

DIPLOMARBEIT

DIMENSIONSREDUKTION IN MULTIVARIATEN AUTOREGRESSIVEN MODELLEN : DER RANGREDUZIERTE FALL

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Diplom-Ingenieurs
unter der Leitung von

Em.O.Univ.Prof. Dipl.-Ing. Dr.techn. Deistler
E105
Institut für Wirtschaftsmathematik

eingereicht an der Technischen Universität Wien
Fakultät für Maschinenwesen und Betriebswissenschaften

von

Alexander Juschitz
e0325828
Afritschgasse 61, 1220 Wien

Wien, am 20. April 2010

Alexander Juschitz

Danksagung

Diese Diplomarbeit entstand am Institut für Wirtschaftsmathematik an der Technischen Universität Wien unter der Leitung von Em.O.Univ.Prof. Dipl.-Ing. Dr.techn. Manfred Deistler. Ihm besonders habe ich die Begeisterung für die Ökonometrie zu verdanken. Im Wintersemester 2008 kam ich nach zwei Jahren Auslandsstudium in Paris an die Technische Universität Wien zurück. Eine der ersten Vorlesungen, die ich nach meiner Rückkehr wieder besuchte, war eine Vorlesung zur Ökonometrie, gehalten von Prof. Deistler. Durch meine beiden Studien an der Tu Wien - Technische Mathematik und Wirtschaftsingenieurwesen-Maschinenbau - kam ich in den Genuss die unterschiedlichsten Arten der Didaktik besser kennen zu lernen. Durch meinen Auslandsaufenthalt, war es mir möglich Unterschiede bei Lehrmethoden zu erkennen und sowohl Stärken als auch Schwächen dieser zu identifizieren.

Trotz der reichen Erfahrung, war es für mich das erste Mal, dass ich bei einem Professor eine so gewinnende Begeisterung für sein Fach vorfinden durfte, wie es bei Prof. Deistler der Fall ist. Zudem kann sich Prof. Deistler in die Denkweise seiner Studenten, die sich zum ersten Mal mit der Materie beschäftigen, einfühlen. So gelingt es ihm seine Studenten in die Vorlesung zu einzubinden und ihnen seine Leidenschaft für das Fach der Ökonomie zu vermitteln. Für diese ansteckende Begeisterungsfähigkeit und seine Unterstützung möchte ich mich ausdrücklich bei Prof. Deistler bedanken! Ich werde versuchen diese auch in meinem zukünftigen Arbeitsleben einfließen zu lassen.

Bei Cécile Deterre möchte ich mich ganz herzlich bedanken. Sie hat viele Stunden für das Korrekturlesen aufgewandt. Dadurch, dass sie sich selbst mit Teilchenphysik beschäftigt, steht sie der Materie der Ökonomie ohne Vorwissen gegenüber. Mit ihrer Hilfe ist es mir hoffentlich gelungen eine Arbeit zu verfassen, die sowohl wissenschaftliche Standards erfüllt als auch im Hinblick auf die Verständlichkeit einen interessierten Leser, der vielleicht nicht über das Wissen einer Mathematik- oder WirtschaftsstudentIn verfügt, an der Hand nimmt und der Materie näher bringt. Über allem natürlich steht meine Familie. Meinen lieben Eltern, ohne die mein Studium nie in ähnlicher Weise möglich gewesen werde, möchte ich an dieser Stelle danken. Durch ihre finanzielle Unterstützung und Großzügigkeit ermöglichten sie mir vier wunderschöne Jahre in Wien, sowie einen unvergesslichen Auslandsaufenthalt in Paris. Sie ermutigten mich stets und waren immer für mich da. Für die moralische Unterstützung gilt meiner Schwester Barbara Juschitz besonderer Dank.

Alexander Juschitz Wien, März 2010

Inhaltsverzeichnis

Einleitung	1
1 Kanonische Korrelation	5
1.1 Einleitung	5
1.2 Grundprinzip	6
1.3 Anzahl der kanonischen Korrelationen	8
1.4 Schätzung der kanonische Korrelationen und der Zufallsvariablen	10
1.5 Signifikanztests	11
1.6 Mathematische Herleitung	11
1.6.1 Herleitung der charakteristischen Gleichung	12
1.6.2 Eigenwerte	15
1.6.3 Eigenvektoren	21
1.7 Beziehung zum Skalarkomponenten-Modellstruktur	23
2 Verschachtelte rangreduzierte autorogressive Modell im stationären Fall	27
2.1 Einleitung	27
2.2 Das verschachtelte rangreduzierte autoregressive Modell	31
2.2.1 Bestimmung und Spezifikation der Ränge	33
2.2.2 Kanonische Form	37
2.2.3 Parametrisierung und Normalisierung	41
3 Rückschluss auf Vektor Autoregressiven Prozessen mit Kointegration und Skalarkomponenten	45
3.1 Einleitung	45
3.2 Das Modell und seine Eigenschaften	46
A Anhang	53
A.1 Grundlagen	53
A.1.1 Prädiktor bzw. vorhersagende Größe	53
A.1.2 Vorherzusagende Größe bzw. Kriteriumsvariable	53
A.1.3 α bzw. Fehler erster Art	53
A.1.4 Stochastischer Prozess	54
A.1.5 Multivariater stochastischer Prozess	54
A.1.6 Kovarianzfunktion	54
A.1.7 Kovarianzmatrix	55
A.1.8 Akaike Informationskriterium (AIC)	55
A.2 Mathematik	55

A.2.1	Laplacescher Erweiterungssatz	55
A.2.2	Laurentreihe	56
A.2.3	Minor	56
	Literatur	59
	Lebenslauf	61

Einleitung

Das Risikomanagement nimmt in einem Unternehmen einen großen Stellenwert in der Ausrichtung für die Zukunft ein. So muss jedes Unternehmen, welches nach den Regeln der International Financial Reporting Standards (IFRS) bilanziert, auf Marktrisiken eingehen und den Value at Risk für unterschiedliche Risikopositionen angeben. Der Ausdruck Value at Risk (VaR), auch Wert im Risiko genannt, bezeichnet jenes Risikomaß, welches den Wert angibt, den eine Risikoposition, wie es zum Beispiel ein Wertpapierportfolio ist, zu einer bestimmten Wahrscheinlichkeit in einem gegebenen Zeitraum verlieren kann. Man verwendet diese Kennzahl um Marktrisiken quantitativ zu messen. Das Marktrisiko hängt von Parameter wie Aktienkurse, Zinsrisiken etc. abhängt. Da es sich bei ihnen um Zeitreihen handelt, empfiehlt sich die Zeitreihenanalyse zur Erarbeitung des Value at Risk.

Die Zeitreihenanalyse beschäftigt sich mit der Extraktion von Information in Form einer Zeitreihe. Dabei gilt zu beachten, dass die Reihenfolge der Beobachtungen als wichtige Information zu betrachten ist. Im Vergleich, dazu lässt eine Permutation von Daten die Resultate in der klassischen Statistik unverändert.

In den letzten Jahren hat die Zeitreihenanalyse in den Wirtschaftswissenschaften, wie zum Beispiel auf dem Gebiet der Finanzmarktökonomie, an Bedeutung gewonnen. Diese Entwicklungsphase setzte sich mit der Verleihung des Nobelpreises im Jahr 2003 an Clive Granger und Robert Engle fort. Neben einer enormen Bandbreite an entwickelten Techniken und Methoden haben die Analyse von integrierten und kointegrierten Prozessen, sowie der Identifikation Vektor-autoregressiver Prozesse am stärksten dazu

beigetragen.

Nun stehen dem Interessierten ein Fülle an entwickelten Modellen gegenüber. Sie alle lassen sich durch ihren universellen Charakter beliebig an Komplexität steigern. Man versucht zum Beispiel durch Hinzunahme von Parameter die Güte eines Modells zu verbessern. So erstellt man ein Modell, welches auf einer immer längeren Vergangenheit basiert. Bei multivariaten Zeitreihen nimmt jedoch die Anzahl der Parameter quadratisch zu mit jedem Schritt in Richtung der ferner zurückliegenden Vergangenheit. Die Anzahl der zur Verfügung stehenden Daten, mit denen das Modell und seine Parameter angepasst werden, sind jedoch beschränkt. Nun stellt sich die Frage, ob der Parameterraum nicht eingeschränkt werden kann und dabei die Güte des Modells erhalten bleibt. Das verschachtelte rangreduzierte autoregressive Modell, wie es von [Ahn and Reinsel, 1988] beschrieben wir, bietet die Möglichkeit für stationäre Zeitreihen mittels der kanonische Korrelationsanalyse Strukturen innerhalb der multivariaten Zeitreihe aufzufinden. Dies erlaubt den Parameterraum einzuschränken. Dadurch kann es zu einer Effizienzsteigerung beim Schätzen der Modellparameter kommen. Durch Transformation auf kanonische Gestalt vereinfacht sich die Struktur des Modells und dessen Interpretation wird vereinfacht.

Da die kanonische Korrelationsanalyse einen großen Bestandteil im verschachtelten rangreduzierten autoregressiven Modell darstellt, werden wir zunächst auf die kanonische Korrelationsanalyse in Kapitel 1 eingehen. Nach einer kurzen Einleitung wollen wir auf das Grundprinzip eingehen und Eigenschaft, sowie Signifikanztests anführen. Danach führen wir die mathematische Herleitung an, die stark zum Verständnis der kanonischen Korrelation beiträgt. Am Ende wir noch ein Beispiel, wie die kanonische Korrelation in einem einfachen Modell zu Verwendung kommt.

Im Kapitel 2 definieren wir das verschachtelte rangreduzierte autoregressive Modell für stationäre Zeitreihen und führen alle Forderungen und Annahmen an. Wir zeigen, wie der Grad des Modells bestimmt werden kann und wie der Rang der einzelnen Parametermatrizen ermittelt werden. Im Anschluss erklären wir, wie ein multivariater autoregressiver Prozess in kanonische Form transformiert werden kann. Wir beenden

das Kapitel mit dem Unterkapitel über die Parametrisierung und Normalisierung der Parametermatrizen.

Die Diplomarbeit endet mit einem kurzen Ausblick, wie das Prinzip aus Kapitel 2 auf teilweise stationäre Prozesse ausgeweitet werden kann.

Kapitel 1

Kanonische Korrelation

1.1 Einleitung

In diesem Kapitel betrachten wir zwei Vektoren von Zufallsvariablen mit gemeinsamer Verteilung und wir analysieren die Korrelation zwischen den Variablen eines Vektors und denen des anderen Vektors. Wir finden ein neues Koordinatensystem im Raum eines Vektors von Zufallsvariablen, sodass die Korrelationen eindeutig dargestellt werden. Wir finden also Linearkombinationen der Zufallsvariablen der einzelnen Vektoren, die maximal zueinander korrelieren. Diese Linearkombinationen sind die ersten Koordinaten des neuen Systems. Danach wird eine zweite Linearkombinationen der Zufallsvariablen jedes Vektors gesucht, sodass die Korrelation maximiert wird. Dabei gilt die Nebenbedingung, dass die zweiten Linearkombinationen unkorreliert zu den vorherigen sind. Dieses Verfahren wird fortgesetzt bis zwei neue Koordinatensystem komplett ermittelt wurden.

Diese statistische Methode ist besonders hilfreich in forschenden Studien. Man hat zwei Vektoren von hoher Dimension mit Zufallsvariablen und möchte die Wechselbeziehung studieren. Wenn die zwei Vektoren von sehr hoher Dimension sind, möchte man vielleicht nur wenige Linearkombinationen betrachten. Dabei werden einem nur jene interessieren, welche die größte Korrelationen aufweisen.

Die zugrunde liegende Theorie wurden von Hotelling ([Hotelling, 1935], [Hotelling, 1936]) entwickelt. Ihre Verwendung in der Analyse von multivariaten Zeitreihen wurde behandelt unter anderem durch [Akaike, 1976], [Cooper and Wood, 1982], [Tiao and Tsay, 1989].

Wir beginnen das Kapitel, indem wir kurz das Grundprinzip der kanonischen Korrelation im Abschnitt 1.2 erklären. Danach gehen wir genauer auf die kanonische Korrelation ein und behandeln, wie man die Anzahl der kanonischen Korrelationen erhält und sie auf Signifikanz testet. Im Unterkapitel über die mathematische Herleitung (siehe 1.6) leiten wir die charakteristische Gleichung her. Später gehen wir auf die genaue Beschaffenheit der Eigenwerte der charakteristischen Gleichung, die für die kanonische Korrelationen benötigt werden, ein. Zum Abschluss bringen wir noch ein Beispiel für ein Modell (das Skalarkomponentenmodell), welches sich der kanonischen Korrelation bedient.

1.2 Grundprinzip

Es sei

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad (1.1)$$

Es soll die kanonische Korrelation zwischen zwei Vektoren von Zufallsvariablen $X_1 = (x_{11}, x_{12}, \dots, x_{1p_1})'$ und $X_2 = (x_{21}, x_{22}, \dots, x_{2p_2})'$ bestimmt werden. Die Dimension von X_1 und X_2 sei

$$p_1 = \dim X_1, \quad (1.2)$$

$$p_2 = \dim X_2, \quad (1.3)$$

wobei wir annehmen, dass $p_1 \leq p_2$ gilt.

Man ermittelt nun die Supermatrix von bivariaten Korrelationen:

$$\Sigma = \left(\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array} \right), \quad (1.4)$$

wobei in dieser Gleichung die Terme folgendes repräsentieren:

- Σ_{11} : Korrelationsmatrix der Zufallsvariablen des Vektors X_1
- Σ_{22} : Korrelationsmatrix der Zufallsvariablen des Vektors X_2
- $\Sigma_{12} = \Sigma'_{21}$: $p_1 \times p_2$ -Matrix der Korrelationen zwischen den einzelnen Zufallsvariablen der Vektoren X_1 und X_2

Wir nehmen an, dass Σ positiv definit ist.

In weiter Folge ist die Vorgehensweise ähnlich der Hauptkomponentenanalyse (Principal Component Analysis *PCA*) [Witte and Horstmann, 1976]. Die *PCA* dient dazu Datensätze durch Strukturierung zu vereinfachen, indem man aus k Variablen die aussagekräftigsten Linearkombinationen (Hauptkomponenten oder Faktoren auch genannt) bestimmt, die sukzessiv die maximale Varianz aufklären. Dabei sind die Linearkombinationen alle orthogonal aufeinander.

Bei der kanonische Korrelationsanalyse entspricht das Verfahren zwei *PCA*, die unabhängig für zwei Vektoren von Zufallsvariablen durchgeführt werden. Im Gegensatz zur *PCA* wird nicht die Varianz als Kriterium herangezogen. Bei der kanonische Korrelationsanalyse erfolgt die Bestimmung der Linearkombination anhand der maximal zu erreichenden Korrelation. Diese wird als kanonische Korrelation bezeichnet. Anders ausgedrückt, werden die Zufallsvariablen des ersten und die des zweiten Vektors getrennt faktorisiert und so rotiert, sodass der Korrelation maximiert wird.

Wenn man die kanonische Korrelationsanalyse formal anschreiben möchte, so geht man wie folgt vor. Es werden aus den Vektoren $X_1 = (x_{11}, x_{12}, \dots, x_{1p_1})'$ und $X_2 = (x_{21}, x_{22}, \dots, x_{2p_2})'$ jene Linearkombinationen $U = \alpha'X_1$ und $V = \gamma'X_2$ gebildet, für die gilt

1. Die Varianz der Zufallsvariablen U und V wird mit 1 normiert,

$$\text{Var}(U) = 1 \tag{1.5}$$

$$\text{Var}(V) = 1 \tag{1.6}$$

2. $\rho = \text{Corr}(U, V) \geq 0$ ist die maximale Korrelation zwischen zwei Linearkombinationen von X_1 und X_2 . Die Bestimmung der Linearkombinationen U und V erfolgt, sodass sie die maximale Korrelation aller möglichen Linearkombinationen realisieren.

Die Gewichtungsvektoren α' und γ' , die die kanonische Korrelation maximieren, erhält man durch Ermittlung der Eigenwerte der Gleichung

$$(\rho_i^2 I - \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}) \alpha_i = 0 \quad i = 1, \dots, p_1. \quad (1.7)$$

Die Wurzel aus dem größten Eigenwert ρ_i^2 gibt die maximale kanonische Korrelation an. Von diesem Eigenwert aus, lassen sich nun die Vektoren α und γ genauer bestimmen.

1.3 Anzahl der kanonischen Korrelationen

Wie man von der Hauptkomponentenanalyse weiß, lässt sich die Gesamtvarianz praktisch niemals durch einen einzigen Faktor bestimmen. Normalerweise bleibt eine gewisse Restvarianz, die groß genug ist um einen zweiten Faktor zu bestimmen. Hierbei muss darauf geachtet werden, dass der zweite Faktor vom ersten unabhängig sein muss.

Ähnliches gilt auch für die kanonischen Korrelationsanalyse. Zuerst extrahieren wir sowohl einen Faktor für den Vektor X_1 als auch für den Vektor X_2 , sodass die Korrelation maximiert wird. Nun verbleibt im Allgemeinen eine Restvarianz in den beiden Variablensätzen. Man extrahiert nun einen zweiten Faktor für die Vektoren X_1 und X_2 , sodass die Korrelation wieder maximiert wird, mit der Bedingung, dass sie unabhängig bezüglich den zuvor ausgewählten sind.

Nach diesem Prinzip kann nun sukzessive Aufklärung der Restvarianz erfolgen, solange, bis einer der beiden Variablensätze erschöpft ist. Wir wissen, dass n wechselseitig korrelierte Variablen in maximal n wechselseitig unkorrelierte Faktoren transformiert werden können. Das bedeutet, dass die Varianz von n Variablen erschöpft ist, wenn n Faktoren ermittelt wurden. Die Anzahl der kanonischen Korrelationen entspricht nun der Anzahl der Variablen im kleineren Variablensatz. In unserem Fall wäre dass p_1 .

Im allgemeinen wird die Anzahl der kanonischen Korrelation mit $m = \min(p_1, p_2)$ angegeben. Durch den Satz von m unabhängigen Faktoren wird die Varianz des kleineren Variablensatzes vollständig erschöpft. Im größeren Variablensatz bleibt somit eine gewisse Restvarianz bestehen, die keine gemeinsame Kovarianz mit dem kleineren hat.

Wie wollen nun die kanonische Korrelationsanalyse neu anschreiben um alle kanonischen Korrelationen zu erfassen. Es sei $X_1 = (x_{11}, x_{12}, \dots, x_{1p_1})'$ und $X_2 = (x_{21}, x_{22}, \dots, x_{2p_2})'$ und $U_i = \alpha_i' X_1$ und $V_i = \gamma_i' X_2$ jene Linearkombinationen, für die gilt

1. Die Zufallsvariablen U_i und V_i sind untereinander unkorreliert und haben die Varianz von Wert 1,

$$\text{Var}(U_i) = \text{Var}(V_i) = 1 \quad (1.8)$$

$$\text{Cov}(U_i, U_j) = \text{Cov}(V_i, V_j) = 0 \quad \forall i \neq j \quad (1.9)$$

$$\text{Cov}(U_i, V_j) = 0 \quad \forall i \neq j \quad (1.10)$$

2. $\rho_1 = \text{Corr}(U_1, V_1) \geq 0$ ist die maximale Korrelation zwischen zwei Linearkombinationen von X_1 und X_2 . Die Bestimmung der Linearkombinationen U_1 und V_1 erfolgt, sodass sie die maximale Korrelation aller möglichen Linearkombinationen realisieren. Weiters gilt somit, dass U_2 und V_2 die Korrelation $\rho_2 = \text{Corr}(U_2, V_2) \geq 0$ haben. Ihre Linearkombination für die die Korrelation maximiert wird, sind unkorreliert zu U_1 und V_1 . Wenn man so fortschreitet, erhält man die U_i und V_i .

1.4 Schätzung der kanonische Korrelationen und der Zufallsvariablen

Sei x_1, \dots, x_N N Beobachtungen von $N(\mu, \Sigma)$. Sei x_α aufgeteilt in zwei Vektoren mit p_1 bzw p_2 Komponente, also

$$x_\alpha = \begin{pmatrix} x_{\alpha.1} \\ x_{\alpha.2} \end{pmatrix} \quad \alpha = 1, \dots, N. \quad (1.11)$$

Der Maximum Likelihoodschätz von Σ ist

$$\hat{\Sigma} = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{pmatrix} = \frac{1}{N} \sum_{\alpha=1}^N (x_\alpha - \bar{x})(x_\alpha - \bar{x})' \quad (1.12)$$

$$= \frac{1}{N} \begin{pmatrix} \sum (x_{\alpha.1} - \bar{x}_1)(x_{\alpha.1} - \bar{x}_1)' & \sum (x_{\alpha.1} - \bar{x}_1)(x_{\alpha.2} - \bar{x}_2)' \\ \sum (x_{\alpha.2} - \bar{x}_2)(x_{\alpha.1} - \bar{x}_1)' & \sum (x_{\alpha.2} - \bar{x}_2)(x_{\alpha.2} - \bar{x}_2)' \end{pmatrix}. \quad (1.13)$$

Um die kanonischen Korrelationen ρ_i zu Schätzen, berechnen wir die Eigenwerte der Gleichung

$$V = \left(r_i^2 I - \hat{\Sigma}_{11}^{-1} \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21} \right) \hat{\alpha}_i = 0 \quad i = 1, \dots, p_1. \quad (1.14)$$

Die Wurzel aus dem Eigenwert r_i^2 gibt die geschätzte kanonische Korrelation an. Von diesem Eigenwert aus, lassen sich nun die Vektoren $\hat{\alpha}$ und $\hat{\gamma}$ bestimmen.

1.5 Signifikanztests

Da man nicht davon ausgehen kann, dass alle m kanonischen Korrelationen der beiden Variablensätzen X_1 und X_2 signifikant ungleich Null sind, müssen diese darauf getestet werden. Zum Beispiel schlug [Tatsuoka, 1971] folgenden Test vor:

$$- \left[N - 3/2 - \frac{(p_1 + p_2)}{2} \right] \sum_{i=1}^m \ln(1 - \rho_i^2). \quad (1.15)$$

Die Gleichung (1.15) ist approximativ χ^2 -verteilt mit $p_1 \cdot p_2$ Freiheitsgraden bezüglich der Nullhypothese. Hat man bereits j kanonische Korrelationen bestimmt, so kann überprüft werden, ob die verbleibende Kovarianz signifikant von Null verschieden ist um ein weiteres α_{j+1} und γ_{j+1} zu bestimmen. Mit

$$V_j = - \left[N - 3/2 - \frac{(p_1 + p_2)}{2} \right] \sum_{i=j+1}^r \ln(1 - \rho_i^2) \quad (1.16)$$

testen wir auf Signifikanz. V_j hat $(p_1 - j) \cdot (p_2 - j)$ Freiheitsgrade.

1.6 Mathematische Herleitung

Es sei

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}. \quad (1.17)$$

Es soll die kanonische Korrelation zwischen zwei Vektoren von Zufallsvariablen $X_1 = (x_{11}, x_{12}, \dots, x_{1p_1})'$ und $X_2 = (x_{21}, x_{22}, \dots, x_{2p_2})'$ bestimmt werden. Über die Dimension von X_1 und X_2 lässt sich folgendes sagen

$$\dim X_1 = p_1, \quad (1.18)$$

$$\dim X_2 = p_2, \quad (1.19)$$

wobei wir annehmen, dass $p_1 \leq p_2$.

Es sei α und γ Gewichtungsvektoren. Durch sie erhalten wir die Linearkombinationen

U von X_1 und V von X_2 folgendermaßen:

$$U = \alpha' X_1 \quad (1.20)$$

$$V = \gamma' X_2. \quad (1.21)$$

Wir suchen nun die Gewichtungsvektoren α und γ so zu bestimmen, dass die Linearkombinationen U und V maximal miteinander korrelieren.

1.6.1 Herleitung der charakteristischen Gleichung

Es seien U und V die Linearkombinationen der beiden Vektoren X_1 und X_2 . Dann sei

$$\rho = \text{Corr}(U, V) = \frac{\text{Cov}(U, V)}{\sqrt{\text{Var}(U)\text{Var}(V)}} \quad (1.22)$$

die Korrelation der Linearkombinationen U und V , die es zu maximieren gilt. Da

$$U = \alpha' X_1 \quad (1.23)$$

$$V = \gamma' X_2 \quad (1.24)$$

erhalten wir

$$\rho = \frac{\text{Cov}(\alpha' X_1, \gamma' X_2)}{\sqrt{\text{Var}(\alpha' X_1)\text{Var}(\gamma' X_2)}}. \quad (1.25)$$

Mit den Eigenschaften

$$\text{Var}(\alpha' X) = \alpha' \text{Var}(X) \alpha, \quad (1.26)$$

$$\text{Cov}(\alpha' X_1, \gamma' X_2) = \alpha' \text{Cov}(X_1, X_2) \gamma \quad (1.27)$$

können wir nun die Korrelation neu anschreiben als

$$\rho = \frac{\alpha' \text{Cov}(X_1, X_2) \gamma}{\sqrt{\alpha' \text{Var}(X_1) \alpha \gamma' \text{Var}(X_2) \gamma}}. \quad (1.28)$$

Die Transformationsvektoren α und γ , die zu einer maximalen Kovarianz zwischen U und V führen, können nicht eindeutig bestimmt werden. Damit wir sie eindeutig bestimmen können, legen wir fest, dass sie die Varianz normalisieren

$$\text{Var}(U) = \text{Var}(\alpha' X_1) = 1 \quad (1.29)$$

$$\text{Var}(V) = \text{Var}(\gamma' X_2) = 1. \quad (1.30)$$

Somit erhalten wir

$$\rho = \alpha' Cov(X_1, X_2) \gamma. \quad (1.31)$$

Anders angeschrieben, erhalten wir

$$\rho = \alpha' \Sigma_{12} \gamma \quad (1.32)$$

mit den Nebenbedingugn

$$\alpha' \Sigma_{11} \alpha = \gamma' \Sigma_{22} \gamma = 1. \quad (1.33)$$

Es sei L die Lagrangefunktion für die zu maximierende Korrelation mit den normalisierten Varianzen als Nebenbedingung. Wir können sie anschreiben als

$$L = \alpha' \Sigma_{12} \gamma - \frac{\lambda}{2} (\alpha' \Sigma_{11} \alpha - 1) - \frac{\mu}{2} (\gamma' \Sigma_{22} \gamma - 1), \quad (1.34)$$

wobei wir die Lagrangemultiplikatoren mit $\frac{\lambda}{2}$ und $\frac{\mu}{2}$ ansetzen.

Wenn wir nun die erste Ableitung nach α und γ bilden und Null setzen, so erhalten wir

$$\frac{\partial L}{\partial \alpha} = \Sigma_{12} \gamma - \lambda \Sigma_{11} \alpha = 0, \quad (1.35)$$

$$\frac{\partial L}{\partial \gamma} = \alpha' \Sigma_{12} - \mu \gamma' \Sigma_{22} = 0. \quad (1.36)$$

Wenn wir die beiden ersten Ableitung Null setzen und die Nebenbedingungen aus Gleichung (1.33) heranziehen, so erhalten wir 4 Gleichung mit den 4 Unbekannten α , γ , λ und μ . Wir erhalten durch Multiplikation von links durch α'

$$\alpha' \Sigma_{12} \gamma - \lambda \alpha' \Sigma_{11} \alpha = 0, \quad (1.37)$$

und von rechts durch γ

$$\alpha' \Sigma_{12} \gamma - \mu \gamma' \Sigma_{22} \gamma = 0. \quad (1.38)$$

Mittels der Normalisierung

$$\alpha' \Sigma_{11} \alpha = \gamma' \Sigma_{22} \gamma = 1 \quad (1.39)$$

vereinfachen sich die Gleichung und wir erhalten das Gleichungssystem

$$\alpha' \Sigma_{12} \gamma - \lambda = 0, \quad (1.40)$$

$$\alpha' \Sigma_{12} \gamma - \mu = 0. \quad (1.41)$$

Wir sehen, dass für λ und μ folgendes gilt

$$\lambda = \mu = \alpha' \Sigma_{12} \gamma. \quad (1.42)$$

Wir schreiben nun die Gleichungen (1.35) und (1.36) neu an

$$\Sigma_{12} \gamma = \lambda \Sigma_{11} \alpha \quad (1.43)$$

$$\alpha' \Sigma_{12} = \mu \gamma' \Sigma_{22} \quad (1.44)$$

Durch transponieren der Gleichung (1.44) auf beiden Seiten erhalten wir

$$(\alpha' \Sigma_{12})' = \Sigma'_{12} \alpha = \mu \Sigma_{22} \gamma = \mu \Sigma'_{22} \gamma = (\mu \gamma' \Sigma_{22})', \quad (1.45)$$

wobei $\Sigma_{22} = \Sigma'_{22}$ gilt.

Mit $\mu = \lambda$ und $\Sigma'_{12} = \Sigma_{21}$ kommen wir zu der Gleichung

$$\Sigma_{21} \alpha = \lambda \Sigma_{22} \gamma. \quad (1.46)$$

Unter der Annahme dass Σ_{22} eine nichtsinguläre quadratische Matrix ist, können wir nun Gleichung (1.46) neu anschreiben

$$\gamma = \frac{1}{\lambda} \Sigma_{22}^{-1} \Sigma_{21} \alpha. \quad (1.47)$$

Wenn wir nun das Ergebnis für γ aus Gleichung (1.47) einsetzen in Gleichung (1.43), so erhalten wir

$$\Sigma_{12} \left(\frac{1}{\lambda} \Sigma_{22}^{-1} \Sigma_{21} \alpha \right) = \lambda \Sigma_{11} \alpha. \quad (1.48)$$

Durch Subtraktion von $\lambda \Sigma_{11} \alpha$ können wir nun zusammenfassend unser Ergebnis anschreiben

$$\Sigma_{12} \left(\frac{1}{\lambda} \Sigma_{22}^{-1} \Sigma_{21} \alpha \right) - \lambda \Sigma_{11} \alpha = 0, \quad (1.49)$$

$$\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \alpha - \lambda^2 I \alpha = 0, \quad (1.50)$$

$$(\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} - \lambda^2 I) \alpha = 0, \quad (1.51)$$

wobei wir annehmen, dass Σ_{11}^{-1} existiert.

Die Produktmatrix

$$\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \quad (1.52)$$

ist eine quadratische, im Allgemeinen nicht symmetrische Matrix der Dimension $p_1 \times p_1$. Einer ihrer Eigenwerte lautet λ^2 und der dazugehörige Eigenvektor ist α . Da nun $\lambda = \alpha'\Sigma_{12}\gamma$ und $\rho = \alpha'\Sigma_{12}\gamma$, ist der Eigenwert λ^2 das Quadrat der maximalen kanonische Korrelation zwischen den beiden Variablensätzen X_1 und X_2 .

1.6.2 Eigenwerte

Die Eigenwerte erhalten wir indem wir die Determinante von

$$|\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} - \lambda^2 I| \quad (1.53)$$

Null setzen und nach λ^2 lösen. Wir schreiben die Gleichungen (1.35) und (1.36) neu an. Durch Vertauschen der Summanden, Ersetzen der Rechtsmultiplikation von α' bzw. γ' durch eine Linksmultiplikation von α bzw. γ und dem Ergebnis das $\lambda = \mu$ aus Gleichung (1.42) erhalten wir

$$-\lambda\Sigma_{11}\alpha + \Sigma_{12}\gamma = 0, \quad (1.54)$$

$$\Sigma_{12}\alpha - \lambda\Sigma_{22}\gamma = 0. \quad (1.55)$$

Wir schreiben das Gleichungssystem in Matrixnotation neu an

$$\begin{bmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & -\lambda\Sigma_{22} \end{bmatrix} \begin{bmatrix} \alpha \\ \gamma \end{bmatrix} = 0. \quad (1.56)$$

Auf Grund der Gleichungen (1.29) und (1.30) suchen wir eine nichttriviale Lösung. Die Matrix muss singular sein

$$\begin{vmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & -\lambda\Sigma_{22} \end{vmatrix} = 0. \quad (1.57)$$

Da Σ die Dimension $(p_1+p_2) \times (p_1+p_2)$ hat, ist die Determinante ein Polynom vom Grad $p_1 + p_2$. Um das zu demonstrieren, verwenden wir den Laplaceschen Erweiterungssatz für die Minoren der ersten p_1 Spalten. Ein Term ist somit

$$|-\lambda\Sigma_{11}| \cdot |-\lambda\Sigma_{22}| = (-\lambda)^{p_1+p_2} |\Sigma_{11}| \cdot |\Sigma_{22}|. \quad (1.58)$$

Alle anderen Terme in der Erweiterung sind von geringerem Grad von λ , weil eine oder mehrere Reihen jedes Minoren in den ersten p_1 Spalten kein λ enthält. Zudem handelt es sich bei der Supermatrix Σ um die Kovarianzmatrix von $[X_1, X_2]$. Wir haben angenommen, dass Σ positiv definit ist, gilt

$$|\Sigma_{11}| \cdot |\Sigma_{22}| \neq 0. \quad (1.59)$$

Damit ist das Polynom aus Gleichung (1.57) von Grad $p_1 + p_2$ und hat $p_1 + p_2$ Wurzeln. Wir definieren die Wurzeln mit $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{p_1+p_2}$. Da wir die maximale Korrelation suchen, setzen wir $\lambda = \lambda_1$. Sei α_1, γ_1 Lösungen für Gleichung (1.56) mit $\lambda = \lambda_1$, dann sei $U_1 = \alpha_1' X_1$ und $V_1 = \gamma_1' X_2$. U_1 und V_1 sind die normalisierten Linearkombinationen von X_1 und X_2 mit maximaler Korrelation.

Wir werden jetzt zeigen, dass wir insgesamt $m = \min(p_1, p_2)$ Eigenwerte, die den quadrierten kanonischen Korrelationen entsprechen, erhalten. Es sei $A = (\alpha_1, \alpha_2, \dots, \alpha_m)$ und $\Gamma_1 = (\gamma_1, \gamma_2, \dots, \gamma_m)$ und

$$\Lambda = \begin{bmatrix} \lambda_{(1)} & 0 & \cdots & 0 \\ 0 & \lambda_{(2)} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_{(m)} \end{bmatrix}. \quad (1.60)$$

A und Γ_1 sind somit Matrizen, in deren Spalten die unterschiedlichen Gewichtsvektoren enthalten sind und es gilt

$$\dim A = p_1 \times m, \quad (1.61)$$

$$\dim \Gamma_1 = p_2 \times m. \quad (1.62)$$

Mit den Gleichungen (1.33) können wir zusammenfassen

$$A' \Sigma_{11} A = I. \quad (1.63)$$

Da Σ_{11} von Rang p_1 ist und I von Rang m , haben wir $m \leq p_1$. Wir werden nun zeigen, dass $m < p_1$ zu einem Widerspruch zu den geforderten Bedingungen führt. Da $A' \Sigma_{11}$ von Dimension $m \times p_1$ ist, existiert eine $p_1 \times (p_1 - m)$ Matrix E (von Rang

$p_1 - m$) sodass $A'\Sigma_{11}E = 0$ gilt. E liegt somit im Kern von $A'\Sigma_{11}$. Ähnlich gibt es auch eine $p_2 \times (p_2 - m)$ Matrix F (von Rang $p_2 - m$) sodass $\Gamma'_1\Sigma_{22}F = 0$ gilt. Wir haben also $\Gamma'_1\Sigma_{21}E = \Lambda A'\Sigma_{11}E = 0$ und $A'\Sigma_{12}F = \Lambda\Gamma'_1\Sigma_{22}F = 0$. Da E von Rang $p_1 - m$, ist $E'\Sigma_{11}E$ nichtsingulär (wenn $m < p_1$). Ähnlich gilt für $F'\Sigma_{22}F$. Somit gibt es zumindestens eine Wurzel von

$$\begin{vmatrix} -\nu E'\Sigma_{11}E & E'\Sigma_{12}F \\ F'\Sigma_{12}E & -\nu F'\Sigma_{22}F \end{vmatrix} = 0, \quad (1.64)$$

da $|E'\Sigma_{11}E| \cdot |F'\Sigma_{22}F| \neq 0$. Es existieren somit die Vektoren a und b sodass

$$E'\Sigma_{12}Fa = \nu E'\Sigma_{11}Ea, \quad (1.65)$$

$$F'\Sigma_{21}Eb = \nu F'\Sigma_{22}Fb. \quad (1.66)$$

Es sei $Ea = g$ und $Fb = h$. Wir wollen nun zeigen dass ν , g und h eine neue Lösung $\lambda_{(m+1)}$, $\alpha_{(m+1)}$ und $\gamma_{(m+1)}$ bilden. Sei $\Sigma_{11}^{-1}\Sigma_{12}h = k$. Da $A'\Sigma_{11}k = A'\Sigma_{12}Fb = 0$, ist k somit orthogonal zu den Zeilen von $A'\Sigma_{11}$ und somit eine Linearkombination der Spalten von E . Ec sei diese Linearkombination. Damit kann die Gleichung $\Sigma_{12}h = \Sigma_{11}k$ neu angeschrieben werden als

$$\Sigma_{12}Fb = \Sigma_{11}Ec. \quad (1.67)$$

Durch Multiplikation von E' erhalten wir

$$E'\Sigma_{12}Fb = E'\Sigma_{11}Ec. \quad (1.68)$$

Da $E'\Sigma_{11}E$ nichtsingulär ist, erhalten wir das Resultat, dass $c = \nu a$ und damit $k = \nu g$.

Damit gilt

$$\Sigma_{12}h = \nu\Sigma_{11}Eg. \quad (1.69)$$

In ähnlicher Weise, zeigen wir

$$\Sigma_{21}g = \nu\Sigma_{22}Eh. \quad (1.70)$$

Damit ist $\nu = \lambda_{(m+1)}$, $g = \alpha_{(m+1)}$ und $h = \gamma_{(m+1)}$ eine weitere Lösung. Dies ist jedoch vereinbar mit der Annahme, dass $\lambda_{(m)}$, $\alpha_{(m)}$ und $\gamma_{(m)}$ die letzte Lösung ist. Somit muss gelten $m = p_1$. Die Bedingungen für die λ , α und γ können zusammengefasst werden

als

$$A'\Sigma_{11}A = I, \quad (1.71)$$

$$A'\Sigma_{12}\Gamma_1 = \Lambda, \quad (1.72)$$

$$\Gamma_1'\Sigma_{22}\Gamma_1 = I. \quad (1.73)$$

Es sei $\Gamma_2 = (\gamma_{(p_1+1)}, \dots, \gamma_{(p_2)})$ eine $p_2 \times (p_2 - p_1)$ Matrix die folgende Eigenschaften erfüllt

$$\Gamma_2'\Sigma_{22}\Gamma_1 = 0, \quad (1.74)$$

$$\Gamma_2'\Sigma_{22}\Gamma_2 = I. \quad (1.75)$$

Jede beliebige Γ_2 kan multipliziert werden von recht mit einer beliebigen $(p_2 - p_1) \times (p_2 - p_1)$ orthogonalen Matrix. Diese Matrix wird den Spalten nach aufgebaut: $\gamma_{(p_1+1)}$ ist ein orthogonaler Vektor zu $\Sigma_{22}\Gamma_1$ und normalisiert sodass gilt $\gamma'_{(p_1+1)}\Sigma_{22}\gamma_{(p_1+1)} = 1$; $\gamma_{(p_1+2)}$ ist ein orthogonaler Vektor zu $\Sigma_{22}(\Gamma_1 \gamma_{(p_1+1)})$ und normalisiert sodass gilt $\gamma'_{(p_1+2)}\Sigma_{22}\gamma_{(p_1+2)} = 1$; und so fort. Es sei $\Gamma = (\Gamma_1 \Gamma_2)$; Diese quadratische Matrix ist nichtsingulär da $\Gamma'\Sigma_{22}\Gamma = I$. Wenn wir die Determinante betrachten

$$\begin{vmatrix} A' & 0 \\ 0 & \Gamma_1' \\ 0 & \Gamma_2' \end{vmatrix} \cdot \begin{vmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{vmatrix} \cdot \begin{vmatrix} A & 0 & 0 \\ 0 & \Gamma_1 & \Gamma_2 \end{vmatrix} \quad (1.76)$$

$$= \begin{vmatrix} -\lambda I & \Lambda & 0 \\ \Lambda & -\lambda I & 0 \\ 0 & 0 & -\lambda I \end{vmatrix} \quad (1.77)$$

$$= (-\lambda)^{p_2-p_1} \begin{vmatrix} -\lambda I & \Lambda \\ \Lambda & -\lambda I \end{vmatrix} \quad (1.78)$$

$$= (-\lambda)^{p_2-p_1} |-\lambda I| \cdot |-\lambda I - \Lambda(-\lambda I)^{-1}\Lambda| \quad (1.79)$$

$$= (-\lambda)^{p_2-p_1} |\lambda^2 I - \Lambda^2| \quad (1.80)$$

$$= (-\lambda)^{p_2-p_1} \prod (\lambda^2 - \lambda_i^2). \quad (1.81)$$

Das obige Polynom entspricht bis auf eine Konstante der Determinanten

$$\begin{vmatrix} -\lambda\Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda\Sigma_{22} \end{vmatrix}. \quad (1.82)$$

Die Wurzeln aus Gleichung (1.57) sind die Wurzeln auch aus Gleichung (1.82) und somit $\lambda = \pm\lambda_i$, $i = 1, \dots, p_1$ und $\lambda = 0$ mit Vielfachheit $p_2 - p_1$. Damit ist $(\lambda_1, \dots, \lambda_{p_1+p_2}) = (\lambda_1, \dots, \lambda_{p_1}, 0, \dots, 0, -\lambda_{p_1}, \dots, -\lambda_1)$. Um zu zeigen, dass $\lambda^{(i)2}$, $i = 1, \dots, p_1$ in $\{\lambda_i^2\}_{i=1, \dots, p_1}$ ist, müssen wir nur zeigen, dass $\lambda^{(i)}$ nichtnegativ ist (und damit eines der λ_i , $i = 1, \dots, p_1$). Wir haben beobachtet, dass gilt

$$\Sigma_{12}\gamma^{(r)} = -\lambda^{(r)}\Sigma_{11}(-\alpha^{(r)}), \quad (1.83)$$

$$\Sigma_{21}(-\alpha^{(r)}) = -\lambda^{(r)}\Sigma_{22}\gamma^{(r)}, \quad (1.84)$$

wobei gilt, dass wenn $\lambda^{(r)}, \alpha^{(r)}, \gamma^{(r)}$ eine Lösung ist, dann auch $-\lambda^{(r)}, -\alpha^{(r)}, \gamma^{(r)}$. Wenn $\lambda^{(r)}$ negativ ist, dann ist $-\lambda^{(r)}$ nicht negativ und $-\lambda^{(r)} \geq \lambda^{(r)}$. Da jedoch $\lambda^{(r)}$ das Maximum ist, muss $\lambda^{(r)} \geq -\lambda^{(r)}$ gelten und wir haben somit $\lambda^{(r)} \geq 0$. Es gilt somit $\lambda_{(i)} = \lambda_i$. Sei

$$U = \begin{pmatrix} U_1 \\ \vdots \\ U_{p_1} \end{pmatrix} = A'X_1, \quad (1.85)$$

$$V_1 = \begin{pmatrix} V_1 \\ \vdots \\ V_{p_1} \end{pmatrix} = \Gamma'_1 X_2, \quad (1.86)$$

$$V_2 = \begin{pmatrix} V_{p_1+1} \\ \vdots \\ V_{p_2} \end{pmatrix} = \Gamma'_2 X_2. \quad (1.87)$$

$$(1.88)$$

Die Komponenten von U sind die Menge der Linarkombinationen, die Komponenten

$V = (V'_1, V'_2)'$ die Linearkombination für X_2 . Wir haben

$$\mathbb{E} \begin{pmatrix} U \\ V_1 \\ V_2 \end{pmatrix} (U', V'_1, V'_2) = \begin{pmatrix} A' & 0 \\ 0 & \Gamma'_1 \\ 0 & \Gamma'_2 \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} A & 0 & 0 \\ 0 & \Gamma_1 & \Gamma_2 \end{pmatrix} = \begin{pmatrix} I_{p_1} & \Lambda & 0 \\ \Lambda & I_{p_1} & 0 \\ 0 & 0 & I_{p_2-p_1} \end{pmatrix} \quad (1.89)$$

wobei

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_{p_1} \end{pmatrix}. \quad (1.90)$$

Wir können nun verifizieren, dass U_1 und V_1 maximal korrelieren. Die Linearkombinationen $a'U = (a'A)X_1$ und $b'V = (b'\Gamma')X_2$ sind normalisiert durch $a'a = 1$ und $b'b = 1$. Da A und Γ nichtsingulär sind, kann jeder Vektor α neu angeschrieben werden als Aa und jeder Vektor γ als Γb . Damit kann jede Linearkombination $\alpha'X_1$ und $\gamma'X_2$ neu angeschrieben werden als $a'U$ und $b'V$. Die Korrelation zwischen ihnen ist

$$a' \begin{pmatrix} \Lambda & 0 \end{pmatrix} b = \sum_{i=1}^{p_1} \lambda_i a_i b_i. \quad (1.91)$$

Sei $\lambda_i a_i / \sqrt{\sum (\lambda_i a_i)^2} = c_i$. Dann ist das Maximum von

$$a' \begin{pmatrix} \Lambda & 0 \end{pmatrix} b = \sqrt{\sum (\lambda_i a_i)^2} \sum c_i b_i \quad (1.92)$$

in Abhängigkeit zu b für $b_i = c_i$, da $\sum c_i b_i$ der Kosinus des Winkels zwischen dem Vektor b und $(c_1, \dots, c_{p_1}, 0, \dots, 0)$ ist. Dann wird (1.91) zu

$$\sqrt{\sum \lambda_i^2 a_i^2} = \sqrt{\sum_2^{p_1} (\lambda_i^2 - \lambda_1^2) a_i^2 + \lambda_1^2}, \quad (1.93)$$

und wird maximiert wenn man $a_i = 0$, $i = 2, \dots, p_1$ setzt. Damit sind die Korrelation maximierenden Linearkombinationen U_1 und V_1 . Indem wir verifizieren, dass u_2 und V_2 das zweite Paar für die kanonische Korrelation bilden, können wir einen Verlust in der Korrelation zwischen U_1 und einer Linearkombination $a'U$ feststellen. Da bedeutet $0 = \mathbb{E}U_1 a'U = \mathbb{E}U_1 \sum_{i=2}^{p_1} a_i U_i = a_1$ und für die Korrelation zwischen V_1 und $b'V$ erhalten wir $0 = b_1$.

Theorem 1.6.1. *Die kanonische Korrelationen sind invariant zu Transformationen $X_i^* = C_i X_i$, wobei C_i nichtsingulär ist, $i = 1, 2$ und jede Funktion von Σ nicht abhängig ist von seinen kanonischen Korrelationen.*

Hinweis Die kanonische Korrelation entspricht dem Winkel zwischen zwei Linearkombination. Da der Winkel unabhängig von der Koordinatendarstellung ist, bleiben die kanonische Korrelationen invariant zu einer Transformation mit einer nichtsingulären Matrix.

Beweis.

$$0 = \begin{vmatrix} -\lambda C_1 \Sigma_{11} C_1' & C_1 \Sigma_{12} C_2' \\ C_2 \Sigma_{21} C_1' & -\lambda C_2 \Sigma_{22} C_2' \end{vmatrix} = \begin{vmatrix} C_1 & 0 \\ 0 & C_2 \end{vmatrix} \cdot \begin{vmatrix} -\lambda \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda \Sigma_{22} \end{vmatrix} \cdot \begin{vmatrix} C_1' & 0 \\ 0 & C_2' \end{vmatrix}, \quad (1.94)$$

□

Wenn wir die Determinante entwickeln, erhalten wir ein Polynom der Ordnung $\max(p_1, p_2)$. Wir bekommen somit $\max(p_1, p_2)$, wobei $\min(p_1, p_2)$ davon, nicht negativ sind. Alle nicht negativen Eigenwerte sind die Quadrate der kanonische Korrelation.

1.6.3 Eigenvektoren

Wenn man mal die Eigenwerte hat, muss man die dazugehörigen Eigenvektoren bestimmen. Zu den $\min(p_1, p_2)$ nicht negative Eigenwerten müssen die Eigenvektoren a_s mit $s = 1, 2, \dots, \min(p_1, p_2)$ ermittelt werden. Bei der Normierung ist die Bedingung

$$a_s^{*'} \Sigma_{11} a_s^* = 1 \quad s = 1, 2, \dots, \min(p_1, p_2) \quad (1.95)$$

zu erfüllen. Für die Normierung, berechnen wir zunächst

$$a_s' \Sigma_{11} a_s = m^2. \quad (1.96)$$

Wenn wir nun von links und von rechts durch m dividieren, erhalten wir

$$\frac{1}{m} a_s' \Sigma_{11} a_s \frac{1}{m} = 1. \quad (1.97)$$

Die von uns gesuchten Eigenvektoren lauten somit

$$a_s^* = \frac{1}{m} a_s. \quad (1.98)$$

Durch Gleichung (1.47) erhalten wir zum a_s^* das passende b_s^* .

1.7 Beziehung zum Skalkomponentenmodellstruktur

Das Wissen über die Existenz von kanonischen Korrelationen von Wert Null kann auch interpretiert werden als Information über die Existenz eines Skalkomponentenmodells (SCM) in der Struktur des Vektor Autoregressive Moving Average Prozesses (VARMA). Die Kenntniss über dieses SCM führt zu Eigenschaften über die besondere Struktur in der Parameterisierung der ARMA Darstellung für den Prozess.

Definition (*Skalkomponentenmodell*) Ein Skalkomponentenmodell (SCM) [Tiao and Tsay, 1989] der Ordnung (p^*, q^*) mit $p^* \leq p$ und $q^* \leq q$, existiert für einen Prozess $\{Y_t\}$, wenn die Linearkombination $z_t = a'Y_t$ existiert, sodass

$$z_t - \sum_{j=1}^{p^*} a'\Phi_j Y_{t-j} = a'\varepsilon_t - \sum_{j=1}^{q^*} a'\Theta_j \varepsilon_{t-j}, \quad (1.99)$$

wobei dann

$$u_t = z_t - \sum_{j=1}^{p^*} a'\Phi_j Y_{t-j} \quad (1.100)$$

Folgende Bedingungen müssen also gelten:

$$a'\Phi_{p^*} \neq 0 \quad \forall 0 \leq p^* \leq p, \quad (1.101)$$

$$a'\Phi_l = 0 \quad \forall l = p^* + 1, \dots, p, \quad (1.102)$$

$$a'\Theta_{q^*} \neq 0 \quad \forall 0 \leq q^* \leq q, \quad (1.103)$$

$$a'\Theta_l = 0 \quad \forall l = 1^* + 1, \dots, q. \quad (1.104)$$

Eine etwas allgemeiner gehaltene Definition ist, dass $z_t = a'Y_t$ die Struktur eines Skalkomponentenmodells von Grad (p^*, q^*) hat, wenn es k -dimensionale Vektoren b_1, \dots, b_{p^*} gibt, sodass die Linearkombination

$$u_t = z_t - \sum_{j=1}^{p^*} b'_j Y_{t-j} \quad (1.105)$$

unkorreliert zu ε_{t-j} für $j > q^*$ aber mit ε_{t-q^*} korreliert. Das bedeutet, dass die Existenz eines Skalarkomponentenmodell der Ordnung (p^*, q^*) mindestens eine kanonische Korrelation von Wert Null zwischen

$$Y_{p^*,t} = \begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p^*} \end{pmatrix} \quad (1.106)$$

und $Y_{n,t-j-1}$ für $j \geq q^*$. Somit kann man durch die kanonische Korrelationsanalyse die Existenz von Skalarmodellen unterschiedlicher Ordnung nachweisen.

[Tiao and Tsay, 1989] präsentierten in ihrem Ansatz zu Spezifizierung eines Vektor ARMA Modells durch Verwendung von SCM. Dabei wird ein Set von k Skalarkomponenten $z_{it} = a'_i Y_t$ der Ordnung (p_i, q_i) für $i = 1, \dots, k$ gesucht, sodass die Ordnung

$$p_i + q_i \quad (1.107)$$

möglichst gering ist. Wenn einmal so ein Set von Skalarkomponenten gefunden ist, so wird die Modellstruktur des ARMA Modells durch

$$AY_t - \sum_{j=1}^p B_j Y_{t-j} = A\varepsilon_t - \sum_{j=1}^q G_j \varepsilon_{t-j} \quad (1.108)$$

präzisiert, wobei gilt

$$A = [a_1, \dots, a_k]' \quad (1.109)$$

$$B_j = A\Phi_j \quad \forall j = 1, \dots, p \quad (1.110)$$

$$G_j = A\Theta_j \quad \forall j = 1, \dots, q \quad (1.111)$$

$$p = \max \{p_i\} \quad (1.112)$$

$$q = \max \{q_i\}. \quad (1.113)$$

Zudem ist in Gleichung (1.108) die i -te Zeile der Matrix B_j Null für $j > p_i$ und die i -te Zeile der Matrix G_j Null für $j > q_i$. Somit gelangt man durch Linksmultiplikation der Gleichung (1.108) mit A^{-1} zu einem ARMA(p, q) Modell für das Y_t die Standardform annimmt und die Koeffizientenmatrizen Φ_j und Θ_j nicht von vollen Rang sind.

Auf der anderen Seite wenn man den Faktor $A^{-1}A$ vor Y_{t-j} und ε_{t-j} einfügt in Gleichung (1.108), so erhält man das Vektor ARMA(p, q) Modell in Standardform für den transformierten Prozess $Z_t = AY_t$

$$Z_t - \sum_{j=1}^p \Phi_j^* Z_{t-j} = e_t - \sum_{j=1}^q \Pi_j^* e_{t-j}, \quad (1.114)$$

wobei gilt

$$\Phi_j^* = B_j A^{-1} = A \Phi_j A^{-1} \quad (1.115)$$

$$\Theta_j^* = G_j A^{-1} = A \Theta_j A^{-1} \quad (1.116)$$

$$e_t = A \varepsilon_t. \quad (1.117)$$

Die ARMA Darstellung aus Gleichung (1.117) für den transformierten Prozess Z_t beinhaltet, dass die i -te Reihe für Φ_j^* Null ist für $j > p_i$ und die i -te Reihe für Θ_j^* Null ist für $j > q_i$. Zusätzlich können Elemente in der i -ten Reihe der Matrix Θ_j^* für $j = 1, \dots, q_i$ den Wert Null annehmen auf Grund von Redundanzen zwischen den Parametern der AR und MA Matrizen in (1.117).

[Tiao and Tsay, 1989] schlagen vor im ersten Schritt den Skalarkomponenten Prozess $Z_t = AY_t$ und dessen Ordnungen (p_i, q_i) durch die kanonische Korrelationsanalyse zu identifizieren und dann das ARMA Modell für den transformierten Prozess Z_t zu schätzen, indem man das Wissen über die Parameterstruktur, also in welchen Zeilen der Wert Null sein muss, für Gleichung (1.117) verwendet.

Kapitel 2

Verschachtelte rangreduzierte autorogressive Modell im stationären Fall

2.1 Einleitung

Es sei Y_t eine m -dimensionale Zeitreihe, die wir wie folgt anschreiben

$$Y_t = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{mt} \end{bmatrix}. \quad (2.1)$$

Wir betrachten ein autoregressives Modell für die Zeitreihe Y_t mit der Form

$$\Phi(L)Y_t = \varepsilon_t \quad (2.2)$$

wobei

$$\Phi(L) = I_m - \Phi_1 L - \Phi_2 L^2 - \dots - \Phi_p L^p. \quad (2.3)$$

Dabei gilt

$$\dim I_m = m \times m, \quad (2.4)$$

$$\dim \Phi_j = m \times m \quad \forall j = 1, \dots, p. \quad (2.5)$$

Definition Wir definieren den backward shift Operator $L \in \mathbb{Z}$ durch

$$L(y_t | t \in \mathbb{Z}) = (y_{t-1} | t \in \mathbb{Z}). \quad (2.6)$$

Der backward shift Operator ist linear und bijektiv.

Hinweis Der backward shift Operator wäre nicht bijektiv, wenn er auf \mathbb{N} definiert wäre.

Als Fehlerterm definieren wir ε als ein m -dimensionales White Noise (WN) mit den Eigenschaften

$$\mathbb{E}(\varepsilon_t) = 0, \quad (2.7)$$

$$\text{cov}(\varepsilon_t) = \Omega, \quad (2.8)$$

wobei Ω positiv definit ist.

Annahme

$$\mathbb{E}(\varepsilon_t^4) < \infty \quad (2.9)$$

und Stationarität für den Prozess, also

$$\det \{\Phi(L)\} \neq 0 \quad \forall |L| \leq 1 \quad L \in \mathbb{C}. \quad (2.10)$$

Definition *Stationarität* Ein stochastischer Prozess heisst stationär im weiteren Sinne, wenn

1. $\mathbb{E}y_t' y_t < \infty \quad \forall t \in \mathbb{Z}$,
2. $\mathbb{E}y_t = m = \text{const} \quad \forall t \in \mathbb{Z}$ und
3. $\gamma(s, t) = \gamma(s + r, t + r)$ gilt für alle $r, s, t \in \mathbb{Z}$.

Hinweis Um die eingeschwungene Lösung zu erhalten, möchten wir die Inverse zu $\Sigma \Phi_j L^j$. Die Potenzreihe in L lässt sich wie die Potenzreihe $\Phi(z) := \sum_{j=0}^{\infty} \Phi_j z^j, z \in \mathbb{C}$

multiplizieren. Wir können nun die Potenzreihe $\Phi(z)$ invertieren und erhalten somit

$$\frac{1}{\Phi(z)} = \frac{1}{c\Pi(z - z_i)} = \sum_{j=0}^{\infty} k_j z^j, \quad (2.11)$$

wobei die z_i die Nullstellen des Polynoms $\Phi(z) = c\Pi(z - z_i)$.

$$\frac{1}{(z - z_i)} = \frac{-\frac{1}{z_i}}{-\frac{z}{z_i} + 1} = -\frac{1}{z_i} \sum_{j=0}^{\infty} \left(\frac{z}{z_i}\right)^j = -\frac{1}{z_i} \sum_{j=0}^{\infty} \frac{1}{z_i^j} z^j \quad (2.12)$$

Mit der zusätzlichen Stationaritätsbedingung $\Phi(z) \neq 0$ für $|z| \leq 1$ folgt $\sum k_j^2 < \infty$

Wir verwenden später die Eigenschaft, dass es sich bei y_t um eine stationäre Zeitreihe handelt, um die asymptotische Verteilungstheorie der Parameterschätzer zu entwickeln. Prinzipiell ist Stationarität keine notwendige Bedingung um die asymptotische Verteilungstheorie zu entwickeln.

Wir möchten hier darauf hinweisen, dass wir L sowohl als shift Operator, als auch als komplexe Variable verwendet haben. Grundsätzlich ist die Laurentreihe

$$a(z) = \sum_{j=-\infty}^{\infty} a_j L^j \quad (2.13)$$

im Lagoperator L und die Laurentreihe¹ in der komplexen Variablen L isomorph bezüglich der Multiplikation. Es ist ein etwas umständlicher Weg auszudrücken, dass für sukzessive Anwendung von linearen Transformationen als auch von Multiplikationen von Laurentreihen in der komplexen Variablen L der folgende Zusammenhang gilt:

$$h(L) = k(L)l(L) = \left(\sum_{i=-\infty}^{\infty} k_i L^i\right) \left(\sum_{j=-\infty}^{\infty} l_j L^j\right) = \sum_{i=-\infty}^{\infty} \left(\sum_{i=-\infty}^{\infty} k_j l_{i-j}\right) L^i = \sum_{i=-\infty}^{\infty} h_i L^i \quad (2.14)$$

Daher kann der Prozess aus Gleichung (2.2) auch angeschrieben werden als ein unendlicher MA Prozess mit der Form

$$Y_t = \Psi(L) \varepsilon_t = \varepsilon_t + \sum_{i=1}^{\infty} \Psi_i \varepsilon_{t-i} \quad (2.15)$$

¹siehe A.2.2

wobei für Ψ gilt

$$\Psi(L) = \Phi(L)^{-1} = I_m + \Psi_1 L + \Psi_2 L^2 + \dots \quad (2.16)$$

Das Problem mit multivariaten Zeitreihen und deren Darstellung ist, dass mit der Anzahl der Lags die Anzahl der zu schätzenden Parametern quadratisch zunimmt. Die Parametermatrizen sind quadratisch von Dimension $m \times m$ auf Grund der multivariaten Zeitreihe. Bei Grad p wächst die Zahl der zu schätzenden Parameter auf pm^2 an. Es stehen aber zur Schätzung von pm^2 Parameter nur mT Daten zur Verfügung. Es kann sich ein Effizienzverlust einstellen, wenn pm^2 Parameter geschätzt werden müssen im Vergleich zu dem NRRAR Modell und seinem eingeschränkten Parameterraum. Zusätzlich kann es zum Overfitting kommen. Dies ist unabhängig von der Größe von p . All diesen Vorteilen steht die Notwendigkeit nach a priori Informationen bzw. des Testens des Modells gegenüber.

Um die Anzahl der Parameter, die in $\Phi(L)$ enthalten sind, zu reduzieren, betrachten [Velu et al., 1986] das reduced-rank autoregressive model; einen ähnlichen Ansatz wurde zuvor schon von [Reinsel, 1983] getätigt. Sein Vorhaben war die Anzahl der Parameter zu reduzieren, aber auch eine detailliertere Beschreibung der Struktur der multivariaten Zeitreihe zu erarbeiten. Wir betrachten hier das NRRAR Modell, welches eine Verallgemeinerung des RRAR Modells ist. Wir werden die Frage behandeln wie kanonische Korrelationsanalyse verwendet werden kann um den Rang der Parametermatrizen des NRRAR Modells zu bestimmen. Wir werden im folgenden über die Darstellung, die Parametrisierung, der Schätzung der Modellparameter und über die asymptotische Verteilung sprechen.

Die Entwicklung dieses Modells basiert auf der Arbeit von [Tiao and Tsay, 1985]. In dieser Arbeit werden wir aber das Hauptaugenmerk auf Modellbildung und Schätzung der Modellparameter für das NRRAR Modell für die Zeitreihe Y_t legen. Es wird nochmals auf die Stationarität des Prozesses Y_t in Gleichung (2.2) hingewiesen. Damit kann die Theorie über die asymptotische Verteilung der Parameterschätzer entwickelt werden.

2.2 Das verschachtelte rangreduzierte autoregressive Modell

Wir gehen nun genauer auf die Parameterstruktur aus Gleichung (2.2).

Definition Wir bezeichnen mit r_j den Rang der Matrix A_j , also

$$\text{rang}(\Phi_j) = r_j \quad \forall j = 1, \dots, p. \quad (2.17)$$

Annahme Wir nehmen an, dass für r_j

$$r_j \geq r_{j+1} \quad \forall j = 1, \dots, p \quad (2.18)$$

gilt.

Somit gilt existiert für jedes j die Matrizen A_j und B_j mit folgenden Eigenschaften

$$\dim A_j = m \times r_j, \quad (2.19)$$

$$\dim B_j = r_j \times m, \quad (2.20)$$

$$\text{rang} A_j = \text{rang} B_j = r_j, \quad (2.21)$$

sodass gilt

$$\Phi_j = A_j B_j \quad (2.22)$$

Annahme Wir nehmen an, dass

$$\text{im}(A_j) \supset \text{im}(A_{j+1}) \quad j = 1, \dots, p-1 \quad (2.23)$$

gilt. Dies impliziert die Forderung nach $r_j \geq r_{j+1}$.

Hinweis Es gilt

$$\text{im}(A_j) \supset \text{im}(\Phi_j) \quad (2.24)$$

da die Φ_j eine Linearkombination der Basisvektoren von A_j sind. Es kann somit nur ein Unterraum von A_j aufgespannt werden kann. Dadurch dass $\text{rang} A_j = \text{rang} \Phi_j$ gilt,

muss auch gelten

$$\text{im}(A_j) = \text{im}(\Phi_j) \quad (2.25)$$

und damit

$$\text{im}(\Phi_j) \supset \text{im}(A_j) \quad (2.26)$$

Mit Gleichung (2.22) kann man Gleichung (2.2) anschreiben als

$$Y_t = \sum_{j=1}^p A_j B_j Y_{t-j} + \varepsilon_t \quad (2.27)$$

Das Model aus Gleichung (2.27) wird auch verschachteltes rangreduziertes autoregressives Modell, oder auch NRRAR Modell (nested reduced-rank autoregressive model) genannt.

Hinweis Das NRRAR Modell ist eine Verallgemeinerung des RRAR Modells, welches schon von [Velu et al., 1986] behandelt wurde. Durch

$$r_j \equiv r < m \quad (2.28)$$

und

$$A_j = A \quad (2.29)$$

erhalten wir das RRAR Modell

$$Y_t = A \sum_{j=1}^p B_j Y_{t-j} + \varepsilon_t \quad (2.30)$$

Ein anderer Spezialfall des Modells in Gleichung (2.27), in dem

$$r_j = r_{k_1} = m \quad j \leq k_1 < p, \quad (2.31)$$

$$r_j = r_{k_2} < m \quad j > k_1 \quad (2.32)$$

gesetzt wurde, wurde schon von [Anderson, 1951] im Zusammenhang mit multivariater Regression betrachtet.

Eine Interpretation the Modells in Gleichung (2.27) wäre, dass jedes Φ_j r_j Kanäle für die Übertragung von Information der vergangenen Y_{t-j} enthält. Durch $r_j \geq r_{j+1}$ wird Information aus jüngerer Vergangenheit durch mehr Kanäle übertragen. $im(A_j) \supset im(A_{j+1})$ zeigt an, dass die Kanäle der länger zurückliegenden Vergangenheit, Unterkanäle der jüngeren sind. Somit hängen Teile von Y_t von der Vergangenheit ab, andere nicht.

Um eindeutige Parameter zu bekommen, ist eine Normalisierung der Koeffizientenmatrizen A_j und B_j von Nöten. Eine Normalisierung ähnlich zu der in [Ahn and Reinsel, 1987] besprochen, wird später entwickelt werden.

2.2.1 Bestimmung und Spezifikation der Ränge

Um Information über den Rang der Matrizen Φ_j im NRRAR Modell aus Gleichung (2.27) zu erhalten, verwenden wir den Ansatz mittels kanonisch Korrelationanalyse, wie schon in [Tiao and Tsay, 1985], [Tsay and Tiao, 1985] und [Cooper and Wood, 1982].

Definition Sei $Y_{k,t} = [Y'_t, Y'_{t-1}, \dots, Y'_{t-k}]'$. Somit ist $Y_{k,t}$ ein Spaltenvektor, in dem die Realisierungen der m -dimensionalen Zeitreihe vom Zeitpunkt t bis zurück bis zum Zeitpunkt $t - k$ enthalten sind.

Wir werden zeigen, dass in dem NRRAR Modell aus Gleichung (2.27) die Anzahl der kanonischen Korrelationen von Wert Null zwischen $Y_{k,t}$ und $Y_{k,t-1}$ mit k steigt und direkt im Zusammenhang steht mit dem Rang der Koeffizientenmatrix Φ_{k+1} . Dabei ist $Y_{k,t-1}$ ein Spaltenvektor, in dem die Werte zum vorangegangenen Zeitpunkt zu $Y_{k,t}$ enthalten sind. Die Anzahl der kanonischen Korrelation zum Wert Null nimmt zu, je länger man in die Vergangenheit zurück geht, bzw je länger die zu vergleichenden Vektoren $Y_{k,t}$ und $Y_{k,t-1}$ werden.

Da $im(A_s) \supset im(A_{s+1})$ gilt, können wir einen Vektor $f \neq 0$ finden für den

$$f' A_s = 0 \implies f' A_k = 0 \quad \forall k \geq s \quad (2.33)$$

gilt. Die Matrix A_s hat die Dimension $m \times r_s$. Durch die Definition, dass A_s von Rang r_s ist, gibt es eine $(m - r_s) \times m$ Matrix F'_s für die gilt

$$F'_s A_s = 0 \quad (2.34)$$

und damit

$$F'_s A_k = 0 \quad \forall k \geq s \quad (2.35)$$

Die Zeilen von F'_s spannen somit den Kern zu A_s auf. Der Rang von F'_s und somit der Defekt von A_s ist $(m - r_s)$. Es gilt

$$F'_s \left(Y_t - \sum_{j=1}^{s-1} \Phi_j Y_{t-j} \right) = F'_s \left(Y_t - \sum_{j=1}^p \Phi_j Y_{t-j} \right) = F'_s \varepsilon_t \quad \forall s = 1, \dots, p \quad (2.36)$$

Beweis.

$$Y_t = \sum_{j=1}^p \Phi_j Y_{t-j} + \varepsilon_t \quad (2.37)$$

$$Y_t = \sum_{j=1}^p A_j B_j Y_{t-j} + \varepsilon_t \quad (2.38)$$

$$Y_t - \sum_{j=1}^p A_j B_j Y_{t-j} = \varepsilon_t \quad (2.39)$$

$$F'_s \left(Y_t - \sum_{j=1}^p A_j B_j Y_{t-j} \right) = F'_s \varepsilon_t \quad (2.40)$$

$$F'_s \left(Y_t - \sum_{j=1}^{s-1} \Phi_j Y_{t-j} \right) = F'_s \varepsilon_t \quad (2.41)$$

Wir haben bei der Gleichungsfolge angenommen, dass gilt

$$F'_s \sum_{j=1}^p A_j B_j Y_{t-j} = F'_s \sum_{j=1}^{s-1} \Phi_j Y_{t-j}. \quad (2.42)$$

Dies gilt auf Grund von

$$F'_s \sum_{j=1}^p A_j B_j Y_{t-j} = \sum_{j=1}^p \overbrace{F'_s A_j}^{=0 \quad \forall p \geq s} B_j Y_{t-j} \quad (2.43)$$

$$= \sum_{j=1}^{s-1} F'_s A_j B_j Y_{t-j} \quad (2.44)$$

$$= F'_s \sum_{j=1}^{s-1} A_j B_j Y_{t-j} \quad (2.45)$$

$$= F'_s \sum_{j=1}^{s-1} \Phi_j Y_{t-j} \quad (2.46)$$

□

Definition

$$Y_{s-1,t-1} = [Y'_{t-1}, \dots, Y'_{t-s}]' \quad (2.47)$$

Y_t kann auch angeschrieben werden als

$$Y_t = \varepsilon_t + \sum_{i=1}^{\infty} \Psi_i \varepsilon_{t-i}. \quad (2.48)$$

Somit ist jedes Y_t nur von ε_i mit $i \leq t$ abhängig. Das bedeutet, dass $Y_{s-1,t-1}$ nur von $\varepsilon_{s-1-i,t-1-i}$ mit $i \geq 0$ abhängt. In Gleichung (2.41) kommt es durch Linearkombination von $Y_{s-1,t}$ zu einer gegenseitigen Aufhebung der Fehlerterme, sodass nur der Fehlerterm ε_t übrig bleibt. Dieser Fehlerterm kommt in keinem der $Y_{s-1,t-1}$ vor. Bei ε_t handelt es sich um Weisses Rauschen. Die Terme ε_i und ε_j sind unabhängig für $i \neq j$. Das Ergebnis in Gleichung (2.41) ist somit unabhängig jeder möglichen Linearkombination von $Y_{s-1,t-1}$. Etwas entscheidendes in Gleichung (2.36) ist die Tatsache, dass dem Y_t kein Faktor bevorsteht. Dies bedeutet, dass die Information aus Y_t immer in Gleichung (2.36) einfließt. Dagegen stehen den Y_{t-j} die Parametermatrizen Φ_j vor.

Es kommt durch die Matrix F_s zu einer linearen Transformation von Rang $m-r_s$, sodass der Erwartungswert der Linearkombination der $Y_{s-1,t}$ aus Gleichung (2.36) unabhängig von $Y_{s-1,t-1}$ ist. Das Ergebnis aus Gleichung (2.41) ist $F'_s \varepsilon_t$, also ein Spaltenvektor der Dimension $(m-r_s) \times 1$. Das bedeutet, dass durch F'_s genau $m-r_s$ Linearkombinationen

entstehen, die unabhängig sind. Es gibt somit mindestens $m - r_s$ kanonische Korrelationen von Wert 0 zwischen $Y_{s-1,t}$ und $Y_{s-1,t-1}$ gibt. Dies ist gleichzusetzen, dass es mindestens $m - r_s$ partielle kanonische Korrelationen von Wert 0 gibt zwischen Y_t und Y_{t-s} gibt bei gegebenen $Y_{t-1}, \dots, Y_{t-s+1}$.

Definition Es sei

$$\Omega_{i,j,k,l} = Cov(Y_{i,j}, Y_{k,l}) \quad (2.49)$$

Hinweis $\Omega_{i,j,k,l}$ ist eine Matrix in der sich die Kreuzkovarianzen befinden.

Wie in Abschnitt 1.2 (siehe Seite 8) beschrieben, müssen wir für die kanonische Korrelationsanalyse zuerst die Matrix V wie folgt bilden Dann gibt es mindestens $m - r_s$ Eigenwert von Wert Null für

$$V(s-1) = \Omega_{s-1,t}^{-1} \Omega_{s-1,t,s-1,t-1} \Omega_{s-1,t-1}^{-1} \Omega_{s-1,t-1,s-1,t}. \quad (2.50)$$

Die kanonischen Korrelationen entsprechen dann den Eigenwerten aus Gleichung (2.50). Da wir gesagt haben, dass es mindestens $m - r_s$ kanonische Korrelationen von Wert 0 geben muss, bedeutet das, dass es mindestens $m - r_s$ Eigenwerte von Wert 0 gibt.

Daher, wenn man nun die kanonische Korrelationsanalyse betrachtet für $Y_{s-1,t}$ und $Y_{s-1,t-1}$ mit $s = 1, 2, \dots$, kann man die NRRAR Modellstruktur erkennen. Anhand von r_s , dem Rang von Φ_s , können wir den Grad p des Polynoms und des AR Modells ermitteln. Da für $s > p$ mindestens m Nullkorrelationen zwischen $Y_{s-1,t}$ und $Y_{s-1,t-1}$, müssen wir das kleinste s ermitteln für das die Gleichung (2.50) m Eigenwert zum Wert 0 hat. Mit $s > p$ ermitteln wir dann den Grad des AR Modells.

Um den Grad und die Struktur des multivariaten AR Prozesses zu bestimmen, müssen wir die Anzahl der kanonischen Korrelationen zwischen $Y_{k,t}$ und $Y_{k,t-1}$ systematisch bestimmen können. Dies erreichen wir durch die Bestimmung der Anzahl der Eigenwerte von Wert Null für $V(k)$, wobei

$$\hat{V}(k) = (\Sigma Y_{k,t} Y_{k,t}')^{-1} (\Sigma Y_{k,t} Y_{k,t-1}') (\Sigma Y_{k,t-1} Y_{k,t-1}')^{-1} (\Sigma Y_{k,t-1} Y_{k,t}'). \quad (2.51)$$

Für die Gleichung (2.51) brauchen nun eine Teststatistik um die Hypothese, dass s der kleinsten Eigenwerte von $V(k)$ den Wert Null entsprechen.

[Tsay and Tiao, 1985] schlugen die Teststatistik

$$C(k, s) = -(T - k - 1) \sum_{i=1}^s \log \{1 - \hat{\lambda}_i(k)\} \quad (2.52)$$

vor, wobei $\hat{\lambda}_i(k)$ der i -t kleinste Eigenwert von $V(\hat{k})$ ist. Für $C(k, s)$ wissen wir, dass sie asymptotisch χ^2 -verteilt ist mit s^2 Freiheitsgraden, wobei s die kleinsten Eigenwerte von Wert Null für $V(k)$ sind ([Tsay and Tiao, 1985]). Dies ist erfüllt, wenn $\text{rang}\Phi_{k+1} \leq m - s$ ist.

Um die Anzahl der Null-Eigenwerte zu bestimmen, muss getestet werden ob die m kleinsten Null sind. Wenn die Hypothese zu verwerfen ist, dann muss für $(m - 1)$ getestet werden, usw. Dieses Vorgehen ist nicht statistisch unabhängig [Anderson, 1984, S. 498]. Für den Anfang liefert dieses Verfahren gute Information zur Identifikation und Spezifizierung des Modells. Man kann aber auch in umgekehrter Richtung vorgehen. Man nimmt an, dass es keinen Eigenwert mit Wert Null gibt. Falls diese Annahme verworfen wird, prüft man ob die Teststatistik für einen Eigenwert zu Wert Null erfüllt wird, usw.

2.2.2 Kanonische Form

Wir wollen in diesem Abschnitt darauf eingehen, wie wir durch Transformation die Struktur des Modells sohingehend verändern können, sodass die Eigenschaften des NRRAR Modells klarer ersichtlich werden. Dies vereinfacht die Interpretation. Die Vorgehensweise zur gezielten Transformation des Modells wird kanonische Transformation genannt.

Annahme Wir nehmen an das gilt

$$r_j = r_k \Rightarrow A_j = A_k \quad (2.53)$$

für Gleichung (2.27).

Definition Sei u die Anzahl der unterschiedlichen Ränge der Φ_j . Dann sei k_i der kleinste Lag sodass gilt

$$\text{rang}(\Phi_j) < k_i \quad \forall j > k_i. \quad (2.54)$$

Dies bedeutet, dass im Fall $u < m$ es mindestens ein Paar r_j und r_i mit $i \neq j$ gibt, für das gilt

$$r_i = r_j. \quad (2.55)$$

Da somit nicht jeder Wert im Intervall $[1, m]$ als Rang angenommen wird, werden manche übersprungen.

Damit haben $\Phi_1, \dots, \Phi_{k_1}$ den Rang r_{k_1} , $\Phi_{k_1+1}, \dots, \Phi_{k_2}$ Rang r_{k_2} usw. Dadurch dass es u unterschiedliche Ränge gibt, gilt

$$r_{k_i} = 0 \quad \forall i > u. \quad (2.56)$$

Hinweis Wir passen die Annahme an, indem wir festlegen, dass gilt

$$r_{k_0} = m \quad (2.57)$$

und

$$k_0 = 0. \quad (2.58)$$

Es sei A_{k_i} eine $m \times r_{k_i}$ Matrix, für die gilt

$$\Phi_j = A_{k_i} B_j \quad \text{mit } k_{i-1} < j \leq k_i \quad (2.59)$$

Aus Gleichung (2.53) können wir nun schließen dass $A_j = A_{k_i}$ mit $k_{i-1} < j \leq k_i$ gilt. Da es u unterschiedliche r_j gibt, gibt es somit u unterschiedliche Matrizen A_j , A_{k_1}, \dots, A_{k_u} im Modell (2.27).

Da

$$\text{im}(A_{k_i}) \supset \text{im}(A_{k_{i+1}}) \quad \forall i = 1, \dots, u - 1 \quad (2.60)$$

gilt, können wir eine $m \times m$ Matrix P bilden, mit $P' = [F_u, F_{u-1}, \dots, F_0]$, wobei F_i eine $m \times (r_{k_i} - r_{k_{i+1}})$. Die Matrix mit den Spalten $[F_i, \dots, F_1, F_0]$ hat die Dimension $m \times r_{k_{i+1}}$ und ist das orthogonale Komplement zu $\text{im}(A_{k_{i+1}})$.

Hinweis Wenn $\text{rang}(A_{k_1}) = r_{k_1} = m$, ist F_0 nicht vorhanden.

Somit gilt

$$[F_i, \dots, F_1, F_0]' A_{k_j} = 0 \quad i = 1, \dots, u-1 \wedge j > i. \quad (2.61)$$

Definition Es sei $Z_t = PY_t$ und $e_t = P\varepsilon_t$.

Wir betrachten nun

$$Z_t = PY_t = P \left(\sum_{j=1}^p A_j B_j Y_{t-j} + \varepsilon_t \right) \quad (2.62)$$

$$= \sum_{j=1}^p P A_j B_j Y_{t-j} + P \varepsilon_t \quad (2.63)$$

$$= \sum_{j=1}^p P A_j B_j P^{-1} P Y_{t-j} + P \varepsilon_t \quad (2.64)$$

$$= \sum_{j=1}^p P A_j B_j P^{-1} Z_{t-j} + e_t. \quad (2.65)$$

Wir zerlegen nun Z_t in $u+1$ Untervektoren, wobei der i -te Untervektor Z_{it} von $Z_t = [z_{1,t}, \dots, z_{m,t}]'$ die Form

$$Z_{it} = F'_{u-i+1} Y_t \quad (2.66)$$

und die Dimension $(r_{k_{u-i+1}} - r_{k_{u-i+2}}) \times 1$ hat. Im Fall $r_{k_1} = m$, gibt es die Komponente $Z_{(u+1)t}$ nicht. Für e_t erfolgt eine ähnliche Unterteilung.

Hinweis Für ein kleines i , wird aus den Y_t durch Linksmultiplikation durch F'_{u-i+1} die Information aus der jüngeren Vergangenheit herausgefiltert. Mit steigendem i nimmt die Information ab. Dies ist aufgrund der Annahmen des NRRAR Modells bezüglich älteren Informationen.

Mit $C_j = B_j P^{-1}$ erhalten wir dann

$$Z_{it} = \sum_{j=1}^p F'_{u-i+1} A_j B_j P^{-1} Z_{t-j} + e_{it} \quad (2.67)$$

$$= F'_{u-i+1} \sum_{j=1}^{k_{u-i+1}} A_j C_j Z_{t-j} + e_{it}. \quad (2.68)$$

Da $F'_{u-i+1}A_j = 0$ für alle $j > k_{u-i+1}$ gilt, verkürzt sich die Summe. Es muss nicht mehr bis zum Grad p aufsummiert werden, sondern es reicht, wenn wir die Summe bis k_{u-i+1} bilden. Wir können aus dem Ergebnis entstehend durch die Transformation folgern, dass für steigendes i zunehmend die ältere Information ausgeblendet wird und dass Z_{it} von kürzlichen Ereignissen abhängt. Für den Fall, dass $R_{k_1} < m$, besteht $Z_{(u+1)t}$ aus reinem weißen Rauschen. Wir lassen ein Beispiel folgen:

Beispiel Es sei Y_t ein 5×1 Vektor der nach Modell (2.27) für $p = 4$ gegeben. Weiters gilt für Φ_i :

$$\text{rang}(\Phi_1) = 5, \quad (2.69)$$

$$\text{rang}(\Phi_2) = \text{rang}(\Phi_3) = 3, \quad (2.70)$$

$$\text{rang}(\Phi_4) = 1. \quad (2.71)$$

Somit ist $u = 3$ und für die k_i gilt :

$$k_1 = 1, \quad (2.72)$$

$$k_2 = 3, \quad (2.73)$$

$$k_3 = 4. \quad (2.74)$$

Da $r_{k_1} = 5 = m$ ist, ist F_0 nicht vorhanden.

Wir erhalten nun für die r_{k_i} :

$$r_{k_1} = 5 \quad (2.75)$$

$$r_{k_2} = 3 \quad (2.76)$$

$$r_{k_3} = 1 \quad (2.77)$$

$$r_{k_i} = 0 \quad \forall i > u. \quad (2.78)$$

Mit $\dim F_i = m \times (r_{k_i} - r_{k_{i+1}})$ können wir nun für die Dimension von F_i schliessen:

$$\dim F_1 = 5 \times (5 - 3), \quad (2.79)$$

$$\dim F_2 = 5 \times (3 - 1), \quad (2.80)$$

$$\dim F_3 = 5 \times (1 - 0). \quad (2.81)$$

Wir erhalten somit die transformierten Reihen für Z_{it} in kanonischer Form

$$\begin{bmatrix} Z_{1t} \\ Z_{2t} \\ Z_{3t} \end{bmatrix} = \begin{bmatrix} F'_3 \{A_{k_1} C_1 Z_{t-1} + A_{k_2} (C_2 Z_{t-2} + C_3 Z_{t-3}) + A_{k_3} C_4 Z_{t-4}\} \\ F'_2 \{A_{k_1} C_1 Z_{t-1} + A_{k_2} (C_2 Z_{t-2} + C_3 Z_{t-3})\} \\ F'_1 A_{k_1} C_1 Z_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{bmatrix} \quad (2.82)$$

Z_{1t} hängt somit von der Vergangenheit durch $Z_{t-1}, Z_{t-2}, Z_{t-3}, Z_{t-4}$ ab, während Z_{3t} alleinig von Z_{t-1} abhängt. Z_{4t} ist nicht präsent da ja F_0 nicht vorhanden ist ($r_{k_1} = m = 5$). Die Koeffizientenmatrix der Z_{t-j} sind in kanonischer Form von folgender Gestalt:

$$\begin{aligned} Z_t = & \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix} Z_{t-1} + \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} Z_{t-2} + \\ & + \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} Z_{t-3} + \begin{bmatrix} * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} Z_{t-4} + e_t, \end{aligned}$$

wobei die $*$ für Werte stehen, die möglicherweise ungleich 0 sein können.

2.2.3 Parametrisierung und Normalisierung

Für jede nichtsinguläre $r_j \times r_j$ Matrize Q , gilt

$$\Phi_j = A_j Q^{-1} Q B_j = \bar{A}_j \bar{B}_j, \quad (2.83)$$

wobei $\bar{A}_j = A_j Q^{-1}$ und $\bar{B}_j = Q B_j$ ist. Wir benötigen somit eine Normalisierung um die Parameter in Gleichung (2.27) eindeutig bestimmen zu können. Wir werden eine bestimmte Normalisierung voraussetzen und darauf aufbauend die Parametrisierung erklären. Die Parameter für die wir uns hier interessieren sind die A_{k_i} für $i = 1, \dots, u$ und die B_j für $j = 1, \dots, p$.

Wir schreiben die A_{k_i} an als

$$A_{k_i} = \begin{bmatrix} A_{1k_i} \\ A_{2k_i} \end{bmatrix}. \quad (2.84)$$

Da die Matrix A_{k_i} von Dimension $m \times r_{k_i}$ ist laut Gleichung (2.19) und wir angenommen haben, dass A_{k_i} von vollen Rang ist, kann durch geschickte Anordnung der Komponenten von Y_t kann gewährleistet werden, dass die A_{1k_i} für $i = 1, \dots, u$ vollen Rang haben. Das bedeutet, dass A_{1k_i} die Dimension $r_{k_i} \times r_{k_i}$ hat und A_{2k_i} von Dimension $(m - r_{k_i}) \times r_{k_i}$ ist. Bezüglich der Normalisierung der A_{k_i} im Modell (2.27) gehen wir wie folgt vor. Zuerst nehmen wir an, dass A_{k_u} die von der Form

$$A_{k_u} = \begin{bmatrix} I_{(k_u)} \\ A_{*k_u} \end{bmatrix}, \quad (2.85)$$

ist, wobei $I_{(k_u)}$ eine $r_{k_u} \times r_{k_u}$ Einheitsmatrix ist, und A_{*k_u} eine $(m - r_{k_u}) \times r_{k_u}$ Parametermatrix ist. Nun beginnt man sukzessive die Form der Matrizen in absteigender Reihenfolge beginnend bei A_{k_u} bis A_{k_1} zu bestimmen. Der Zusammenhang ist wie folgt

$$A_{k_i} = [A_{k_{i+1}}, M_{k_i}]. \quad (2.86)$$

Über die Form von M_{k_i} können wir folgendes aussagen

$$M_{k_i} = \begin{bmatrix} O_{k_i} \\ I_{(k_i)} \\ A_{*k_i} \end{bmatrix}, \quad (2.87)$$

wobei O_{k_i} eine $r_{k_{i+1}} \times (r_{k_i} - r_{k_{i+1}})$ Nullmatrix, $I_{(k_i)}$ eine $(r_{k_i} - r_{k_{i+1}}) \times (r_{k_i} - r_{k_{i+1}})$ Einheitsmatrix, und A_{*k_i} eine $(m - r_{k_i}) \times (r_{k_i} - r_{k_{i+1}})$ Parametermatrix ist. Somit wäre $A_{k_{u-1}}$ von der Gestalt

$$A_{k_{u-1}} = [A_{k_u}, M_{k_{u-1}}] = \begin{bmatrix} A_{1k_u} & O_{k_{u-1}} \\ A_{2k_u} & I_{(k_{u-1})} \\ & A_{*k_{u-1}} \end{bmatrix}. \quad (2.88)$$

Hinweis Im fall $r_{k_1} = m$ wird zu Erstellung von A_{k_1} nur die Matrix $[O'_{k_1}, I_{(k_1)}]'$ hinzugefügt, keinen zusätzlichen Parameter.

Beispiel Wir nehmen hier wieder das Beispiel aus Abschnitt 2.2.2.

$$A_{k_u} = A_{k_3} = \begin{bmatrix} I^{(k_u)} \\ A_{*k_u} \end{bmatrix} = \begin{bmatrix} 1 \\ * \\ * \\ * \\ * \end{bmatrix} \quad (2.89)$$

Für die $A_{k_{u-1}}$ gilt

$$A_{k_{u-1}} = [A_{k_u}, M_{k_{u-1}}], \quad (2.90)$$

mit

$$M_{k_{u-1}} = \begin{bmatrix} O_{k_{u-1}} \\ I^{(k_{u-1})} \\ A_{*k_{u-1}} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ * & * \\ * & * \end{bmatrix}. \quad (2.91)$$

Somit erhalten wir für A_{k_2}

$$A_{k_2} = \begin{bmatrix} 1 & 0 & 0 \\ * & 1 & 0 \\ * & 0 & 1 \\ * & * & * \\ * & * & * \end{bmatrix}. \quad (2.92)$$

Wenn wir in ähnlicher Weise fortschreiten, erhalten wir für A_{k_1}

$$A_{k_1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ * & 1 & 0 & 0 & 0 \\ * & 0 & 1 & 0 & 0 \\ * & * & * & 1 & 0 \\ * & * & * & 0 & 1 \end{bmatrix}. \quad (2.93)$$

Es kann nun leicht gezeigt werden, dass die A_{k_i} die selbe Struktur wie die Koeffizientenmatrizen haben. Zusätzlich ist die Parameterisierung eindeutig und es kann gezeigt werden dann für eine nichtsinguläre Matrix Q mit $\Phi_j = A_j B_j = A_j Q^{-1} Q B_j = \bar{A}_j \bar{B}_j$, wobei $\bar{A}_j = A_j Q^{-1}$ und $\bar{B}_j = Q B_j$ eine Einheitsmatrix sein muss damit \bar{A} die gleiche Struktur wie A_j hat.

Es sei $D_j = [I_{r_j}, O'_{(r_1-r_j) \times r_j}]'$ für $j = 1, \dots, p$ mit $D_1 = I_{r_1}$. Dann gilt $A_{k_i} = A_{k_1} D_{k_i}$. Dadurch dass $D_j = D_{k_i}$ für $k_{i-1} < j \leq k_i$ gilt, können wir unser Model (2.27) neu anschreiben als

$$Y_t = A_{k_1} \sum_{j=1}^p D_j B_j Y_{t-j} + \varepsilon_t. \quad (2.94)$$

Die Anzahl der Parameter in A_{k_1} ist $a \equiv \sum_{i=1}^u (m - r_{k_i})(r_{k_i} - r_{k_{i+1}})$, da ja A_{*k_i} $(m - r_{k_i})(r_{k_i} - r_{k_{i+1}})$ Parameter hat. Dadurch das B_j mr_j Parameter hat, ist die Gesamtanzahl der Parameter in Model (2.27) $a + m \sum_{j=1}^p r_j$.

Dadurch dass $imA_{k_i} \supset imA_{k_{i+1}}$ gilt, können die ersten $r_{k_{i+1}}$ Spalten von A_{k_i} von der Matrix $A_{k_{i+1}}$ genommen werden, also $A_{k_i} = A_{k_1} * D_{k_i}$. Somit reduziert sich die Normalisierung der A_{k_i} auf die Normalisierung der Spalten von A_{k_1} . Die Spalten der Matrix A_{k_1} des allgemeinen (noch nicht normalisierten) Modells (2.94) können bestimmt werden durch Verwendung der Eigenvektoren aus der kanonischen Korrelationsanalyse (2.50). Die Eigenvektoren mit dazugehörigen kleinen Eigenwerten der Produktmatrix aus Gleichung (2.51) können dazu verwendet werden eine erste Schätzung der Transformationsmatrix P aus Abschnitt 2.2.2 zu bestimmen und damit auch A_{k_1} zu schätzen in der allgemeinen Form. Wenn man mal eine allgemeine $m \times r_{k_1}$ Matrix hat, kann man beginnen sie zu normalisieren indem man A_{k_1} zu einer Matrix reduziert, die $I_{(k_u)}$, eine $r_{k_u} \times r_{k_u}$ Einheitsmatrix, im oberen linken Block enthält. Diese Matrix kann dann weiter reduziert werden durch weitere Spaltenmanipulationen zu einer Matrix die $[O'_{(k_{u-1})}, I_{(k_{u-1})}]'$ in den nächsten $r_{k_{u-1}} - r_{k_u}$ Spalten und so weiter. Bis jetzt wurde angenommen dass die Komponenten von Y_t so angeordnet waren, dass eine Reduktion stets möglich war. Jedoch ist die Reduktion nicht notwendiger Weise möglich bei einer zufälligen Anordnung der Komponenten von Y_t . Daher, muss man die betroffenen Reihen von A_{k_1} vertauschen sobald Probleme bei der Reduzierung von A_{k_1} auftauchen (z.B. Die Matrix im linken oberen Block hat nicht vollen Rang und kann somit nicht zu einer Einheitsmatrix transformiert werden). Das Tauschen der Reihen der Matrix A_{k_1} entspricht einer Neuordnung der Komponenten von Y_t im Modell (2.94).

Kapitel 3

Rückschluss auf Vektor Autoregressiven Prozessen mit Kointegration und Skalarkomponenten

3.1 Einleitung

Multivariate Zeitreihen finden eine breite Anwendung, im Speziellen in der Wirtschaft. Sie beschreiben die Beziehung zwischen Variablen die häufig zusammen gemessen werden um eine bessere Vorhersage für die Reihen tätigen zu können. [Hotelling, 1935] [Hotelling, 1936] [Witte and Horstmann, 1976].

Dabei wird besonderes Augenmerk auf die strukturelle Parameterisierung des Modells gelegt um somit das Modell einfach zu halten. Dies führt zu einem besseren Verständnis der Zusammenhänge und ermöglicht genauere Vorhersagen für die Zeitreihen. Dabei werden wir versuchen durch Verringerung der Ränge, Kronecker Indizes und Skalarkomponenten¹ die Struktur des Modells zu vereinfachen. Das in [Tiao and Tsay, 1989]

¹siehe 1.7

vorgestellte Skalar-Komponenten Modell ermöglicht die Struktur einer Komponente der Vektorzeitreihe zu erfassen. Das Modell hat starke Ähnlichkeiten zu dem Ansatz zur Reduzierung der Ränge der Parametermatrizen [Reinsel, 1993].

Bei der Kointegration wird das Langzeitverhalten von teilweise nichtstationären Zeitreihen betrachtet und gibt Hinweis auf einen gemeinsamen Trend [Stock and Watson, 1988]. Die Reduzierung der Rangmatrizen, wie sie in [Ahn and Reinsel, 1990], [Johansen, 1988] und [Phillips, 1991] für autoregressive Vektormodelle in Fehlerdarstellung vorgestellt wird, betrifft die strukturelle Parameterisierung des gemeinsamen Trends. Das Ziel ihrer Arbeiten ist equivalent zur Kointegration.

Für stationäre Zeitreihen haben [Vahid and Engle, 1993] gezeigt, dass durch Linearkombination der Komponenten zur Minimierung des Rangs des gleitenden Durchschnittes man Teilabhängigkeiten auffinden kann. Eine Form von Teilabhängigkeit ist die Autokorrelation [Engle and Kozicki, 1993], wobei die Linearkombination von Zeitreihen ein weißes Rauschen ergeben soll. Das Skalar-Komponenten Modell des Rangs $(0, 0)$ existiert für Zeitreihen die Autokorrelation als gemeinsame Eigenschaft aufweisen. Beispiele für Autokorrelation als gemeinsame Charakteristika und das Skalar-Komponenten Modell beinhalten das autoregressive Vektorindexmodell von [Reinsel, 1983] und das nested reduced-rank autoregressive model von [Ahn and Reinsel, 1988].

3.2 Das Modell und seine Eigenschaften

Folgend [Ahn and Reinsel, 1990] betrachten wir einen m -dimensionalen teilweise nichtstationären autoregressiven Prozess Y_t , gegeben durch

$$\Phi(L) Y_t = \left(I_m - \sum_{j=1}^p \Phi_j L^j \right) Y_t = \epsilon_t. \quad (3.1)$$

Annahme Die Eigenschaft, dass der Prozess teilweise stationär ist, bezieht sich auf die Annahme, dass die nichtstationären Wurzeln des autoregressiven Operators $\Phi(L)$ alle den Wert 1 annehmen. Sie können somit als Einheitswurzeln bezeichnet werden. Eine

weitere Annahme ist dass es $d < m$ Einheitswurzeln gibt und alle restlichen Wurzeln von $\det \Phi(L) = 0$ sich ausserhalb des Einheitskreises befinden.

Hinweis Da d Nullstellen des charakteristischen Polynoms von Wert 1 sind, so ist $\det(\Phi(1)) = 0$ und

$$\Phi(1) = I - \sum_{j=1}^p \Phi_j \quad (3.2)$$

nicht mehr von vollen Rang und wir erhalten

$$\text{Rang} \{ \Phi(1) \} = r = m - d. \quad (3.3)$$

$\Phi(1)$ hat d Eigenwerte von Wert Null und dies bedeutet, dass durch d -maliges Differenzieren der Prozess stationär wird. Der stochastische Prozess ist nichtstationär, da er Einheitswurzel besitzt. Dies wird auch als stochastischer Trend bezeichnet.

ε m -dimensionales White Noise (WN) mit den Eigenschaften

$$\mathbb{E}(\varepsilon_t) = 0 \quad (3.4)$$

$$\text{cov}(\varepsilon_t) = \Omega \quad (3.5)$$

Der autoregressive Prozess von Ordnung p aus Gleichung (3.1) kann auch in seiner Fehlerkorrekturform dargestellt werden [Engle and Granger, 1987]

$$Z_t = CY_{t-1} + \sum_{j=1}^{p-1} \Phi_j^* Z_{t-j} + \varepsilon_t, \quad (3.6)$$

wobei gilt

$$Z_t = Y_t - Y_{t-1} \quad (3.7)$$

$$\Phi_j^* = - \sum_{k=j+1}^p \Phi_k \quad \forall j = 1, \dots, p-1 \quad (3.8)$$

$$C = -\Phi(1) = - \left(I_m - \sum_{j=1}^p \Phi_j \right). \quad (3.9)$$

Beweis.

$$Z_t = CY_{t-1} + \sum_{j=1}^{p-1} \Phi_j^* Z_{t-j} + \varepsilon_t \quad (3.10)$$

$$Z_t = CY_{t-1} + \sum_{j=1}^{p-1} \Phi_j^* (Y_{t-j} - Y_{t-j-1}) + \varepsilon_t \quad (3.11)$$

$$Z_t = CY_{t-1} + \sum_{j=1}^{p-1} \Phi_j^* Y_{t-j} - \sum_{j=1}^{p-1} \Phi_j^* Y_{t-j-1} + \varepsilon_t \quad (3.12)$$

$$Z_t = CY_{t-1} + \sum_{j=1}^{p-1} \Phi_j^* Y_{t-j} - \sum_{j=2}^p \Phi_{j-1}^* Y_{t-j} + \varepsilon_t \quad (3.13)$$

$$Z_t = CY_{t-1} + \Phi_1^* Y_{t-1} + \sum_{j=2}^{p-1} \Phi_j^* Y_{t-j} - \sum_{j=2}^{p-1} \Phi_{j-1}^* Y_{t-j} - \Phi_{p-1}^* Y_{t-p} + \varepsilon_t \quad (3.14)$$

$$Z_t = CY_{t-1} + \Phi_1^* Y_{t-1} + \sum_{j=2}^{p-1} (\Phi_j^* - \Phi_{j-1}^*) Y_{t-j} - \Phi_{p-1}^* Y_{t-p} + \varepsilon_t \quad (3.15)$$

$$Z_t = CY_{t-1} + \Phi_1^* Y_{t-1} + \sum_{j=2}^{p-1} \left(- \sum_{k=j+1}^p \Phi_k + \sum_{k=j}^p \Phi_k \right) Y_{t-j} - \Phi_{p-1}^* Y_{t-p} + \varepsilon_t \quad (3.16)$$

$$Z_t = CY_{t-1} + \Phi_1^* Y_{t-1} + \sum_{j=2}^{p-1} \Phi_j Y_{t-j} - \Phi_{p-1}^* Y_{t-p} + \varepsilon_t \quad (3.17)$$

$$Z_t = - \left(I_m - \sum_{j=1}^p \Phi_j \right) Y_{t-1} - \sum_{k=2}^p \Phi_k Y_{t-1} + \sum_{j=2}^{p-1} \Phi_j Y_{t-j} + \sum_{k=p}^p \Phi_k Y_{t-p} + \varepsilon_t \quad (3.18)$$

$$Z_t = -Y_{t-1} + \sum_{j=1}^p \Phi_j Y_{t-1} - \sum_{k=2}^p \Phi_k Y_{t-1} + \sum_{j=2}^{p-1} \Phi_j Y_{t-j} + \Phi_p Y_{t-p} + \varepsilon_t \quad (3.19)$$

$$Y_t - Y_{t-1} = -Y_{t-1} + \Phi_1 Y_{t-1} + \sum_{j=2}^p \Phi_j Y_{t-j} + \varepsilon_t \quad (3.20)$$

$$Y_t = \sum_{j=1}^p \Phi_j Y_{t-j} + \varepsilon_t \quad (3.21)$$

□

Die Gleichung aus (3.6) bekommt man also einfach durch die Algebra für Matrizen, da der autoregressive Operator $\Phi(L)$ immer dargestellt werden kann als

$$\Phi(L) = \Phi^*(L)(1 - L) + \Phi(1) \quad (3.22)$$

mit

$$\Phi^*(L) = I - \sum_{j=1}^{p-1} \Phi_j^* L^j. \quad (3.23)$$

Beispiel In einem autoregressive Prozess zweiter Ordnung haben wir

$$\Phi(L) = I - \Phi_1 L - \Phi_2 L^2 = (I - IL) + (I - \Phi_1 - \Phi_2)L + \Phi_2(I - IL)L \quad (3.24)$$

$$= (I + \Phi_2 L)(I - IL) + (I - \Phi_1 - \Phi_2)L \equiv \Phi^*(L)(1 - L) + \Phi(1)L, \quad (3.25)$$

wobei $\Phi^*(L) = I + \Phi_2 L \equiv I - \Phi_1^*(L)$ mit $\Phi_1^* = -\Phi_2$. Unter den Annahmen für $\Phi(L)$, kann man zeigen dass $\Phi^*(L)$ ein stationärer autoregressiver Operator ist mit alle Wurzeln von $\det\{\Phi^*(L)\} = 0$ ausserhalb des Einheitskreises.

Hinweis Die Fehlerkorrekturdarstellung aus Gleichung (3.6) ist deshalb so praktisch, da die Einheitswurzeln des autoregressiven Operators $\Phi(L)$ in den Term CY_{t-1} eingebunden werden können, sodass die nichtstationäre Eigenschaft des Modells in der Matrix C konzentriert wird.

Da $C = -\Phi(1)$ gilt, wissen wir dass gilt

$$\text{Rang}(C) = \text{Rang}(-\Phi(1)) = r = m - d. \quad (3.26)$$

Da somit der Rang von C nicht voll ist, existiert ein Kern zu C . Der Defekt von C ist

$$\text{Defekt}(C) = m - \text{Rang}(C) = m - r = d. \quad (3.27)$$

Der Kern von C besteht aus d linear unabhängigen Vektoren. Für d linear unabhängige Vektoren f mit $f'C = 0$ gilt

$$f'Z_t = f'CY_{t-1} + f' \sum_{j=1}^{p-1} \Phi_j^* Z_{t-j} + f'\varepsilon_t \quad (3.28)$$

$$f'Z_t = \sum_{j=1}^{p-1} f'\Phi_j^* Z_{t-j} + f'\varepsilon_t. \quad (3.29)$$

Dann existiert eine linear unabhängige Gruppe der d Skalarkomponenten vom Grad $(p-1, 0)$ für Z_t .

Da C nicht vollen Rang hat, wird C oft angeschrieben als $C = AB$ wobei gilt

$$\dim A = m \times r \quad (3.30)$$

$$\dim B = r \times m \quad (3.31)$$

und beide von vollen Rang sind. Die Φ_j^* und die A werden stationäre Parameter genannt wegen ihrer Verbindung zum stationären Prozess Z_{t-j} und BY_{t-1} . B wird nichtstationärer Parameter genannt auf Grund seiner Verbindung zum nichtstationären Prozess Y_{t-1} .

Annahme Um die Struktur der stationären Komponenten der nichtstationären multivariaten Zeitreihe zu beschreiben und zu vereinfachen, nehmen wir die Struktur des NRRAR Modells für die Parametermatrizen an

$$\text{im}(\Phi_j^*) \supset \text{im}(\Phi_{j+1}^*) \quad \forall j = 1, \dots, p-2, \quad (3.32)$$

wobei

$$\text{Rang}\Phi_j^* = r_j \quad \forall j = 1, \dots, p-1. \quad (3.33)$$

Weiters nehmen wir an dass der Vektor α_k existiert sodass

$$\alpha_k' A = 0 \quad (3.34)$$

und

$$\alpha_k' \Phi_j^* = 0 \quad \forall j > k \quad (3.35)$$

gilt.

Diese Annahme impliziert nicht notwendigerweise eine eingeschränkte Struktur zwischen C und Φ_j^* . Die Existenz eines solchen Vektor α_k steht im starken Zusammenhang zum Skalarkomponentenmodell von [Tiao and Tsay, 1989] für die Serie der ersten Differenzen Z_t . Eigentlich ist $\alpha_k' Z_t$ eine Skalarkomponente vom Grad $(k, 0)$ für Z_t , da

$$\alpha_k' Z_t = \sum_{j=1}^k \alpha_k' \Phi_j^* Z_{t-j} + \alpha_k' \varepsilon_t. \quad (3.36)$$

Die Existenz von α_0 bedeutet, dass Autokorrelation in Z_t vorliegt, da $\alpha_0'Z_t$ ein weißes Rauschen ist, was wiederum eine Skalarkomponente von Grad $(0, 0)$ ist. Anzumerken ist, dass die α_k zu einer Vereinfachung der Struktur von Z_t durch die Verwendung des Skalarkomponentenmodells führt.

Beim Modell aus Gleichung (3.6) ist der erste Schritt den Grad p der Autoregression, den Rang r_j von Φ_j^* für $j = 1, \dots, p - 1$ und den Kointegrationsrang r von C zu bestimmen. Für die Bestimmung des Grades der Autoregression kann man entweder den Ansatz über die kanonische Korrelation wählen [Tiao and Tsay, 1985] oder ein Informationskriterium, wie zum Beispiel das Akaike Informations Kriterium (AIC), wählen. Zu einem früheren Zeitpunkt, kann man die Bestimmung des Ranges r_j für $j = 1, \dots, p - 1$ auch durch den Ansatz mittels kanonischer Korrelation erzielen, da $m - r_k$ partielle kanonische Korrelationen von Wert 0 zwischen $X_t = Z_t - CY_{t-1}$ und Z_{t-k} , da diese auf $Z_{t-1}, \dots, Z_{t-k+1}$ regressiert wurden, existieren. Da C im Regelfall unbekannt ist, wird zuerst \tilde{C} ermittelt. \tilde{C} ist der Kleinstquadratschätzer von C für die Regression von Z_t auf $Y_{t-1}, Z_{t-1}, \dots, Z_{t-p+1}$. Man verwendet dann die partielle kanonische Korrelation als einen Indikator für den Rang von Φ_j^* . [Ahn, 1997] schlägt eine Verfahren vor, wie es in ähnlicher Weise in der Hauptkomponentenanalyse für multivariate Statistik verwendet wird [Rencher, 1995]. Man soll allen partiellen kanonischen Korrelationen, deren Quadratsumme weniger als 80% der gesamten Quadratsumme darstellt, den Wert Null zuweisen. Eine andere Möglichkeit wäre die Gruppierung kleiner und großer kanonischer Korrelationen (scree plot).

Um den Kointegrationsrang zu bestimmen, und damit auch den Rang von C , berechnen wir die partielle kanonische Korrelation zwischen Z_t und Y_{t-1} nach Regression auf $Z_{t-1}, \dots, Z_{t-p+1}$. [Reinsel and Ahn, 1982] zeigten, dass

$$\Lambda = -(T - p) \sum_{i=1}^d \log \{1 - \hat{\lambda}_i\}, \quad (3.37)$$

wobei $\hat{\lambda}_i$ die i -te kleinste quadratische partielle kanonische Korrelation ist. $d = m - r$ ist asymptotisch verteilt mit

$$tr \left[\left\{ \int_0^1 dW_d(u) W_d(u)' \right\} \left\{ \int_0^1 W_d(u) W_d(u)' \right\}^{-1} \left\{ \int_0^1 W_d(u) dW_d(u)' \right\} \right], \quad (3.38)$$

wobei $W_d(u)$ ein d -dimensionale standard Brownsche Bewegung ist.

Hinweis Wenn die partielle kanonische Korrelation mit mittelwertbereinigte Stichprobe der Zeitreihe erhalten wurde, dann nimmt die asymptotische Verteilung von Λ die selbe Form an wie in Gleichung (3.38) an nur mit dem Unterschied, dass die Brownsche Bewegung $W_d(u)$ ersetzt wird durch $\bar{W}_d(u)$, welche gegeben ist als

$$\bar{W}_d(u) = W_d(u) - \int_0^1 W_d(u). \quad (3.39)$$

Nach Identifizierung der Ordnung p der Autoregression, des Ranges r für die Kointegration, der reduzierten Ränge r_j für die Parametermatrizen Φ_j^* , kann man nun die Annahme für die stationären Parameter testen einschließlich der nested structure und der Existenz von Skalarkomponenten niedrigen Grades mittels Likelihood ratio Test. Die Teststatistik ist eine asymptotisch χ^2 -verteilte Zufallsvariable. Die Anzahl ihrer Freiheitsgrade ist gleich der unterschiedlichen Anzahl der Parameter unter der Nullhypothese und der Alternativhypothese.

Anhang A

Anhang

A.1 Grundlagen

A.1.1 Prädiktor bzw. vorhersagende Größe

Eine Variable wird als Prädiktor bzw. vorhersagende Größe bezeichnet, wenn sich ein Modell auf sie stützt. Durch den Prädiktor soll mit einer bestimmten Wahrscheinlichkeit eine Vorhersage bzw. Prognose gemacht werden.

A.1.2 Vorherzusagende Größe bzw. Kriteriumsvariable

Die vorherzusagende Größe wird auch als Kriteriumsvariable, als Prognosevariable oder auch als Regressand bezeichnet. Der Ausdruck bezeichnet eine Variable die es zu erklären gilt und stellt eine abhängige Größe dar.

A.1.3 α bzw. Fehler erster Art

In der Statistik wird beim Testen einer Hypothese das Verwerfen der eigentlich richtigen Hypothese auf Grund von irreführenden Daten als Fehler erster Art bezeichnet, also

die Nullhypothese H_0 wird abgelehnt obwohl sie zutreffend ist. Der Fehler erster Art wird auch mit α -Fehler bezeichnet.

A.1.4 Stochastischer Prozess

Definition Eine Familie $(u_t)_{t \in T}$ von Zufallsgrößen $u_t : \Omega \rightarrow \mathbb{R}$ heißt stochastischer Prozess. Hier bezeichnete T immer die Zeitachse, also \mathbb{N} oder \mathbb{Z}

A.1.5 Multivariater stochastischer Prozess

Definition Es sei $(X_{it})_{t \in T}$ mit $i = 1, \dots, m$ ein stochastischer Prozess. Der Vektor $X_t = (X_{1t}, \dots, X_{mt})$ wird dann eine m -dimensionale oder m -variate Zeitreihe genannt.

Hinweis *Trajektorie* $(u_t(\omega)|t \in T)$, $\omega \in \Omega$ heißt Pfad oder Trajektorie.

A.1.6 Kovarianzfunktion

Definition Es sei $\mathbb{E}x_t'x_t < \infty$, $t \in T$, dann wird die Funktion

$$\gamma : \mathbb{T} \times \mathbb{T} : \longrightarrow \mathbb{C}^{n \times n} \tag{A.1}$$

$$(s, t) \longmapsto \gamma(s, t) = \mathbb{E} (x_s - \mathbb{E}x_s) (x_t - \mathbb{E}x_t)' \tag{A.2}$$

die Kovarianzfunktion von $(x_t|t \in T)$ genannt.

γ kann auch als Matrix $(\gamma_{ij})_{i,j=1,\dots,m}$ interpretiert werden, deren Einträge die Funktionen $\gamma_{ij} : \mathbb{T} \times \mathbb{T} \longrightarrow \mathbb{C}$ sind. Die Funktion γ kann somit als ein Adressbuch für die linearen Abhängigkeiten angesehen werden, zum Beispiel $\gamma_{ij} = Cov(x_s^{(i)}, x_t^{(j)}) = \mathbb{E} \left(x_s^{(i)} - \mathbb{E}x_s^{(i)} \right) \left(x_t^{(j)} - \mathbb{E}x_t^{(j)} \right)'$ beschreibt die lineare Abhängigkeit zwischen $x_s^{(i)}$ und $x_t^{(j)}$.

A.1.7 Kovarianzmatrix

Definition Es sei X ein stochastischer Vektor und X_i eine Komponente von $X = (X_1, X_2, \dots, X_n)$. Die Komponenten X_i sollen endliche Varianz besitzen. Dann wird die Matrix $\Sigma = (\sigma_{ij})_{i,j}$ mit $\sigma_{ij} = \text{Cov}(X_i, X_j)$ der Dimension $n \times n$ als Kovarianzmatrix von X bezeichnet.

Hinweis Die Matrix Σ ist immer positiv semidefinit und symmetrisch. Die Diagonale enthält die Varianzen der einzelnen Komponenten.

A.1.8 Akaike Informationskriterium (AIC)

Beim Versuch die Güte der Anpassung zu maximieren und dabei die Komplexität niedrig zu halten (hier messen wir die Komplexität durch die Anzahl der Parameter p), entwickelte [Akaike, 1973] das Akaike Informationskriterium (AIC).

Das Kriterium lautet

$$AIC(p) = \log \hat{\sigma}^2(p) + \frac{2p}{T}. \quad (\text{A.3})$$

Das Ziel ist das die Gleichung (A.3) minimierende p zu bestimmen.

A.2 Mathematik

A.2.1 Laplacescher Erweiterungssatz

Mit dem Laplaceschen Erweiterungssatz [Gawronski, 1996] kann die Determinante einer Matrix entwickelt werden. Wir nehmen an, dass die Matrix A die Dimension $n \times n$ hat. Dann kann die Determinante sowohl nach der Zeile

$$\det A = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det A_{ij} \quad (\text{A.4})$$

als auch nach der Spalte

$$\det A = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det A_{ij} \quad (\text{A.5})$$

entwickelt werden, wobei a_{ij} das Element in der i -ten Zeile und der j -ten Spalte ist. A_{ij} ist eine $(n-1) \times (n-1)$ Untermatrix von A . Man erhält A_{ij} indem man die i -te Zeile und die j -te Spalte der Matrix A streicht. Das Produkt

$$(-1)^{i+j} \det A_{ij} \quad (\text{A.6})$$

wird auch Kofaktor \tilde{a}_{ij} genannt.

A.2.2 Laurentreihe

Die Laurent-Reihe, benannt nach dem französischen Mathematiker Pierre Alphonse Laurent, ist eine unendliche Reihe. Die Laurentreihe hat im Allgemeinen in x mit Entwicklungspunkt c die Gestalt [Freitag and Busam, 2000]

$$f(x) = \sum_{-\infty}^{\infty} a_n (x-c)^n \quad x, a_n, c \in \mathbb{C}. \quad (\text{A.7})$$

A.2.3 Minor

Als Minor [Havlicek, 2006] bzw. Unterdeterminante, wird die Determinante einer quadratischen Untermatrix genannt. Wir nehmen an, dass die Matrix A die Dimension $n \times n$ hat. A_{ij} ist eine $(n-1) \times (n-1)$ Untermatrix von A . Man erhält A_{ij} indem man die i -te Zeile und die j -te Spalte der Matrix A streicht. Der Minor M_{ij} ist dann die Determinante der Untermatrix A_{ij} , also

$$M_{ij} = \det A_{ij}. \quad (\text{A.8})$$

Den Kofaktor \tilde{a}_{ij} erhält man dann durch

$$\tilde{a}_{ij} = (-1)^{i+j} M_{ij}. \quad (\text{A.9})$$

Literaturverzeichnis

- [Ahn, 1997] Ahn, S. K. (1997). Inference of Vector Autoregressive Models With Cointegration and Scalar Components. *Journal of the American Statistical Association*, 92:350–356.
- [Ahn and Reinsel, 1987] Ahn, S. K. and Reinsel, G. C. (1987). Distribution of Residual Autocovariances and Prediction Mean Square Error Properties for the Multivariate Reduced Rank Autoregressive Model. *Communications in Statistics, Part A*, 16:61–78.
- [Ahn and Reinsel, 1988] Ahn, S. K. and Reinsel, G. C. (1988). Nested Reduced-Rank Autoregressive Models for Multiple Time Series. *Journal of the American Statistical Association*, 83:849–856.
- [Ahn and Reinsel, 1990] Ahn, S. K. and Reinsel, G. C. (1990). Estimation for Partially Nonstationary Multivariate Autoregressive Models. *Journal of the American Statistical Association*, 85:813–823.
- [Akaike, 1973] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the 2nd International Symposium on Information Theory*, pages 267–281.
- [Akaike, 1976] Akaike, H. (1976). Canonical correlation analysis of time series and the use of an information criterion. In Mehra, R. K. and Lainiotie, D. G., editors, *System Identification: Advances and Case Studies*, pages 27–96. New York Academic Press.
- [Anderson, 1951] Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics*, 22:327–351.
- [Anderson, 1984] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. John Wiley, New York, 2 edition.
- [Barlett, 1947] Barlett, M. S. (1947). Multivariate Analysis. *Journal of the Royal Statistical Society*, B(9):176–197.
- [Cooper and Wood, 1982] Cooper, D. M. and Wood, E. F. (1982). Identifying multivariate time series models. *Journal of Time Series Analysis*, 3:153–164.

- [Engle and Granger, 1987] Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55:251–276.
- [Engle and Kozicki, 1993] Engle, R. F. and Kozicki, S. (1993). Testing for Common Features. *Journal of Business and Economic Statistics*, 11:369–380.
- [Freitag and Busam, 2000] Freitag, E. and Busam, R. (2000). *Funktionentheorie 1*. Springer-Verlag, Berlin, 3 edition.
- [Fujikoshi, 1974] Fujikoshi, Y. (1974). The likelihood ratio tests for the dimensionality of regression coefficients. *Journal of Multivariate Statistics*, 4:327–340.
- [Gawronski, 1996] Gawronski, W. (1996). *Grundlagen der Linearen Algebra*. Aula-Verlag, Wiesbaden, 2 edition.
- [Havlicek, 2006] Havlicek, H. (2006). *Lineare Algebra für Technische Mathematiker*. N. Heldermann, Berlin, 2 edition.
- [Hotelling, 1935] Hotelling, H. (1935). The most predictable criterion. *J. Educ. Psychol.*, 26:139–142.
- [Hotelling, 1936] Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28:321–377.
- [Johansen, 1988] Johansen, S. (1988). Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control*, 12:231–254.
- [Phillips, 1991] Phillips, P. C. B. (1991). Optimal Inference in Cointegrated System. *Econometrica*, 59:283–306.
- [Reinsel, 1983] Reinsel, G. C. (1983). Some Results on Multivariate Autoregressive Index Models. *Biometrika*, 70:145–156.
- [Reinsel, 1993] Reinsel, G. C. (1993). *Elements of Multivariate Time Series Analysis*. Springer-Verlag, New York.
- [Reinsel and Ahn, 1982] Reinsel, G. C. and Ahn, S. K. (1982). Vector Autoregressive Models With Unit Roots and Reduced Rank Structure: Estimation, Likelihood Ratio Test, and Forecasting. *Journal of Time Series Analysis*, 13:353–375.
- [Rencher, 1995] Rencher, A. C. (1995). *Methods of Multivariate Analysis*. Wiley, New York.
- [Stock and Watson, 1988] Stock, J. H. and Watson, M. W. (1988). Testing for Common Trends. *Journal of the American Statistical Association*, 83:1097–1107.
- [Tatsuoka, 1971] Tatsuoka, M. M. (1971). *Multivariate Analysis*. John Wiley, New York.
- [Tiao and Tsay, 1985] Tiao, G. C. and Tsay, R. S. (1985). *A Canonical Correlation Approach to Modeling Multivariate Time Series*. American Statistical Association.

- [Tiao and Tsay, 1989] Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series (with Discussion). *Journal of the Royal Statistical Society*, B(51):157–213.
- [Tsay and Tiao, 1985] Tsay, R. S. and Tiao, G. C. (1985). Use of Canonical Analysis in Time Series Model Identification. *Biometrika*, 72:299–315.
- [Vahid and Engle, 1993] Vahid, F. and Engle, R. F. (1993). Common Trends and Common Cycles. *Journal of Applied Econometrics*, 8:341–360.
- [Velu et al., 1986] Velu, R. P., Reinsel, G. C., and Wichern, D. W. (1986). Reduced Rank Models for Multiple Time Series. *Biometrika*, 73:105–118.
- [Witte and Horstmann, 1976] Witte, E. H. and Horstmann, H. (1976). Kanonische korrelationsanalyse: Ihre Ähnlichkeit zu anderen Verfahren und zwei Anwendungsbeispiele aus dem Bereich Graphometrie-Persönlichkeit. *Psychol. Beiträge*, 18:553–570.

Lebenslauf

Ausbildung

Name Alexander Juschitz
Adresse Afritschgasse 61, 1220 Wien
Geboren 28.04.1984 in Wien, Österreich

Ausbildung

2008–2010 Wiederaufnahme der Studien an der **Technischen Universität Wien**, mit Spezialisierung in **Wettbewerb und Unternehmensführung** und **Wirtschaftsmathematik**
2006–2008 Studium an der **Ecole Centrale Paris**, einer führenden Ingenieurhochschule Frankreichs, im Rahmen des Austauschprogramms „TI-ME“ („Top Industrial Managers for Europe“)
2003–2006 Studium an der **Technischen Universität Wien**, in **Wirtschaftsingenieurwesen-Maschinenbau** und **Technischer Mathematik**
2002–2003 Absolvierung der Grundwehrdienstes
2002 Matura am Gymnasium **Theresianische Akademie**

Beruflicher Werdegang und Praktika

2009 **Cornell University** Ithaca, NY USA
Visiting Non-Degree Student an der School of Civil and Environmental Engineering, Stipendium von der Marshall Plan Stiftung
2007 **Bauman Universität** Moskau, Russland
Lehrwerkstätte in Verbindung mit einem Sprachkurs

