen wird. In der Praxis wird der Hyperparameter  $\alpha$  dazu genutzt, die Stärke der Regularisierung zu bestimmen. Wenn das Whitening verwendet wurde liegt der Fall der verallgemeinerten Tikhonov-Regularisierung vor.

Als optimales  $\alpha$  wird nun dasjenige gewählt, das die Generalized Cross-Validation-Funktion, GCV, minimiert. Siehe Bartel, Harald (2017), Kapitel 11. 4 für einen Vergleich verschiedener Selektionskriterien. Dabei sind die *effektiven* Freiheitsgrade einzusetzen. Je stärker mittels  $\alpha$  geshrinkt wird, desto stärker gehen die effektiven Freiheitsgrade gegen null. Dies reduziert die GCV, während sie durch die Residuenquadratsumme erhöht wird:

$$\text{GCV} = \frac{\text{SSE}}{\tau} \left( \frac{\tau}{\tau - df_{\text{eff}}} \right)^2.$$

Die Anzahl der effektiven Freiheitsgrade,  $df_{\text{eff}}$ , lässt sich nicht leicht bestimmen. Zum einen ist das Shrinkage durch die Tikhonov-Regularisierung zu beachten, die zu nicht-ganzzahligen effektiven Freiheitsgraden führt. Zum anderen lässt sich das System nicht als in den Parametern lineares Regressionsmodell,  $\hat{y} = X\hat{\beta}$ , schreiben: Durch Vektorisierung erhält man die  $(p\tau \times 1)$ -dimensionale Darstellung  $\text{vec}(d\hat{Y}_m) = (dX^T \otimes I_p) \text{vec}(F_{y_m} E_x)$ , die zwar linear in den Effekten, nicht aber linear in den strukturellen Parametern  $\boldsymbol{\theta}$  ist. Wir verwenden daher als Schätzer die allgemeine Formel von Stein für die effektiven Freiheitsgrade unter Normalverteilungsannahme, siehe Tibshirani (2014), Formel (19):  $df_{\text{eff}} = \sum_{i=1}^p \partial \hat{y}_i / \partial y_i$ . Für jede der  $p$  mittelwertbereinigten manifesten Variablen ist also die Sensitivität der Prognose in Bezug auf das beobachtbare Datum zu bestimmen und diese

sind dann aufzuaddieren. Da das Whitening bereits in der Zielfunktion berücksichtigt ist, können die Größen  $\hat{y}_i$  und  $\hat{y}_i$  betrachtet werden anstatt der gewhiteten Größen  $\hat{z}_i$  und  $\hat{z}_i$ . Die Sensitivitäten werden selbst wiederum als Summe über alle  $\tau$  beobachteten Leverage Scores bestimmt, siehe [https://en.wikipedia.org/wiki/Leverage\\_\(statistics\)](https://en.wikipedia.org/wiki/Leverage_(statistics)):  $\partial \hat{y}_i / \partial y_i = \sum_{t=1}^{\tau} \partial \hat{Y}_{it} / \partial Y_{it}$ . Wir ignorieren die Tatsache, dass im Rahmen der Cross Validation mehrere effektive Freiheitsgrade bestimmt werden, wodurch sich deren Wert noch erhöht, siehe Dijkstra, T. K. (2013). Insgesamt erhält man:

$$df_{\text{eff}} = \sum_{i=1}^p \sum_{t=1}^{\tau} \frac{\partial \hat{Y}_{it}}{\partial Y_{it}}$$

Für jeden der  $p\tau$  Leverage Scores sind mehrere NLS-Schätzungen durchzuführen, da die Ableitung numerisch bestimmt wird. Daher verwenden wir eine Approximation: Es werden nur einige wenige Leverage Scores berechnet und zwar die zu den größten absoluten Residuen  $|\hat{Y}_{it} - Y_{it}|$  gehörenden und deren Summe wird auf die Summe aller absoluten Residuen hochskaliert. Von den wenigen so nacheinander hochgerechneten Schätzungen für  $df_{\text{eff}}$  wird zur Robustifizierung der Median verwendet.

Die effektiven Freiheitsgrade  $df_{\text{eff}}$  haben keinen direkten Einfluss auf den Optimierungsalgorithmus – abgesehen davon, dass sie über die Kreuzvalidierung das spätere optimale Shrinkage  $\alpha$  bestimmen, das in die Zielfunktion eingeht. In der Praxis wird regelmäßig das kleinste  $\alpha$  gewählt, bei dem die Hessesche Matrix wohlkonditioniert ist. Nach der Schätzung bei gegebenem Shrinkage wird die Konditionierung der Hesses überprüft. Zusammen mit dem Shrinkage-Parameter fließen hier die effektiven Freiheitsgrade über die Varianz-Kovarianzmatrix der Schätzer in den Tikhonov-Term für die Hessesche ein. Bei sehr schwachem Shrinkage werden die effektiven Freiheitsgrade groß, was zu reduziertem Rang der Hesseschen führen kann.

Wie oben dargestellt, fließen die effektiven Freiheitsgrade auch in die geschätzte Residuenvarianz  $\hat{\sigma}^2$  ein. Somit haben Sie auch Einfluss auf die Varianz-Kovarianzmatrix  $C$  der Schätzer und deren Konfidenzintervalle: Ein stärkeres Shrinkage führt zu kleineren effektiven Freiheitsgraden, was die geschätzte Residuenvarianz und die Konfidenzintervalle reduziert, während sie gleichzeitig durch eine größere Residuenquadratsumme erhöht werden. Der Gesamteffekt hängt also vom konkreten Fall ab. Die Optimierungsalgorithmen benötigen bei stärkerem Shrinkage weniger Iterationen, da die Zielfunktion glatter ist.

Die gesamte Zielfunktion lautet damit:

$$v_{\text{SSE}} = \text{SSE} + \text{Tikh}$$

$$= \left\| d\hat{Y}_m - dY_m \right\|_{\Sigma_{dy_m}^{-1}}^2 + \alpha \left\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \right\|_{\Sigma_{\text{SSE},\boldsymbol{\theta}}^{-1}}^2$$

$$= \text{tr} \left( \left( d\hat{Y}_m - dY_m \right)^T \Sigma_{dy_m}^{-1} \left( d\hat{Y}_m - dY_m \right) \right) + \alpha \left( \boldsymbol{\theta} - \boldsymbol{\theta}_0 \right)^T \Sigma_{\text{SSE},\boldsymbol{\theta}}^{-1} \left( \boldsymbol{\theta} - \boldsymbol{\theta}_0 \right).$$

Die beiden Summanden der Zielfunktion sind unterschiedlich formuliert: Im ersten Summanden verbergen sich die Gewichte in den Koeffizientenmatrizen  $M_x$  und  $M_y$  und im zweiten Summanden sind sie kompakt im Vektor  $\boldsymbol{\theta}$  zusammengefasst. Die Bildung des Gradienten wird entsprechend

unterschieden. Die Gradienten der Zielfunktion in Bezug auf die Gewichte lauten mittels Matrix-Calculus:

$$\frac{\partial \text{SSE}}{\partial \mathbf{M}_y} = (\nabla d\hat{\mathbf{Y}}^T) \circ \text{ID}_y,$$

(n×n)

$$\frac{\partial \text{SSE}}{\partial \mathbf{M}_x} = (\nabla d\mathbf{X}^T) \circ \text{ID}_x,$$

(n×m)

mit dem Term

$$\nabla = 2 \left( \mathbf{F}_{y_m} (\mathbf{I}_n - \mathbf{M}_y)^{-1} \right)^T \Sigma_{d\mathbf{y}_m}^{-1} (d\hat{\mathbf{Y}}_m - d\mathbf{Y}_m)$$

(n×τ)

und der Gradient der Regularisierung lautet:

$$\frac{\partial \text{Tikh}}{\partial \boldsymbol{\theta}} = 2\alpha \Sigma_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0).$$

(q×1)

In dem Term  $\nabla$  ist der  $(p \times \tau)$  Gradient  $\partial \text{SSE} / \partial d\hat{\mathbf{Y}}_m = 2 \Sigma_{d\mathbf{y}_m}^{-1} (d\hat{\mathbf{Y}}_m - d\mathbf{Y}_m)$  enthalten. Das System besitzt aufgrund seiner Nichtlinearität keine geschlossene Lösung für das Nullsetzen des Gradienten und muss iterativ bestimmt werden.

Der gesamte Gradient kann wie folgt zusammengestellt werden. Hierfür bezeichnen wir mit  $V(\mathbf{Y}, \mathbf{X}, \text{ID}_y, \text{ID}_x)$  oder kurz  $V(\mathbf{Y}, \mathbf{X})$  die eindeutige Vektorisierungs- und Selektionsfunktion, die die Elemente der Matrixdarstellung mit Hilfe von Selektionsmatrizen spaltenweise in die  $(q \times 1)$  Vektordarstellung überführt:

$$\begin{aligned} \boldsymbol{\theta} &= V(\mathbf{M}_y, \mathbf{M}_x) \\ &= \text{vec} \left( \mathbf{R}_y \text{vec}(\mathbf{M}_y), \mathbf{R}_x \text{vec}(\mathbf{M}_x) \right). \end{aligned}$$

Hierfür definieren wir die Selektionsmatrizen  $\mathbf{R}_x$ ,  $\mathbf{R}_y$  und  $\mathbf{R}$ , die aus der Einheitsmatrix entstehen, indem man die Zeilen löscht, die nicht den freien Gewichten entsprechen. Also gilt beispielsweise:

$$\boldsymbol{\theta} = V(\mathbf{M}_y, \mathbf{M}_x),$$

$$\frac{\partial \text{SSE}}{\partial \boldsymbol{\theta}} = V \left( \frac{\partial \text{SSE}}{\partial \mathbf{M}_y}, \frac{\partial \text{SSE}}{\partial \mathbf{M}_x} \right).$$

Mit der umgekehrten Funktion  $(V_y^{-1}(\mathbf{v}), V_x^{-1}(\mathbf{v})) = V^{-1}(\mathbf{v})$  werden aus der Vektorform die Matrizen der Matrixform erstellt, also beispielsweise:

$$(M_y, M_x) = V^{-1}(\boldsymbol{\theta}),$$

$$\left( \frac{\partial \text{Tikh}}{\partial M_y}, \frac{\partial \text{Tikh}}{\partial M_x} \right) = \left( V_y^{-1} \left( \frac{\partial \text{Tikh}}{\partial \boldsymbol{\theta}} \right), V_x^{-1} \left( \frac{\partial \text{Tikh}}{\partial \boldsymbol{\theta}} \right) \right) = V^{-1} \left( \frac{\partial \text{Tikh}}{\partial \boldsymbol{\theta}} \right)$$

Der gesamte Gradient lautet damit in Vektorform:

$$\begin{aligned} \frac{\partial v}{\partial \boldsymbol{\theta}} &= \frac{\partial \text{SSE}}{\partial \boldsymbol{\theta}} + \frac{\partial \text{Tikh}}{\partial \boldsymbol{\theta}} \\ (q \times 1) & \\ &= V \left( \frac{\partial \text{SSE}}{\partial M_y}, \frac{\partial \text{SSE}}{\partial M_x} \right) + 2\alpha \Sigma_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta} - \boldsymbol{\theta}_0), \end{aligned}$$

bzw. in Matrixform:

$$\frac{\partial v}{\partial M_y} = \frac{\partial \text{SSE}}{\partial M_y} + V_y^{-1} \left( \frac{\partial \text{Tikh}}{\partial \boldsymbol{\theta}} \right),$$

$(n \times n)$

$$\frac{\partial v}{\partial M_x} = \frac{\partial \text{SSE}}{\partial M_x} + V_x^{-1} \left( \frac{\partial \text{Tikh}}{\partial \boldsymbol{\theta}} \right).$$

$(n \times m)$

Die Hessesche Matrix, also die Matrix der zweiten Ableitungen der Zielfunktion nach den Gewichten, siehe [https://en.wikipedia.org/wiki/Hessian\\_matrix](https://en.wikipedia.org/wiki/Hessian_matrix), setzt sich ebenfalls aus zwei Termen zusammen:

$$H_{\text{SSE}+\text{Tikh}} = H_{\text{SSE}} + H_{\text{Tikh}}.$$

Die Bestimmung der Hesseschen Matrix der Zielfunktion ist für den SSE-Term ist nicht einfach möglich, da dieser nicht algebraisch als Vektor vorliegt und die Ableitung der Matrix nach einer anderen Matrix zu einem vierdimensionalen Tensor führen würden. Für den zweiten Summanden ist die Bestimmung der Hesseschen Matrix möglich:

$$H_{\text{Tikh}} = \frac{\partial \text{Tikh}^2}{\partial \boldsymbol{\theta} \boldsymbol{\theta}^T} = 2\alpha \Sigma_{\boldsymbol{\theta}}^{-1}.$$

$(q \times q)$

Wir wollen die Zielfunktion, den Gradienten und auch die Hessesche Matrix als Funktion von  $\boldsymbol{\theta}$  darstellen. Alle nicht selektierten Elemente sind auf null restringiert. Ferner werden die freien Ge-

wichte vektorisiert und in den Vektoren  $\theta_y$ ,  $\theta_x$  und  $\theta$  mit den Dimensionen  $q_y + q_x = q$  zusammengefasst:

$$\theta_{(q \times 1)} = \begin{bmatrix} \theta_y \\ \theta_x \end{bmatrix} = \begin{bmatrix} R_y \text{vec}(M_y) \\ R_x \text{vec}(M_x) \end{bmatrix} = \begin{bmatrix} R_y \\ R_x \end{bmatrix} \begin{bmatrix} \text{vec}(M_y) \\ \text{vec}(M_x) \end{bmatrix} = R \text{vec}(M_y, M_x).$$

Auch die Gradienten der Zielfunktion werden entsprechend vektorisiert und selektiert, wobei eine noch nicht vorliegende Formulierung des Gradienten einheitlich in Form von  $\theta$  oder  $M_y$ ,  $M_x$  vorausgesetzt wird:

$$\frac{\partial v}{\partial \theta}_{(q \times 1)} = \begin{bmatrix} \frac{\partial v}{\partial \theta_y} \\ \frac{\partial v}{\partial \theta_x} \end{bmatrix} = \begin{bmatrix} R_y \text{vec} \left( \frac{\partial v}{\partial M_y} \right) \\ R_x \text{vec} \left( \frac{\partial v}{\partial M_x} \right) \end{bmatrix} = \begin{bmatrix} R_y \\ R_x \end{bmatrix} \begin{bmatrix} \text{vec} \left( \frac{\partial v}{\partial M_y} \right) \\ \text{vec} \left( \frac{\partial v}{\partial M_x} \right) \end{bmatrix} = R \text{vec} \left( \frac{\partial v}{\partial M_y}, \frac{\partial v}{\partial M_x} \right).$$

Die Hessesche Matrix ergibt sich dann durch erneute Ableitung des Gradienten nach  $\theta$ . Die Elemente von Zielfunktion, Gradient und Hessescher Matrix können jedoch nicht direkt algebraisch in Matrixschreibweise bestimmt werden, da sie nicht als Funktionen von  $\theta_x$  bzw.  $\theta_y$  dargestellt werden können. Ein anderer Ansatz besteht darin, die Hessesche Matrix des SSE-Terms approximativ zu bestimmen, mittels Gauß-Newton-Verfahren für die Lösung nichtlinearer Kleinster-Quadrate-Probleme, NLS, siehe [https://en.wikipedia.org/wiki/Gauss-Newton\\_algorithm](https://en.wikipedia.org/wiki/Gauss-Newton_algorithm). Dieses ist anwendbar, wenn Fehlerquadratsummen minimiert werden und hat den Vorteil, dass die Hessesche Matrix nicht über die zweiten Ableitungen bestimmt werden muss. Die Approximation besteht in der Taylorentwicklung erster Ordnung des Gradienten. Die Hessesche Matrix des SSE-Terms ergibt sich demnach wie folgt aus der ersten Ableitung der Prognosen nach  $\theta$ :

$$\begin{aligned} H_{\text{SSE}} &= \frac{\partial \text{SSE}^2}{\partial \theta \theta^T} = \begin{bmatrix} \frac{\partial \text{SSE}^2}{\partial \theta_y \partial \theta_y^T} & \frac{\partial \text{SSE}^2}{\partial \theta_y \partial \theta_x^T} \\ \frac{\partial \text{SSE}^2}{\partial \theta_x \partial \theta_y^T} & \frac{\partial \text{SSE}^2}{\partial \theta_x \partial \theta_x^T} \end{bmatrix} = \begin{bmatrix} \frac{\partial R_y \text{vec} \left( \frac{\partial \text{SSE}}{\partial M_y} \right)}{\partial \theta_y^T} & \frac{\partial R_y \text{vec} \left( \frac{\partial \text{SSE}}{\partial M_y} \right)}{\partial \theta_x^T} \\ \frac{\partial R_x \text{vec} \left( \frac{\partial \text{SSE}}{\partial M_x} \right)}{\partial \theta_y^T} & \frac{\partial R_x \text{vec} \left( \frac{\partial \text{SSE}}{\partial M_x} \right)}{\partial \theta_x^T} \end{bmatrix} \\ &\approx \frac{1}{2} \left[ \frac{\partial d\hat{y}_{(1)}}{\partial \theta}, \dots, \frac{\partial d\hat{y}_{(p)}}{\partial \theta} \right] (\Sigma_{\text{dym}}^{-1} \otimes I_\tau) \left[ \frac{\partial d\hat{y}_{(1)}}{\partial \theta}, \dots, \frac{\partial d\hat{y}_{(p)}}{\partial \theta} \right]^T \\ &= \frac{1}{2} J (\Sigma_{\text{dym}}^{-1} \otimes I_\tau) J^T, \end{aligned}$$

für die allgemeine GLS-Gewichtung siehe Judge et al. (1988), Formel (12.4.3). Dabei ist  $J$  der Gradient der Prognosen:

$$J_{(q \times n\tau)} = \left[ \frac{\partial d\hat{y}_{(1)}}{\partial \theta}, \dots, \frac{\partial d\hat{y}_{(p)}}{\partial \theta} \right]$$

$$= \begin{bmatrix} \frac{\partial d\hat{y}_{(1)}}{\partial \theta_y} & \cdots & \frac{\partial d\hat{y}_{(p)}}{\partial \theta_y} \\ \frac{\partial d\hat{y}_{(1)}}{\partial \theta_x} & \cdots & \frac{\partial d\hat{y}_{(p)}}{\partial \theta_x} \end{bmatrix}.$$

Jedoch besteht auch hier das Problem, dass die Prognosen nicht als Funktionen von  $\theta_y$  bzw.  $\theta_x$  dargestellt werden können.

Zur Bestimmung der Hesseschen Matrix  $H_{SSE}$  muss daher eine Ableitung der *einzelnen* Elemente erfolgen. Für die Hessesche Matrix werden nur die durch die Identifikationsmatrizen bestimmten die Elemente benötigt. Also muss sowohl das Element, das abgeleitet wird ungleich null sein, als auch das Element nach dem abgeleitet wird. Bei den nachfolgenden Ableitungen von Matrix zu Skalar wird das Numerator-Layout und dann das  $kl$ -te Element verwendet. Die Hessesche wird algebraisch abgeleitet, aber an der Stelle der beobachteten Daten  $dX$ ,  $dY_m$  ausgewertet. Dadurch wird berücksichtigt, dass zwar algebraische Identifikation vorliegen kann, aber keine numerische Identifikation, da beispielsweise Daten oder Effekte gleich null sein können. Es handelt sich um einen Schätzer für den Erwartungswert der Hesseschen, bzw. die negative Fisher Information Matrix, siehe [https://en.wikipedia.org/wiki/Fisher\\_information#Matrix\\_form](https://en.wikipedia.org/wiki/Fisher_information#Matrix_form):

$$\begin{aligned} \frac{\partial \left( \frac{\partial SSE}{\partial M_x} \right)_{kl}}{\partial m_{xij}} &= \left( \frac{\partial}{\partial m_{xij}} 2 \left( F_{ym} (I_n - M_y)^{-1} \right)^T \Sigma_{dy_m}^{-1} (d\hat{Y}_m - dY_m) dX^T \right)_{kl} (ID_x)_{kl} (ID_x)_{ij} \\ &= 2 \frac{\partial}{\partial m_{xij}} \left( A^T \Sigma_{dy_m}^{-1} (d\hat{Y}_m) dX^T \right)_{kl} (ID_x)_{kl} (ID_x)_{ij} \\ &= 2 \frac{\partial}{\partial m_{xij}} \left( A^T \Sigma_{dy_m}^{-1} \left( F_{ym} (I_n - M_y)^{-1} M_x dX \right) dX^T \right)_{kl} (ID_x)_{kl} (ID_x)_{ij} \\ &= 2 \left( A^T \Sigma_{dy_m}^{-1} F_{ym} (I_n - M_y)^{-1} J_x^{ij} dX dX^T \right)_{kl} (ID_x)_{kl} (ID_x)_{ij} \\ &= 2 (A^T B dX dX^T)_{kl} (ID_x)_{kl} (ID_x)_{ij} \end{aligned}$$

$$\begin{aligned} \frac{\partial \left( \frac{\partial SSE}{\partial M_x} \right)_{kl}}{\partial m_{yij}} &= \left( \frac{\partial}{\partial m_{yij}} 2 \left( F_{ym} (I_n - M_y)^{-1} \right)^T \Sigma_{dy_m}^{-1} (d\hat{Y}_m - dY_m) dX^T \right)_{kl} (ID_x)_{kl} (ID_y)_{ij} \\ &= 2 \left( \frac{\partial}{\partial m_{yij}} \left( F_{ym} (I_n - M_y)^{-1} \right)^T \Sigma_{dy_m}^{-1} \left( F_{ym} (I_n - M_y)^{-1} M_x dX \right) dX^T \right. \\ &\quad \left. - (I_n - M_y)^{-1T} J_y^{ji} (I_n - M_y)^{-1T} F_{ym}^T \Sigma_{dy_m}^{-1} dY_m dX^T \right)_{kl} (ID_x)_{kl} (ID_y)_{ij} \end{aligned}$$

$$= 2 \left( \left( A^T \Sigma_{\hat{y}_m}^{-1} \frac{\partial F_{y_m} (I_n - M_y)^{-1}}{\partial m_{yij}} + \frac{\partial (F_{y_m} (I_n - M_y)^{-1})^T}{\partial m_{yij}} \Sigma_{\hat{y}_m}^{-1} A \right) M_x dX dX^T - C^T dY_m dX^T \right)_{kl} (ID_x)_{kl} (ID_y)_{ij}$$

$$= 2 \left( \left( A^T \Sigma_{\hat{y}_m}^{-1} F_{y_m} (I_n - M_y)^{-1} J_y^{ij} (I_n - M_y)^{-1} + (I_n - M_y)^{-1} J_y^{ji} (I_n - M_y)^{-1} F_{y_m}^T \Sigma_{\hat{y}_m}^{-1} A \right) M_x dX dX^T - C^T dY_m dX^T \right)_{kl} (ID_x)_{kl} (ID_y)_{ij}$$

$$= 2 \left( (A^T C + C^T A) M_x dX dX^T - C^T dY_m dX^T \right)_{kl} (ID_x)_{kl} (ID_y)_{ij}$$

$$\frac{\partial \left( \frac{\partial SSE}{\partial M_y} \right)_{kl}}{\partial m_{xij}} = \left( \frac{\partial}{\partial m_{xij}} 2 (F_{y_m} (I_n - M_y)^{-1})^T \Sigma_{\hat{y}_m}^{-1} (d\hat{Y}_m - dY_m) d\hat{Y}^T \right)_{kl} (ID_y)_{kl} (ID_x)_{ij}$$

$$= 2 \left( \frac{\partial}{\partial m_{xij}} A^T \Sigma_{\hat{y}_m}^{-1} (A M_x dX - dY_m) dX^T M_x^T I^T \right)_{kl} (ID_y)_{kl} (ID_x)_{ij}$$

$$= 2 \left( A^T \Sigma_{\hat{y}_m}^{-1} \left( \frac{\partial}{\partial m_{xij}} A M_x dX dX^T M_x^T - \frac{\partial}{\partial m_{xij}} dY_m dX^T M_x^T \right) I^T \right)_{kl} (ID_y)_{kl} (ID_x)_{ij}$$

$$= 2 \left( A^T \Sigma_{\hat{y}_m}^{-1} \left( A \left( M_x dX dX^T \frac{\partial M_x^T}{\partial m_{xij}} + \frac{\partial M_x}{\partial m_{xij}} dX dX^T M_x^T \right) - dY_m dX^T J_x^{ijT} \right) I^T \right)_{kl} (ID_y)_{kl} (ID_x)_{ij}$$

$$= 2 \left( A^T \Sigma_{\hat{y}_m}^{-1} \left( A (M_x dX dX^T J_x^{ii} + J_x^{ij} dX dX^T M_x^T) - dY_m dX^T J_x^{ijT} \right) I^T \right)_{kl} (ID_y)_{kl} (ID_x)_{ij}$$

$$= 2 \left( A^T \Sigma_{\hat{y}_m}^{-1} \left( A (D^T + D) - dY_m dX^T J_x^{ijT} \right) I^T \right)_{kl} (ID_y)_{kl} (ID_x)_{ij}$$

$$\frac{\partial \left( \frac{\partial SSE}{\partial M_y} \right)_{kl}}{\partial m_{yij}} = \left( \frac{\partial}{\partial m_{yij}} 2 (F_{y_m} (I_n - M_y)^{-1})^T \Sigma_{\hat{y}_m}^{-1} (d\hat{Y}_m - dY_m) d\hat{Y}^T \right)_{kl} (ID_y)_{kl} (ID_y)_{ij}$$

$$\begin{aligned}
&= 2 \left( \frac{\partial}{\partial m_{yij}} (I_n - M_y)^{-1T} F_{y_m}^T \Sigma_{d_{y_m}}^{-1} (d\hat{Y}_m - dY_m) dX^T M_x^T (I_n - M_y)^{-1T} \right)_{kl} (ID_y)_{kl} (ID_y)_{ij} \\
&= 2 \left( \frac{\partial}{\partial m_{yij}} (I_n - M_y)^{-1T} F_{y_m}^T \Sigma_{d_{y_m}}^{-1} F_{y_m} (I_n - M_y)^{-1} M_x dX dX^T M_x^T (I_n - M_y)^{-1T} \right. \\
&\quad \left. - \frac{\partial}{\partial m_{yij}} (I_n - M_y)^{-1T} F_{y_m}^T \Sigma_{d_{y_m}}^{-1} dY_m dX^T M_x^T (I_n - M_y)^{-1T} \right)_{kl} (ID_y)_{kl} (ID_y)_{ij} \\
&= 2 \left( \frac{\partial}{\partial m_{yij}} (I_n - M_y)^{-1T} G (I_n - M_y)^{-1} H (I_n - M_y)^{-1T} \right. \\
&\quad \left. - \frac{\partial}{\partial m_{yij}} (I_n - M_y)^{-1T} E (I_n - M_y)^{-1T} \right)_{kl} (ID_y)_{kl} (ID_y)_{ij} \\
&= 2 \left( \left( \frac{\partial (I_n - M_y)^{-1T}}{\partial m_{yij}} G (I_n - M_y)^{-1} H (I_n - M_y)^{-1T} \right. \right. \\
&\quad + (I_n - M_y)^{-1T} G \frac{\partial (I_n - M_y)^{-1}}{\partial m_{yij}} H (I_n - M_y)^{-1T} \\
&\quad + (I_n - M_y)^{-1T} G (I_n - M_y)^{-1} H \frac{\partial (I_n - M_y)^{-1T}}{\partial m_{yij}} \left. \right) \\
&\quad - \left( (I_n - M_y)^{-1T} E \frac{\partial (I_n - M_y)^{-1T}}{\partial m_{yij}} \right. \\
&\quad \left. \left. + \frac{\partial (I_n - M_y)^{-1T}}{\partial m_{yij}} E (I_n - M_y)^{-1T} \right) \right)_{kl} (ID_y)_{kl} (ID_y)_{ij} \\
&= 2 \left( (F^T G I H I^T + I^T G F H I^T + I^T G I H F^T) - (I^T E F^T + F^T E I^T) \right)_{kl} (ID_y)_{kl} (ID_y)_{ij}
\end{aligned}$$

mit der  $(n \times n)$  Elementarmatrix  $J_y^{ij}$ , der  $(n \times m)$  Elementarmatrix  $J_x^{ij}$  und den Abkürzungen:

$$A = F_{y_m} I$$

$$B = \Sigma_{d_{y_m}}^{-1} A J_x^{ij}$$

$$C = \Sigma_{d_{y_m}}^{-1} A J_y^{ij} I$$

$$D = J_x^{ij} dX dX^T M_x^T$$

$$E = F_{y_m}^T \Sigma_{dy_m}^{-1} dY_m dX^T M_x^T$$

$$F = I_{y_m}^{ij} I$$

$$G = F_{y_m}^T \Sigma_{dy_m}^{-1} F_{y_m}$$

$$H = M_x dX dX^T M_x^T$$

$$I = (I_n - M_y)^{-1}.$$

Die Hessesche Matrix kann insgesamt für folgende Zwecke verwendet werden:

- Identifikation: Bei vollem Rang der Hesseschen ist das System lokal identifiziert.
- Die Kovarianzmatrix der Gewichte folgt direkt aus der Hesseschen:  $\hat{\Sigma}_{SSE, \theta} = 2\hat{\sigma}^2 H_{SSE}^{-1}$ .
- Newton-Raphson- oder Gauß-Newton-Algorithmus: Die iterative Schätzung der Gewichte kann durch die Krümmungsinformation der Hesseschen verbessert werden, indem sie invertiert von links an den Gradienten multipliziert wird.

Das Optimierungsverfahren unter Beachtung der Identifikationsrestriktionen, der Tikhonov-Regularisierung und mit Lernrate  $r$  lautet also:

$$\theta \leftarrow \theta - r H_{SSE+Tikh}^{-1} \frac{\partial v_{SSE}}{\partial \theta}$$

bzw. wenn  $H_{SSE+Tikh} = I_q$  ist auch die Matrixdarstellung möglich:

$$M_y \leftarrow M_y - r \frac{\partial v_{SSE}}{\partial M_y} \circ ID_y,$$

$$M_x \leftarrow M_x - r \frac{\partial v_{SSE}}{\partial M_x} \circ ID_x.$$

Als Startwerte werden die Größen  $M_y$  und  $M_x$  aus dem theoretischen Modell genommen. Im Fall  $H = I_q$  ergibt sich das Gradientenabstiegsverfahren, auch Verfahren des steilsten Abstiegs genannt. Im Fall einer Approximation der Hesseschen Matrix ergibt sich das Gauß-Newton-Verfahren. Wird die Hessesche Matrix, wie oben, ohne Approximation verwendet, so resultiert der Newton-Raphson Algorithmus.

Die Parameterschätzung, also die Optimierung der Zielfunktion erfolgt in Python mittels `scipy.optimize.minimize`. Dort stehen verschiedene Optimierungsalgorithmen zur Verfügung. Wir verwenden: BFGS, Nelder-Mead, Powell, CG, dogleg, trust-ncg, Newton-CG, trust-krylov und trust-exact. Wir folgen dabei der Empfehlung, den Gradienten und die Hessesche als geschlossene Formel zu übergeben, sofern verfügbar: [http://www.scipy-lectures.org/advanced/mathematical\\_optimization](http://www.scipy-lectures.org/advanced/mathematical_optimization).

Eine Alternative besteht darin, die Gradienten mittels AD, algorithmischer Differenzierung, zu bestimmen. Dies kann durch das Python-Modus PyTorch und seinem integrierten Modul autograd erfolgen. Dabei wird der Gradient auf Ebene des Programmcodes bestimmt und numerisch exakt ausgewertet. Dieser Ansatz wird im Rahmen der Künstlichen Intelligenz genutzt, um sehr tiefe neuronale Netze schnell und exakt mittels Backpropagation zu schätzen. In unserem Fall werden die Strukturellen Parameter-Matrizen  $M_x$  und  $M_y$  iterativ geschätzt, indem die Ableitung der Kleinste-Quadrate-Zielfunktion nach diesen Parameters algorithmisch bestimmt wird.

Nach erfolgter Schätzung des linearisierten mittelwertbereinigten Systems sind die Koeffizienten des nichtlinearen und nicht mittelwertbereinigten Ausgangsmodells  $y_j = M^j(\mathbf{y}, \mathbf{x})$ ,  $j = 1 \dots, n$  geeignet anzupassen, so dass dessen partielle Ableitungen besser mit den geschätzten Koeffizienten und den beobachteten Daten übereinstimmen. Hierfür gibt es keine eindeutige Zuordnung, so dass die Anpassung im Bemessen des Modellierers liegt. Zur Unterstützung wird für jede Gleichung eine Konstante als Modifikationsindikator bestimmt. Da diese im Allgemeinen nicht simultan identifiziert sind, werden die Modifikationsindikatoren jeweils *einzel*n so bestimmt, dass die Zielfunktion optimiert wird. Es wird eine numerische Optimierung ohne Gradient durchgeführt. Für den Fall der simultanen Schätzung ist in Anhang 7 der Gradient der Zielfunktion in Bezug auf die Prognosen dargestellt, jedoch nur für den Spezialfall einer diagonalen Gewichtungsmatrix  $\Sigma_{\mathbf{d}y_m}^{-1}$ . Für die Schätzung und die Berechnung der Modifikationsindikatoren verwenden wir PyTorch. Die Effekte werden also mittels algorithmischer Differentiation bestimmt.

Die Modifikationsindikatoren sind Indizien, wie man die Modellgleichungen anpassen könnte, um einen besseren Fit zu erhalten. Die Konstanten sind unabhängig von den Gewichten, da sie beim Ableiten entfallen. Die Indikatoren werden für die Gleichungen des Ausgangsmodells bestimmt, also passend zu den Startwerten der Gewichte, da ja gerade das Problem gelöst werden soll, wie man von den geschätzten Gewichten zu den Modellgleichungen kommt. Die Indikatoren zeigen für eine Gleichung, welche Konstante oder welcher Faktor geeignet ist, um die entsprechende Spalte der Matrix  $M_y$  durch einen additiven Term oder einen Faktor zu modifizieren. Wendet man die Modifikationsindikatoren auf nur eine einzelne Beobachtung an, so kann deren Besonderheit untersucht werden, dies ist aufgrund des Freiheitsgrades von eins nur für jeweils eine einzelne Größe möglich.

### 5.3 Kovarianzstrukturmodelle (psychometrische SEM)

Kovarianzstrukturmodelle sind strukturelle simultane Mehrgleichungsmodelle mit latenten Variablen. Unter der Annahme der Normalverteilung können die strukturellen Modellparameter aus den vom Modell induzierten Verteilungsparametern geschätzt werden, die als suffiziente Statistik dienen. Neben dem Mittelwert ist insbesondere die Kovarianzmatrix relevant. Diese SEM Modelle können aufgrund dieses Ansatzes auch Koeffizienten latenter Variablen schätzen. Diese kommen in der Psychometrie häufig vor, zum Beispiel im Bereich der Intelligenzmessung. Psychometrische Typischerweise erfolgt eine Maximum-Likelihood (ML)-Schätzung der Parameter, so dass die durch das Modell implizierte Kovarianzmatrix möglichst gut die beobachtete Kovarianzmatrix trifft. SEM verfügen aufgrund der bekannten asymptotischen Verteilung über eine Testtheorie und sind somit empirisch überprüfbar. Eine umfassende Darstellung findet sich in Bollen (1989), ein Überblick in Magnus and Neudecker (2007), Kapitel 17, Abschnitte 12-15.

Die einfachste, jedoch vollkommen allgemeine, Darstellung eines Kovarianzstrukturmodells ist das

Reticular Action Model (RAM), McArdle (1980), McArdle and McDonald (19984). Wir verwenden die Variante des sogenannten kompakten RAMs. Andere Darstellungen sind LISREL, Bentler-Weeks und EzPath, siehe Steiger.

Aufgrund der Existenz der latenten Variablen ist der Identifikation der Koeffizienten von besonderer Bedeutung. Unter der Annahme der Normalverteilung und bei Mittelwerten von null, kann das RAM ausschließlich über die Struktur der Kovarianz bestimmt werden, siehe Bekker et al (1994).

Das kompakte RAM lautet:

$$d\mathbf{v} = A d\mathbf{v} + \mathbf{r}.$$

Die Koeffizientenmatrix folgt der Form des strukturellen Modells. Sämtliche gerichteten Pfadkoeffizienten werden in der asymmetrischen Matrix mit Hauptdiagonale null zusammengefasst:

$$A_{(n+m) \times (n+m)} = \begin{bmatrix} M_y & M_x \\ 0_{n,n+m} \end{bmatrix}.$$

Die reduzierte Form lautet:

$$d\mathbf{v} = (I_{n+m} - A)^{-1} \mathbf{r}$$

mit der Matrix der totalen Effekte:

$$(I_{n+m} - A)^{-1} = \begin{bmatrix} E_y & E_x \\ 0_{m,n} & I_m \end{bmatrix}.$$

Für die Spezifikation dieser Modelle ist es erforderlich, zu verstehen, welche Variablen endogen bzw. exogen sind und welche beobachtbar (manifest) und welche latent sind: Die  $n$  endogenen Variablen  $d\mathbf{y}$  werden, aufgeteilt in  $l$  latente endogene Variablen,  $d\mathbf{y}_l$ , und  $p$  manifeste endogene Variablen,  $d\mathbf{y}_m$ , also  $n = l + p$ . Außerdem sind in unserem Fall sämtliche exogenen Variablen,  $d\mathbf{x}$ , manifest und gehören zu den manifesten Variablen  $d\mathbf{m}$ . Das RAM lässt jedoch auch latente exogene Variable zu. Alle Variablen werden in einen gemeinsamen Vektor geschrieben. Zur einfacheren Darstellung nehmen wir an, dass die manifesten endogenen Variablen unten im Vektor  $d\mathbf{y}$  stehen:

$$\begin{aligned} d\mathbf{v}_{(n+m \times 1)} &= \begin{bmatrix} d\mathbf{y}_l \\ d\mathbf{y}_m \\ d\mathbf{x} \end{bmatrix} = \begin{bmatrix} d\mathbf{y} \\ d\mathbf{x} \end{bmatrix} \\ &= \begin{bmatrix} d\mathbf{y}_l \\ d\mathbf{y}_m \\ d\mathbf{x} \end{bmatrix} = \begin{bmatrix} d\mathbf{y}_l \\ d\mathbf{m} \end{bmatrix}. \end{aligned}$$

Die exogenen Variablen werden im erweiterten Vektor erfasst. Dieser enthält auch den  $(n \times 1)$  Fehlerterm  $\mathbf{u}$ . Dessen Komponente  $\mathbf{u}_m$  sind die Messfehler der manifesten endogenen Variablen, wo-

hingegen die latenten Fehler gleich null sind. Wir nehmen für manifeste endogene Variable stets einen Messfehler an, während wir latente Variablen stets ohne Messfehler annehmen:

$$\begin{aligned} \mathbf{r} &= \begin{bmatrix} \mathbf{0}_l \\ \mathbf{u}_m \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{u} \\ \mathbf{x} \end{bmatrix} \\ (n+m \times 1) & \\ &= \begin{bmatrix} \mathbf{0}_l \\ \mathbf{u}_m \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{0}_l \\ \mathbf{r}_m \end{bmatrix}. \end{aligned}$$

Wir zerlegen die Einheitsmatrix in verschiedenen Filtermatrizen:

$$\begin{aligned} I_{n+m} &= \begin{bmatrix} I_l & 0_{l,p} & 0_{l,m} \\ 0_{p,l} & I_p & 0_{p,m} \\ 0_{m,n} & 0_{m,l} & I_m \end{bmatrix} = \begin{bmatrix} F_{y_l} & 0_{l,m} \\ F_{y_m} & 0_{p,m} \\ 0_{m,n} & I_m \end{bmatrix} \\ &= \begin{bmatrix} I_l & 0_{l,p+m} \\ 0_{p+m,l} & I_{p+m} \end{bmatrix} = \begin{bmatrix} F_l \\ F_m \end{bmatrix} \\ &= \begin{bmatrix} I_n & 0_{n,m} \\ 0_{m,n} & I_m \end{bmatrix} = \begin{bmatrix} F_y \\ F_m \end{bmatrix}. \end{aligned}$$

Wir verwenden also zum Beispiel für die Selektion der manifesten Variablen  $\mathbf{dm}$  in  $\mathbf{dv}$  die  $(p+m \times n+m)$  Filtermatrix  $F_m$ , die sich aus der  $(n+m)$ -dimensionalen Einheitsmatrix ergibt, wenn man die Zeilen löscht die nicht zu manifesten endogenen Variablen gehören. Für  $F_m$  gilt:

$$\begin{aligned} \mathbf{dm} &= F_m \mathbf{dv}. \\ (p+m \times 1) & \end{aligned}$$

Die Kovarianzmatrix der exogenen Variablen,  $\Sigma_{dx}$ , ist eigentlich im Modell per Schätzung zu bestimmen. Da die exogenen Variablen durch keine Größe im Modell beeinflusst werden, ist jedoch direkt klar, dass ihr Maximum-Likelihood-Schätzer resultiert. Um die Anzahl der freien Parameter möglichst klein zu halten, wird direkt ihr ML-Schätzer

$$\hat{\Sigma}_{dx} = \frac{(X - \bar{X})(X - \bar{X})^T}{\tau - 1} = \frac{dXdX^T}{\tau - 1}$$

verwendet. Die Datenmatrizen  $dY$  ( $p \times \tau$ ),  $d\hat{Y}$  ( $p \times \tau$ ) und  $dX$  ( $m \times \tau$ ) sind wie im vorangehenden Abschnitt definiert. Analog gehen wir für die Kovarianzmatrix der Fehler der manifesten endogenen Variablen,  $\Sigma_{u_m}$ , vor. Die manifesten endogenen Variablen sind in der Regel nur Proxies für die entsprechenden Modellvariablen und werden daher nur mit korrelierten Fehlern gemessen und werden daher geschätzt. Da deren Schätzung jedoch die Konvergenz des Maximum-Likelihood-Verfahrens

erschwert, verwenden wir auch hier direkt ihren resultierenden ML-Schätzer, der die aktuellen Prognosen auf Basis der aktuellen Koeffizienten nutzt:

$$\hat{\Sigma}_{\mathbf{u}_m} = \frac{(\hat{Y} - Y)(\hat{Y} - Y)^T}{\tau - 1} = \frac{d\hat{Y}d\hat{Y}^T}{\tau - 1}.$$

Die latenten endogenen Variablen werden im Modell exakt bestimmt, die Kovarianzmatrix ihrer Fehler ist null,  $\Sigma_{\mathbf{u}_l} = \mathbf{0}_{l,l}$ . Die  $(n \times n)$  Kovarianzmatrix  $\Sigma_{\mathbf{u}}$  weist also nur Elemente der  $(p \times p)$  Kovarianzmatrix  $\hat{\Sigma}_{\mathbf{u}_m}$  der manifesten latenten Variablen auf und ist ansonsten mit Nullen aufgefüllt. Die Invertierbarkeit wird nur für die Kovarianzmatrix  $\Sigma_{\mathbf{d}_m}(\boldsymbol{\theta})$  der manifesten Variablen benötigt und ist durch  $\Sigma_{\mathbf{u}_m} \neq 0$  sichergestellt; also dürfen  $\Sigma_{\mathbf{u}}$  und  $\Sigma_{\mathbf{r}}$  Nullzeilen bzw. -Spalten aufweisen. Bei der Interpretation ist die Besonderheit zu beachten, dass auch die endogenen latenten Variablen eine Varianz von Null haben können. Dies ist der Fall für die rein endogenen latenten Variablen, die also nicht durch exogene Variablen beeinflusst werden. Siehe die vierte Zeile in der nachfolgenden Darstellung für  $\Sigma_{\mathbf{d}_v}$ . Es ergibt sich die Kovarianzmatrix der Residuen  $\mathbf{r}$ :

$$\begin{aligned} \Sigma_{\mathbf{r}} &= E(\mathbf{r}\mathbf{r}^T) \\ (n + m \times n + m) & \\ &= \begin{bmatrix} \boxed{0_{l,l} & 0_{l,p}} & 0_{l,m} \\ 0_{p,l} & \Sigma_{\mathbf{u}_m} & 0_{p,m} \\ 0_{m,l} & 0_{m,p} & \Sigma_{\mathbf{x}} \end{bmatrix} = \begin{bmatrix} \Sigma_{\mathbf{u}} & 0_{n,m} \\ 0_{m,n} & \Sigma_{\mathbf{d}_x} \end{bmatrix} \\ &= \begin{bmatrix} 0_{l,l} & 0_{l,p} & 0_{l,m} \\ 0_{p,l} & \boxed{\Sigma_{\mathbf{u}_m} & 0_{p,m}} \\ 0_{m,l} & 0_{m,p} & \Sigma_{\mathbf{d}_x} \end{bmatrix} = \begin{bmatrix} 0_{l,l} & 0_{l,p+m} \\ 0_{p+m,l} & \Sigma_{\mathbf{r}_m} \end{bmatrix}. \end{aligned}$$

mit

$$\Sigma_{\mathbf{u}} = \begin{bmatrix} 0_{l,l} & 0_{l,p} \\ 0_{p,l} & \Sigma_{\mathbf{u}_m} \end{bmatrix}$$

und

$$\Sigma_{\mathbf{u}_m} = F_{y_m} \Sigma_{\mathbf{u}} F_{y_m}^T$$

und da die latenten Variablen einen Fehler von null haben, kann man aus den Fehlern der manifesten Variablen die gesamte Kovarianzmatrix der endogenen Fehler rekonstruieren:

$$\Sigma_{\mathbf{u}} = F_{y_m}^T \Sigma_{\mathbf{u}_m} F_{y_m}.$$

Die Kovarianzmatrix aller Variablen  $\mathbf{d}_v$  beträgt:

$$\begin{aligned}
\Sigma_{dv} &= E(dvdv^T) \\
(n + m \times n + m) & \\
&= (I_{n+m} - A)^{-1} E(\mathbf{r}\mathbf{r}^T) (I_{n+m} - A)^{-1T} \\
&= (I_{n+m} - A)^{-1} \Sigma_r (I_{n+m} - A)^{-1T} \\
&= \begin{bmatrix} E_y & E_x \\ 0_{m,n} & I_m \end{bmatrix} \begin{bmatrix} 0_{l,l} & 0_{l,p+m} \\ 0_{p+m,l} & \Sigma_{r_m} \end{bmatrix} \begin{bmatrix} E_y & E_x \\ 0_{m,n} & I_m \end{bmatrix}^T \\
&= \begin{bmatrix} E_y & E_x \\ 0_{m,n} & I_m \end{bmatrix} \begin{bmatrix} \Sigma_u & 0_{n,m} \\ 0_{m,n} & \Sigma_{dx} \end{bmatrix} \begin{bmatrix} E_y & E_x \\ 0_{m,n} & I_m \end{bmatrix}^T \\
&= \begin{bmatrix} E_y \Sigma_u E_y^T + E_x \Sigma_{dx} E_x^T & E_x \Sigma_{dx} \\ (E_x \Sigma_{dx})^T & \Sigma_{dx} \end{bmatrix} \\
&= \begin{bmatrix} \Sigma_{dy} & E_x \Sigma_{dx} \\ (E_x \Sigma_{dx})^T & \Sigma_{dx} \end{bmatrix} \\
&= \begin{bmatrix} \boxed{F_{y_1} \Sigma_{dy} F_{y_1}^T} & \boxed{F_{y_1} \Sigma_{dy} F_{y_m}^T} & F_{y_1} E_x \Sigma_{dx} \\ \boxed{(F_{y_1} \Sigma_{dy} F_{y_m}^T)^T} & \boxed{\Sigma_{dy_m}} & F_{y_m} E_x \Sigma_{dx} \\ \boxed{(F_{y_1} E_x \Sigma_{dx})^T} & \boxed{(F_{y_m} E_x \Sigma_{dx})^T} & \boxed{\Sigma_{dx}} \end{bmatrix} \\
&= \begin{bmatrix} F_{y_1} \Sigma_{dy} F_{y_1}^T & F_{y_1} \Sigma_{dy} F_{y_m}^T & F_{y_1} E_x \Sigma_{dx} \\ \boxed{(F_{y_1} \Sigma_{dy} F_{y_m}^T)^T} & \boxed{\Sigma_{dy_m}} & \boxed{F_{y_m} E_x \Sigma_{dx}} \\ \boxed{(F_{y_1} E_x \Sigma_{dx})^T} & \boxed{(F_{y_m} E_x \Sigma_{dx})^T} & \boxed{\Sigma_{dx}} \end{bmatrix} \\
&= \begin{bmatrix} F_{y_1} \Sigma_{dy} F_{y_1}^T & F_{y_1} \Sigma_{dy} F_{y_m}^T & F_{y_1} E_x \Sigma_{dx} \\ \boxed{(F_{y_1} \Sigma_{dy} F_{y_m}^T)^T} & & \\ \boxed{(F_{y_1} E_x \Sigma_{dx})^T} & \boxed{\Sigma_{dm}(\boldsymbol{\theta})} & \end{bmatrix}
\end{aligned}$$

mit

$$\begin{aligned}
\Sigma_{dy} &= E(dydy^T) \\
(n \times n) & \\
&= F_y E(dvdv^T) F_y^T \\
&= F_y \Sigma_{dv} F_y^T
\end{aligned}$$

$$\begin{aligned}
&= E_y \Sigma_u E_y^T + E_x \Sigma_{dx} E_x^T \\
&= \begin{bmatrix} F_{y_1} \Sigma_{dy} F_{y_1}^T & F_{y_1} \Sigma_{dy} F_{y_m}^T \\ (F_{y_1} \Sigma_{dy} F_{y_m}^T)^T & \Sigma_{dy_m} \end{bmatrix}
\end{aligned}$$

und

$$\begin{aligned}
\Sigma_{dy_m} &= F_{y_m} \Sigma_{dy} F_{y_m}^T \\
(p \times p) \\
&= F_{y_m} (E_y \Sigma_u E_y^T + E_x \Sigma_{dx} E_x^T) F_{y_m}^T \\
&= (F_{y_m} E_y F_{y_m}^T) \Sigma_{u_m} (F_{y_m} E_y F_{y_m}^T)^T + (F_{y_m} E_x) \Sigma_{dx} (F_{y_m} E_x)^T.
\end{aligned}$$

Bezeichne  $\boldsymbol{\theta}$  den  $(q \times 1)$  Vektor sämtlicher nicht redundanter Modellparameter, also die freien Parameter von  $A$  (bzw.  $M_y$  und  $M_x$ ) und  $\text{vech}(\Sigma_r)$ , wobei  $\text{vech}$  der halbe Vektorisierungsoperator für symmetrische Matrizen ist, der das Dreieck oberhalb der Hauptdiagonale ignoriert. In unserem Fall hat  $\Sigma_r$  jedoch keine freien Parameter, da wir die empirischen Kovarianzen von  $d\hat{Y}$  und  $d\hat{X}$  und sonst nur Null-Blöcke verwenden. Dies rechtfertigt die Verwendung der Bezeichnung  $\boldsymbol{\theta}$  wie im vorangegangenen Abschnitt für die gleichen Parameter im neuronalen Netz. Die durch das Modell implizierte Kovarianzmatrix der manifesten Variablen  $d\mathbf{m}$  lautet damit, als Funktion von  $\boldsymbol{\theta}$ :

$$\begin{aligned}
\Sigma(\boldsymbol{\theta}) &= \Sigma_{dm}(\boldsymbol{\theta}) \\
(p + m \times p + m) \\
&= E(d\mathbf{m}d\mathbf{m}^T) \\
&= F_m E(d\mathbf{v}d\mathbf{v}^T) F_m^T \\
&= F_m \Sigma_{dv} F_m^T \\
&= F_m (I_{n+m} - A)^{-1} \Sigma_r (I_{n+m} - A)^{-1T} F_m^T \\
&= \begin{bmatrix} \Sigma_{dy_m} & F_{y_m} E_x \Sigma_{dx} \\ (F_{y_m} E_x \Sigma_{dx})^T & \Sigma_{dx} \end{bmatrix}.
\end{aligned}$$

Die ML-Schätzung freier Parameter erfolgt so, dass die Likelihood der manifesten Variablen maximiert wird, unter der Nebenbedingung, dass deren Kovarianzmatrix die Modellstruktur  $\Sigma(\boldsymbol{\theta})$  besitzt. Dies entspricht einer Wahl der Parameter, so dass die Modell-Kovarianzmatrix möglichst gut die empirische Kovarianzmatrix der manifesten Variablen

$$\hat{\Sigma} = \hat{\Sigma}_{dm}$$

$(p + m \times p + m)$

$$\begin{aligned}
&= \frac{1}{\tau - 1} \mathbf{F} \begin{bmatrix} \mathbf{Y} - \bar{\mathbf{Y}} \\ \mathbf{X} - \bar{\mathbf{X}} \end{bmatrix} \begin{bmatrix} \mathbf{Y} - \bar{\mathbf{Y}} \\ \mathbf{X} - \bar{\mathbf{X}} \end{bmatrix}^T \mathbf{F}^T \\
&= \frac{1}{\tau - 1} \mathbf{F} \begin{bmatrix} d\mathbf{Y} \\ d\mathbf{X} \end{bmatrix} \begin{bmatrix} d\mathbf{Y} \\ d\mathbf{X} \end{bmatrix}^T \mathbf{F}^T
\end{aligned}$$

erklärt. Dies ist äquivalent zur Maximierung der die Likelihood bzw. zur Minimierung der für Testzwecke geeigneten ML-Zielfunktion, siehe Bollen (1989), Anhang 4A:

$$ML(\boldsymbol{\theta}) = \ln|\Sigma(\boldsymbol{\theta})| + \text{tr}(\hat{\Sigma}\Sigma(\boldsymbol{\theta})^{-1}) - \ln|\hat{\Sigma}| - (p + m).$$

Für die Maximum-Likelihood Schätzung wird also die Invertierbarkeit von  $\Sigma(\boldsymbol{\theta})$  benötigt. Die ML-Schätzung hat den Vorteil, dass sie skalenunabhängig ist, so dass eine Gewichtung der Fehler nicht erforderlich ist.

Auch das Kovarianzstrukturmodell wird durch Hinzufügen des Tikhonov-Terms regularisiert, wie wir ihn im vorangegangenen Abschnitt 5.2. für die ökonometrischen Strukturmodelle verwendet haben. Für Regularisierung im psychometrischen SEM siehe van Kesteren and Oberski (2019) Abschnitt 3.3. Die gesamte Zielfunktion lautet damit, siehe auch Jacobucci et. al. (2016) Gleichung (6) für strukturelle Modelle mit latenten Variablen:

$$v_{ML} = ML + \text{Tikh.}$$

Als optimales  $\alpha$  wird dasjenige gewählt, das die im vorangegangenen Abschnitt dargestellte Cross-Validation-Funktion CV minimiert. (ToDo:) Dieses Vorgehen wird gewählt, obwohl CV auf die Optimierung der Fehlerquadratsumme SSE und nicht die ML-Schätzung ausgerichtet ist. Ein entsprechendes Kriterium für den Maximum-Likelihood-Ansatz liegt nicht vor. Für eine ganzzahlige auf dem Rang basierende Anzahl der Freiheitsgrade im RAM-Modell siehe Jacobucci et. al. (2016) Gleichung (7), die hier aber nicht verwendet wird.

Das gesamte Modell kann unter Verwendung der ML-Zielfunktion, also ohne Tikhonov-Term, mittels  $\chi^2$ -Test getestet werden, siehe Bollen (1989), Seite 110. Die verwendeten Freiheitsgrade  $df$  unterscheiden sich von den effektiven Freiheitsgraden  $df_{\text{eff}}$  im Rahmen der Cross-Validation des vorangehenden Abschnitts. Es gilt die asymptotische Verteilung:

$$(\tau - 1)ML \sim \chi^2 \left( df = \frac{(p + m)(p + m + 1)}{2} - q \right).$$

Auch im Fall der Kovarianzstrukturmodelle gibt es keine für alle Fälle gültige notwendige und hinreichende algebraische Methode zur Identifikation. Wir verwenden daher den allgemeinen den *empirischen Ansatz für die lokale Identifikation*: Die Hessesche Matrix  $H_{ML+\text{Tikh}}$  muss vollen Rang haben. Siehe Bollen (1989) Seite 248, Bollen and Bauldry (2010).

Ähnlich wie im vorangegangenen Abschnitt zu den neuronalen Netzen, gibt es einen direkten Zusammenhang zwischen der geschätzten Kovarianzmatrix der Parameter und der Hesseschen Matrix. Die Hessesche ist in diesem Fall die Matrix der zweiten Ableitungen der ML-Zielfunktion nach den Parametern. Bei Maximierung der Log-Likelihood würde gelten  $\hat{\Sigma}_{\text{LogL},\boldsymbol{\theta}} = -H_{\text{LogL}}^{-1}$ , man würde also

die Information Matrix verwenden, siehe Bollen (1998), Formel (4.B3). Bei Minimierung der ML-Zielfunktion, ist für die Schätzung der Kovarianzmatrix die Korrektur um den Faktor  $-2/(\tau - 1)$  erforderlich, siehe Bollen (1998), Formel (4.B5). Judge et al. (1988) zeigen, dass die asymptotischen Kovarianzmatrizen der geschätzten Koeffizienten für die Zielfunktionen SSE und ML identisch sind, siehe Formel (12.2.87). Der ML-Schätzer für die Kovarianzmatrix der Parameter lautet:

$$\hat{\Sigma}_{ML,\theta} = \frac{2}{\tau - 1} H_{ML}^{-1}.$$

Auf eine algebraische Ableitung der ML-Zielfunktion wurde verzichtet. Stattdessen werden Gradient und Hessesche Matrix numerisch bestimmt. Dies kann in der Praxis bei großen Systemen jedoch inakzeptabel lange dauern. Wir verwenden nur Schätzungen mit vollem Rang der Hesseschen.

Nachfolgend wird ergänzend die hinreichende Zwei-Schritt-Regel zur Identifikation dargestellt. Hierfür ist zuerst die Betrachtung einer alternativen Notation hilfreich, da Bollen (1989) die Identifikation in der LISREL-Notation darstellt. Wir verbinden die weit verbreitete LISREL-Notation von Jöreskog und Sörbom, siehe Bollen (1989), mit unserer auf dem RAM-Modell basierenden Darstellung. Die LISREL-Notation besteht aus dem strukturellen Modell für die latenten Variablen und dem strukturellen Modell für die manifesten Variablen:

**latente Variablen**

$$\eta = B\eta + \Gamma\xi + \zeta$$

mit

$$\Phi = \Sigma_{\xi}$$

$$\Psi = \Sigma_{\zeta}$$

**manifeste Variablen**

$$dx = \Lambda_x \xi + \delta$$

$$dy = \Lambda_y \eta + \epsilon$$

mit

$$\Theta_{\delta} = \Sigma_{\delta}$$

$$\Theta_{\epsilon} = \Sigma_{\epsilon}.$$

In unserem Fall gilt folgende Spezifikation:

## latente Variablen

$$\boldsymbol{\eta} = F_{y_1} \mathbf{y} = y_1$$

$$\boldsymbol{\xi} = \mathbf{x}$$

$$\boldsymbol{\zeta} = \mathbf{0}_1$$

$$B = F_{y_1} M_y F_{y_1}^T$$

$$\Gamma = F_{y_1} M_x$$

$$\Phi = \Sigma_x$$

$$\Psi = 0_{1,1}$$

## manifeste Variablen

$$\Lambda_x = I_m$$

$$\Lambda_y = F_{y_m} M_y F_{y_1}^T$$

$$\boldsymbol{\delta} = \mathbf{0}_m$$

$$\boldsymbol{\epsilon} = \mathbf{u}_m$$

$$\Theta_\epsilon = \Sigma_{\mathbf{u}_m}$$

$$\Theta_\delta = 0_{m,m}$$

Allerdings ist die LISREL-Darstellung nicht so allgemein, wie die RAM-Darstellung. So finden Teile der Matrizen  $M_y$  und  $M_x$  keine Berücksichtigung. Ausgedrückt in der LISREL-Notation fehlen die Wirkungen von  $d\mathbf{y}$  auf sich selber, sowie von  $d\mathbf{y}$  auf  $\boldsymbol{\eta}$  und von  $d\mathbf{x}$  auf  $d\mathbf{y}$ . In der RAM-Notation sind dies die Wirkungen von  $d\mathbf{y}_m$  auf sich selber, sowie von  $d\mathbf{y}_m$  auf  $d\mathbf{y}_1$  und von  $d\mathbf{x}$  auf  $d\mathbf{y}_m$ . Dies sind die rechten manifesten Spalten von  $M_y$  sowie die unteren manifesten Zeilen von  $M_x$ . (ToDo:) In Bollen (1989) finden sich auf den Seiten 17 und 311 Hinweise zu speziellen LISREL-Darstellungen. Diese Einschränkungen erschweren die Prüfung der Identifikation, sofern sich diese auf die LISREL-Formulierung bezieht, wie im nachfolgend dargestellten ersten Schritt der Zwei-Schritt-Regel.

Bollen (1989) stellt die Zwei-Schritt-Regel zur Identifikation der allgemeinen Kovarianzstrukturmodelle dar. Die Erfüllung beider Schritte ergibt zusammen eine hinreichende globale Identifikationsbedingung.

Schritt eins erfordert eine Umformulierung in die Form der konfirmatorischen Faktoranalyse und eine Anwendung der entsprechenden Identifikationsregeln. Damit wird die Identifikation der manifesten Variablen überprüft. Wir stellen die hinreichende Zwei-Indikatoren-Regel dar. Die Darstellung als konfirmatorische Faktoranalyse lautet:

$$\begin{matrix} \begin{bmatrix} d\mathbf{y}_m \\ d\mathbf{x} \end{bmatrix} \\ (p + m \times 1) \end{matrix} = \begin{matrix} \begin{bmatrix} F_{y_1} M_y F_{y_1}^T & F_{y_m} M_x \\ 0_{m,1} & I_m \end{bmatrix} \\ (p + m \times 1 + m) \end{matrix} \begin{matrix} \begin{bmatrix} d\mathbf{y}_1 \\ d\mathbf{x} \end{bmatrix} \\ (1 + m \times 1) \end{matrix} + \begin{matrix} \begin{bmatrix} \mathbf{u}_m \\ \mathbf{0}_m \end{bmatrix} \end{matrix}$$

mit der Fehler-Kovarianzmatrix

$$\begin{matrix} \begin{bmatrix} \Sigma_{\mathbf{u}_m} & 0_{p,m} \\ 0_{m,p} & 0_{m,m} \end{bmatrix} \\ (p + m \times p + m) \end{matrix}$$

Die zusammen hinreichenden Bedingungen für Schritt eins lauten:

0.  $m > 1$ .
1. Die Koeffizientenmatrix hat in jeder Zeile genau ein Element ungleich null.
2. Die Koeffizientenmatrix hat in jeder Spalte mindestens zwei Werte ungleich null. Zudem hat jede Spalte einen festen Wert ungleich null, zum Beispiel eins, zur Skalierung.
3. Die Kovarianzmatrix von  $[\mathbf{dy}_1, \mathbf{dx}]^T$  hat in jeder Zeile mindestens ein nicht-Diagonalelement ungleich null.
4. Die Fehler-Kovarianzmatrix ist diagonal sein. Diese Bedingungen sind in unserem Fall für die Variablen  $\mathbf{dx}$  erfüllt. Somit verbleibt nur die Prüfung der manifesten Variablen  $\mathbf{dy}_m$ .

Insbesondere die Bedingungen 1 und 2 sind in unserem Fall im Allgemeinen nicht erfüllt. Die Skalierungsanforderung resultiert vermutlich jedoch aus der Tatsache, dass die latenten Variablen in der konfirmatorischen Faktoranalyse selbst nicht durch andere Variablen beeinflusst werden. Insofern ist diese Identifikationsanforderung wohl nicht auf unseren Fall übertragbar.

Schritt zwei ist die Behandlung der latenten Variablen als manifest. Damit wird die Identifikation der latenten Variablen überprüft. Hierfür verwenden wir das notwendige und hinreichende Rangkriterium für psychometrische Kovarianzstrukturmodelle, bezogen auf die Matrix

$$M_1 = [I_1 + F_{y_1} M_y F_{y_1}^T, F_{y_1} M_x].$$

Wie im Kapitel zu den ökonometrischen Strukturmodellen dargestellt, bezeichnet  $M_{[j,-j]}$  die Matrix  $M$ , bei der die  $j$ -te Zeile herausgeschnitten wird und nur die Spalten behalten werden, bei denen in der  $j$ -ten Zeile eine null stand. Das Rangkriterium ist für die  $j$ -te latente Gleichung erfüllt, wenn voller Zeilenrang vorliegt:

$$\text{rk}(M_{[j,-j]}) = l - 1.$$

Die Rangbedingung (sowohl die psychometrische als auch die ökonometrische) ist nur notwendig und hinreichend, wenn alle Fehler des Systems ohne Restriktionen voll korreliert sind, siehe Bollen and Bauldry (2010), Bollen (1989) Seite 103, Magnus and Neudecker (2007) Kapitel 16. Dies ist beim RAM gerade nicht der Fall.

Neben den oben dargestellten Ansätzen zur Identifikation gibt es noch weitere: Unter der Annahme der Normalverteilung ist die Kovarianzmatrix der manifesten Variablen bei mittelwertbereinigten Daten eine suffiziente Statistik. Eine globale Identifikation liegt vor, wenn die Kovarianz-Bestimmungsgleichung  $\Sigma = \Sigma(\boldsymbol{\theta})$  algebraisch nach den Parametern  $\boldsymbol{\theta}$  aufgelöst werden kann. Alternativ kann die parametrisierte Kovarianzmatrix  $\Sigma(\boldsymbol{\theta})$  nach den freien Parametern  $\boldsymbol{\theta}$  abgeleitet werden und geprüft werden, ob diese vollen Rang hat. Diese Ansätze sind jedoch nur bei einfachen Modellen problemlos möglich. Für den hier nicht relevanten Spezialfall, dass alle Variablen manifest sind, gibt Toyoda (1994) eine einfache notwendige Bedingung für die Identifikation des RAMs.

## 6) Zusammenfassung

Die graphentheoretischen Effekte, bei denen mittels do-Calculus Variablen kontrafaktisch auf einen festen Wert gesetzt werden, entsprechen den totalen Ableitungen nach dieser Variablem im Gleichungssystem. Die Aufteilung der Effekte auf die direkt nachfolgenden Mediatorvariablen führt zu den finalen Effekten.

Die exogenen Effekte sind im Allgemeinen eine voll besetzte Matrix:

$$E_x = \frac{dy}{dx^T} = (I_n - M_y)^{-1} M_x.$$

(n×m)

Die endogenen Effekte lauten:

$$E_y = \frac{dy}{dy^T} = (I_n - M_y)^{-1} \left( I_n \circ (I_n - M_y)^{-1} \right)^{-1}$$

(n×n)

und können im DAG vereinfacht als strikte untere Dreiecksmatrix geschrieben werden:

$$E_y = (I_n - M_y)^{-1} = I_n + \sum_{k=1}^{n-1} M_y^k.$$

(n×n)

Die exogenen finalen Effekte über die direkt nachfolgenden Mediatorvariablen auf eine gegebene finale Zielvariable  $y_j$  lauten:

$$E_{y_j x} = \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(m)} \right) \circ M_x$$

(n×m)

und sind im DAG eine Matrix mit allen Zeilen größer als  $j$  gleich null, da es keinen Einfluss von nachfolgenden endogenen Variablen auf  $y_j$  gibt. Ist  $y_j$  jedoch die letzte Variable, so ist die Matrix voll besetzt.

Die endogenen finalen Effekte über die direkt nachfolgenden Mediatorvariablen auf eine gegebene finale Zielvariable  $y_j$  lauten:

$$E_{y_j y} = (1_{nn} - I_n) \circ \left[ \left( (I_n - M_{y_{11}})^{-1} \right)_{(j)}^T, \dots, \left( (I_n - M_{y_{nn}})^{-1} \right)_{(j)}^T \right] \circ M_y$$

(n×n)

und können im DAG vereinfacht als strikte untere Dreiecksmatrix geschrieben werden, bei der alle Spalten größer als  $j$  gleich null sind:

$$E_{y_j y} = \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(n)} \right) \circ M_y.$$

Die endogenen Variablen ergeben sich aus dem System der Modellgleichungen:

$$\mathbf{y} = \mathbf{M}(\mathbf{y}, \mathbf{x}).$$

Die linear approximierten endogenen Variablen lauten:

$$\mathbf{y} = \boldsymbol{\mu} + d\mathbf{y}$$

mit den linear approximierten Änderungen der endogenen Variablen in der strukturellen Form:

$$d\mathbf{y} = \mathbf{M}_y d\mathbf{y} + \mathbf{M}_x d\mathbf{x}$$

bzw. in der reduzierten Form:

$$d\mathbf{y} = \mathbf{E}_x d\mathbf{x}.$$

## 7) Beispiele

Wir betrachten einige Beispiele und Spezialfälle. Die Beispiele sind so gestaltet, dass sie weitgehend direkt rechnerisch nachvollzogen werden können. Im Fall des DAG sind zusätzlich die einfacheren speziellen Formeln angegeben. Eine weitere Überprüfung ist im DAG möglich, indem die Pfadeffekte direkt auf Basis des partiellen Graphen bestimmt werden. Einen Pfadeffekt erhält man, indem man die partiellen Ableitungen auf dem Pfad zwischen zwei Variablen miteinander multipliziert. Die Summe aller möglichen Pfadeffekte zwischen zwei Variablen ergibt den Effekt. Damit können in diesen einfachen Beispielen einzelne Effekte bestimmt werden, ohne das gesamte System zu lösen. Zur besseren Nachrechenbarkeit werden die exogenen Variablen an der Stelle ihrer Erwartungswerte und die endogenen Variablen deterministisch an dieser Stelle betrachtet.

### Beispiel 1 Nichtlineares System

In diesem Beispiel liegt ein nichtlineares Gleichungssystem vor, so dass die Ableitung nicht konstant ist, sondern vom Niveau der Variablen abhängt. Die Ableitung ist also an der Stelle  $\mathbf{y}$  zu bestimmen. Da sich dieses System in strikter Dreiecksform schreiben lässt, also ein DAG vorliegt, ist die Lösung einfach bestimmbar. Im Allgemeinen ist das nichtlineare simultane Gleichungssystem aber mit einem Optimierungsverfahren zu lösen.

Die Effekte können überprüft werden, indem man die Pfadeffekte als Produkt über alle partiellen Ableitungen eines Pfades berechnet und über alle möglichen Pfade aufaddiert. So beträgt der Effekt von  $x_1$  auf  $y_3$  über die beiden Pfade  $x_1 \rightarrow y_1 \rightarrow y_3$  und  $x_1 \rightarrow y_1 \rightarrow y_2 \rightarrow y_3$ :  $1 * 1 + 1 * 12 * 1 = 13$ . Der totale Graph ist im Allgemeinen ein kompletter Graph und hat somit auch Kanten, wo die partiellen Ableitungen null sind. Für die finalen Effekte wird insbesondere die letzte Zeile von  $E_y$  benötigt, die zur „finalen“ Variable  $y_3$  gehört. Im finalen Graph für  $y_3$  wird deutlich, dass sich ein Knoten in die Summe seiner ausgehenden Kanten zerlegt, jedoch im Allgemeinen nicht der Summe seiner eingehenden Kanten entspricht.

Modellgleichungen:

Exogene Variablen:

$$y_1 = x_1$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$y_2 = 2y_1^2 + x_2$$

$$y_3 = y_1 + y_2$$

Ableitungen:

Lösung:

$$M_x = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 20 \\ 23 \end{bmatrix}$$

$$M_y = \begin{bmatrix} 0 & 0 & 0 \\ 4y_1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 12 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

Weder die Linearität noch die Homogenität ist gegeben, da die Variable  $y_1$  in der Matrix  $M_y$  vorkommt. Das System ist identifiziert, da die Hessesche Matrix vollen Rang hat. Die Identifikation erkennt man auch daran, dass man die 5 partiellen Ableitungen für die Nicht-Null-Elemente in  $M_y$  und  $M_x$  aus den 5 nicht-trivialen Gleichungen für die exogenen totalen Ableitungen  $E_x$  berechnen kann. Das System global identifiziert, da die Anzahl der Variablen und Gleichungen übereinstimmen und eine eindeutige Lösung vorliegt. Siehe auch Bollen und Bauldry (2010), die Computer-Algebra-Systeme zur Identifikation komplexerer Systeme vorschlagen:

$$E_x = (I_n - M_y)^{-1} M_x$$

$$\begin{aligned} \begin{bmatrix} 1 & 0 \\ 12 & 1 \\ 13 & 1 \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 \\ -m_{y21} & 1 & 0 \\ -m_{y31} & -m_{y32} & 1 \end{bmatrix}^{-1} \begin{bmatrix} m_{x11} & 0 \\ 0 & m_{x22} \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ m_{y21} & 1 & 0 \\ m_{y21}m_{y32} + m_{y31} & m_{y32} & 1 \end{bmatrix} \begin{bmatrix} m_{x11} & 0 \\ 0 & m_{x22} \\ 0 & 0 \end{bmatrix} \end{aligned}$$

$$\Rightarrow 1 = m_{x11}, 1 = m_{x22} \text{ und } 12 = m_{y21}m_{x11}, 13 = m_{y21}m_{y32}m_{x11} + m_{y31}m_{x11}, 1 = m_{y32}m_{x22}$$

$$\Rightarrow 12 = m_{y21}, 1 = m_{y32}, 1 = m_{y31}$$

Effekte:

$$E_x = (I_n - M_y)^{-1} M_x = \begin{bmatrix} 1 & 0 & 0 \\ 12 & 1 & 0 \\ 13 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 4y_1 & 1 \\ 4y_1 + 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 12 & 1 \\ 13 & 1 \end{bmatrix}$$

$$E_y = (I_n - M_y)^{-1} (I_n \circ (I_n - M_y)^{-1})^{-1} = I_3 + M_y + M_y^2$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 12 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 12 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 4y_1 & 1 & 0 \\ 4y_1 + 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 12 & 1 & 0 \\ 13 & 1 & 1 \end{bmatrix}$$

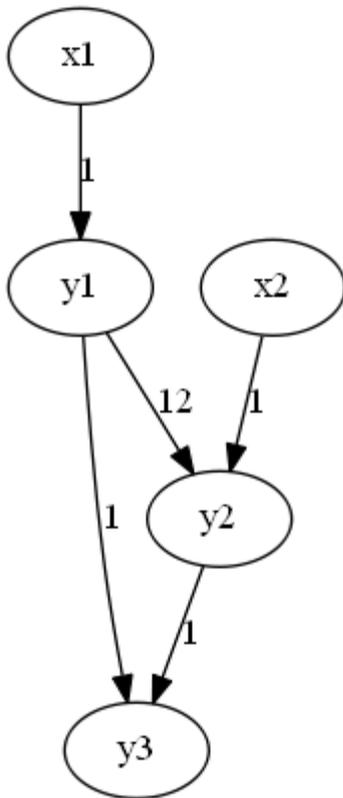
Finale Effekte:

$$E_{y_j x} = \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(m)} \right) \circ M_x = \begin{bmatrix} 13 & 13 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \circ \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 13 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

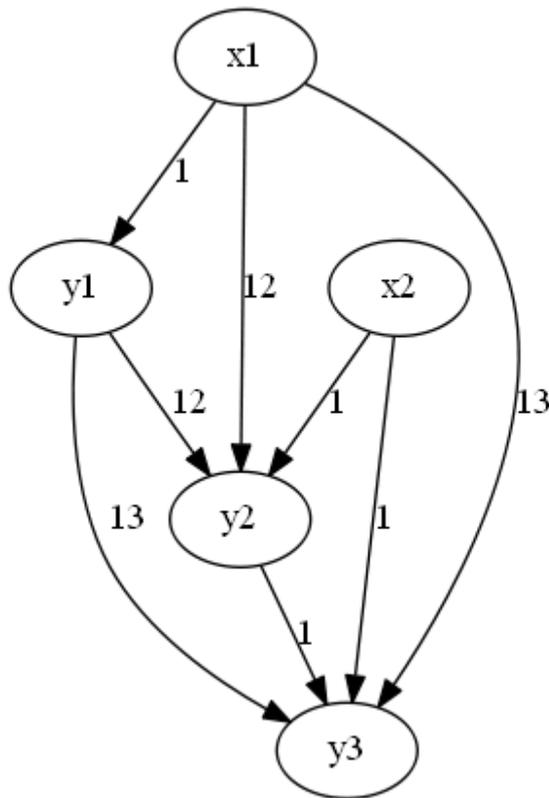
$$\begin{aligned} E_{y_j y} &= (1_{nn} - I_n) \circ \left[ \left( (I_n - M_{y11})^{-1} \right)_{(j)}^T, \dots, \left( (I_n - M_{y_{nn}})^{-1} \right)_{(j)}^T \right] \circ M_y \\ &= \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(n)} \right) \circ M_y \end{aligned}$$

$$= \begin{bmatrix} 13 & 13 & 13 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \circ \begin{bmatrix} 0 & 0 & 0 \\ 12 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 12 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

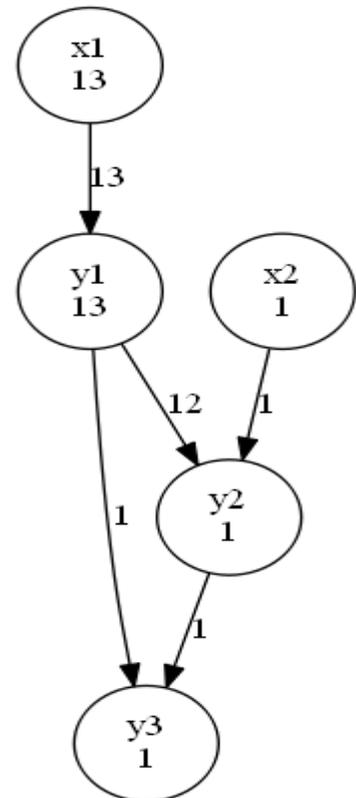
Partieller Graph:



Totaler Graph:



Finaler Graph  $y_3$ :



## Beispiel 2 Unabhängige Teilsysteme

In diesem Beispiel besteht das DAG-System aus zwei unabhängigen DAG-Teilgraphen. Man sieht, dass die Effekte daher unabhängig voneinander für jeden Teilgraphen einzeln bestimmt werden können. Der finale Graph für  $y_3$  wird nicht durch die Variablen des anderen Teilgraphen beeinflusst.

Modellgleichungen:

$$y_1 = 2x_1, \quad y_4 = x_2$$

$$y_2 = 3y_1, \quad y_5 = 4y_4$$

$$y_3 = y_2$$

Ableitungen:

Exogene Variablen:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Lösung:

$$M_x = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_4 \\ y_2 \\ y_5 \\ y_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 6 \\ 4 \\ 6 \end{bmatrix}$$

$$M_y = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Das System ist identifiziert, da die Hessesche Matrix vollen Rang hat.

Effekte:

$$E_x = (I_n - M_y)^{-1} M_x = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 \\ 0 & 4 & 0 & 1 & 0 \\ 3 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 6 & 0 \\ 0 & 4 \\ 6 & 0 \end{bmatrix}$$

$$E_y = (I_n - M_y)^{-1} (I_n \circ (I_n - M_y)^{-1})^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 \\ 0 & 4 & 0 & 1 & 0 \\ 3 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Finale Effekte:

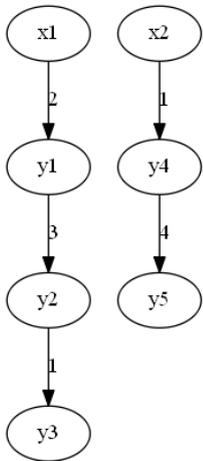
$$E_{y_j x} = \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(m)} \right) \circ M_x = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 4 & 4 \\ 1 & 1 \end{bmatrix} \circ \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 6 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$E_{y_j y} = (1_{nn} - I_n) \circ \left[ \left( (I_n - M_{y_{11}})^{-1} \right)_{(j)}^T, \dots, \left( (I_n - M_{y_{nn}})^{-1} \right)_{(j)}^T \right] \circ M_y$$

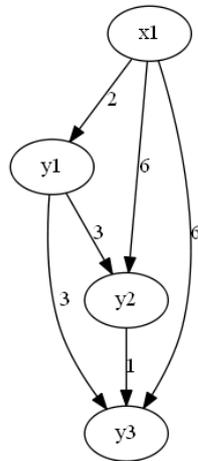
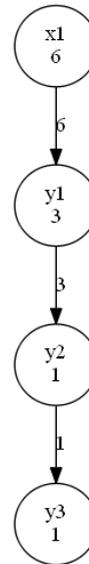
$$= \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(n)} \right) \circ M_y$$

$$= \begin{bmatrix} 3 & 3 & 3 & 3 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \circ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Partieller Graph:



Totaler Graph:

Finaler Graph  $y_3$ :

### Beispiel 3 Fehlende Bilanzgleichung

In der Praxis wenden wir die Effekte auf Modelle an, die Bilanzdaten von Unternehmen als exogene Inputgrößen verwenden. Eine Bilanz ist per Konstruktion dadurch gekennzeichnet, dass die Summe aller Aktiva der Summe aller Passiva entspricht. Wir untersuchen nun, wie diese Bilanzgleichung aufzunehmen ist. Die Zielvariable  $y_1$  hängt, hier vereinfachend als Summe, von den Aktiva,  $x_1$ , und den Passiva,  $x_2$ , ab. Eine Änderung der Aktiva bzw. Passiva wirkt in dieser Modellierung eins zu eins auf die Zielgröße. Durch eine separate Änderung einer Bilanzseite wird jedoch die Bilanzgleichung, Aktiva = Passiva, bzw.  $x_1 = x_2$  verletzt: der Effekt von  $x_1$  auf  $y_1$  beträgt nur 1, anstatt der gewünschten 2; gleiches gilt für den Effekt von  $x_2$  auf  $y_1$ . Daher betrachten wir Erweiterungen in den nachfolgenden Beispielen 4 und 5.

Modellgleichung:

$$y_1 = x_1 + x_2$$

Ableitungen:

$$M_x = [1 \quad 1]$$

$$M_y = 0$$

Exogene Variablen:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Lösung:

$$y_1 = 2$$

Das System ist identifiziert, da die Hessesche Matrix vollen Rang hat.

Effekte:

$$E_x = (I_n - M_y)^{-1} M_x = 1[1 \quad 1] = [1 \quad 1]$$

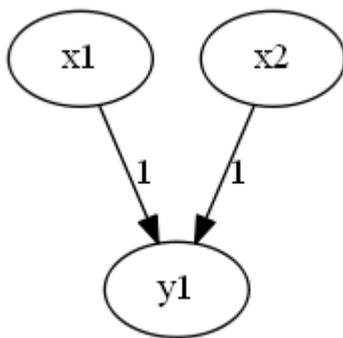
$$E_y = (I_n - M_y)^{-1} (I_n \circ (I_n - M_y)^{-1})^{-1} = 1$$

Finale Effekte:

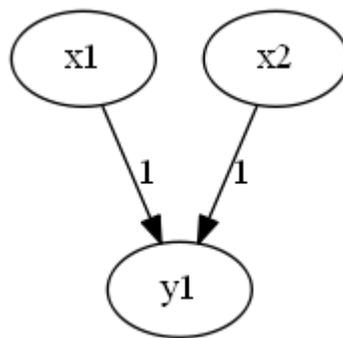
$$E_{y_j x} = \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(m)} \right) \circ M_x = [1 \quad 1] \circ [1 \quad 1] = [1 \quad 1]$$

$$\begin{aligned} E_{y_j y} &= (1_{nn} - I_n) \circ \left[ \left( (I_n - M_{y_{11}})^{-1} \right)_{(j)}^T, \dots, \left( (I_n - M_{y_{nn}})^{-1} \right)_{(j)}^T \right] \circ M_y \\ &= \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(n)} \right) \circ M_y \\ &= 1 * 0 = 0 \end{aligned}$$

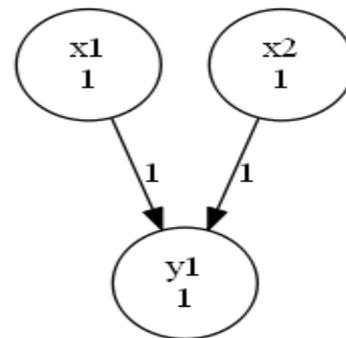
Partieller Graph:



Totaler Graph:



Finaler Graph  $y_1$ :



#### Beispiel 4 Asymmetrische Bilanzgleichung

Die Lösung des Problems der Bilanzdarstellung in Beispiel 3 kann jedoch auch nicht in folgendem System bestehen:  $y_1 = x_1$ ,  $y_2 = x_2$ ,  $y_1 = y_2$ , denn dann ist das System mit nur zwei endogenen Variablen aber drei Modellgleichungen überbestimmt. Die Variable  $y_1$  hat zwei, sich im Allgemeinen widersprechende, Bestimmungsgleichungen.

Wir nehmen nun die Bilanzgleichung in das Modell auf. Sie lautet hier  $x_1 = y_1$ , da die Passiva  $y_1$  nun eine endogene Variable sind. Die Aktiva  $x_1$  haben nun den korrekten Effekt in Höhe von 2 auf die Zielvariable. Dieser setzt sich aus einem direkten Effekt und einem indirekten Effekt über die Passiva, jeweils in Höhe von 1, zusammen. Graphisch ergibt sich dies, indem man auf jedem Pfad, der von den Aktiva zur Zielgröße führt, alle partiellen Ableitungen aufmultipliziert und dann die Summe über dies Pfadefekte bildet. Der Effekt der Passiva beträgt jedoch weiterhin nur eins. Da die Bilanzgleichung nicht in die Aktiva eingesetzt wurde, gibt es keinen Effekt von den Passiva auf die Aktiva. Die Bilanzgleichung wirkt asymmetrisch, sie ist nur bei Änderungen der Aktiva erfüllt. Die Lösung wird in nachfolgendem Beispiel 5 dargestellt.

Modellgleichungen:

Exogene Variablen:

$$y_1 = x_1$$

$$x_1 = 1$$

$$y_2 = x_1 + y_1$$

Ableitungen:

Lösung:

$$M_x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$M_y = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

Das System ist identifiziert, da die Hessesche Matrix vollen Rang hat.

Effekte:

$$E_x = (I_n - M_y)^{-1} M_x = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$E_y = (I_n - M_y)^{-1} (I_n \circ (I_n - M_y)^{-1})^{-1} = I_2 + M_y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

Finale Effekte:

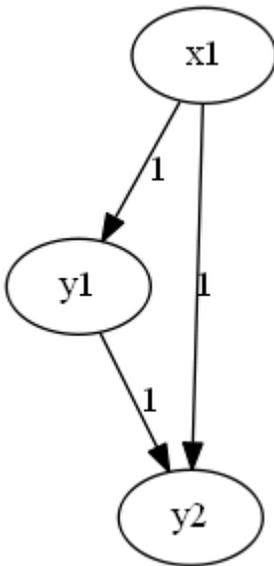
$$E_{y_j x} = \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(m)} \right) \circ M_x = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \circ \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$E_{y_j y} = (1_{nn} - I_n) \circ \left[ \left( (I_n - M_{y_{11}})^{-1} \right)_{(j)}^T, \dots, \left( (I_n - M_{y_{nn}})^{-1} \right)_{(j)}^T \right] \circ M_y$$

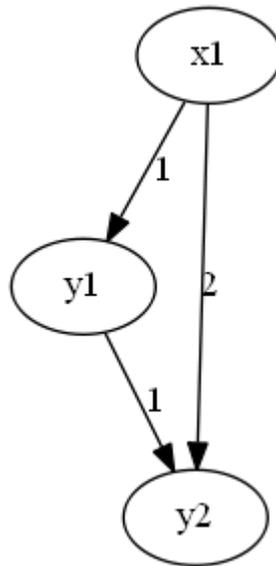
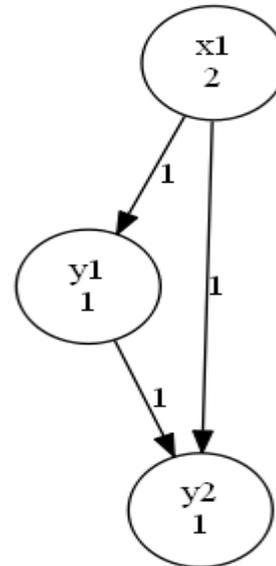
$$= \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(n)} \right) \circ M_y$$

$$= \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \circ \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

Partieller Graph:



Totaler Graph:

Finaler Graph  $y_2$ :

### Beispiel 5 Symmetrische Bilanzänderungen

Wir nehmen die Bilanzgleichung nun symmetrisch in das Modell auf. Die Aktiva  $y_1$  und die Passiva  $y_2$  sind nun endogene Variablen. Sie setzen sich jeweils zusammen aus ihren exogen gegebenen Ursprungswerten  $x_1$  und  $x_4$  und ihren ebenfalls exogen gegebenen Änderungen  $x_2$  und  $x_3$ . Diese wirken symmetrisch auf beide Bilanzseiten, wodurch die Bilanzidentität gewährleistet ist. Ihr Effekt auf die Zielvariable beträgt in diesem Beispiel, wie beabsichtigt, jeweils 2. Hingegen beträgt der Effekt der Aktiva  $y_1$  und der Passiva  $y_2$  auf die Zielgröße  $y_3$  nur eins, da es hier an der symmetrischen Wirkung der Bilanzgleichung fehlt.

Nach diesem Ansatz sind also alle Bilanzpositionen zu verdreifachen, aufgeteilt in den exogenen Ausgangswert, die exogene Änderung und den endogenen Gesamtwert. Die Änderungen sind insofern nur Dummyvariablen. Sie werden auf null gesetzt und haben keinen Einfluss auf die Lösung des Systems, aber ihre Existenz ermöglicht die Bestimmung der korrekten kausalen Effekte. Die Effekte der exogenen Ausgangswerte und der endogenen Gesamtwerte dürfen jedoch *nicht* interpretiert werden, da sie nicht die Bilanzgleichung erfüllen. Bei einer feineren Aufgliederung der Bilanzstruktur muss zudem *inhaltlich* in der Modellierung entschieden werden, wie genau die Bilanzgleichung erfüllt wird. So kann die Erhöhung einer Passivposition durch eine entsprechende Erhöhung von Aktivpositionen (Bilanzverlängerung) oder durch die Reduktion von anderen Passivpositionen (Passivtausch) erfolgen. Ferner können neben der Bilanzgleichung auch Gleichungen zur Gewinn- und Verlustrechnung vorliegen, die entsprechend abgebildet werden müssten.

Die symmetrische Abbildung der Bilanzgleichung über zusätzliche Veränderungsvariablen hat den Vorteil, dass keine Zyklen aufgenommen werden. Allerdings erhöht sich die Zahl der Variablen deutlich und es sind zahlreiche inhaltliche Annahmen zu treffen, auf welche Bilanzpositionen sich die Änderungen auswirken. Es kann auch sein, dass die exakte Abbildung keine große Wirkung auf die Effekte hat. Eine Alternative hierzu besteht daher darin, eine Stratifizierung durchzuführen oder das verwandte Propensity Score Matching durchzuführen, um aus den beobachteten Daten eine

möglichst gut randomisierte kontrollierte Studie zu konstruieren. Siehe [https://en.wikipedia.org/wiki/Propensity\\_score\\_matching](https://en.wikipedia.org/wiki/Propensity_score_matching). Dann können möglichst ähnliche Unternehmen – abgesehen von der interessierenden Effekt-Ausgangsvariable – gebündelt werden und auf dieser Basis der konsistente, real beobachtete Effekt numerisch bestimmt werden.

Der finale Graph ist zudem ein Beispiel für einen non-conserving Flow: für die endogenen Variablen  $y_1$  und  $y_2$  beträgt die Summe der eingehenden finalen Effekte drei, während der ausgehende finale Effekt nur eins beträgt.

Modellgleichungen:

Exogene Variablen:

$$y_1 = x_1 + x_2 + x_3$$

$$y_2 = x_2 + x_3 + x_4$$

$$y_3 = y_1 + y_2$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Ableitungen:

Lösung:

$$M_x = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

$$M_y = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

Das System ist identifiziert, da die Hessesche Matrix vollen Rang hat.

Effekte:

$$E_x = (I_n - M_y)^{-1} M_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 \end{bmatrix}$$

$$E_y = (I_n - M_y)^{-1} (I_n \circ (I_n - M_y)^{-1})^{-1} = I_3 + M_y + M_y^2$$

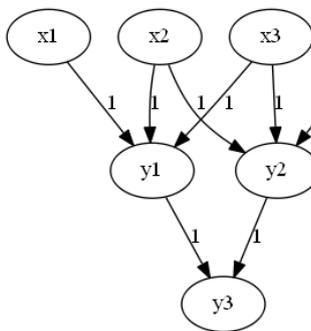
$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Finale Effekte:

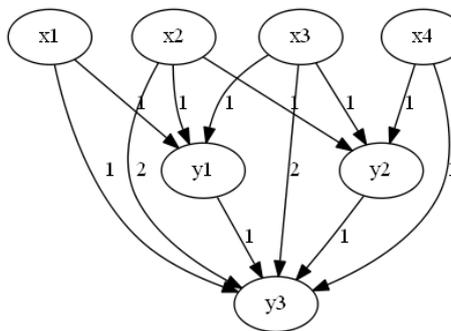
$$E_{y_jx} = \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(m)} \right) \circ M_x = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \circ \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{aligned}
E_{y|y} &= (1_{nn} - I_n) \circ \left[ \left( (I_n - M_{y11})^{-1} \right)_{(j)}^T, \dots, \left( (I_n - M_{y_{nn}})^{-1} \right)_{(j)}^T \right] \circ M_y \\
&= \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(n)} \right) \circ M_y \\
&= \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \circ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}
\end{aligned}$$

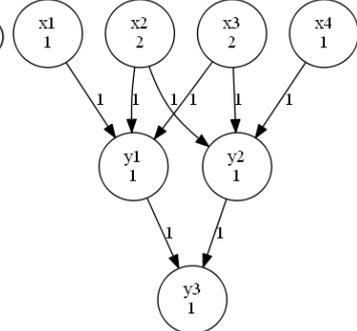
Partieller Graph:



Totaler Graph:



Finaler Graph  $y_3$ :



### Beispiel 6 Zyklisches System

Wir betrachten ein System mit drei endogenen Variablen, drei Modellgleichungen und einer exogenen Variablen:  $y_2 = y_1$ ,  $y_3 = x_1 - y_1$ ,  $y_2 = y_3$ . Das System kann ökonomisch interpretiert den:  $y_1$  ist der Preis eines Gutes,  $y_2$  ist die Angebotsfunktion und  $y_3$  die Nachfragefunktion. Das Angebot steigt mit dem Preis und die Nachfrage sinkt mit dem Preis, wobei  $x_1$  das exogen gegebene Nachfragepotential bei einem Preis von  $y_1 = 0$  ist. Die Lösung des Systems ist der Preis, der Angebot und Nachfrage ausgleicht, also das Marktgleichgewicht darstellt. Diese Angebots-Nachfrage-Systeme sind in der Ökonomie sehr verbreitet. Wir schreiben das System so um, dass jeweils genau eine Variable auf der linken Seite steht. Zur Unabhängigkeit von der Darstellung siehe auch Beispiel 7.

Modellgleichungen:

Exogene Variable:

$$y_1 = y_2$$

$$x_1 = 2$$

$$y_2 = y_3$$

$$y_3 = x_1 - y_1$$

Dieses System kann nicht in Dreiecksform geschrieben werden, es ist ein zyklisches System, es ist kein DAG. Seine Lösung findet man zum Beispiel durch Einsetzen der zweiten und dritten Gleichung in die erste, so dass  $y_1 = x_1 - y_1$  und somit  $y_1 = x_1/2$  und schließlich auch  $y_2 = y_3 = x_1/2$ . Da das System linear ist, kann es mittels Matrixinversion gelöst werden, wobei sich die Inverse im

zyklischen Fall nicht mehr als endliches Polynom darstellen lässt. Die Lösung des linearen Systems in Matrixschreibweise lautet:

$$\mathbf{y} = M_y \mathbf{y} + M_x \mathbf{x}$$

$$\Rightarrow \mathbf{y} = (I_n - M_y)^{-1} M_x \mathbf{x} = E_x \mathbf{x},$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} x_1$$

$$\Rightarrow \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 & 0.5 \\ -0.5 & 0.5 & 0.5 \\ -0.5 & -0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} * 2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Der exogene Effekt von  $x_1$  auf die  $y_j$  kann im zyklischen Fall nicht mehr einfach über die Pfadefekte nachvollzogen werden. Er ist aber aus obiger Beziehung  $y_3 = y_2 = y_1 = x_1/2$  herleitbar:  $dy_j/dx_1 = 1/2$ . Hier ist zu beachten, dass wir die Ableitungssymbole  $d$  für die Ableitung totale Ableitung des reduzierten Systems verwenden, obwohl wir mathematisch die partielle Ableitung  $\partial$  nach einer Variablen gebildet haben. Siehe hierzu die Erläuterungen in Abschnitt 5.1. Dieses Ergebnis bestätigt, dass die Formel für  $E_x$  auch im zyklischen Fall richtig ist.

Für die endogenen Effekte wurde jedoch eine allgemeine Formel hergeleitet, die sich vom derjenigen für den DAG-Fall unterscheidet. Wir verwenden nachfolgend sowohl die DAG-Formel als auch die allgemeine Formel für die endogenen Effekte, um zu zeigen, welchen Fehler man mit der Anwendung der einfacheren DAG-Formel macht, obwohl es sich um einen zyklischen Graphen handelt.

Bei folgenden Größen sind die Formeln für den allgemeinen Fall und den DAG-Fall gleich:

Ableitungen:

Lösung:

$$M_x = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$M_y = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \end{bmatrix}$$

Das System ist identifiziert, da die Hessesche Matrix vollen Rang hat.

Exogene Effekte:

$$E_x = (I_n - M_y)^{-1} M_x = \begin{bmatrix} 0.5 & 0.5 & 0.5 \\ -0.5 & 0.5 & 0.5 \\ -0.5 & -0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \end{bmatrix}$$

$$E_y = (I_n - M_y)^{-1} (I_n \circ (I_n - M_y)^{-1})^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ -1 & -1 & 1 \end{bmatrix}$$

Finale Effekte:

$$E_{y_j x} = \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(m)} \right) \circ M_x = \begin{bmatrix} -0.5 \\ -0.5 \\ 0.5 \end{bmatrix} \circ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0.5 \end{bmatrix}$$

$$E_{y_j y} = (1_{nn} - I_n) \circ \left[ \left( (I_n - M_{y_{11}})^{-1} \right)_{(j)}^T, \dots, \left( (I_n - M_{y_{nn}})^{-1} \right)_{(j)}^T \right] \circ M_y$$

$$= \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \circ \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \circ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}$$

Die endogenen Effekte unter *falscher* Verwendung der Formeln für den DAG-Fall sind:

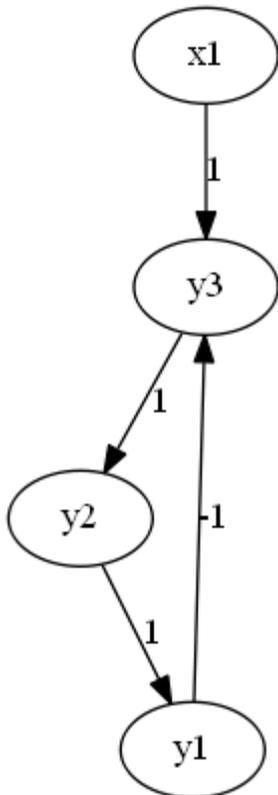
$$E_y = (I_n - M_y)^{-1} = \begin{bmatrix} 0.5 & 0.5 & 0.5 \\ -0.5 & 0.5 & 0.5 \\ -0.5 & -0.5 & 0.5 \end{bmatrix}$$

Die endogenen Effekte unter korrekter Verwendung der allgemeinen Formeln, also nach spaltenweiser Normierung der DAG-Lösung mit ihren Hauptdiagonalelementen, lauten:

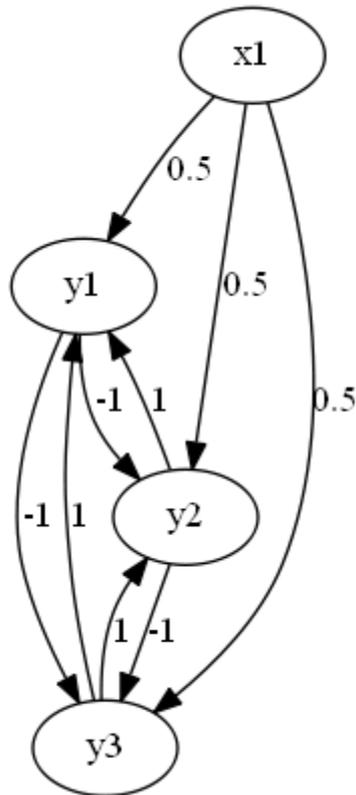
$$E_y = (I_n - M_y)^{-1} (I_n \circ (I_n - M_y)^{-1})^{-1} = \begin{bmatrix} 0.5 & 0.5 & 0.5 \\ -0.5 & 0.5 & 0.5 \\ -0.5 & -0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ -1 & -1 & 1 \end{bmatrix}$$

Die zyklischen Graphen mit den korrekten endogenen Effekten sind:

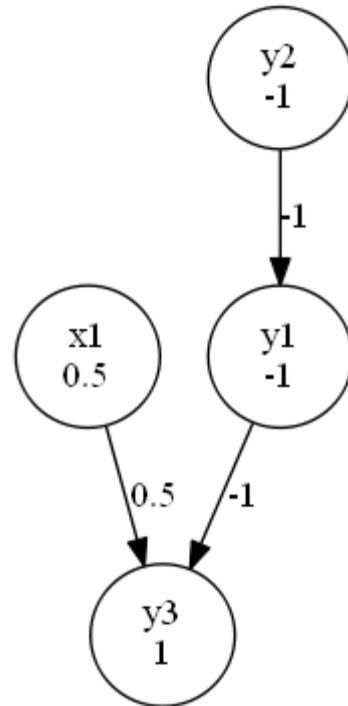
Partieller Graph:



Totaler Graph:



Finaler Graph  $y_3$ :



### Beispiel 7 Unabhängigkeit von der Darstellung

Wir zeigen nun dass die Effekte unabhängig sind von der Art der Darstellung der partiellen Ableitungen des Systems. Hierzu verwenden wir das zyklische Beispiel, wobei  $M_y$  ohne Normierung Nicht-Null-Elemente auf der Hauptdiagonale hätte:

Modellgleichungen:

Exogene Variablen:

$$y_1 = 3y_1 + y_2 + y_3$$

$$x_1 = 1$$

$$y_2 = y_1 + 2y_2 + y_3$$

$$y_3 = y_1 + 2y_2 + 2y_3 + x_1$$

Wir schreiben dieses System so, dass die endogenen Variablen nur auf der linken Seite der Modellgleichungen vorkommen. Beide Systeme führen zu den gleichen Effekten:

Modellgleichungen:

$$y_1 = -y_2/2 - y_3/2$$

$$y_2 = -y_1 - y_3$$

$$y_3 = -y_1 - 2y_2 - x_1$$

Ableitungen:

$$M_x = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$$

Lösung:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

$$M_y = \begin{bmatrix} 0 & -1/2 & -1/2 \\ -1 & 0 & -1 \\ -1 & -2 & 0 \end{bmatrix}$$

Das System ist identifiziert, da die Hessesche Matrix vollen Rang hat.

Effekte:

$$E_x = (I_n - M_y)^{-1} M_x = \begin{bmatrix} 2 & -1 & 0 \\ 0 & -1 & 1 \\ -2 & 3 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

$$E_y = (I_n - M_y)^{-1} (I_n \circ (I_n - M_y)^{-1})^{-1} = \begin{bmatrix} 2 & -1 & 0 \\ 0 & -1 & 1 \\ -2 & 3 & -1 \end{bmatrix} \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ -1 & -3 & 1 \end{bmatrix}$$

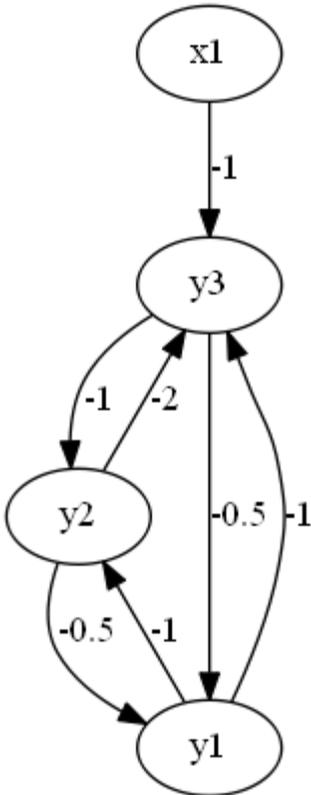
Finale Effekte:

$$E_{y_j x} = \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(m)} \right) \circ M_x = \begin{bmatrix} -2 \\ 3 \\ -1 \end{bmatrix} \circ \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

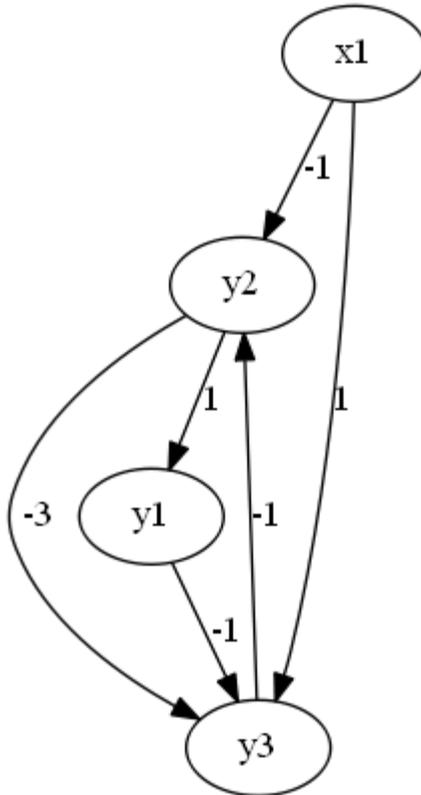
$$E_{y_j y} = (1_{nn} - I_n) \circ \left[ \left( (I_n - M_{y_{11}})^{-1} \right)_{(j)}^T, \dots, \left( (I_n - M_{y_{nn}})^{-1} \right)_{(j)}^T \right] \circ M_y$$

$$= \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \circ \begin{bmatrix} 0 & -2 & 0 \\ 2 & 0 & 0 \\ -1 & 2 & 1 \end{bmatrix} \circ \begin{bmatrix} 0 & -1/2 & -1/2 \\ -1 & 0 & -1 \\ -1 & -2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -2 & 0 & 0 \\ 1 & -4 & 0 \end{bmatrix}$$

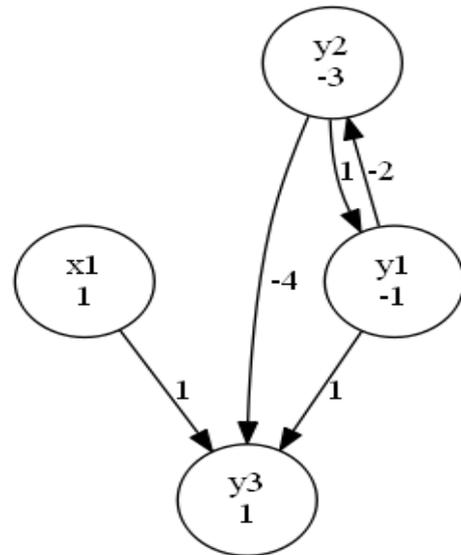
Partieller Graph:



Totaler Graph:



Finaler Graph  $y_3$ :



Dieses Beispiel der Unabhängigkeit von der Darstellung kann noch weiter verallgemeinert werden. So kann man von dem allgemeinen Fall ausgehen, dass die Modellgleichungen als homogene Gleichungen gegeben sind. Es sind also alle Terme der Modellgleichungen auf eine Seite gebracht. In diesem Fall also:  $0 = 2y_1 + y_2 + y_3$ ,  $0 = y_1 + y_2 + y_3$  und  $0 = y_1 + 2y_2 + y_3 + x_1$ . Dies ist eine allgemeine Darstellung, in die jede Modellgleichung gebracht werden kann. Gemäß unserer Annahmen muss das System in die Form  $y_j = M^j(\mathbf{y}, \mathbf{x})$  gebracht werden, wobei  $y_j$  auch auf der rechten Seite auftauchen kann. Wir addieren also einfach für  $j = 1, \dots, n$  den Term  $y_j$  auf beiden Seiten der  $j$ -ten Modellgleichung – eine Erweiterung, die die inhaltliche Bedeutung des Systems nicht ändert. Die Reihenfolge der Gleichungen ist beliebig. Durch diese Anpassung liegt genau die Formulierung des Beispiels 7 vor und somit erhält man die gleiche Lösung. Die Art der Darstellung ist also egal, wenn die Gleichungen zuerst in die homogene Form gebracht werden und dann die jeweils  $j$ -te Gleichung um  $y_j$  erweitert wird.

### Beispiel 8 Stark nichtlineares System

Wir betrachten ein stark nichtlineares Modell. Anders als in Beispiel 1, sind auch die partiellen Ableitungen noch nichtlinear. Zudem ist das System zyklisch und alle endogenen Variablen sind miteinander verbunden. Die numerischen Werte sind zum Teil gerundet angegeben.

Modellgleichungen:

Exogene Variablen:

$$y_1 = y_2^3 y_3 + 2y_3 + \frac{1}{10} x_1$$

$$y_2 = \frac{1}{10} y_1 + \frac{1}{10} \ln\left(y_1^2 + \frac{1}{2}\right) + \frac{1}{10} y_3 + \frac{1}{10} x_2$$

$$y_3 = \frac{1}{10} y_1 + \frac{1}{4} y_2 + \frac{1}{5} x_3$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Ableitungen:

$$M_x = \begin{bmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{10} & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix}$$

Lösung:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \approx \begin{bmatrix} 0.77 \\ 0.22 \\ 0.33 \end{bmatrix}$$

$$M_y = \begin{bmatrix} 0 & 0.05 & 2.01 \\ 0.24 & 0 & \frac{1}{10} \\ \frac{1}{10} & \frac{1}{4} & 0 \end{bmatrix}$$

Das System ist identifiziert, da die Hessesche Matrix vollen Rang hat.

Effekte:

$$E_x = (I_n - M_y)^{-1} M_x = \begin{bmatrix} 0.15 & 0.09 & 0.63 \\ 0.04 & 0.12 & 0.18 \\ 0.02 & 0.04 & 0.31 \end{bmatrix}$$

$$E_y = (I_n - M_y)^{-1} (I_n \circ (I_n - M_y)^{-1})^{-1} = \begin{bmatrix} 1 & 0.69 & 2.04 \\ 0.26 & 1 & 0.59 \\ 0.16 & 0.32 & 1 \end{bmatrix}$$

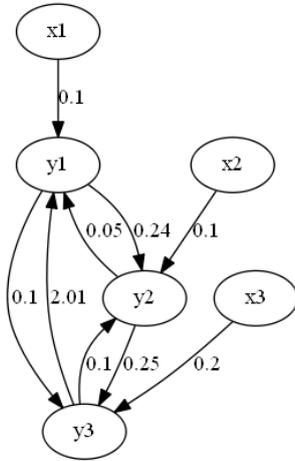
Finale Effekte:

$$E_{y_j x} = \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(m)} \right) \circ M_x = \begin{bmatrix} 0.03 & 0 & 0 \\ 0 & 0.04 & 0 \\ 0 & 0 & 0.31 \end{bmatrix}$$

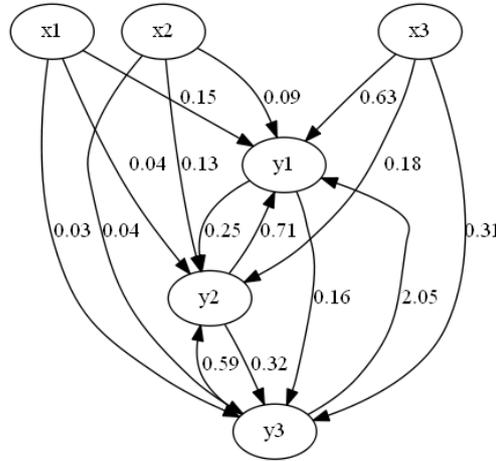
$$E_{y_j y} = (1_{nn} - I_n) \circ \left[ \left( (I_n - M_{y_{11}})^{-1} \right)_{(j)}^T, \dots, \left( (I_n - M_{y_{nn}})^{-1} \right)_{(j)}^T \right] \circ M_y$$

$$= \begin{bmatrix} 0 & 0.01 & 0 \\ 0.06 & 0 & 0 \\ 0.10 & 0.31 & 0 \end{bmatrix}$$

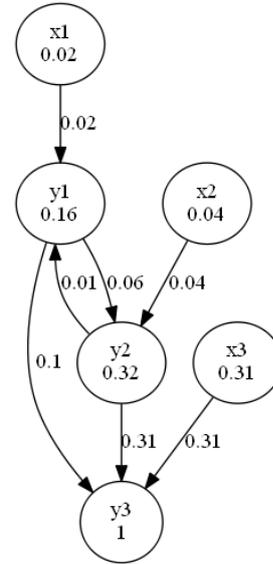
Partieller Graph:



Totaler Graph:



Finaler Graph  $y_3$ :



### Beispiel 9 Kausalität

In diesem Beispiel werden sowohl die Kausalität als auch Besonderheiten bei der Interpretation der endogenen Effekte erläutert.

In der Praxis kommt es vor, dass Variablen als beobachtbare exogene Variable vorliegen, die gemäß dem inhaltlichen Kausalitätsmodell gar nicht exogen, sondern endogen sind. Das sogenannte Bewertungsmodell, mit dem alle Größen des Modells bestimmt werden können, weicht also von Kausalitätsmodell ab.

Im nachfolgenden Beispiel steht EK für das Eigenkapital, dass sich als Summe aus Eigenkapital A und Eigenkapital B, also EKA und EKB ergibt. Die Assets A sind die Summe aus Eigenkapital und Liabilities L. Die Eigenkapitalquote  $q$  setzt das Eigenkapital B ins Verhältnis zu den Assets.

In diesem Beispiel seien die Größen EK, EKA und A als beobachtbar gegeben. Sie müssen also als exogen behandelt werden, um die anderen Größen daraus abzuleiten. Das Bewertungsmodell der Gleichungen ist – im Gegensatz zu dem nachfolgend dargestellten Kausalitätsmodell – „rückwärts“ konstruiert. So werden EKB und L als Residualgrößen bestimmt:

Modellgleichungen Bewertungsmodell

$$\begin{aligned} \text{EKB} &= \text{EK} - \text{EKA} \\ \text{L} &= \text{A} - \text{EK} \\ q &= \text{EKB}/\text{A} \end{aligned}$$

Ableitungen:

Exogene Variablen:

$$\mathbf{x} = \begin{bmatrix} \text{EK} \\ \text{EKA} \\ \text{A} \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 100 \end{bmatrix}$$

Lösung:

$$M_x = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & -\frac{EKB}{A^2} \end{bmatrix} \quad y = \begin{bmatrix} EKB \\ L \\ q \end{bmatrix} = \begin{bmatrix} 1 \\ 98 \\ 0.01 \end{bmatrix}$$

$$M_y = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{1}{A} & 0 & 0 \end{bmatrix}$$

Das System ist identifiziert, da die Hessesche Matrix vollen Rang hat.

Effekte:

$$E_x = (I_n - M_y)^{-1} M_x = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 0 & 1 \\ 0.01 & -0.01 & -0.0001 \end{bmatrix}$$

$$E_y = (I_n - M_y)^{-1} (I_n \circ (I_n - M_y)^{-1})^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.01 & 0 & 1 \end{bmatrix}$$

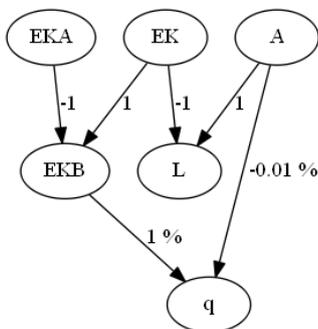
Finale Effekte:

$$E_{y_j x} = \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(m)} \right) \circ M_x = \begin{bmatrix} 0.01 & -0.01 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -0.0001 \end{bmatrix}$$

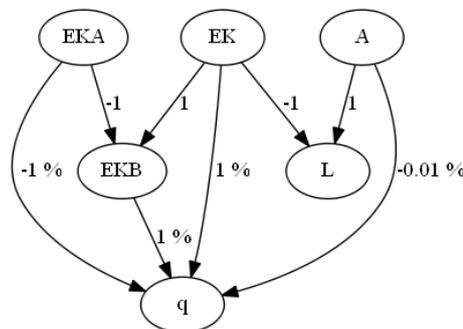
$$E_{y_j y} = (1_{nn} - I_n) \circ \left[ \left( (I_n - M_{y_{11}})^{-1} \right)_{(j)}^T, \dots, \left( (I_n - M_{y_{nn}})^{-1} \right)_{(j)}^T \right] \circ M_y$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.01 & 0 & 0 \end{bmatrix}$$

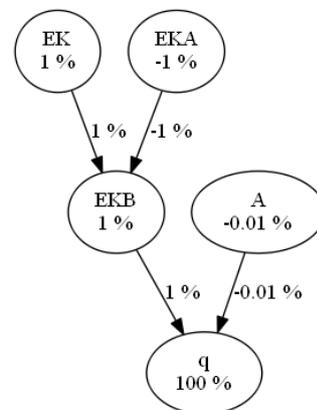
Partieller Graph:



Totaler Graph:



Finaler Graph q:



Man sieht die unerwünschte Kausalität anhand des EKA. Eine Erhöhung des EKA reduziert entsprechend das residual bestimmte EKB bei gegebenem EK und reduziert somit fälschlicherweise die EKB-Quote  $q$  entsprechend stark.

Um die gewünschten Kausalitäten zu erreichen, werden die residual bestimmten Größen im Kausalitätsmodell nun als exogene Variablen aufgefasst. EKB wird also exogen und EK wird endogen. Dies entspricht der Auflösung der ursprünglichen EKB-Gleichung nach EK. Im Modell verbleiben nur Gleichungen mit inhaltlich gewünschter Kausalität und jeder endogenen Variablen wird nur eine Kausalität zugewiesen. Dabei hat die mathematische Auflösung gar keinen Effekt, sondern dient lediglich der direkten Ablesbarkeit der Kausalität. Das Gleichheitszeichen  $=$  kann für die Zwecke der Kausalität als  $\leftarrow$  gelesen werden. Eine Erhöhung von EKA bzw. EKB erhöht also das EK entsprechend. Entscheidend für die mathematische Kodierung der Kausalität ist die Tatsache, dass dieser Gleichung die endogene Variable EK zugewiesen wird. Um diese Zuordnung direkt sichtbar zu machen, schreiben wir die Gleichungen so, dass die jeweils endogene Variable auf der linken Seite der Gleichung steht. So wird sichergestellt, dass die Kausalität korrekt in den Matrizen  $M_x$  und  $M_y$  der partiellen Ableitungen abgebildet wird. Für die Lösung des Gleichungssystems hingegen ist keine Zuordnung von endogenen Variablen und Gleichungen erforderlich. Die Definition der exogenen Variablen ändert sich durch die Auflösung der Gleichungen entsprechend. Die Exogenisierung der residual berechneten Variablen wird durch einen Zwischenschritt ermöglicht, indem das Modell simultan gelöst wird und so die gewünschten exogenen Variablen numerisch bestimmt werden.

Modellgleichungen Kausalitätsmodell:

$$\begin{aligned} EK &= EKA + EKB \\ A &= EK + L \\ q &= EKB/A \end{aligned}$$

Ableitungen:

$$M_x = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & \frac{1}{A} & 0 \end{bmatrix}$$

$$M_y = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -\frac{EKB}{A^2} & 0 \end{bmatrix}$$

Exogene Variablen:

$$\mathbf{x} = \begin{bmatrix} EKA \\ EKB \\ L \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 98 \end{bmatrix}$$

Lösung:

$$\mathbf{y} = \begin{bmatrix} EK \\ A \\ q \end{bmatrix} = \begin{bmatrix} 2 \\ 100 \\ 0.01 \end{bmatrix}$$

Das System ist identifiziert, da die Hessesche Matrix vollen Rang hat.

Effekte:

$$E_x = (I_n - M_y)^{-1} M_x = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ -0.0001 & 0.0099 & -0.0001 \end{bmatrix}$$

$$E_y = (I_n - M_y)^{-1} (I_n \circ (I_n - M_y)^{-1})^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -0.0001 & -0.0001 & 1 \end{bmatrix}$$

Finale Effekte:

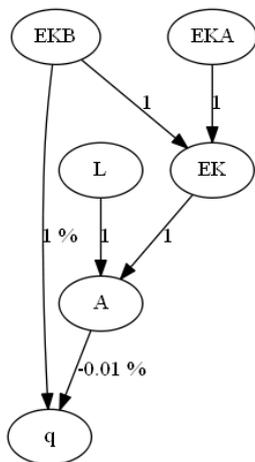
$$E_{y_j x} = \left( \left( (I_n - M_y)^{-1} \right)_{(j)}^T \mathbf{1}_{(m)} \right) \circ M_x = \begin{bmatrix} -0.0001 & -0.0001 & 0 \\ 0 & 0 & -0.0001 \\ 0 & 0.01 & 0 \end{bmatrix}$$

$$E_{y_j y} = (1_{nn} - I_n) \circ \left[ \left( (I_n - M_{y11})^{-1} \right)_{(j)}^T, \dots, \left( (I_n - M_{y1nn})^{-1} \right)_{(j)}^T \right] \circ M_y$$

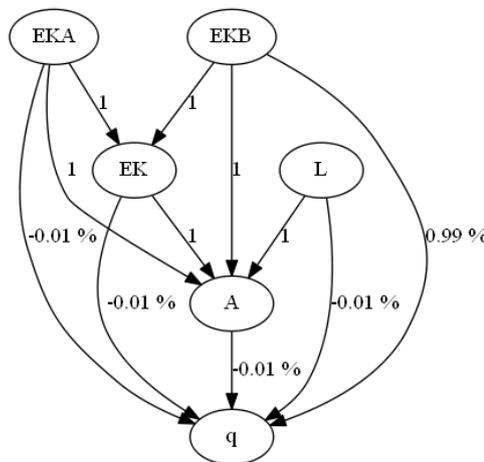
$$= \begin{bmatrix} 0 & 0 & 0 \\ -0.0001 & 0 & 0 \\ 0 & -0.0001 & 0 \end{bmatrix}$$

Mit gewünschter Kausalität

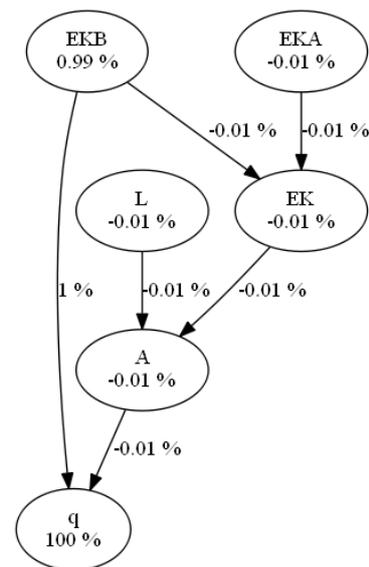
Partieller Graph:



Totaler Graph:



Finaler Graph q:



Man sieht, dass der starke negative Effekt des EKA auf die Quote q nun nicht mehr vorliegt. Nun ist der Effekt des EKA wie gewünscht leicht negativ, da er die Bilanzsumme erhöht, während das EKB unverändert bleibt.

Nachfolgende Überlegungen zeigen Besonderheiten bei der Interpretation der endogenen Effekte.

Ein für den Modellierer unerwünschter Effekt kann weiterhin jedoch darin liegen, dass bei dem Effekt einer endogenen Variablen nicht berücksichtigt wird, durch welche anderen Größen ihre Änderung bewirkt wurde. Dies ist ein Nachteil des do-Calculus, bei dem eine endogene Größe gesetzt wird und dabei eine Verletzung der Systemgleichungen in Kauf genommen wird. So führt in diesem Beispiel die Erhöhung des Eigenkapitals zu einer geringen Reduktion der EKB-Quote q, da sich die Assets als Bezugsgröße erhöhen. Wenn man entgegen der graphentheoretischen Effekt-Definition

vorangehende Variablen verantwortlich machen will, so bedeutet dies implizit, dass das EK nicht über das EKB, sondern über das EKA erhöht wurde. Ist die graphentheoretischen Interpretation nicht gewünscht, so sind die Ursprünge der Wirkungen zusätzlich über die modifizierten Matrizen  $\bar{M}_x$  und  $\bar{M}_y$  festzulegen. Um die gewünschte Interpretation zu erreichen, werden in diesem Beispiel die endogenen totalen Effekte und somit auch die finalen Effekte durch den Modellierer per Hand vorgegeben. Wir beheben den unerwünschten EK-Effekt durch folgende Spezifikation:

$$\bar{M}_x = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} / \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ -0.0001 & 0.0099 & -0.0001 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & \frac{A^2}{A - EKB} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 101.01 & 0 \end{bmatrix}$$

$$\bar{M}_y = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} / \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -0.0001 & -0.0001 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Dabei werden relative Zuordnungsmatrizen verwendet. Diese sind durch eine Zeilensumme von eins gekennzeichnet und werden stets noch elementweise durch die Effektmatrix  $E_x$  bzw.  $E_y$  dividiert. An den relativen Zuordnungsmatrizen wird deutlich, dass eine Änderung von EK nun auf eine gleich große Änderung von EKB zurückgeführt wird; eine Änderung von A wird auf L zurückgeführt und eine Änderung von q wird ebenfalls auf EBK zurückgeführt.

Die modifizierten endogenen Effekte ergeben sich, abweichend von der üblichen Formel, aus dem Produkt der unmodifizierten und modifizierten exogenen Effekte. Es erfolgt dann noch eine Normierung, so dass  $\bar{E}_y$  Einsen auf der Hauptdiagonalen hat. Die Normierung funktioniert nur, wenn die Hauptdiagonalelemente vorher ungleich null sind; eine notwendige Bedingung hierfür ist, dass  $E_x$  und  $\bar{E}_x$  keine Nullzeilen haben, jede endogene Variable muss also von mindestens einer exogenen Variablen abhängen. Schließlich werden noch die Effekte ausgenullt, die in  $E_y$  nicht vorhanden waren. So werden keine neuen Effekte eingeführt werden, bereits vorhandene Effekte werden lediglich skaliert. Man könnte im Sinne dieses Ansatzes insgesamt folgende Formeln für die modifizierten endogenen Effekte verwenden:

$$\begin{aligned} \bar{E}_y &= (E_x \bar{E}_x^T (I_n \circ E_x \bar{E}_x^T)^{-1}) \circ E D_y = (E_x \bar{E}_x^T) \circ E D_y \\ (n \times n) & \\ &= \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ -0.0001 & 0.0099 & -0.0001 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & \frac{1}{0.0099} \\ 0 & 1 & 0 \end{bmatrix} \circ \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \frac{1}{0.0099} \\ 1 & 1 & \frac{1}{0.0099} \\ 0.0099 & -0.0001 & 1 \end{bmatrix} \circ \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0.0099 & -0.0001 & 1 \end{bmatrix} \end{aligned}$$

mit

$$\bar{E}_x = (I_n - \bar{M}_y)^{-1} \bar{M}_x = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & \frac{A^2}{A - EKB} & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & \frac{1}{0.0099} & 0 \end{bmatrix}.$$

Die finale Variable q entspricht der dritten Zeile von  $\bar{E}_y$  bzw.  $E_y$ . Das relative Verhältnis der endogenen Effekte auf die finale Variable q beträgt mit elementweiser Division:

$$(\bar{E}_y/E_y)_{(3)} = \begin{bmatrix} 0.0099 \\ -0.0001 \\ 1 \end{bmatrix} / \begin{bmatrix} -0.0001 \\ -0.0001 \\ 1 \end{bmatrix} = \begin{bmatrix} -99 \\ 1 \\ 1 \end{bmatrix}.$$

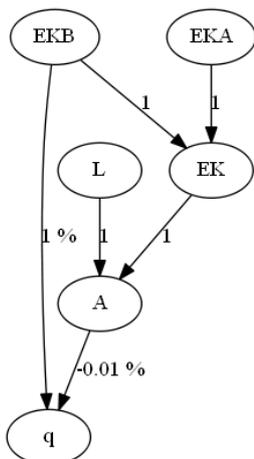
und die modifizierten endogenen finalen Effekte auf die finale Variable q lauten mittels spaltenweiser Multiplikation:

$$\begin{aligned} \bar{E}_{y_{3y}} = E_{y_{3y}} \text{diag}((\bar{E}_y/E_y)_{(3)}) &= \begin{bmatrix} 0 & 0 & 0 \\ -0.0001 & 0 & 0 \\ 0 & -0.0001 & 0 \end{bmatrix} \begin{bmatrix} -99 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ (n \times n) &= \begin{bmatrix} 0 & 0 & 0 \\ 0.0099 & 0 & 0 \\ 0 & -0.0001 & 0 \end{bmatrix}. \end{aligned}$$

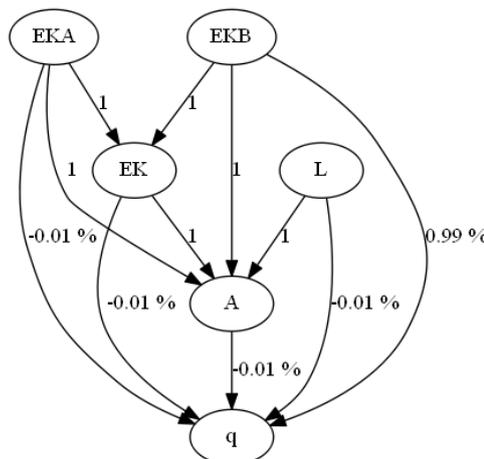
Es wird also lediglich die Wirkung von EK auf q modifiziert, und zwar um den Faktor -99. So erhöht sich der Effekt, wie gewünscht, von -0.01% auf 0.99%. Die modifizierten totalen und finalen Effekte werden nur im finalen Graphen angezeigt:

Mit modifizierten Effekten

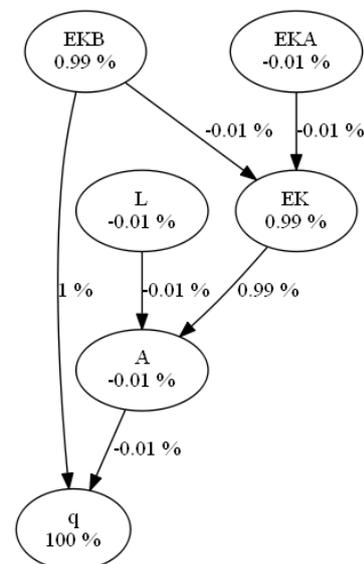
Partieller Graph:



Totaler Graph:



Finaler Graph q:



In dieser Darstellung ist sowohl gewünschte die Kausalität als auch deren gewünschte modifizierte Höhe erreicht. So reduziert das EKA die Quote q leicht und das EKB und EK erhöhen die Quote entsprechend stark. Auch die Zerlegung der totalen Effekte über die finalen Effekte funktioniert: So

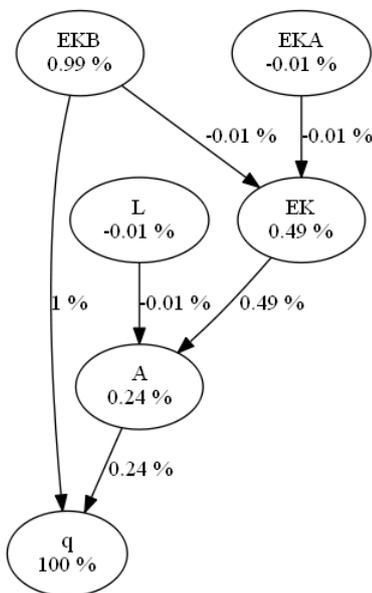
ist ersichtlich, dass eine Erhöhung des EKB um 1 die EKB-Quote  $q$  direkt um 1 Prozentpunkt erhöht. Jedoch gibt es über den indirekten Effekt der Erhöhung des EK und somit des Nenners der Quote einen leichten negativen Effekt in Höhe von 0.01 Prozentpunkten. Der modifizierte finale Effekt des EK beträgt damit 0,99 Prozentpunkte.

Es gibt weitere Alternativen, die relativen Zuordnungsmatrizen festzulegen. Man kann endogene kontrafaktische Änderungen auf die Vorgängervariablen zurückführen, letztendlich auf die exogenen Variablen. Dort führen Interventionen nicht zu Widersprüchen. Die Vorgänger können Confounder sein, also sowohl auf die Ausgangs- als auch auf die Zielvariable wirken. Die Modifikation wirkt dann besonders stark. Aber man erhält dann nicht mehr die Kausalität des Systems. Diese erhält man nur, wenn man die endogene Gleichung durch einen festen Wert ersetzt, denn so sind Effekte definiert. Verteilt man die Effekte proportional auf alle eingehenden Variablen, so erhält man:

$$\text{exogene relative Zuordnungsmatrix} = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{bmatrix}$$

$$\text{endogene relative Zuordnungsmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \end{bmatrix}$$

Finaler Graph  $q$ :



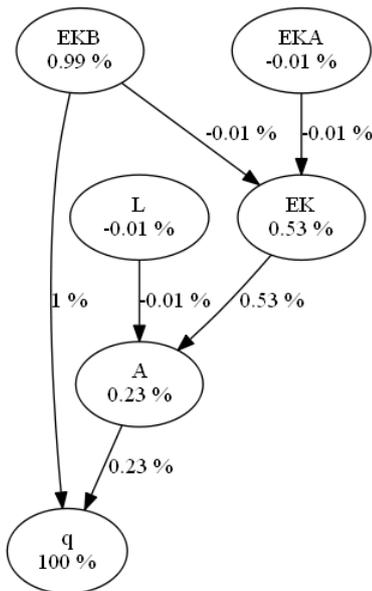
Der modifizierte finale Effekt des EK beträgt damit 0,49 Prozentpunkte.

Verteilt man die Effekte entsprechend der Standardabweichungen der eingehenden Variablen, so erhält man:

$$\text{exogene relative Zuordnungsmatrix} = \begin{bmatrix} 0.455 & 0.545 & 0 \\ 0 & 0 & 0.560 \\ 0 & 0.554 & 0 \end{bmatrix}$$

$$\text{endogene relative Zuordnungsmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0.440 & 0 & 0 \\ 0 & 0.446 & 0 \end{bmatrix}$$

Finaler Graph q:



Der modifizierte finale Effekt des EK beträgt damit 0,53 Prozentpunkte.

Ein anderes Beispiel zur Interpretation der endogenen Effekte: Wenn die Gewinne eines Unternehmens hoch sind, müssen auch viele Steuern gezahlt werden. Die Gewinne wirken positiv auf die Finanzstärke, die Steuern negativ. Bei den modifizierten endogenen Effekten würden die Steuern positiv wirken, da sie auf positive Gewinne zurückgeführt werden.

Ein weiteres Interpretationsbeispiel ist ein Fondsversicherer, der im Vergleich zum Markt eine geringe klassische Deckungsrückstellung aufweist. Dies führt im Niedrigzinsumfeld zu einem sehr positiven Effekt auf das ökonomische Eigenkapital. Der Effekt wird jedoch schon beim direkten Nachfolger, den zukünftigen passivischen Zinsüberschüssen weitgehend kompensiert, da diese sich aus der Differenz von Deckungsrückstellung und Deckungsrückstellung nach Zinsbewertung ergeben. Insgesamt verbleibt also nur der Effekt der Zinsbewertung auf die relativ kleine Deckungsrückstellung übrig. Eine nur ausschnittsweise Betrachtung der Effekte kann also irreführend sein.

## Anhang

Es werden weitere, nicht verwendete, Zusammenhänge dargestellt.

### A.1 Elemente von Potenzen triangulärer Matrizen

Shur (2014) gibt eine geschlossene Formel für die Elemente von Potenzen von triangulären Matrizen. Dabei ist  $0^0 \equiv 1$  definiert. Dies führt dazu, dass nach dem rekursiven einsetzen der Faktoren  $p_{ijs}$  nur diejenigen Terme  $x^0 \equiv 1$  mit Exponent null verbleiben, alle anderen fallen weg. Dann wird der Ausdruck an der Stelle  $x = 0$  ausgewertet. Da das Theorem für obere Dreiecksmatrizen formuliert ist, betrachten wir das transponierte Problem und vertauschen Zeilen- und Spaltenindex bei den Elementen  $m_{ij}$ . Für  $1 \leq i, j \leq n$  und Potenz  $k \geq 1$  gilt:

$$m_{ji,k} = \begin{cases} \left[ \sum_{s=i}^j p_{ijs} x^{s(k-1)} \right]_{x=0} & j \geq i + k - 1, \\ 0 & \text{sonst} \end{cases},$$

mit den rekursiv definierten Faktoren:

$$p_{ijs} = \begin{cases} \frac{\sum_{t=s}^{j-1} p_{its} m_{jt}}{x^s - x^j} & i \leq s < j \\ m_{ji} - \sum_{t=i}^{j-1} p_{ijt} & i < s = j \\ x^j & i = s = j \\ 0 & j < i \end{cases}.$$

Diese Darstellung ist nicht leicht interpretierbar, da die Faktoren rekursiv definiert sind.

### A.2 Elemente der Lösung linearer Gleichungssysteme

Wir wenden die Cramersche Regel zur Lösung eines linearen Gleichungssystems beispielhaft auf den endogenen Effekt an. Das Subsystem lautet

$$\mathbf{e}_{y\iota} = (\mathbf{I}_n - \mathbf{M}_{y\iota})^{-1} \mathbf{m}_{y\iota}, \quad \iota = 1, \dots, n,$$

und der einzelne Effekt lässt sich im Fall des DAG darstellen als:

$$e_{y\iota} = \frac{|(\mathbf{I}_n - \mathbf{M}_{y\iota})_j|}{|\mathbf{I}_n - \mathbf{M}_{y\iota}|} \\ = |(\mathbf{I}_n - \mathbf{M}_{y\iota})_j|, \quad j, \iota = 1, \dots, n, \quad j \neq \iota,$$

wobei  $(I_n - M_{y_{\iota\iota}})_j$  die Matrix  $I_n - M_{y_{\iota\iota}}$  ist, bei der die  $j$ -te Spalte durch den Spaltenvektor  $\mathbf{m}_{y_{\iota\iota}}$  ersetzt wurde. Dieser Vektor ist die  $\iota$ -te Spalte von  $M_y$ , bei der die  $\iota$ -te Zeile ausgenullt wurde.

Die Matrix  $I_n - M_{y_{\iota\iota}}$  ist eine unitrianguläre untere Dreiecksmatrix, hat also Determinante eins. Die Matrix  $(I_n - M_{y_{\iota\iota}})_j$ ,  $j = 1, \dots, n - 1$  ist im Allgemeinen keine Dreiecksmatrix, da der Spaltenvektor  $\mathbf{m}_{y_{\iota\iota}}$  eingesetzt wird. Im Fall  $j \leq \iota$  ist sie eine (nicht strikte, nicht unitrianguläre) untere Dreiecksmatrix und hat im  $j$ -ten Hauptdiagonalelement eine null und somit eine Determinante von null. Dies ist klar, da nachfolgende Variablen im azyklischen System keinen Einfluss auf vorangehend haben.

### A.3 Inversion von Matrixsummen

Miller (1981) liefert eine rekursive Formel für die Inversion von Matrixsummen, siehe auch Magidin (2011). Das Theorem von Miller verwendet in seinem Lemma die Sherman–Morrison Formel, die wiederum ein Spezialfall der Woodbury-Matrix Identität ist. Wir betrachten hier beispielhaft die endogene Effektmatrix im DAG-Fall,  $E_y = (I_n - M_y)^{-1}$ . Die zu invertierende Summe besteht also nur aus zwei Summanden und wir nehmen an, dass  $r = \text{Rang}(M_y) = n - 1$ . Die Miller-Inversion beruht auf einem iterativen Verfahren. Es benötigt die Zerlegung von  $M_y$  in  $r$  Summanden von jeweils Rang 1. Wir können in unserem Fall also direkt die ersten  $r$  Spalten von  $M_y$  verwenden. Dies führt bei einer unteren Dreiecksstruktur, also wenn letzte Spalte null ist, stets zur gesuchten Inversen der Summe. Danach gilt:

$$(I_n - M_y)^{-1} = C_r^{-1} - g_r C_r^{-1} M_{y_{0,r}} C_r^{-1},$$

mit

$$C_{k+1}^{-1} = C_k^{-1} - g_k C_k^{-1} M_{y_{0,k}} C_k^{-1},$$

$$g_k = \frac{1}{\text{tr}(C_k^{-1} M_{y_{0,k}})},$$

wobei  $\text{tr}(\cdot)$  die Spur bezeichnet und die  $M_{y_{0,k}}$  Null-Matrizen sind, deren  $k$ -te Spalte durch die  $k$ -te Spalte von  $M_y$  ersetzt wurde:

$$M_{y_{0,k}} = \begin{bmatrix} 0 & \dots & \frac{\partial M^1}{\partial y_k} & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \dots & \frac{\partial M^n}{\partial y_k} & \dots & 0 \end{bmatrix}$$

(n×n)

$$= [0 \ \dots \ \mathbf{m}_{y_{\iota\iota}} \ \dots \ 0].$$

Aufgrund der Rekursion sind die einzelnen Elemente der invertierten Summe nicht direkt ablesbar und somit nicht direkt interpretierbar.

#### A.4 Charakteristisches Polynom

Das charakteristische Polynom von  $M_y$  ist definiert als:

$$p_{M_y}(t) = |tI_n - M_y|.$$

Die Determinante einer Matrix ist also ihr charakteristisches Polynom an der Stelle eins.

Ferner gilt mit der Adjazenzmatrix  $A = M_y^T$ :

$$p_{M_y}(t) = p_{M_y^T}(t) = p_A(t).$$

Die Nullstellen  $t_i, i = 1, \dots, n$  des charakteristischen Polynoms einer Matrix sind die Eigenwerte der Matrix:

$$p_{M_y}(t) = |tI_n - M_y| = 0.$$

Die Determinante einer Matrix ist das Produkt ihrer Eigenwerte.

Im DAG-Fall einer strikten Dreiecksmatrix  $M_y$  gilt:

$$|I_n - M_y| = p_{M_y}(1) = 1.$$

Bei Dreiecksmatrizen sind die Hauptdiagonalelemente die Eigenwerte und die Determinante ist das Produkt der Hauptdiagonalelemente. Da es sich im DAG-Fall bei  $M_y$  um eine strikte Dreiecksmatrix handelt, sind alle ihre Eigenwerte und ihre Determinante null und da  $E_y = (I_n - M_y)^{-1}$  eine unitrianguläre Matrix ist, sind alle ihre Eigenwerte und ihre Determinante eins.

#### A.5 Determinanten von Subsystemen

Jacobis Formel zeigt, dass die Determinante des Subsystems bis auf das Vorzeichen der Ableitung der Determinante des ursprünglichen Systems nach dem im Subsystem gelöschten Element entspricht:

$$\frac{\partial |(I_n - M_y)^{-1}|}{\partial \left( (I_n - M_y)^{-1} \right)_{jh}} = (-1)^{j+h} |I_n - M_{yjh}|.$$

Die Determinante von Subsystemen taucht auch bei der Definition der Inversen auf:

$$\begin{aligned}
(I_n - M_y)^{-1} &= \frac{\text{ADJ}(I_n - M_y)}{|I_n - M_y|} \\
&= \frac{\text{COF}(I_n - M_y)^T}{|I_n - M_y|} \\
&= \frac{[(-1)^{j+h}|I_n - M_{y_{hj}}|]_{j,h=1,\dots,n}}{|I_n - M_y|},
\end{aligned}$$

wobei  $|\cdot|$  die Determinante bezeichnet,  $\text{ADJ}(\cdot)$  ist die Adjunkte, also die transponierte Kofaktormatrix  $\text{COF}(\cdot)$ . Die Kofaktormatrix ist die Matrix der mit dem entsprechenden Vorzeichen multiplizierten Minoren. Ein Minor ist die Determinante der Submatrix, bei der die entsprechende Zeile und Spalte gelöscht wurden. Also folgt die erste Darstellung:

$$\left((I_n - M_y)^{-1}\right)_{jh} = \frac{(-1)^{j+h}|I_n - M_{y_{hj}}|}{|I_n - M_y|},$$

und die zweite Darstellung für  $j, h \rightarrow \iota$ :

$$\frac{|I_n - M_{y_{\iota\iota}}|}{|I_n - M_y|} = \left((I_n - M_y)^{-1}\right)_{\iota\iota}$$

und entsprechend gilt für eine Submatrix:

$$(I_n - M_{y_{\iota\iota}})^{-1} = \frac{[(-1)^{j+h}|I_n - M_{y_{\{h\}\{j\}}}|]_{j,h=1,\dots,n}}{|I_n - M_{y_{\iota\iota}}|},$$

also folgt die erste Darstellung für  $i \rightarrow \iota$ :

$$\left((I_n - M_{y_{\iota\iota}})^{-1}\right)_{jh} = \frac{(-1)^{j+h}|I_n - M_{y_{\{h\}\{j\}}}|}{|I_n - M_{y_{\iota\iota}}|}$$

und die zweite Darstellung für  $j, h \rightarrow \iota$  und  $\iota \rightarrow h$  und  $\iota \rightarrow j$ :

$$\frac{|I_n - M_{y_{\{h\}\{j\}}}|}{|I_n - M_{y_{hj}}|} = \frac{\left((I_n - M_{y_{hj}})^{-1}\right)_{\iota\iota}}{(-1)^{\iota+\iota}},$$

wobei die geschweiften Klammern Mengen von gelöschten, also ausgenullten, Zeilen bzw. Spalten angeben. Sollen die Elemente der Inversen identisch sein, wie für den Beweis benötigt, so muss ihr Quotient gemäß der ersten Darstellungen gleich eins sein. Dabei sind der Zähler und der Nenner definiert, da die Invertierbarkeit der Matrizen gegeben ist. Der Quotient ist jedoch nur definiert, wenn das entsprechende Element der Submatrix ungleich null ist:

$$\begin{aligned}
\frac{\left((I_n - M_y)^{-1}\right)_{jh}}{\left((I_n - M_{y\iota})^{-1}\right)_{jh}} &= \frac{(-1)^{j+h} |I_n - M_{yhj}|}{|I_n - M_y|} \frac{|I_n - M_{y\iota}|}{(-1)^{j+h} |I_n - M_{y\{h\}\{j\}}|}, \quad j, h \neq \iota \\
&= \frac{|I_n - M_{y\iota}|}{|I_n - M_y|} \frac{(-1)^{j+h}}{(-1)^{+h}} \frac{|I_n - M_{yhj}|}{|I_n - M_{y\{h\}\{j\}}|} \\
&= \left((I_n - M_y)^{-1}\right)_{\iota} \frac{(-1)^{j+h}}{(-1)^{j+h}} \frac{(-1)^{\iota+\iota}}{\left((I_n - M_{yhj})^{-1}\right)_{\iota}} \\
&= \frac{\left((I_n - M_y)^{-1}\right)_{\iota}}{\left((I_n - M_{yhj})^{-1}\right)_{\iota}} (-1)^{j+h+\iota+\iota-j-h} \\
&= \frac{\left((I_n - M_y)^{-1}\right)_{\iota}}{\left((I_n - M_{yhj})^{-1}\right)_{\iota}}.
\end{aligned}$$

Dabei haben wir die Zähler vertauscht und die zweiten Darstellungen eingesetzt.

## A.6 Geometrische Reihen von Matrizen

In Abschnitt 2.2 haben wir für die Darstellung der Inversen  $(I_n - M_y)^{-1}$  die als konvergent angenommene Neumann-Reihe verwendet, siehe [https://en.wikipedia.org/wiki/Neumann\\_series](https://en.wikipedia.org/wiki/Neumann_series):

$$(I_n - M_y)^{-1} = I_n + \sum_{k=1}^{\infty} M_y^k.$$

Im DAG-Fall bricht die Reihe nach spätestens  $n$  Schritten ab und der endogene Effekt ist die Summe der endogenen Pfadefekte über alle Wege und Längen für die Pfade von  $y_\iota$  nach  $y_j$ . Setzt man anstatt  $M_y$  die ungewichtete Identifikationsmatrix  $ID_y^T$  ein, die nur aus Einsen und Nullen besteht, so erhält man die Anzahlen der  $k$ -Schritt-Pfade im gerichteten Graphen.

Stanley (2013) verwendet in seinem Theorem 3.4 folgende Formel einer geometrischen Reihe, deren Konvergenz durch Annahme, dass die Frobenius Norm bzw. alle Eigenwerte von  $M$  kleiner als eins sind, gewährleistet ist:

$$(I_n - M)^{-2} = I_n + \sum_{k=1}^{\infty} (k+1)M^k = \sum_{k=0}^{\infty} (k+1)M^k.$$

Er wendet diese auf die Adjazenzmatrix  $A = M^T$  der Subsysteme an, um die erwartete Hitting Time  $h_{ji}$  zu bestimmen, also die mittlere Anzahl an Schritten, bei der ausgehend vom  $j$ -ten Knoten zum ersten Mal der  $i$ -te Knoten erreicht wird. Der Erwartungswert wird über alle Längen  $k$  berechnet, bei den der Endknoten noch nicht erreicht wurde und im  $k + 1$ -ten Schritt wird das Ziel erreicht. Dabei hat  $M$  eine spezielle Struktur: Sie ist eine Übergangsmatrix, die angibt mit welcher Wahrscheinlichkeit sich der Random Walk im Markovsystem von einem Knoten zum nächsten bewegt. Die Zeilensumme ist daher eins. Es ergibt sich eine Formel, die den hier betrachteten Subsystemen scheinbar recht ähnlich ist:

$$H = [h_{ji}]_{j,i=1,\dots,n} = \sum_{k=0}^{\infty} (k+1) \mathbf{a}_{(i)} A_{ii}^k = \mathbf{a}_{(i)} (\mathbf{I}_n - A_{y_{ii}})^{-2}.$$

$$= \sum_{k=0}^{\infty} (k+1) M_{ii}^k \mathbf{m}_i = (\mathbf{I}_n - M_{y_{ii}})^{-2} \mathbf{m}_i.$$

Bei genauerer Betrachtung wird jedoch deutlich, dass lediglich die Definition der geometrischen Reihe verwendet wurde und keine Verbindung zum Ausgangssystem hergestellt wurde.

Da alle Eigenwerte von strikten Dreiecksmatrizen gleich null sind, konvergieren die Reihen in unserem Fall stets und die entsprechenden Inversen existieren. Die Elemente  $j$  des  $i$ -ten Vektors:

$$(\mathbf{I}_n - M_{y_{ii}})^{-2} \mathbf{m}_{y_i} = (\mathbf{I}_n - M_{y_{ii}})^{-1} (\mathbf{I}_n - M_{y_{ii}})^{-1} \mathbf{m}_{y_i}$$

$$= (\mathbf{I}_n - M_{y_{ii}})^{-1} \mathbf{e}_{y_{ii}}$$

sind daher die Summe der mit der Länge  $(k + 1)$  des Pfades gewichteten Pfadeffekte.

## A.7 Spezielle Zielfunktion

In Anlehnung an die Gradienten  $\partial \text{SSE} / \partial M_y$  und  $\partial \text{SSE} / \partial M_x$  in Abschnitt 5.2 kann auch der Gradient in Bezug auf die Prognosen mit der dort verwendeten  $(n \times \tau)$  Matrix  $\nabla$  bestimmt werden. Der Gradient kann für die Bestimmung der in Abschnitt 5.2 behandelten Modifikationsindikatoren nützlich sein:

$$\frac{\partial \text{SSE}}{\partial \hat{\mathbf{y}}} = 2 \left( F_{y_m} (\mathbf{I}_n - M_y)^{-1} \right)^T \Sigma_{d\hat{\mathbf{y}}_m}^{-1} (d\hat{\mathbf{Y}}_m - dY_m) \mathbf{1}_\tau = \nabla \mathbf{1}_\tau.$$

$(n \times 1)$

Dieser Gradient kann jedoch nicht analytisch hergeleitet werden. Nachfolgend wird daher der Gradient für den Spezialfall einer diagonalen Gewichtungsmatrix  $\Sigma_{d\hat{\mathbf{y}}_m}^{-1}$  abgeleitet.

Wir betrachten den Spezialfall der Zielfunktion mit diagonalen  $(n \times n)$  Selektions- und Skalierungsmatrix  $S$ , anstatt der Cholesky-Zerlegung und Dimension  $(p \times n)$ . Zudem fehlt die Tikhonov-

Regularisierung. S ergibt sich aus der Matrix L, indem die Zeilen ausgenullt werden, die nicht zu den Output-Variablen gehören und  $LL = \text{diag}(\text{diag}(\Sigma))^{-1}$ . Die Output-Variablen werden gleich stark gewichtet, indem durch ihre Standardabweichungen geteilt wird. S ist eine  $(n \times n)$  Diagonalmatrix, deren Hauptdiagonalelemente die Kehrwerte der Standardabweichungen sind, wenn die Variablen zu den Output-Variablen gehören, ansonsten sind die Hauptdiagonalelemente null.

Die Zielfunktion ist die gewichtete Fehlerquadratsumme der Output-Variablen:

$$\begin{aligned} v &= \text{tr} \left( S(\hat{Y} - Y)(\hat{Y} - Y)^T S \right) \\ &= \sum_{t=1}^{\tau} \sum_{j=1}^n \left( s_{jj}(\hat{y}_{jt} - y_{jt}) \right)^2 \\ &= \sum_{t=1}^{\tau} v_t \\ &= \mathbf{v}^T \mathbf{1}_{\tau} \end{aligned}$$

mit  $\mathbf{v} = [v_1, \dots, v_{\tau}]^T$  und dem Quadratfehler für eine einzelne Beobachtung:

$$v_t = \sum_{j=1}^n \left( s_{jj}(\hat{y}_{jt} - y_{jt}) \right)^2.$$

Die Zielfunktion ist additiv über die Beobachtungen und die Output-Variablen. Ihre Ableitung lässt sich daher ebenfalls als Summe darstellen.

Wir betrachten zuerst die Gradienten der speziellen Zielfunktion in Bezug auf die *Prognosen* zum jeweiligen Zeitpunkt. Dabei sind die Prognosen die abzuleitenden Funktionen und die Beobachtungen sind fix gegeben:

$$\begin{aligned} \frac{\partial v_t}{\partial \hat{y}_{ht}} &= \sum_{j=1}^n \frac{\partial \left( s_{jj}(\hat{y}_{jt} - y_{jt}) \right)^2}{\partial \hat{y}_{ht}} \\ &= 2 \sum_{j=1}^n s_{jj}(\hat{y}_{jt} - y_{jt}) \frac{\partial s_{jj}(\hat{y}_{jt} - y_{jt})}{\partial \hat{y}_{ht}} \\ &= 2 \sum_{j=1}^n s_{jj}^2(\hat{y}_{jt} - y_{jt}) \frac{\partial \hat{y}_{jt}}{\partial \hat{y}_{ht}} \end{aligned}$$

$$\begin{aligned}
&= 2 \sum_{j=1}^n s_{jj}^2 (\hat{y}_{jt} - y_{jt}) \mathbf{e}_{yjh} \\
&= 2 \mathbf{e}_{yh}^T \mathbf{S} \mathbf{S} (\hat{\mathbf{y}}_t - \mathbf{y}_t)
\end{aligned}$$

und in Matrixschreibweise:

$$\begin{aligned}
\nabla_{\mathbf{y}\tau} &= \frac{\partial \mathbf{v}^T}{\partial \hat{\mathbf{y}}_t} = \left[ \frac{\partial \mathbf{v}_t}{\partial \hat{y}_{jt}} \right]_{\substack{j=1, \dots, n \\ t=1, \dots, \tau}} \\
(n \times \tau) & \\
&= 2 \mathbf{E}_y^T \mathbf{S} \mathbf{S} (\hat{\mathbf{Y}} - \mathbf{Y}),
\end{aligned}$$

$$\begin{aligned}
\nabla_{\mathbf{y}} &= \frac{\partial \mathbf{v}}{\partial \hat{\mathbf{y}}} = \nabla_{\mathbf{y}\tau} \mathbf{1}_\tau \\
(n \times 1) &
\end{aligned}$$

und analog, jedoch nicht benötigt, für die exogenen Variablen:

$$\begin{aligned}
\nabla_{\mathbf{x}\tau} &= \frac{\partial \mathbf{v}^T}{\partial \mathbf{x}} = \left[ \frac{\partial \mathbf{v}_t}{\partial x_{it}} \right]_{\substack{i=1, \dots, m \\ t=1, \dots, \tau}} \\
(m \times \tau) & \\
&= 2 \mathbf{E}_x^T \mathbf{S} \mathbf{S} (\hat{\mathbf{Y}} - \mathbf{Y}),
\end{aligned}$$

$$\begin{aligned}
\nabla_{\mathbf{x}} &= \frac{\partial \mathbf{v}}{\partial \mathbf{x}} = \nabla_{\mathbf{x}\tau} \mathbf{1}_\tau. \\
(m \times 1) &
\end{aligned}$$

Die Gradienten der speziellen Zielfunktion in Bezug auf die *Gewichte* lauten mittels Matrix-Calculus:

$$\begin{aligned}
\nabla_{\mathbf{M}_y} &= \frac{\partial \mathbf{v}}{\partial \mathbf{M}_y} = \left[ \frac{\partial \mathbf{v}}{\partial m_{yjt}} \right]_{j,t=1, \dots, n} \\
(n \times n) & \\
&= 2 \mathbf{E}_y^T \mathbf{S} \mathbf{S} (\hat{\mathbf{Y}} - \mathbf{Y}) \hat{\mathbf{Y}}^T \\
&= \nabla_{\mathbf{y}\tau} \hat{\mathbf{Y}}^T
\end{aligned}$$

und analog für die exogenen Gewichte, wobei zu beachten ist, dass ebenfalls der Gradient  $\nabla_{\mathbf{y}\tau}$  zu verwenden ist:

$$\begin{aligned}
\nabla_{\mathbf{M}_x} &= \frac{\partial \mathbf{v}}{\partial \mathbf{M}_x} = \left[ \frac{\partial \mathbf{v}}{\partial m_{xji}} \right]_{\substack{j=1, \dots, n \\ i=1, \dots, m}} \\
(n \times m) &
\end{aligned}$$

$$= \nabla_{y\tau} X^T.$$

## Literatur

- Avin, C., Shpitser, I. and Pearl, J. (2005) "Identifiability of path-specific effects", Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence IJCAI-05, Morgan-Kaufmann Publishers, Edinburgh, UK.
- Bartel, Harald (2017) "Prognosen unter Nebenbedingungen", unveröffentlichtes Manuskript.
- Bekker, P. A., Merckens, A. and Wansbeck, T. (1994) "Identification, Equivalent Models and Computer Algebra", Boston: Academic Press.
- Bollen, K. A. (1989) "Structural Equations with Latent Variables", Wiley.
- Bollen, K. A., Bauldry, S. (2010) "Model Identification and Computer Algebra", Social Methods Res., 39(2): 127–156.  
[https://www.researchgate.net/publication/51502218\\_Model\\_Identification\\_and\\_Computer\\_Algebra](https://www.researchgate.net/publication/51502218_Model_Identification_and_Computer_Algebra).
- Bollen, K. A., Bauldry, S. (2010b) "A Note on Algebraic Solutions to Identification" The Journal of Mathematical Sociology, Volume 34, Issue 2.  
[https://www.researchgate.net/profile/Kenneth\\_Bollen/publication/44636388\\_A\\_Note\\_on\\_Algebraic\\_Solutions\\_to\\_Identification/links/00b4952416e01b0a46000000/A-Note-on-Algebraic-Solutions-to-Identification.pdf](https://www.researchgate.net/profile/Kenneth_Bollen/publication/44636388_A_Note_on_Algebraic_Solutions_to_Identification/links/00b4952416e01b0a46000000/A-Note-on-Algebraic-Solutions-to-Identification.pdf).
- Brualdi, R. A., Cvetkovic, D. (2008) "A Combinatorial Approach to Matrix Theory and Its Applications", Chapman and Hall/CRC.
- Chen, B., Pearl, J. and Kline, R. B. (2016) "Graphical Tools for Linear Path Models", UCLA Cognitive Systems Laboratory, Technical Report (R-469), September 2016. Forthcoming, Psychometrika.
- Chiang, A. C. (1984) "Fundamental Methods of Mathematical Economics", McGraw-Hill, Third Edition, 1984.
- Dijkstra, T., K. (2013) "Ridge regression and its degrees of freedom", Quality and Quantity.
- Hendry, D. F. (?) "Chapter 11 Simultaneous equations systems", <https://www.nuffield.ox.ac.uk/users/hendry/book/ans11.pdf>.
- Huber, W. A. (2015), "Why does inversion of a covariance matrix yield partial correlations between random variables?", <https://stats.stackexchange.com/a/156812/173285>.
- Israel, R. (2011) "Taking Powers of a Triangular Matrix", <https://math.stackexchange.com/q/69937>.
- Jacobucci, R., Grimm, K. J., McArdle, J. J. (2016) "Regularized Structural Equation Modeling", Structural Equation Modeling, 23(4): 555–566. [www.ncbi.nlm.nih.gov/pmc/articles/PMC4937830](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4937830).
- Judge, G. J, Hill, R. C., Griffiths, W. E., Lütkepohl, H., Lee, T.-C (1998) "Introduction to the Theory and Practice of Econometrics", Wiley.

- Kawaguchi, K. (2016) “Deep Learning without Poor Local Minima“, Massachusetts Institute of Technology, MIT-CSAIL-TR-2016-005, <https://arxiv.org/abs/1605.07110>.
- van Kesteren, E.-J., Oberski, D. L. (2019) “Structural Equation Modeling using Computation Graphs“, <https://arxiv.org/abs/1905.04492>.
- Khaki-Sedigh, A., Moaveni, B. (2009) “Control Configuration Selection for Multivariable Plants“, Springer.
- Koster, J. T. A. (1995) “Markov properties of non-recursive causal models”, To appear in Annals of Statistics, November 1995.
- Magnus, J. R., Neudecker, H. (2007) “Matrix Differential Calculus with Applications in Statistics and Econometrics“, John Wiley, Third Edition.
- Magidin, A. (2011), “Inverse of the Sum of Matrices“, URL (version: 2011-01-16): <https://math.stackexchange.com/q/17780>.
- Miller, K. S. (1981) “On the Inverse of the Sum of Matrices“, Mathematics Magazine, Vol. 54, No. 2 (Mar., 1981), pp. 67-72.
- Pearl, J. (2001) “Direct and Indirect Effects“, Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, San Francisco, CA: Morgan Kaufmann, 411-420, 2001.
- Pearl, J. (2009) “Causal inference in statistics: An overview“, Statistics Surveys, 3:96--146, 2009.
- Pearl, J. (2017) “A Linear Microscope for Interventions and Counterfactuals“, Journal of Causal Inference, 5(1):1-15, March 2017.
- Pearl, J. (2018) “Causal and Counterfactual Inference”, The Handbook of Rationality, MIT press.
- Pearl and Dechter (2013) “Identifying Independencies In Causal Graphs with Feedback”, Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence (UAI1996).
- Rojas, R. (1996): “Neural Networks – A Systematic Introduction“, Springer-Verlag, Berlin, 1996.
- Rothenberg, T. J. (1971) “Identification in Parametric Models”, Econometrica, Vol. 39, No. 3.
- Shpitser, I., Pearl, J. (2006) ”Identification of Conditional Interventional Distributions”, Proceeding UAI'06 Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence Pages 437-444.
- Shur, W. (2014) “A Generalized Closed Form for Triangular Matrix Powers“, 19 May 2014, arXiv: 1301.6820v2.
- Skogestad, S., Postlethwaite, I. (2007) “Multivariable Feedback Control: Analysis and Design“, Wiley.

Spirtes, P. (1994) “Conditional independence in directed cyclic graphical models for feedback”, Technical Report CMU-PHIL-53, Carnegie-Mellon University, Department of Philosophy, Pittsburgh, PA, 1994. To appear in Networks.

Stanley, R. P. (2013), “Topics in Algebraic Combinatorics“, 1 February 2013, <http://www-math.mit.edu/~rstan/algcomb/algcomb.pdf>.

Steiger, J. H. “Statistical Models in Structural Equation Modelling“, [http://www.statpower.net/Content/GCM/Handouts/Statistical Models in Structural Equation Modeling.pdf](http://www.statpower.net/Content/GCM/Handouts/Statistical%20Models%20in%20Structural%20Equation%20Modeling.pdf).

Susmel, R. “Lecture 16 SEM“, <http://www.bauer.uh.edu/rsusmel/phd/ec1-16.pdf>.

Tibshirani, R. J. (2014), “Degrees of Freedom and Model Search“, *Statistica Sinica* 25(3) · February 2014.

Toyoda, H. (1994) “A New Identification Rule and Estimator for the Simultaneous Equation Model Using the Notation of the Reticular Action Model“, *Behaviormetrika*, Vol. 21, No. 2, 163 – 175.