

Zeitstetige Modellbildung am Beispiel einer volkswirtschaftlichen Produktionsstruktur

Dipl.-Kfm. Dipl.-Volksw. Thomas Mazzoni*

April 2005

Zusammenfassung

Dieser Beitrag zeigt wie die Modellparameter einer klassischen volkswirtschaftlichen Produktionsstruktur mit Hilfe eines in sich geschlossenen, zeitstetigen Zustandsraummodells geschätzt werden können. Das mathematisch etwas aufwendigere Verfahren führt dabei qualitativ und quantitativ zu besseren Ergebnissen und bietet neben einem eleganten Modellansatz auch funktionale Einblicke in die zugrundeliegenden Prozesse.

Ferner ermöglicht der verwendete Ansatz eine Inter- bzw. Extrapolation von Werten zu beliebigen Zeitpunkten, ohne dabei ein ad-hoc Postulat formulieren zu müssen, mit dem das Informationsvakuum zwischen den Beobachtungszeitpunkten gefüllt wird, wie es bei den konventionellen ökonometrischen Modellen erforderlich ist. Daneben werden auch andere grundlegende Probleme der zeitdiskreten Modellbildung, wie die gemeinsame Behandlung von Stock- und Flowdaten oder die maximale Informationsausbeute, durch den vorgestellten Ansatz gelöst.

Schlüsselbegriffe: Bedingte Wahrscheinlichkeit; *Kalman*-Filter/Glätter; Maximum-Likelihood Schätzung; Stochastische Differentialgleichung; Zustandsraummodell.

1. Einleitung

Dynamische makroökonomische Modelle werden häufig in diskreter Zeit formuliert. Oft handelt es sich bei den modellierten Vorgängen aber um in der Zeit stetige Abläufe, die nur zu bestimmten Zeitpunkten gemessen werden. Dadurch wird ein Spezifikationsfehler in das Modell „eingepflanzt“, der weitreichende Folgen haben kann.

Im Folgenden wird das in Geldeinheiten gemessene Brutto-Inlandsprodukt der Bundesrepublik Deutschland im Zeitraum 1991 bis 2004 resp. seine Entstehung in Abhängigkeit von den Faktoren Arbeit und Kapital untersucht. Es zeigen sich dabei eklatante Unterschiede zwischen zeitstetigen und zeitdiskreten Modellvarianten. Verwendet werden dabei zum einen herkömmliche ökonometrische Modelle, deren Parameter nach dem Prinzip der kleinsten Quadrate geschätzt werden und zum anderen ein Zustandsraummodell, das dem kontinuierlichen Verlauf der Zeit auf der Systemebene Rechnung trägt und durch ein zeitdiskretes Messmodell ergänzt wird. Weiterhin wird gezeigt, dass dieser Ansatz eine umfassendere Modellklasse darstellt, die auf struktureller Ebene eine exaktere Behandlung der Parameter erlaubt und aufgrund geringerer Verzerrungen auch zu besseren Schätzungen führt.

*Wissenschaftlicher Mitarbeiter am Lehrstuhl für angewandte Statistik und Methoden der empirischen Sozialforschung (Univ.-Prof. Dr. Hermann Singer) der Fernuniversität in Hagen, Universitätsstraße 41, 58084 Hagen, Tel.: 02331/9872106, E-Mail: Thomas.Mazzoni@Fernuni-Hagen.de

2. Die Produktionsstruktur

Die volkswirtschaftliche Produktion stellt die Summe vieler einzelner Produktionsprozesse dar, die sich alle durch in der Regel verschiedene Produktionsfunktionen charakterisieren lassen. Aggregation führt jedoch auf homogene Funktionstypen, wie sich sogar empirisch belegen lässt¹. Eine allgemeine Form homogener Funktionen erhält man, wenn man die Produktionsstruktur durch die Gleichung

$$Y = \|\Theta \mathbf{x}\|_p^r \quad (1)$$

ausdrückt. Die Produktion Y wird hier durch die p -Norm des Matrix-Vektor-Produkts der $(m \times m)$ Koeffizientenmatrix Θ und des $(m \times 1)$ Faktorenvektors \mathbf{x} geschrieben. Etwas unscharf kann Y anschaulich als Metrik im Faktorenraum aufgefasst werden. Der Parameter r gibt den Homogenitätsgrad (Skalenelastizität) an. Gleichung (1) enthält zum einen keine Restriktionen für die Koeffizientenmatrix² und zum anderen bekannte Produktionstypen als Spezialfälle.

Lässt sich die Menge der Produktionsfaktoren in m disjunkte einelementige Teilmengen zerlegen, ist Θ eine Diagonalmatrix, was bedeutet, dass jeder Faktor als Elementarfaktor in die Produktion eingeht. Für die Klasse linear-homogener Funktionen ($r = 1$) erhält man beispielsweise für $p = 1$ den Fall der perfekten Substitution

$$Y = \sum_{i=1}^m |\theta_i x_i|. \quad (2)$$

Wird der Grenzwert für $p \rightarrow -\infty$ betrachtet, führt (1) auf eine limitationale Produktionsfunktion (*Leontieff*-Funktion)

$$Y = \lim_{p \rightarrow -\infty} \|\Theta \mathbf{x}\|_p = \min \theta_i x_i, \quad i = 1, \dots, m. \quad (3)$$

In der klassischen volkswirtschaftlichen Theorie werden häufig zwei unabhängige Produktionsfaktoren betrachtet, Arbeit und Kapital. Die Beachtung dieser zusätzlichen Einschränkung führt auf eine weitere, noch speziellere Klasse von Produktionsfunktionen. Aus (1) erhält man ohne weitere Annahmen

$$Y = [(\theta_1 A)^p + (\theta_2 K)^p]^{\frac{1}{p}} \quad (4a)$$

bzw. mit Hilfe der Substitution $\tilde{\theta}_i = \theta_i^p$

$$Y = \left(\tilde{\theta}_1 A^p + \tilde{\theta}_2 K^p \right)^{\frac{1}{p}}, \quad (4b)$$

die CES³-Funktion⁴. Die Substitutionselastizität der beiden Produktionsfaktoren kann durch $\sigma = -(1 - p)^{-1}$ unmittelbar berechnet werden⁵. Im Grenzwert $p \rightarrow 0$ ergibt sich demzufolge die substitutionale *Cobb-Douglas*-Produktionsfunktion⁶

$$Y = a A^{\frac{1}{2}} K^{\frac{1}{2}}, \quad \text{mit } a = \sqrt{\theta_1 \theta_2} \quad (5a)$$

¹Die *Cobb-Douglas*-Produktionsfunktion stellt beispielsweise ein gewogenes geometrisches Mittel der Faktorinputs dar (*Samuelson* [29] Kap. 27).

²Durch die Nicht-Diagonalelemente θ_{ij} , $i \neq j$ können laterale oder mehrstufige Strukturen beschrieben werden, in denen mehrere Faktoren in die Produktion oder die Erzeugung anderer Faktoren eingehen.

³„Constant Elasticity of Substitution“

⁴Die CES-Funktion wird häufig auch mit negativem Vorzeichen des Substitutionsparameters notiert (vgl. *Krelle* [21] Kap. 7.4.2).

⁵ $\sigma = -d \frac{A}{K} / d \text{GRS} \cdot \text{GRS} / \frac{A}{K}$

⁶Logarithmieren und Anwendung der Regel von *l'Hospital* führen zum Ergebnis.

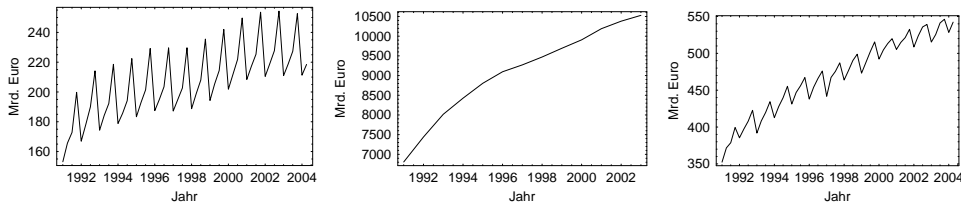


Abbildung 1: Löhne & Gehälter, Anlagevermögen und BIP in Mrd. Euro

bzw.

$$Y = A^\alpha K^{1-\alpha}, \text{ mit } \alpha = \frac{\tilde{\theta}_1}{\tilde{\theta}_1 + \tilde{\theta}_2}, \quad (5b)$$

die die Substitutionselastizität minus eins aufweist. Die Varianten (4a) und (5a) sind etwas restriktiver als die üblich definierte Form, was aus der engen Verknüpfung der Produktionsfaktoren mit ihren jeweiligen Koeffizienten resultiert. Der durch Gleichung (1) beschriebene Ansatz ist im Hinblick auf die Produktivität der einzelnen Faktoren statisch, d.h. die Koeffizienten sind feste Parameter und keine Funktionen der Zeit. Es ist jedoch durchaus denkbar, dass die Produktivität der einzelnen Faktoren infolge äußerer Bedingungen oder technischer Innovationen schwankt. Man kann erwarten, dass sich solche Effekte, solange sie keine globale Niveaushiftung darstellen, volkswirtschaftlich vorwiegend in der Beschäftigung der Produktionsfaktoren ausdrücken. Aus diesem Grund sind zumindest bei der Arbeitsbeschäftigung Schwankungen zu erwarten, da hier jahreszeitliche Bedingungen eine Rolle spielen. Diese Annahme deckt sich mit den Beobachtungen.

3. Konventionelle Modellbildung

Als Datengrundlage werden die vom statistischen Bundesamt⁷ erhobenen Werte für Brutto-Inlandsprodukt, Bruttolöhne und -gehälter und Brutto-Anlagevermögen des Zeitraums 1/1991 bis 2/2004 verwendet. Abbildung 1 zeigt die Beobachtungswerte, die zur Verdeutlichung der zeitlichen Reihenfolge mit einer Linie verbunden wurden. Der zu beobachtende Aufwärtstrend resultiert aus dem Wachstum der Volkswirtschaft und dem Anstieg des Preisniveaus. Da alle Beträge in Preisen der jeweiligen Periode, bzw. das Anlagevermögen zu Wiederbeschaffungspreisen, angegeben sind, kann der Wachstumstrend als vollständig exogen betrachtet werden.

Es treten zunächst zwei Probleme im Zusammenhang mit den Daten auf. Erstens handelt es sich beim BIP und den Löhnen/Gehältern um kumulative Größen, die im Zeitraum zwischen zwei Beobachtungen aufintegriert werden (Flow-Größen), das Anlagevermögen stellt hingegen einen Bestand (Stock-Größe) dar. Zweitens liegen für das Anlagevermögen nur jährliche Messungen vor, für BIP und Arbeitsbeschäftigung werden vierteljährlich Daten erhoben.

Im Hinblick auf die gemeinsame Behandlung von Stock- und Flowdaten bietet ein statischer oder zeitdiskreter Ansatz in der Regel kein Instrumentarium, um die unterschiedlichen Charakteristika adäquat zu unterscheiden. Durch dieses Manko werden mit der Wahl einer solchen Modellklasse bereits Spezifikationsfehler in Kauf genommen. *A.R. Bergstrom*⁸ schreibt zu dieser Problematik:

„Standard procedures for the estimation of econometric models formulated in discrete time make no distinction between the treatment of these two types

⁷<http://www.destatis.de>

⁸*Bergstrom* [5] S. 1 f.

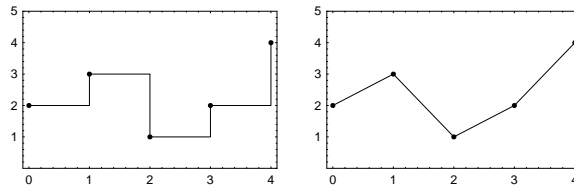


Abbildung 2: *Sprungfunktion und Polygonzug*

of variable, and, consequently, the estimates can be seriously biased because of the specification error resulting from the aggregation over time implicit in the definition of the flow variables.“

Auch das zweite Problem der unterschiedlich langen Messintervalle lässt sich in zeitdiskreten Modellen nur durch ad-hoc Annahmen über die Zwischenwerte der nicht gemessenen Variablen lösen, da die Theorie keine Aussagen über Zeitpunkte machen kann, die nicht mit einem Beobachtungszeitpunkt übereinstimmen. Abbildung 2 zeigt zwei typische Interpolationsverfahren. Wird der letzte Beobachtungswert einfach fortgeschrieben, bis ein neuer Wert eintrifft, entsteht das charakteristische Bild der Sprungfunktion. Der Polygonzug verbindet einfach die Beobachtungswerte durch gerade Linien. In beiden Fällen sind die interpolierten Werte an jeder Stelle leicht zu berechnen.

Es besteht auch die Möglichkeit, intervallweise ein Polynom anzupassen, wodurch eine differenzierbare Kurve erzeugt werden kann. Der Nachteil dieser sog. Spline-Interpolation⁹ ist die aufwendigere Berechnung eines beliebigen Punktes im Messintervall. Allen Methoden ist gemein, dass sie willkürlich sind und keinen inneren Bezug zum Theorierahmen aufweisen.

Für die Parameterschätzung werden im Folgenden Sprungfunktion und Polygonzug betrachtet, wobei die zu interpolierende Größe, das Anlagevermögen, immer zum Ende einer Periode, in diesem Fall zum Ende des Jahres, gemessen wird. Dadurch befinden sich die Beobachtungszeitpunkte jeweils am Ende des Messintervalls und nicht wie Abbildung 2 suggeriert, am Anfang.

Die Anforderung, die die gewählte Interpolationsfunktion an die Daten stellt, führt darüberhinaus dazu, dass nicht alle verfügbaren Messzeitpunkte in die Parameterschätzung einbezogen werden können. Die Sprungfunktion-Interpolation benötigt Kenntnisse über den rechten Rand des Messintervalls, wodurch die BIP- und Gehaltsdaten der ersten beiden Quartale 2004 keine Verwendung finden. Die Polygonzug-Interpolation setzt sogar die Kenntnis beider Intervallränder voraus, was die verwertbare Datenmenge noch weiter reduziert. Besonders bei kurzen Zeitreihen, bedeutet das einen beträchtlichen Informationsverlust, in diesem Fall von mehr als 11% der verfügbaren Quartalsdaten.

Die Parameterschätzung selbst kann nach einem beliebigen Prinzip durchgeführt werden, in der Praxis wird meistens Least Squares (LS) oder Maximum-Likelihood (ML) verwendet (vgl. *Fahrmeir, Künstler, Pigeot und Tutz* [9] Kap. 9). Ausgangspunkt ist das Modell

$$\mathbf{Y} = \mathbf{f}(\boldsymbol{\theta}, \mathbf{X}) + \boldsymbol{\epsilon}, \quad (6)$$

wobei \mathbf{Y} in diesem Fall einen Vektor darstellt und $\boldsymbol{\epsilon}$ ein $N(\mathbf{0}, \sigma^2 \mathbf{I})$ -verteilter Zufallsvektor ist. Der Kleinste-Quadrate-Schätzer ist dann als

$$\hat{\boldsymbol{\theta}}_{LS} = \arg \min_{\boldsymbol{\theta}} \boldsymbol{\epsilon}' \boldsymbol{\epsilon} = \sum_{t=1}^T [Y_t - f(\boldsymbol{\theta}, \mathbf{x}_t)]^2 \quad (7)$$

⁹Siehe bspw. *Bronstein, Semendjajew, Musiol und Mühlig* [6] S. 955 ff.

definiert. Bezeichne \mathbf{y}_t die gegebenen Daten zum Zeitpunkt t , dann ist die Likelihood-Funktion durch $L(\mathbf{y}_1, \dots, \mathbf{y}_T; \boldsymbol{\theta}) = p(\mathbf{z}_1, \dots, \mathbf{z}_T | \boldsymbol{\theta})$ gegeben. Der ML-Schätzer kann damit als

$$\hat{\boldsymbol{\theta}}_{ML} = \arg \max_{\boldsymbol{\theta}} p(\mathbf{y}_1, \dots, \mathbf{y}_T | \boldsymbol{\theta}) \quad (8)$$

definiert werden. Hier bezeichnet $p(\dots)$ die Dichtefunktion. Das ML-Verfahren wählt durch Maximierung der Likelihood den Parametervektor, der am besten zu den gegebenen Daten passt. Aufgrund der Unabhängigkeit der einzelnen Störterme kann die Likelihood-Funktion als Produkt der einzelnen Dichten geschrieben werden

$$L(\mathbf{y}_1, \dots, \mathbf{y}_T; \boldsymbol{\theta}) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2} \left(\frac{y_t - f(\boldsymbol{\theta}, \mathbf{x}_t)}{\sigma} \right)^2}. \quad (9)$$

Es ist leicht zu sehen, dass die Likelihood an keiner Stelle negativ werden kann, wodurch die logarithmierte Likelihood eine umkehrbare Abbildung nach \mathbb{R} darstellt. Ferner gilt

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{ML} &= \arg \max_{\boldsymbol{\theta}} \log p(\mathbf{y}_1, \dots, \mathbf{y}_T | \boldsymbol{\theta}) \\ &= \arg \max_{\boldsymbol{\theta}} -\frac{1}{2} \sum_{t=1}^T [Y_t - f(\boldsymbol{\theta}, \mathbf{x}_t)]^2 \\ &= \arg \min_{\boldsymbol{\theta}} \sum_{t=1}^T [Y_t - f(\boldsymbol{\theta}, \mathbf{x}_t)]^2 = \hat{\boldsymbol{\theta}}_{LS}, \end{aligned} \quad (10)$$

ML- und LS-Schätzer führen zum selben Parametervektor¹⁰. Da, ausgehend von Gleichung (1), die zu minimierende bzw. maximierende Funktion nicht linear ist, muss die Lösung im Allgemeinen durch numerische Verfahren approximiert werden. Im Falle der Kleinsten-Quadrate-Schätzung spricht man auch von NLLS-Schätzung¹¹.

Im vorliegenden Fall wurden die NLLS-Schätzer mit Hilfe des BFGS-Algorithmus¹² bestimmt. Es ergeben sich für die beiden besprochenen Interpolationsvarianten die Parameterschätzungen

$$\hat{\boldsymbol{\theta}}_{Spr.} = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{p} \\ \hat{r} \end{pmatrix} = \begin{pmatrix} 0.4301 \\ 0.4067 \\ 0.1626 \\ 0.9662 \end{pmatrix} \quad \text{bzw.} \quad \hat{\boldsymbol{\theta}}_{Pol.} = \begin{pmatrix} 0.435 \\ 0.3679 \\ 0.2579 \\ 0.9234 \end{pmatrix} \quad (11)$$

im Zusammenhang mit der Produktionsfunktion (4b), unter Einbeziehung der Skalenelastizität. Abbildung 3 zeigt einige Höhenlinien (Isoquanten) der Produktionsfunktion für beide Schätzungen.

Um die Qualität der Schätzungen beurteilen zu können, kann die Summe der Residuenquadrate untersucht werden. Dabei muss jedoch berücksichtigt werden, dass

¹⁰Die Äquivalenz der beiden Schätzverfahren lässt sich darauf zurückführen, dass σ hier ein freier Parameter ist, der keinen Einfluss auf die Lokalisation des Maximums in Abhängigkeit von $\boldsymbol{\theta}$ hat. σ kann auch separat geschätzt und dann in die Likelihood-Funktion eingesetzt werden (Konzentrierte Likelihood).

¹¹„Nonlinear Least Squares“

¹²Das BFGS-Verfahren ist ein numerischer Algorithmus des *Newton*-Typs zur Bestimmung lokaler Extrema. Im Gegensatz zum *Newton-Raphson*-Verfahren kommt der BFGS-Algorithmus ohne *Hesse*-Matrix aus und verwendet statt dessen in jeder Iteration ein positiv-definites Sekantenupdate. Die Bezeichnung BFGS steht für *Broyden*, *Fletcher*, *Goldfarb*, *Shanno* (für eine ausführliche Diskussion siehe *Dennis* und *Schnabel* [7] Kap. 9).

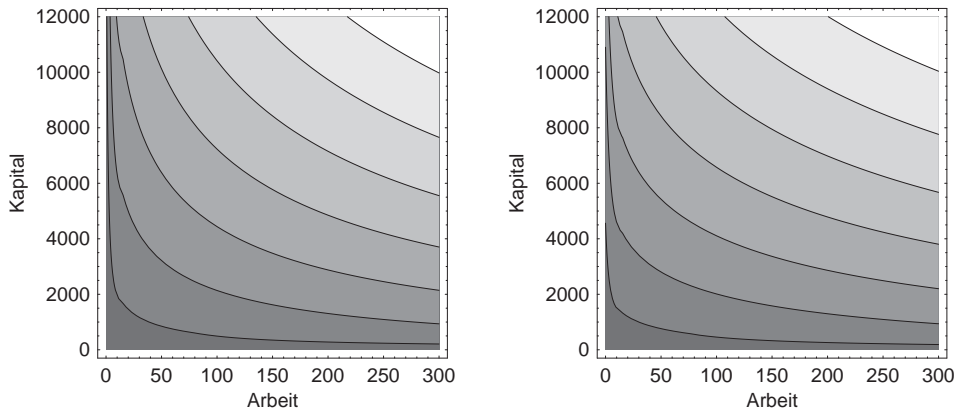


Abbildung 3: *Isoquanten der Sprungfunktion-Interpolation (links) und Polygonzug-Interpolation (rechts)*

beiden Modellen ein unterschiedlicher Stichprobenumfang zugrunde liegt. Interpretiert man weiterhin die Residuen als Schätzer für die Störterme, erhält man auch eine Schätzung für die Varianz der Störterme

$$\hat{\sigma}^2 = \frac{1}{T_{\text{eff}}} \hat{\epsilon}' \hat{\epsilon}, \quad (12)$$

wobei T_{eff} von der Interpolationsmethode abhängt. Mit dieser Größe ist es möglich abzuschätzen, wie groß der stochastische Anteil im Modell sein muss, damit es zu den Daten passt. Natürlich ist es wünschenswert, nur einen geringen Teil durch Zufallsfluktuationen „auspolstern“ zu müssen, da im Gegenzug der aus dem Modell erklärte Anteil steigt. Für die beiden geschätzten Modelle errechnet man

$$\hat{\sigma}_{Spr.}^2 = 49.28 \quad \text{bzw.} \quad \hat{\sigma}_{Pol.}^2 = 45.51. \quad (13)$$

Offensichtlich führt die Interpolation mit Polygonzug zu einer besseren Anpassung. Jedoch wird damit nichts über die generelle Eignung des Modells zur Beschreibung des vorliegenden Prozesses gesagt. Abbildung 4 zeigt eine eingehende Analyse der Residuen zur Überprüfung der Modellspezifikation.

Die Residuen scheinen annahmegemäß normalverteilt zu sein, sind jedoch mit starken Autokorrelationseffekten behaftet. Das bedeutet, dass die Fehler im Rahmen der gegenwärtigen Modellstruktur von den Fehlern der Vorperioden abhängen würden, was im Ergebnis einen klaren Hinweis auf Fehlspezifikation darstellt. Die optische Inspektion der Residuen vermittelt zudem nicht den Eindruck, als würde es sich hier um weißes Rauschen handeln, da Gebiete existieren, in denen auffällig viele aufeinanderfolgende Werte tendenziell größer oder kleiner null sind.

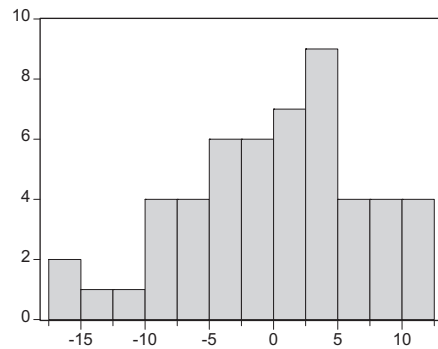
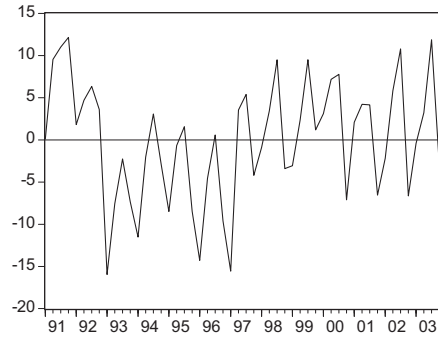
4. Zustandsraummodelle und *Kalman*-Filter

Die Vermutung liegt nahe, dass die beobachteten Unzulänglichkeiten in der Modellspezifikation stark mit der fehlerhaften Behandlung der Variable „Zeit“ zusammenhängen. Das aus diesem Umstand resultierende Problem der mangelnden Unterscheidung zwischen Stock- und Flowvariablen wurde schon angesprochen. Der modellspezifische Fehler ist jedoch viel grundlegender.

Durch einen zeitdiskreten Modellrahmen kann lediglich ein Prozess beschrieben werden, der nur zu diskreten Zeitpunkten stattfindet. Ein Beobachter, der die Auslenkung einer Schaukel jeweils zu definierten Beobachtungszeitpunkten misst, würde

Sample: 1991Q1 2003Q4
Included observations: 52

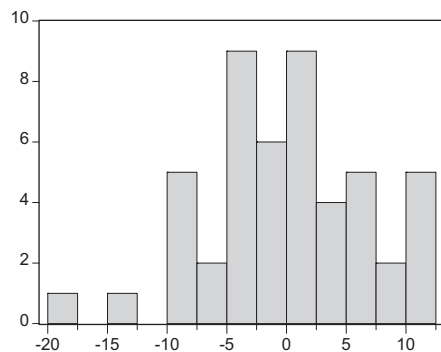
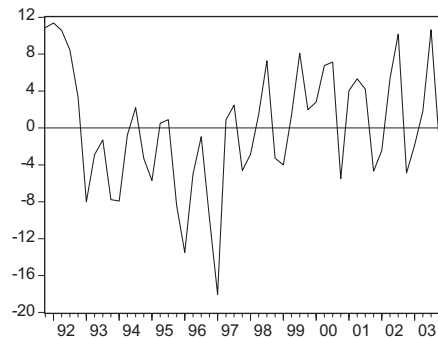
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.328	0.328	5.9368	0.015
		2 -0.077	-0.208	6.2725	0.043
		3 0.188	0.336	8.3018	0.040
		4 0.587	0.479	28.454	0.000
		5 -0.006	-0.514	28.456	0.000
		6 -0.349	-0.144	35.892	0.000
		7 -0.037	0.029	35.978	0.000
		8 0.354	0.153	43.995	0.000
		9 -0.089	0.003	44.516	0.000
		10 -0.272	0.150	49.455	0.000
		11 0.083	0.056	49.930	0.000
		12 0.360	-0.102	59.006	0.000
		13 -0.134	-0.259	60.297	0.000
		14 -0.324	-0.136	68.052	0.000
		15 0.049	0.119	68.232	0.000
		16 0.169	-0.131	70.461	0.000



Series: RESIDUEN1	
Sample 1 52	
Observations 52	
Mean	0.012859
Median	0.878022
Maximum	12.13850
Minimum	-15.95390
Std. Dev.	7.088818
Skewness	-0.332796
Kurtosis	2.544415
Jarque-Bera	1.409569
Probability	0.494215

Sample: 1991Q4 2003Q4
Included observations: 49

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.414	0.414	8.9214	0.003
		2 0.001	-0.206	8.9214	0.012
		3 0.192	0.349	10.926	0.012
		4 0.478	0.321	23.597	0.000
		5 0.046	-0.411	23.716	0.000
		6 -0.254	-0.062	27.476	0.000
		7 -0.019	0.062	27.497	0.000
		8 0.304	0.169	33.114	0.000
		9 0.020	-0.051	33.139	0.000
		10 -0.169	0.071	34.974	0.000
		11 0.041	-0.010	35.086	0.000
		12 0.260	-0.030	39.666	0.000
		13 -0.100	-0.242	40.357	0.000
		14 -0.320	-0.076	47.675	0.000
		15 -0.073	0.088	48.064	0.000
		16 0.018	-0.252	48.088	0.000



Series: RESIDUEN2	
Sample 1991Q4 2003Q4	
Observations 49	
Mean	0.012347
Median	0.506440
Maximum	11.37590
Minimum	-18.04490
Std. Dev.	6.590329
Skewness	-0.258498
Kurtosis	2.890969
Jarque-Bera	0.569980
Probability	0.752022

Abbildung 4: Residualanalyse der Sprungfunktion-Interpolation (oben) und Polygonzug-Interpolation (unten)

im zeitdiskreten Kontext einen sprunghaft alternierenden Prozess beschreiben. Diese Vorstellung ist absurd, da man intuitiv weiß, dass die Schaukel sich stetig bewegt, ihre Position aber nur an bestimmten Punkten gemessen wird. Diese Überlegung kann im Rahmen eines Zustandsraummodells¹³ formalisiert werden. Das Zustandsraummodell besteht aus zwei Gleichungen, der Zustandsgleichung und der Messgleichung.

$$d\mathbf{y}(t) = \mathbf{f}(\mathbf{y}, t)dt + \mathbf{G}(\mathbf{y}, t)d\mathbf{W}(t) \quad t \in [t_0, t_T] \quad (14a)$$

$$\mathbf{z}(t_i) = \mathbf{h}(\mathbf{y}, t_i) + \epsilon_i \quad i = 0, \dots, T \quad (14b)$$

Die Zustandsgleichung (14a) wird dabei als zeitstetige stochastische Differentialgleichung (SDGL) formuliert, die Messgleichung (14b) bildet die Beobachtungen zu diskreten Zeitpunkten ab. In diesem Fall wurde ein lineares Messmodell

$$\mathbf{z}(t_i) = \mathbf{H}\mathbf{y}(t_i) + \mathbf{d} + \epsilon_i \quad i = 0, \dots, T \quad (14c)$$

gewählt. Die Schreibweise $d\mathbf{y}(t)$, dt bzw. $d\mathbf{W}(t)$ ist formal als Kurzform der korrespondierenden Integralgleichung aufzufassen. Insbesondere ist $\mathbf{y}(t)$ nicht differenzierbar, da $\mathbf{W}(t)$ einen n -dimensionalen *Wiener*-Prozess darstellt, der durch normalverteilte, unabhängige Zuwächse charakterisiert ist¹⁴.

Als erstes wird nun die SDGL (14a) betrachtet. Es kann gezeigt werden, dass im Falle der Existenz einer eindeutigen Lösung dieselbe durch einen *Markov*-Prozess gegeben ist (für einen ausführlichen Beweis siehe *Arnold* [2] S. 118 ff., 158 ff.). Weiterhin kann gezeigt werden, dass der Lösungsprozess ein Diffusionsprozess¹⁵ ist, sofern \mathbf{f} und \mathbf{G} stetig in t sind (*Arnold* [2] S. 163 ff.). Die zugehörige Diffusionsgleichung für die Wahrscheinlichkeitsdichte $p(\mathbf{y}, t)$ ist mit der *Fokker-Planck*-Gleichung gegeben (*Arnold* [2] S. 56 ff., *Friedrich* und *Lange* [12] S. 92 ff., *Haken* [16] S. 171 ff., *Risken* [28] S. 63 ff., *Singer* [31] S. 44 f., [32] S. 48 ff.)

$$\dot{p}(\mathbf{y}, t) = - \sum_{i=1}^n \frac{\partial}{\partial y_i} f_i(\mathbf{y}, t)p(\mathbf{y}, t) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2}{\partial y_i \partial y_j} \Omega_{ij}(\mathbf{y}, t)p(\mathbf{y}, t), \quad (15)$$

wobei $\mathbf{\Omega} = \mathbf{G}\mathbf{G}'$ ist¹⁶. Gleichung (15) gilt analog für die Übergangsdichte $p(\mathbf{y}, t|\mathbf{x}, s)$, mit der Anfangsbedingung $p(\mathbf{y}, s|\mathbf{x}, s) = \delta(\|\mathbf{y} - \mathbf{x}\|)$ (*Diracsche* Deltafunktion). Mit Hilfe der *Fokker-Planck*-Gleichung gelingt es, Formeln für die zeitabhängigen Momente abzuleiten (*Jazwinski* [19] Kap. 6.4). Für das erste Moment gilt

$$\begin{aligned} \dot{\boldsymbol{\mu}}(t) &= \int \mathbf{y}\dot{p}(\mathbf{y}, t|\mathbf{x}, s)d\mathbf{y} \\ &= \int \mathbf{f}(\mathbf{y}, t)p(\mathbf{y}, t|\mathbf{x}, s)d\mathbf{y} \\ &= E[\mathbf{f}(\mathbf{y}, t)]. \end{aligned} \quad (16a)$$

¹³Engl. „state space model“.

¹⁴Handelt es sich bei der Diffusionsmatrix $\mathbf{G}(\mathbf{y}(t), t)$ um eine Konstante oder eine nicht stochastische Funktion, kann das stochastische Integral $\int_{t_0}^t \mathbf{G}(\mathbf{y}(s), s)d\mathbf{W}(s)$ als gewöhnliches *Riemann-Stieltjes*-Integral aufgefasst und im herkömmlichen Sinne als Summengrenzwert ausgewertet werden (*Singer* [31] S. 34). Stellt \mathbf{G} eine stochastische Funktion dar, hängt der Grenzwert der Summe von der Lokalisation der Zwischenpunkte ab, wodurch das stochastische Integral nur im Rahmen der *Itô*- oder *Stratonovich*-Interpretation einer Lösung zuzuführen ist (*Arnold* [2] S. 72 ff., *Gard* [14] S. 35 ff., *Singer* [32] S. 27 ff.)

¹⁵Diffusionsprozesse stellen spezielle *Markov*-Prozesse mit stetigen Realisierungen dar.

¹⁶Handelt es sich bei \mathbf{G} um eine stochastische Funktion, gilt die Gleichheit nur für die *Itô*-Lösung des stochastischen Integrals. Die *Stratonovich*-Lösung kann jedoch generell mit Hilfe der *Wong-Zakai*-Korrektur (*Arnold* [2] S. 180, *Singer* [31] S. 41) in die *Itô*-Lösung überführt werden. Für nicht stochastische \mathbf{G} stimmen beide Interpretationen überein.

Die Ableitungen wurden dabei durch partielle Integration auf \mathbf{y} abgewälzt, wobei ausgenutzt wurde, dass alle Randterme für $|y_i| \rightarrow \infty$ gegen null gehen. Für das zweite Moment erhält man auf analogem Weg

$$\dot{\Sigma}(t) = \text{Cov}[\mathbf{f}(\mathbf{y}, t), \mathbf{y}] + \text{Cov}[\mathbf{y}, \mathbf{f}(\mathbf{y}, t)] + E[\Omega(\mathbf{y}, t)]. \quad (16b)$$

Treten keine übermäßigen Nichtlinearitäten auf, lassen sich die Übergangsdichten für kleine Zeiträume als approximativ *gaußsch* behandeln und sind infolge dessen durch die ersten beiden Momente vollständig spezifiziert.

Es handelt sich bei Gleichung (16a) und (16b) formal nicht um Differentialgleichungen, da auf der rechten Seite Erwartungswerte stehen. Ferner ist der Erwartungswert (16a) nicht die deterministische Gleichung, die durch Weglassen des Fluktuationsterms in (14a) entstehen würde, wie man naiv hätte vermuten können. Es ist jedoch möglich, eine *Taylor*-Entwicklung um den Erwartungswertvektor von \mathbf{y} durchzuführen und man erhält

$$\dot{\boldsymbol{\mu}}(t) = \mathbf{f}(\boldsymbol{\mu}, t) + \frac{1}{2} \mathbf{f}_{\mathbf{y}\mathbf{y}}(\boldsymbol{\mu}, t) \Sigma(t) + O(\|\mathbf{y} - \boldsymbol{\mu}\|^3) \quad (17a)$$

$$\begin{aligned} \dot{\Sigma}(t) &= \mathbf{A}(\boldsymbol{\mu}, t) \Sigma(t) + \Sigma(t) \mathbf{A}'(\boldsymbol{\mu}, t) + \Omega(\boldsymbol{\mu}, t) \\ &+ \frac{1}{2} \Omega_{\mathbf{y}\mathbf{y}}(\boldsymbol{\mu}, t) \Sigma(t) + O(\|\mathbf{y} - \boldsymbol{\mu}\|^3), \end{aligned} \quad (17b)$$

wobei die *Jacobi*-Tensoren durch

$$A_{jk} = \frac{\partial f_j(\mathbf{y}, t)}{\partial y_k}, (f_{\mathbf{y}\mathbf{y}})_{jkl} = \frac{\partial^2 f_j(\mathbf{y}, t)}{\partial y_k \partial y_l}, (\Omega_{\mathbf{y}\mathbf{y}})_{jklm} = \frac{\partial^2 \Omega_{jk}(\mathbf{y}, t)}{\partial y_l \partial y_m} \quad (17c)$$

definiert sind. Die *Taylor*-Reihe wurde in (17a) und (17b) bis zum Auftreten von Momenten höherer als zweiter Ordnung entwickelt¹⁷, die in Gleichung (17b) schon mit dem quadratischen Glied entstehen. Wird die Entwicklung nach dem linearen Glied abgebrochen, erhält man die geschlossenen Momentengleichungen

$$\dot{\boldsymbol{\mu}} = \mathbf{f}(\boldsymbol{\mu}) \quad (18a)$$

$$\dot{\Sigma} = \mathbf{A}(\boldsymbol{\mu}) \Sigma + \Sigma \mathbf{A}'(\boldsymbol{\mu}) + \Omega(\boldsymbol{\mu}), \quad (18b)$$

wobei die Zeitvariable zur Vereinfachung nicht notiert wurde. Diese Gleichungen sind jedoch nur für lineare Drift- und Diffusionsterme exakt und gelten ansonsten lediglich approximativ. Im Falle kleiner Zeitschritte und nicht übermäßig starker Nichtlinearitäten liefert die Näherung jedoch gute Ergebnisse. Die approximativen Momentengleichungen bilden den Ausgangspunkt für den erweiterten *Kalman*-Filter (EKF).

Für nichtlineare Strukturen ist generell nicht zu erwarten, dass eine analytische Lösung der Momentengleichungen angegeben werden kann. Der Systemzustand kann aber zwischen den Messungen in beliebig feine Zeitintervalle zerlegt werden, für die Gleichungen (18a) und (18b) numerisch integriert werden können.¹⁸

Als nächstes wird eine Schätzung für den Zustandsvektor benötigt, die aufgrund der Korrelation von Systemzustand und Beobachtungswerten eine Funktion der Daten $\hat{\mathbf{y}} = \mathbf{g}(\mathbf{z})$ sein wird. Man kann zeigen (*Jazwinski* [19] Kap. 5.2), dass der bedingte Erwartungswert $\mathbf{g}(\mathbf{z}) = E[\mathbf{y}|\mathbf{z}]$ unabhängig von den Eigenschaften der bedingten Dichtefunktion $p[\mathbf{y}|\mathbf{z}]$ ein erwartungstreuer¹⁹ und effizienter²⁰ Schätzer ist. Sind \mathbf{y}

¹⁷Die Entwicklung in Gleichung (17a) und (17b) bildet die Grundlage des Second-order Nonlinear Filters (SNF).

¹⁸Zu numerischen Integrationsverfahren siehe *Bronstein, Semendjajew, Musiol und Mühlig* [6] Kap. 19.3.

¹⁹ $E[E[\mathbf{y}|\mathbf{z}]] = E[\mathbf{y}]$

²⁰Effizient bedeutet hier varianzminimal (*Greene* [15] S. 886).

und \mathbf{z} multivariat normalverteilt²¹

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix} \sim N \left(\begin{pmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_z \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yz} \\ \boldsymbol{\Sigma}_{zy} & \boldsymbol{\Sigma}_{zz} \end{pmatrix} \right), \quad (19)$$

gilt der Satz über die Normal-Korrelation²² (*Fahrmeir, Hamerle und Tutz* [10] Satz 3.6, *Singer* [32] Theorem 5.4.1):

Die bedingte Verteilungsfunktion von \mathbf{y} gegeben \mathbf{z} ist eine multivariate Normalverteilung mit der Dichte

$$p(\mathbf{y}|\mathbf{z}) = \phi_{\mathbf{y}|\mathbf{z}} \left(\boldsymbol{\mu}_{\mathbf{y}|\mathbf{z}}, \boldsymbol{\Sigma}_{\mathbf{y}|\mathbf{z}} \right), \text{ mit} \quad (20a)$$

$$E[\mathbf{y}|\mathbf{z}] = \boldsymbol{\mu}_{\mathbf{y}|\mathbf{z}} = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yz} \boldsymbol{\Sigma}_{zz}^{-1} (\mathbf{z} - \boldsymbol{\mu}_z) \quad (20b)$$

$$\text{Var}[\mathbf{y}|\mathbf{z}] = \boldsymbol{\Sigma}_{\mathbf{y}|\mathbf{z}} = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yz} \boldsymbol{\Sigma}_{zz}^{-1} \boldsymbol{\Sigma}_{zy}. \quad (20c)$$

Dieser Satz gilt auch für die generalisierte Inverse²³ (*Liptser und Shiryaev* [23] Kap. 13). Im Rahmen des linearen Messmodells (14c) erhält man die folgenden Größen

$$\boldsymbol{\mu}_z = \mathbf{H} \boldsymbol{\mu}_y + \mathbf{d} \quad (21a)$$

$$\boldsymbol{\Sigma}_{yz} = \boldsymbol{\Sigma}_{yy} \mathbf{H}' \quad (21b)$$

$$\boldsymbol{\Sigma}_{zz} = \mathbf{H} \boldsymbol{\Sigma}_{yy} \mathbf{H}' + \mathbf{R}. \quad (21c)$$

Wird der Zustand auf alle vorherigen Messungen bedingt $\mathbf{y}_i | \mathbf{z}_i, \dots, \mathbf{z}_0 = \mathbf{y}_i | \mathbf{z}^i$, der Index i bezeichnet den Zustand bzw. die Messung zum Zeitpunkt t_i , und spaltet man den letzten Messwert ab $\mathbf{y}_i | \mathbf{z}^i = \mathbf{y} | \mathbf{z}_i, \mathbf{z}^{i-1}$, dann gilt in Erweiterung des Satzes über die Normalkorrelation

$$E[\mathbf{y}_{i+1} | \mathbf{z}_{i+1}, \mathbf{z}^i] = E[\mathbf{y}_{i+1} | \mathbf{z}^i] + \text{Cov}[\mathbf{y}_{i+1}, \mathbf{z}_{i+1} | \mathbf{z}^i] \times \text{Var}[\mathbf{z}_{i+1} | \mathbf{z}^i]^{-1} (\mathbf{z}_{i+1} - E[\mathbf{z}_{i+1} | \mathbf{z}^i]) \quad (22a)$$

$$\text{Var}[\mathbf{y}_{i+1} | \mathbf{z}_{i+1}, \mathbf{z}^i] = \text{Var}[\mathbf{y}_{i+1} | \mathbf{z}^i] - \text{Cov}[\mathbf{y}_{i+1}, \mathbf{z}_{i+1} | \mathbf{z}^i] \times \text{Var}[\mathbf{z}_{i+1} | \mathbf{z}^i]^{-1} \text{Cov}[\mathbf{z}_{i+1}, \mathbf{y}_{i+1} | \mathbf{z}^i]. \quad (22b)$$

Bezeichne im weiteren $\boldsymbol{\mu}_{i|j}$ den Erwartungswertvektor $E[\mathbf{y}_i | \mathbf{z}^j]$ und $\boldsymbol{\Sigma}_{i|j}$ die Kovarianzmatrix $\text{Cov}[\mathbf{y}_i | \mathbf{z}^j]$, lassen sich die bedingten Momente mit Hilfe von (14c), (20a) bis (20c) und (22a), (22b) in der Form

$$\boldsymbol{\mu}_{i+1|i+1} = \boldsymbol{\mu}_{i+1|i} + \boldsymbol{\Sigma}_{i+1|i} \mathbf{H}' (\mathbf{H} \boldsymbol{\Sigma}_{i+1|i} \mathbf{H}' + \mathbf{R}_{i+1})^{-1} (\mathbf{z}_{i+1} - \mathbf{H} \mathbf{y}_{i+1} - \mathbf{d}) \quad (23a)$$

$$\boldsymbol{\Sigma}_{i+1|i+1} = \boldsymbol{\Sigma}_{i+1|i} - \boldsymbol{\Sigma}_{i+1|i} \mathbf{H}' (\mathbf{H} \boldsymbol{\Sigma}_{i+1|i} \mathbf{H}' + \mathbf{R}_{i+1})^{-1} \mathbf{H} \boldsymbol{\Sigma}_{i+1|i} \quad (23b)$$

schreiben. Diese Gleichungen lassen sich rekursiv lösen. Der zugehörige Algorithmus wird als *Kalman-Schmidt-* oder *Extended Kalman-Filter* bezeichnet (*Harvey* [17] S. 160 ff., *Jazwinski* [19] Kap. 9, *Singer* [32] Theorem 5.9.1, [33]).

²¹Die Übergangsdichten des Systemzustands sind approximativ *gaußsch* und der Messfehler wird in der Regel als *gaußsches weißes Rauschen* formuliert, womit $\mathbf{z} \sim N(\mathbf{H}\mathbf{y} + \mathbf{d}, \mathbf{R})$ gilt.

²²Beweis siehe *Fahrmeir, Hamerle und Tutz* [9] S. 27.

²³ $\boldsymbol{\Sigma}^- = (\boldsymbol{\Sigma}' \boldsymbol{\Sigma})^{-1} \boldsymbol{\Sigma}'$

Extended *Kalman-Filter*

Rekursion $i = 0, \dots, T - 1$

Extrapolation $t_i \leq t \leq t_{i+1}$

$$\boldsymbol{\mu}_{t|i} = \boldsymbol{\mu}_{i|i} + \int_{t_i}^t \mathbf{f}(\boldsymbol{\mu}_{\tau|i}) d\tau \quad (24a)$$

$$\boldsymbol{\Sigma}_{t|i} = \boldsymbol{\Sigma}_{i|i} + \int_{t_i}^t \mathbf{A}(\boldsymbol{\mu}_{\tau|i}) \boldsymbol{\Sigma}_{\tau|i} + \boldsymbol{\Sigma}_{\tau|i} \mathbf{A}'(\boldsymbol{\mu}_{\tau|i}) + \boldsymbol{\Omega}(\boldsymbol{\mu}_{\tau|i}) d\tau \quad (24b)$$

Messung $t = t_{i+1}$

$$\mathbf{K}_{i+1} = \boldsymbol{\Sigma}_{i+1|i} \mathbf{H}' \boldsymbol{\Gamma}_{i+1|i}^{-1} \quad (\text{Kalman-Gain}) \quad (24c)$$

$$\boldsymbol{\mu}_{i+1|i+1} = \boldsymbol{\mu}_{i+1|i} + \mathbf{K}_{i+1} \boldsymbol{\nu}_{i+1} \quad (24d)$$

$$\boldsymbol{\Sigma}_{i+1|i+1} = (\mathbf{I} - \mathbf{K}_{i+1} \mathbf{H}) \boldsymbol{\Sigma}_{i+1|i} \quad (24e)$$

Prognosefehler

$$\boldsymbol{\nu}_{i+1} = \mathbf{z}_{i+1} - \mathbf{H} \boldsymbol{\mu}_{i+1|i} - \mathbf{d} \quad (\text{Innovation}) \quad (24f)$$

$$\boldsymbol{\Gamma}_{i+1|i} = \mathbf{H} \boldsymbol{\Sigma}_{i+1|i} \mathbf{H}' + \mathbf{R}_{i+1} \quad (24g)$$

Die Dynamik zwischen den Messzeitpunkten entsteht durch die Differentialgleichungen bzw. deren Lösungen für die Momente (24a) und (24b). Die Entwicklung in der Zeit anhand dieser Gleichungen kann als „Time-Update“ bezeichnet werden. Das „Mess-Update“ hingegen importiert am Messzeitpunkt die neue Information entlang der Filtergleichungen (24c) bis (24g). Die a posteriori Dichte ist dabei eine Superposition der a priori Dichte und des Messfehlers. Im EKF wird davon ausgegangen, dass Übergangsdichten und Messfehler approximativ *gaußsch* sind, wodurch die a posteriori Dichte wieder eine approximative *Gauß-Dichte* ist. Das führt dazu, dass das Mess-Update eine Korrektur der Momentengleichungen durch die neue Messinformation bewirkt.

Der Filter muss lediglich mit den Startwerten $\boldsymbol{\mu}_{0|0}$ und $\boldsymbol{\Sigma}_{0|0}$ initialisiert werden, um die bedingte Dichte $p(\mathbf{y}_T | \mathbf{z}^T)$ sequentiell berechnen zu können. Dies wäre prinzipiell auch ohne Filteralgorithmus durch Anwendung des Satzes über die Normalkorrelation möglich, es würden jedoch sehr schnell immens große Matrizen auftreten, die nicht ohne weiteres einer numerischen Kalkulation zuzuführen wären.

Die sequentielle Berechnung mit dem *Kalman-Filter* hat weiterhin den Vorteil, dass die benötigten Terme für die Parameterschätzung mit der Maximum-Likelihood Methode quasi als Nebenprodukt anfallen. Es kann mit Hilfe der Normalkorrelation gezeigt werden, dass die Gesamtstreuung $\boldsymbol{\Gamma}_{i+1|i}$ als Varianz des Prognosefehlers $\boldsymbol{\nu}_{i+1}$ aufgefasst werden kann²⁴. Darüber hinaus kann die endlichdimensionale Verteilung der Messzeitreihe durch iterierte Anwendung der *Bayes-Formel* $P(A, B) = P(A|B)P(B)$ als Produkt von Übergangswahrscheinlichkeiten geschrieben werden

$$P(\mathbf{z}^T) = \prod_{i=1}^T P(\mathbf{z}_i | \mathbf{z}^{i-1}) P(\mathbf{z}_0). \quad (25)$$

Gleichung (25) gilt analog für die Dichtefunktion. Damit kann die Likelihood-Funktion im Rahmen der Prognose-Fehler-Zerlegung²⁵ (*Schweppe* [30]) für einen

²⁴ $\text{Var}(\mathbf{z}_{i+1} - E[\mathbf{z}_{i+1} | \mathbf{z}_i]) = \text{Var}(\mathbf{z}_{i+1} - E[\mathbf{z}_{i+1}] - \text{Cov}[\mathbf{z}_{i+1}, \mathbf{z}_i] \text{Var}[\mathbf{z}_i]^{-1} (\mathbf{z}_i - E[\mathbf{z}_i])) = \text{Var}(\mathbf{z}_{i+1}) - \text{Cov}(\mathbf{z}_{i+1}, \mathbf{z}_i) \text{Var}(\mathbf{z}_i)^{-1} \text{Cov}(\mathbf{z}_i, \mathbf{z}_{i+1}) = \text{Var}(\mathbf{z}_{i+1} | \mathbf{z}_i)$

²⁵Der Prognosefehler $\boldsymbol{\nu}_i$ ist eine approximativ normalverteilte Zufallsvariable mit Erwartungswert null und Kovarianzmatrix $\boldsymbol{\Gamma}_{i|i-1}$.

beliebigen Parametervektor $\boldsymbol{\theta}$ in der Form

$$L(\mathbf{z}^T; \boldsymbol{\theta}) = \prod_{i=1}^T \frac{1}{\sqrt{(2\pi)^q |\boldsymbol{\Gamma}_{i|i-1}|}} \cdot e^{-\frac{1}{2} \boldsymbol{\nu}'_i \boldsymbol{\Gamma}_{i|i-1}^{-1} \boldsymbol{\nu}_i} \cdot p(\mathbf{z}_0) \quad (26)$$

geschrieben werden. Der Parameter q gibt dabei die Dimension des Innovationsvektors an, kann aber vernachlässigt werden, da er nicht von den Daten abhängt.

Im Falle autonomer, linearer Driftfunktionen gelten die abgeleiteten Formeln exakt. Es lassen sich dann auch explizite Lösungen für die Momente (24a) und (24b) angeben (*Harvey* [17] Kap. 3.2, *Kalman* [20], *Lütkepohl* [24] Kap. 13.3, *Singer* [31] Kap. 4.4, [32] Kap. 5.6). Höhere *Taylor*-Entwicklungen (17a), (17b) um den Erwartungswert führen auf Second-order Nonlinear Filter- (SNF) oder Higher-order Nonlinear Filtergleichungen (HNF) (*Singer* [33], [34]).

5. Maximierung der Likelihood-Funktion, Missing Data und *Kalman*-Glätter

Die Maximierung der Likelihood-Funktion erfolgt numerisch mit Hilfe von Iterationsverfahren des *Newton*-Typs. In diesem Zusammenhang ist es günstiger, die logarithmierte Likelihood-Funktion, bzw. den Teil, der von den Daten abhängt, zu betrachten

$$l(\mathbf{z}^T; \boldsymbol{\theta}) = \sum_{i=1}^T l_{i|i-1} \sim -\frac{1}{2} \sum_{i=1}^T \left(\log \det [\boldsymbol{\Gamma}_{i|i-1}] + \boldsymbol{\nu}'_i \boldsymbol{\Gamma}_{i|i-1}^{-1} \boldsymbol{\nu}_i \right). \quad (27)$$

Der Gradient nach $\boldsymbol{\theta}$, der auch als Likelihood-Score $\mathbf{s}(\boldsymbol{\theta})$ bezeichnet wird, wird im ersten Schritt numerisch bestimmt. Ist seine Vektornorm gleich null, ist eine notwendige Bedingung für das Vorliegen eines Maximums erfüllt. Gleichzeitig muss die *Hesse*-Matrix $\mathbf{H}(\boldsymbol{\theta})$ an dieser Stelle negativ-definit sein. Da die numerische Berechnung der *Hesse*-Matrix sehr aufwendig ist, wurde im vorliegenden Beispiel mit dem BFGS-Algorithmus (*Dennis* und *Schnabel* [7] S. 198 ff.) gearbeitet, der Sekanten-Updates verwendet und dessen Konvergenzeigenschaften annähernd denen des *Newton-Raphson*-Verfahrens entsprechen.

Gleichzeitig gilt die Beziehung $\mathbf{F}(\hat{\boldsymbol{\theta}}) = E[\mathbf{ss}' | \hat{\boldsymbol{\theta}}]$ (*Greene* [15] S. 890, *Mardia, Kent* und *Bibby* [25] S. 98, *Singer* [31] S. 107), wobei die *Fisher*-Informationsmatrix $\mathbf{F}(\hat{\boldsymbol{\theta}})$ für $T \rightarrow \infty$ gegen die Inverse der Kovarianzmatrix des Schätzers $\hat{\boldsymbol{\theta}}$ geht. Da $\mathbf{s}(\hat{\boldsymbol{\theta}})$ eine Summe abhängiger aber unkorrelierter Zufallsvektoren $\mathbf{s}_{i|i-1}(\hat{\boldsymbol{\theta}})$ für $i = 1, \dots, T$ ist²⁶, kann die *Fisher*-Information durch Summation der Erwartungswerte des Dya-denproduktes geschrieben, bzw. für große Stichproben durch Weglassen des Erwartungswertes geschätzt werden

$$\hat{\mathbf{F}}(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^T \mathbf{s}_{i|i-1} \mathbf{s}'_{i|i-1}(\hat{\boldsymbol{\theta}}). \quad (28)$$

Damit gewinnt man im Konvergenzpunkt des Verfahrens die asymptotische Kovarianzmatrix des Schätzers $\hat{\boldsymbol{\theta}}$.

Herkömmliche numerische Optimierungsverfahren können lediglich lokale Extrema identifizieren. Welches Maximum aufgesucht wird, sollte die Funktion über mehrere verfügen, hängt in hohem Maße von der Wahl der Startwerte ab²⁷. Deshalb ist es

²⁶Genauer handelt es sich bei $\mathbf{s}_{i|i-1}(\hat{\boldsymbol{\theta}})$ um eine Martingaldifferenz (zum Begriff des Martingals siehe *Arnold* [2] Kap. 1.9).

²⁷Für eine ausführliche Diskussion siehe *Dennis* und *Schnabel* [7].

sinnvoll, einzelne Modellteile mit verschiedenen systematisch oder zufällig gewählten Startwerten zu schätzen und die gewonnenen Parameterwerte als Startwerte für die simultane Schätzung des gesamten Modells zu verwenden.

Die Behandlung fehlender Daten kann im vorliegenden Modellkontext besonders elegant gelöst werden. Zum einen wird ohnehin durch die Normalkorrelation Messinformation auf alle Variablen verteilt, die mit der oder den gemessenen Variablen korreliert sind, zum anderen werden durch Streichung der betreffenden Zeilen in \mathbf{H} und \mathbf{d} in der Messgleichung (14c) nur die gemessenen Prädiktionsfehler in die Loglikelihood-Funktion (27) überführt. Dadurch wird die Parameterkonstellation ermittelt, die am besten zu den gemessenen Daten passt, ohne dass eine Interpolationsannahme wie in Abschnitt 3 gemacht werden muss.

Ist der optimale Schätzer für den Parametervektor ermittelt worden, kann die Interpolation noch durch das Ausnutzen vorgreifender Messinformation verbessert werden. Der in der Zeit vorwärts laufende *Kalman*-Filter erlaubt die Berechnung der bedingten Erwartungswerte $E[\mathbf{y}_{i+1}|\mathbf{z}^i]$. Diese nutzen jedoch nur die vergangene Information aus (Echtzeitprognose). Bei Kenntnis der gesamten Messreihe kann aber der Erwartungswert $E[\mathbf{y}_{i+1}|\mathbf{z}^T]$ herangezogen werden, der für alle Messungen, außer der letzten, mehr Information beinhaltet. Ein entsprechender Algorithmus, der nur noch die Filtergrößen enthält und in der Zeit rückwärts iteriert, ist der RTS- oder *Kalman*-Glätter (*Rauch, Tung und Striebel* [27]). Für jedes beliebige Diskretisierungsintervall Δt , das im Rahmen der numerischen Integration der Filtergleichungen (24a) und (24b) gewählt wird, gilt das folgende Schema.

Rauch-Tung-Striebel-Glätter

Rekursion $i = T - 1, \dots, 0$

Interpolation $t_{i+1-\Delta t} \geq t \geq t_i$

$$\mathbf{F}_t = \boldsymbol{\Sigma}_{t|i} \mathbf{A}'(\boldsymbol{\mu}) \boldsymbol{\Sigma}_{t+\Delta t|i}^{-1} \quad (\text{Glätter-Gain}) \quad (29a)$$

$$\boldsymbol{\mu}_{t|T} = \boldsymbol{\mu}_{t|i} + \mathbf{F}_t (\boldsymbol{\mu}_{t+\Delta t|T} - \boldsymbol{\mu}_{t+\Delta t|i}) \quad (29b)$$

$$\boldsymbol{\Sigma}_{t|T} = \boldsymbol{\Sigma}_{t|i} + \mathbf{F}_t (\boldsymbol{\Sigma}_{t+\Delta t|T} - \boldsymbol{\Sigma}_{t+\Delta t|i}) \mathbf{F}_t' \quad (29c)$$

Wie leicht zu sehen ist, stimmen Filter- und Glätterlösung in den Messzeitpunkten überein, wenn kein Messfehler berücksichtigt wird und alle Systemzustände gemessen werden. In diesem Fall kann der Glätter intervallweise operieren. Werden Messfehler aufgenommen oder bleiben Zustände unbeobachtet, muss der Glätteralgorithmus mit jeder neuen Messung wieder die gesamte Trajektorie rückwärts iterieren.

6. Das stetig/diskrete Zustandsraummodell

Um ein passendes Modell für den volkswirtschaftlichen Produktionszyklus zu finden, wird zuerst die Beschäftigung des Faktors Arbeit untersucht. In Abschnitt 2 wurde bereits angedeutet, dass die Produktivität aufgrund jahreszeitlicher Bedingung schwankt. Diese Schwankungen sind auf wechselnde Temperaturniveaus zurückzuführen, die eine verrauschte Funktion²⁸ der Position der Erde auf ihrer Kreisbahn um die Sonne darstellen²⁹. Damit ist die Arbeitsbeschäftigung in Abhängigkeit der Zeit offenbar durch eine Differentialgleichung zweiter Ordnung zu beschreiben, die

²⁸Verrauscht bedeutet mit Zufallsfluktuationen überlagert, die durch idealisiertes weißes Rauschen dargestellt werden.

²⁹Das Baugewerbe oder die Landwirtschaft sind Beispiele für stark jahreszeitabhängige Produktionszweige.

um einen Rauschterm erweitert wird

$$\ddot{A}(t) + \omega^2 A(t) = g\zeta(t). \quad (30)$$

Die Variable ω bezeichnet hier die Frequenz der ungedämpften Schwingung³⁰, $\zeta(t)$ ist ein *gaußsches* weißes Rauschen. Zu (30) wird jedoch noch ein Dämpfungsterm addiert, der sowohl wirtschaftlich wie auch technisch motiviert werden kann. Es ist einleuchtend, dass ein kontinuierlich linearer Verlauf der Beschäftigung vorteilhaft für die Disposition sowohl des Einkommens, wie auch der Produktion wäre. Deshalb könnte die Volkswirtschaft versuchen, die Fluktuationen innerhalb gewisser Grenzen mit geeigneten Instrumenten systematisch zu kompensieren. Technisch betrachtet wird in (30) permanent stochastische Energie zugeführt. Diese muss kompensiert werden, um zu verhindern, dass das System explosionsartig über alle Grenzen expandiert. Damit erhält man

$$\ddot{A}(t) + \gamma\dot{A}(t) + \omega^2 A(t) = g\zeta(t), \quad (31)$$

den stochastischen Oszillator. Wie ebenfalls in Abschnitt 2 erläutert, wurden die verarbeiteten Daten nicht preis-, saison- oder wachstumsbereinigt, um diese Einflüsse vollständig zu exogenisieren. Diese Trendeinflüsse müssen in Form von Inhomogenitäten³¹ in Gleichung (31) ergänzt werden. Es wäre also wünschenswert, die ursprüngliche Zeitreihe als Summe zweier Zeitreihen darstellen zu können, von denen eine nur den Schwingungsanteil enthält und die andere nur den Trend. Ein entsprechendes Verfahren wurde von *Hodrick* und *Prescott* [18] entwickelt und ist unter dem Begriff *Hodrick-Prescott*- oder HP-Filter bekannt. Ausgehend von der Separation

$$y_t = g_t + c_t, \quad t = 1, \dots, T \quad (32)$$

schlagen *Hodrick* und *Prescott* vor, den zyklischen Anteil c_t (cyclic) vom Wachstumsteil g_t (growth) durch Lösung des Minimierungsproblems

$$\min_{\{g_t\}_{t=1}^T} \left[\sum_{t=1}^T (y_t - g_t)^2 + \lambda \sum_{t=2}^{T-1} (\nabla^2 g_{t+1})^2 \right] \quad (33)$$

zu trennen. Der Parameter λ gewichtet dabei einen „Strafterm“, ∇ ist der Rückwärts-Differenzenoperator. Wird (33) in Matrixschreibweise notiert, nach \mathbf{g} abgeleitet und gleich null gesetzt, erhält man nach elementaren Umformungen³²

$$\mathbf{g} = (\lambda \mathbf{F}' \mathbf{F} + \mathbf{I}_T)^{-1} \mathbf{y} \quad (34a)$$

$$\mathbf{c} = \mathbf{y} - \mathbf{g} \quad (34b)$$

mit

$$\mathbf{F} = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & -2 & 1 \end{pmatrix}. \quad (34c)$$

Hodrick und *Prescott* empfehlen für Quartalsdaten einen λ -Wert von 1600. Abbildung 5 zeigt die isolierten Komponenten der Arbeitsbeschäftigung. Der reine

³⁰Eine Differentialgleichung zweiter Ordnung wird auch als Schwingungsgleichung bezeichnet. Die genaue Motivation der Terme ist der entsprechenden Literatur zu entnehmen, z.B. *Kuchling* [22] Kap. 13.

³¹Zum Begriff der inhomogenen Differentialgleichung vgl. bspw. *Arnold* [3] Kap. 1.

³² $\frac{\partial}{\partial \mathbf{g}} (\mathbf{y} - \mathbf{g})' (\mathbf{y} - \mathbf{g}) + \lambda (\mathbf{F}\mathbf{g})' (\mathbf{F}\mathbf{g}) \stackrel{!}{=} 0 \Rightarrow \lambda \mathbf{F}' \mathbf{F} \mathbf{g} = \mathbf{y} - \mathbf{g} \Leftrightarrow \mathbf{g} = (\lambda \mathbf{F}' \mathbf{F} + \mathbf{I})^{-1} \mathbf{y}$

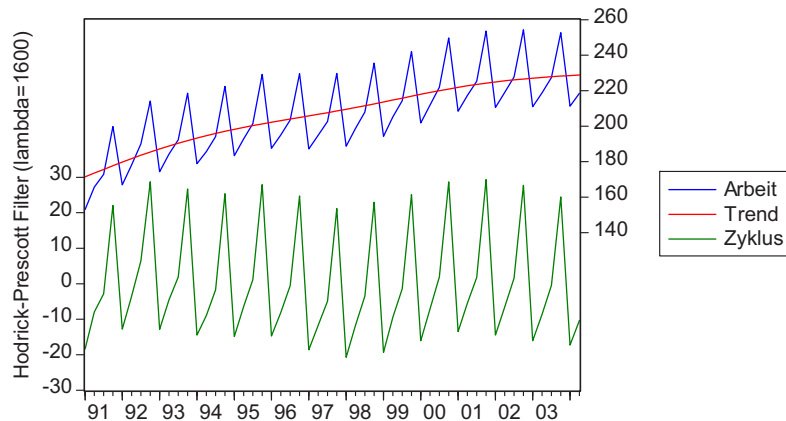


Abbildung 5: Arbeitsbeschäftigung mit Trend und zyklischen Anteilen

Wachstumstrend, der eine Kombination aus Beschäftigungswachstum und Lohnpreisanstieg darstellt, scheint linear oder schwach exponentiell zu verlaufen. Die zyklische Schwankung ist deutlich zu erkennen, hat jedoch nicht den sinusförmigen Charakter, den man hätte erwarten können. Es scheint eher, als würde die Arbeitsbeschäftigung von einer weiteren zyklischen Größe überlagert.

Ökonomisch kann dieser Sachverhalt dadurch erklärt werden, dass in den Sommermonaten die Arbeitsproduktivität offensichtlich hoch ist, der Arbeitseinsatz jedoch weniger stark ansteigt, da viele Arbeitnehmer Urlaub machen. An dieser Stelle ist es auch notwendig, sehr genau zu differenzieren welche Größen im Modell endogen und welche exogen sind. Endogen wird lediglich die Arbeitsproduktivität erklärt, beobachtet wird aber die Arbeitsbeschäftigung, weshalb ein „Urlaubszyklus“ auch als exogene periodische Inhomogenität in die Modellgleichung einfließen muss.

Betrachtet man zunächst nur die ungedämpfte deterministische Differentialgleichung, addiert eine periodische Inhomogenität³³ in der Form

$$\ddot{A} + \omega^2 A = a \sin(2\omega t) \quad (35)$$

und substituiert $b_2 = -\frac{a}{3}\omega^{-2}$, lässt sich durch zweifaches Differenzieren und Einsetzen zeigen, dass

$$\begin{aligned} A &= b_1 \sin(\omega t) + b_2 \sin(2\omega t) \\ &= \sum_{k=1}^2 b_k \sin(k\omega t) \end{aligned} \quad (36)$$

eine Lösung von (35) ist. Es handelt sich bei (36) um eine nach dem zweiten Glied abgebrochene *Fourier*-Reihenentwicklung einer ungeraden³⁴ periodischen Funktion. Abbildung 6 zeigt eine solche Entwicklung bis zum zweiten Glied für die Funktion $f(t) = t$ im periodischen Intervall $[-\frac{1}{2}, \frac{1}{2}]$. Die *Fourier*-Koeffizienten sind dabei $b_k = \frac{1}{k\pi}(-1)^{k+1}$. Ein solcher Verlauf der Arbeitsbeschäftigung ist in mehrererlei Hinsicht schlüssig. Zum einen beschreibt er die Beschäftigungsdämpfung im Sommer, zum anderen zeigt er ein Beschäftigungsmaximum nahe des Jahreswechsels. Das würde mit dem Auftreten konsumintensiver Feiertage und dem Abschlussdruck jährlich disponierter Projekte übereinstimmen.

Berücksichtigt man, dass es sich bei der gemessenen Arbeitsbeschäftigung in Abbildung 5 um eine über das Quartal aufkumulierte Größe handelt, scheint die periodische Inhomogenität die Beschäftigungsstruktur hinreichend genau zu erklären. Die

³³Eine etwaige Phasenverschiebung wird hier zunächst vernachlässigt.

³⁴Für ungerade Funktionen gilt $f(-t) = -f(t)$.

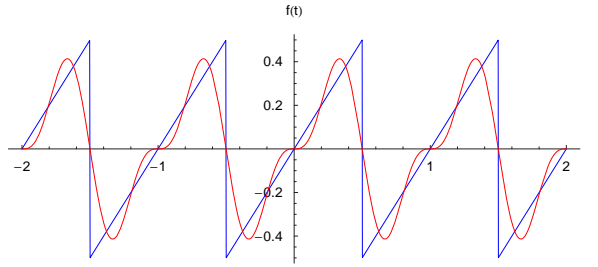


Abbildung 6: Fourier-Entwicklung einer ungeraden Funktion

gesamte stochastische Differentialgleichung lautet damit

$$\ddot{A}_{st.}(t) + \gamma \dot{A}_{st.}(t) + \omega^2 A_{st.}(t) = a \sin[2\omega(t + \varphi)] + b_1 e^{b_2 t} + c + g\zeta(t). \quad (37)$$

Der Affix „st.“ signalisiert, dass es sich um eine Stock-Variable handelt, während die beobachtete Arbeitsbeschäftigung eine Flow-Variable darstellt, die durch Integration der korrespondierenden Stock-Größe jeweils über den Zeitraum eines Quartals entstanden ist.

Gleichung (37) stellt eine sehr allgemeine Form dar. Die jeweils spezielleren Formen entstehen durch Nullsetzen einzelner Parameter. Diese Notation hat Vorteile, da die späteren Schätzungen nicht durch vorher gemachte Annahmen determiniert werden. Wäre beispielsweise der Wachstumstrend nicht exponentiell, sondern linear, würde die Schätzung für den Parameter b_2 nahe bei null liegen und die Inhomogenität auf eine lineare Trendstruktur beschränken.

Im Hinblick auf die Kapitalbeschäftigung fehlen sämtliche Information bezüglich möglicher saisonaler Schwankungen, da jeweils nur ein jährlicher Beobachtungswert vorliegt. Es lassen sich jedoch mit Hilfe der Investitionen Rückschlüsse auf das Kapital ziehen, da diese Ausgaben zur Erhaltung oder Erhöhung des Kapitalstocks sind. *Dornbusch* und *Fischer* unterscheiden in Anlagen-, Bau- und Lagerinvestitionen (*Dornbusch* und *Fischer* [8] Kap. 9). Von letzteren kann man erwarten, dass sie sich der Arbeitsbeschäftigung anpassen, da Lagerinvestitionen Kapital binden, das anderweitig Zinseinkünfte erwirtschaften könnte. Daher kann für die Kapitalbeschäftigung eine zu (37) analoge Gleichung formuliert werden. Es ist lediglich zu beachten, dass es sich beim gemessenen Anlagevermögen bereits um eine Stock-Variable handelt.

Die Produktionsfunktion erhält man dann gemäß (1) bzw. (4b) durch

$$\begin{aligned} Y_{fl.}(t_i) &= \int_{t_{i-1}}^{t_i} Y_{st.}(\tau) d\tau = \int_{t_{i-1}}^{t_i} \|\Theta \mathbf{x}(\tau)\|_p^r d\tau \\ &= \int_{t_{i-1}}^{t_i} \left[\tilde{\theta}_1 A_{st.}^p(\tau) + \tilde{\theta}_2 K_{st.}^p(\tau) \right]^{\frac{r}{p}} d\tau, \end{aligned} \quad (38)$$

für $t_{k-1} < t \leq t_k$. Die Integrationsintervalle für das Brutto-Inlandsprodukt betragen jeweils ein Quartal. Um die Berechnung zu vereinfachen, kann eine kumulierte Größe verwendet werden, die einfach über alle Integrationszeiträume aufaddiert.

$$\begin{aligned} Y_{cum.}(t) &= \sum_{i=1}^{k^*} Y_{fl.}(t_i) + \int_{t_{k^*}}^t Y_{st.}(\tau) d\tau \quad \text{mit} \quad k^* = \arg \max_{t_k \leq t} t_k \\ &= \int_{t_0}^t Y_{st.}(\tau) d\tau \end{aligned} \quad (39)$$

Diese Überlegung gilt analog für die Arbeitsbeschäftigung, da auch hier die Flow-Größe beobachtet wird und auch die Integrationsintervalle dieselben sind.

Die Differentialgleichungen zweiter Ordnung für Arbeits- und Kapitalbeschäftigung lassen sich durch Variablensubstitution in ein System von Differentialgleichungen erster Ordnung überführen. Für den gesamten Zustandsvektor (14a) erhält man

$$d \begin{pmatrix} A_{cum.} \\ A_{st.} \\ L_{st.} \\ K_{st.} \\ NI_{st.} \\ Y_{cum.} \end{pmatrix} = \begin{pmatrix} A_{st.} \\ L_{st.} \\ -\omega_1^2 A_{st.} - \gamma_1 L_{st.} + a_1 \sin[2\omega_1(t + \varphi_1)] + b_{11} e^{b_{12}t} + c_1 \\ NI_{st.} \\ -\omega_2^2 K_{st.} - \gamma_2 NI_{st.} + a_2 \sin[2\omega_2(t + \varphi_2)] + b_{21} e^{b_{22}t} + c_2 \\ \left(\tilde{\theta}_1 A_{st.}^p + \tilde{\theta}_2 K_{st.}^p \right)^{\frac{1}{p}} \end{pmatrix} dt + \text{diag}[0, 0, g_1, 0, g_2, g_3] d(W_1, \dots, W_6)' \quad (40a)$$

Die Zeitvariable wurde aus Übersichtsgründen unterdrückt. Die Integralschreibweise (40a) und die korrespondierende Differentialschreibweise (37) legen nahe, dass das *gaußsche* weiße Rauschen die Ableitung des *Wiener*-Prozesses darstellt $\dot{W}(t) = \zeta(t)$, was natürlich unmöglich ist, da der *Wiener*-Prozess mit Wahrscheinlichkeit eins an jeder beliebigen Stelle nicht differenzierbar ist. Dennoch kann der funktionale Zusammenhang so aufgefasst werden, da beide nur als verallgemeinerte stochastische Prozesse existieren (*Arnold* [2] Kap. 3).

Die Messgleichung (14b) wird hier als identische Funktion der Beobachtungswerte ohne Messfehler formuliert

$$\begin{pmatrix} A_{cum.} \\ K_{st.} \\ Y_{cum.} \end{pmatrix} = \begin{pmatrix} \mathbf{e}'_1 \\ \mathbf{e}'_4 \\ \mathbf{e}'_6 \end{pmatrix} (A_{cum.}, \dots, Y_{cum.})', \quad (40b)$$

wobei \mathbf{e}'_i den i -ten transponierten kanonischen Einheitsvektor bezeichnet. Damit ist das State-Space-Modell komplett spezifiziert, und die Parameter können geschätzt werden. Wie aus (40a) leicht zu erkennen ist, können auch Schätzungen der Netto-Investitionen gewonnen werden, da diese gerade der Änderung des Kapitalstocks entsprechen.

7. Parameterschätzung

Die Schätzung der Parameter selbst ist ein sehr rechenintensives Verfahren, bei dem die Filter- und Maximierungselemente schleifenförmig durchlaufen werden. Abbildung 7 zeigt ein Schema des gesamten Prozesses.

Wie bereits angedeutet, kann die Wahl der Startwerte einen vitalen Einfluss auf das Ergebnis der Parameterschätzung haben, da die Likelihood-Funktion über mehrere lokale Maxima, im Falle periodischer Funktionen sogar über unendlich viele, verfügen kann. Beim vorliegenden Modell wurden einzelne Modellteile im Vorfeld geschätzt, um geeignete Startwerte für die simultane Parameterschätzung zu erhalten. Ferner muss aber auch der *Kalman*-Filter mit einem Erwartungswertvektor und einer Kovarianzmatrix initialisiert werden, die im

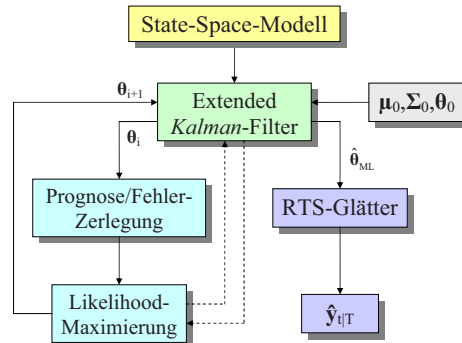


Abbildung 7: *ML*-Schätzung

Allgemeinen nicht bekannt sind, da das State-Space-Modell auch latente Größen abbilden kann, die nicht beobachtbar sein müssen. In diesem Fall kann eine diffuse Anfangsverteilung für die unbekanntenen Startwerte angenommen werden (*Ansley* und *Kohn* [1]). Konkret wurden folgende Werte verwendet

$$\begin{aligned} \boldsymbol{\theta}_0 &= (\omega_1^2, \gamma_1, a_1, \varphi_1, b_{11}, b_{12}, c_1, \omega_2^2, \gamma_2, a_2, \varphi_2, b_{21}, b_{22}, c_2, \\ &\quad \tilde{\theta}_1, \tilde{\theta}_2, p, r, g_1, g_2, g_3)' \\ &= \text{diag} [10^0, 10^{-2}, 10^2, 10^0, 10^0, 10^{-3}, 10^2, 10^0, 10^{-2}, 10^2, \\ &\quad 10^{-1}, 10^2, 10^{-3}, 10^4, 10^1, 10^{-4}, 10^0, 10^{-1}, 10^0, 10^1, 10^0] \times \\ &\quad (2.5, 5.9, 1, 3.8, 3.8, -7.3, 4.4, 2.7, 7.5, 9, 4.8, 3.7, -8.6, 1.8, \\ &\quad 3.9, 5.6, 3.2, 8.9, 7.8, 8.8, 3.9)' \end{aligned} \quad (41a)$$

$$\boldsymbol{\mu}_0 = (0, 160, 0, 6200, 0, 0)' \quad (41b)$$

$$\boldsymbol{\Sigma}_0 = \text{diag} [10^{-6}, 10^4, 10^4, 10^6, 10^6, 10^{-6}]. \quad (41c)$$

Die Erwartungswerte der Diffusoren wurden durch eine grobe Interpolation ermittelt, die Standardabweichungen entsprechen derselben Größenordnung wie die Erwartungswerte. Das numerische Maximierungsverfahren konvergiert nach 141 Iterationen hinreichend. Tabelle 1 zeigt die Ergebnisse der Parameterschätzung. Der Variationskoeffizient setzt die asymptotische Standardabweichung ins Verhältnis zur Größenordnung des geschätzten Parameters. Offenbar liegen keine pathologischen Schätzungen vor, was zunächst aufgrund der geringen Datendichte und der hohen Parameterzahl des Modells zu überprüfen war. Lediglich die beiden nichtlinearen Parameter $\tilde{\theta}_{1/2}$ zeigen eine etwas breitere asymptotische Streuung, was offenbar damit zusammenhängt, dass sie funktional mit dem Parameter p gemäß (4a), (4b) zusammenhängen.

$\hat{\theta}_i \times \text{Skalierung}$	$\hat{\sigma}_i$ (asympt.)	Variationskoeff.
ω_1^2	2.488614×10^0	0.015841
γ_1	5.851422×10^{-2}	0.066943
a_1	1.023391×10^2	2.996754
φ_1	3.824651×10^0	0.076471
b_{11}	3.826755×10^0	0.342050
b_{12}	-7.26182×10^{-3}	0.001818
c_1	4.355059×10^2	3.569347
ω_2^2	2.710988×10^0	0.057419
γ_2	7.514428×10^{-2}	0.141500
a_2	8.906729×10^2	293.5299
φ_2	5.274380×10^{-1}	0.324257
b_{21}	3.366174×10^2	35.25924
b_{22}	-8.61336×10^{-3}	0.001448
c_2	1.812832×10^4	785.5769
$\tilde{\theta}_1$	3.909168×10^1	114.7769
$\tilde{\theta}_2$	5.514963×10^{-4}	0.002681
p	3.184787×10^0	1.996467
r	8.947418×10^{-1}	0.047173
g_1	7.840242×10^0	1.299873
g_2	8.844370×10^1	34.04030
g_3	3.906615×10^0	0.688998

Tabelle 1: Ergebnisse der ML-Schätzung

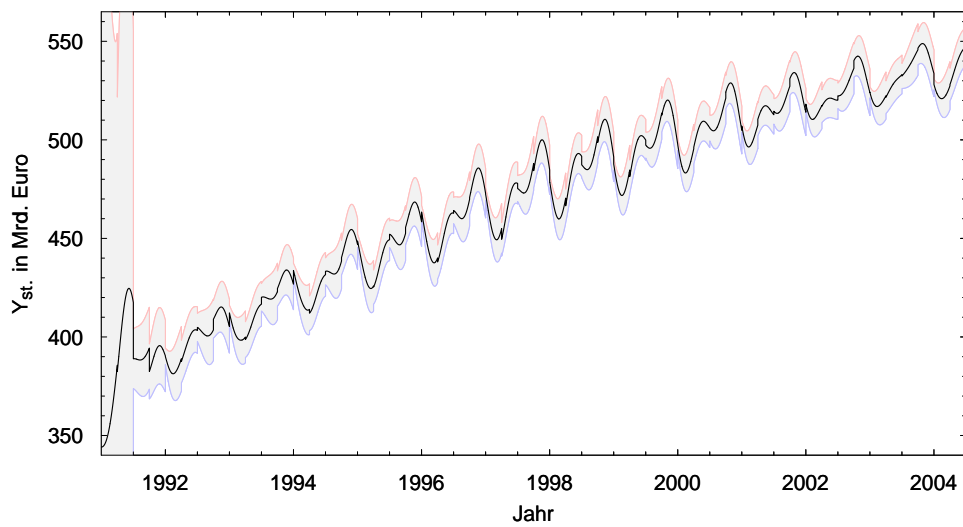
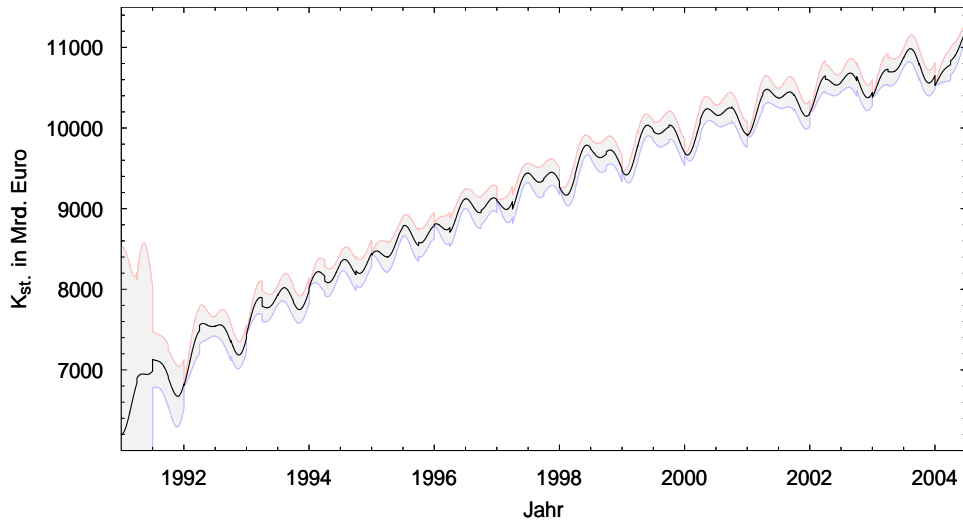
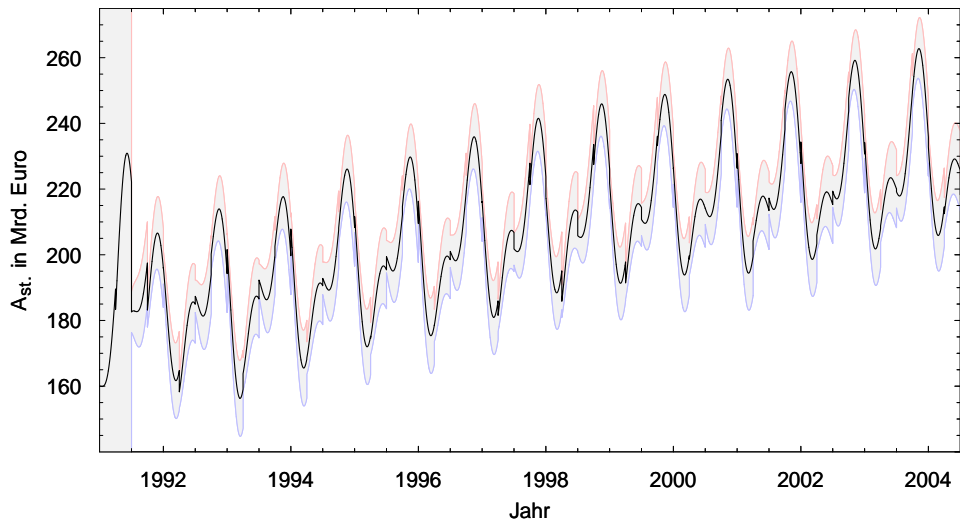


Abbildung 8: Filterlösung und 99%-Konfidenzband

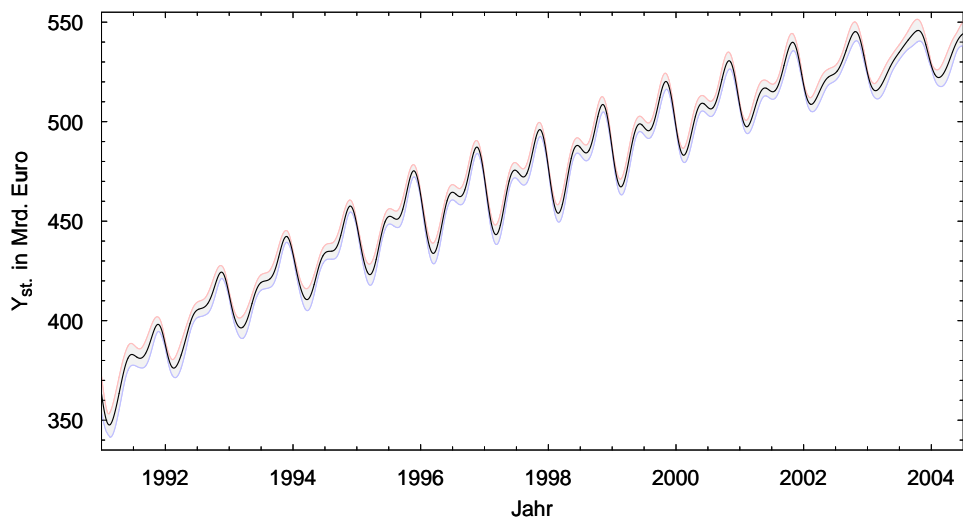
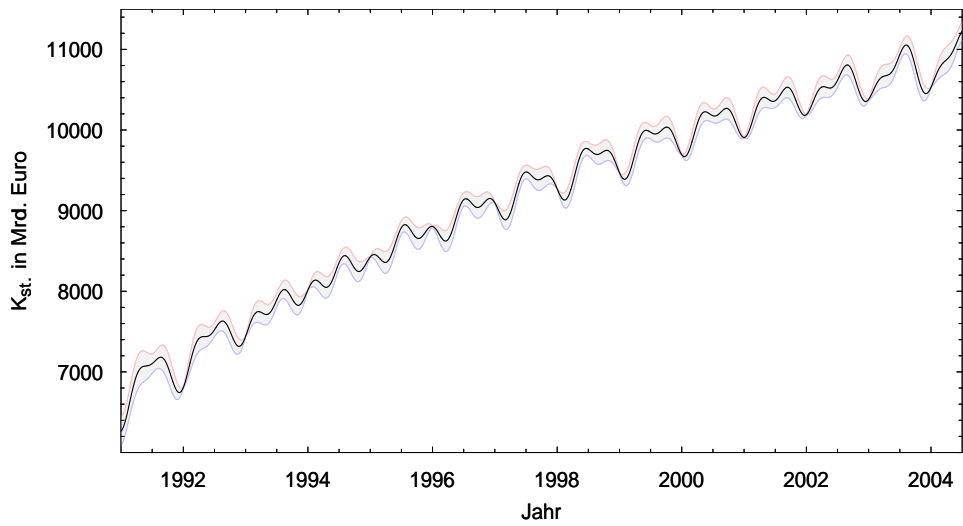
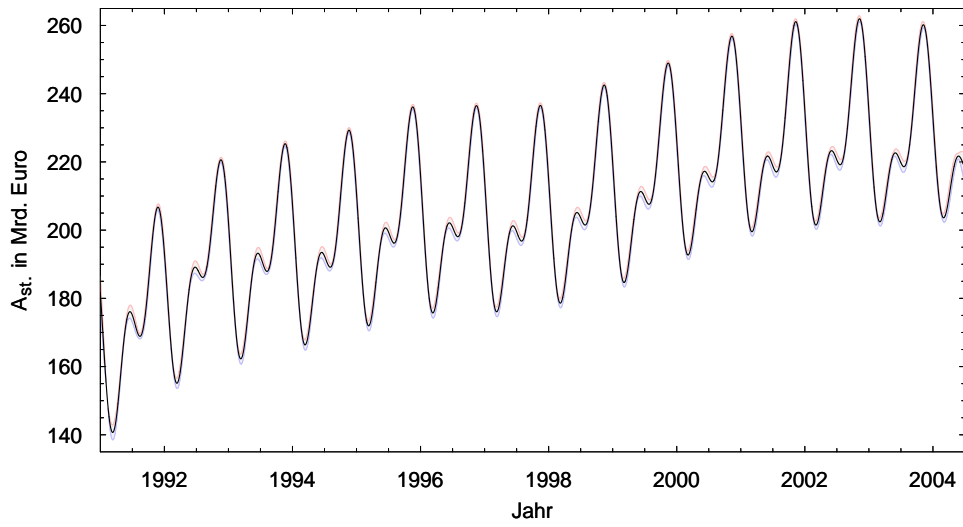


Abbildung 9: Glätterlösung und 99%-Konfidenzband

Mit Hilfe des geschätzten Parametervektors können nun Filter- und Glätterlösungen für jedes beliebige Diskretisierungsintervall berechnet werden. Abbildung 8 und 9 zeigen die Zustandsschätzungen bei einer Diskretisierungsbreite von 1/90, also auf täglicher Basis. Aufgrund der approximativen Normalverteilung der Übergangswahrscheinlichkeiten lassen sich auch Konfidenzbänder berechnen. Es ist deutlich zu erkennen, dass quartalszyklische Messinformation in die Schätzung der Kapitalbeschäftigung einfließt, obwohl das Anlagevermögen nur jährlich beobachtet wird. Es ist auch gut zu sehen, dass der Filterfehler an den Messzeitpunkten des Kapitals vollständig auf null kolabiert. Da die Arbeitsbeschäftigung und das BIP nur über die entsprechenden Flow-Variablen gemessen werden, führt das Eintreffen neuer Information nicht zur völligen Beseitigung der Unsicherheit.

Abbildung 11 zeigt die Filter- und Glättfehler in Form einer Kovarianzmatrix. Es ist gut zu erkennen, dass die einzelnen Variablen auf komplizierte Weise miteinander korreliert sind. Um die Güte der Schätzung zu beurteilen, können, ausgehend von der Modellgleichung (38) in Verbindung mit den Messungen für Y und dem geschätzten Zustands- und Parametervektor, die Residuen ermittelt werden. Abbildung 10 zeigt die Analyse hinsichtlich Normalverteilung und Autokorrelation. Offenbar ist die Zeitreihe frei von signifikanten Autokorrelationseffekten. Auch die Partialisierung der zeitlichen Korrelationen liegt nun unterhalb der Signifikanzschranken. Die Normalverteilungsannahme muss gemäß der *Jarque-Bera*-Statistik ebenfalls nicht verworfen werden. Somit eignet sich die getroffene Modellspezifikation offenbar besser zur Beschreibung des untersuchten Sachverhaltes als die Vorhergehenden.

Um die Güte der Spezifikation zu beurteilen, kann gemäß (12) eine Schätzung für die Varianz des Fehlerterms des ursprünglichen Modells

$$\mathbf{Y}_{fl.} = \int_Q \mathbf{f}(\boldsymbol{\theta}, \mathbf{X}_{st.}) + \boldsymbol{\epsilon} \quad (42)$$

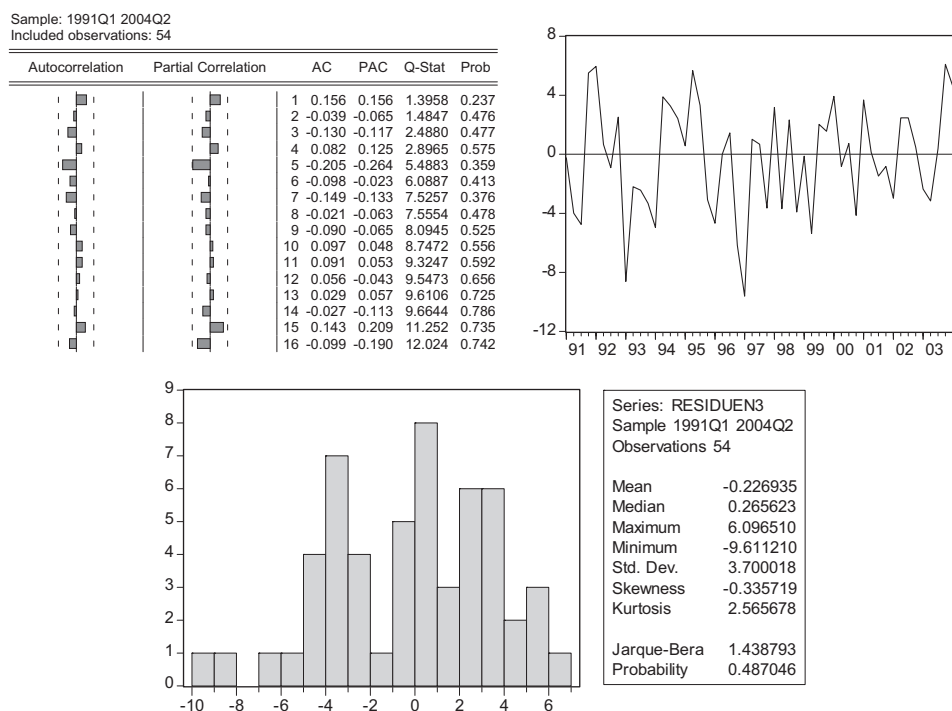


Abbildung 10: Residualanalyse der integrierten Glätterlösung

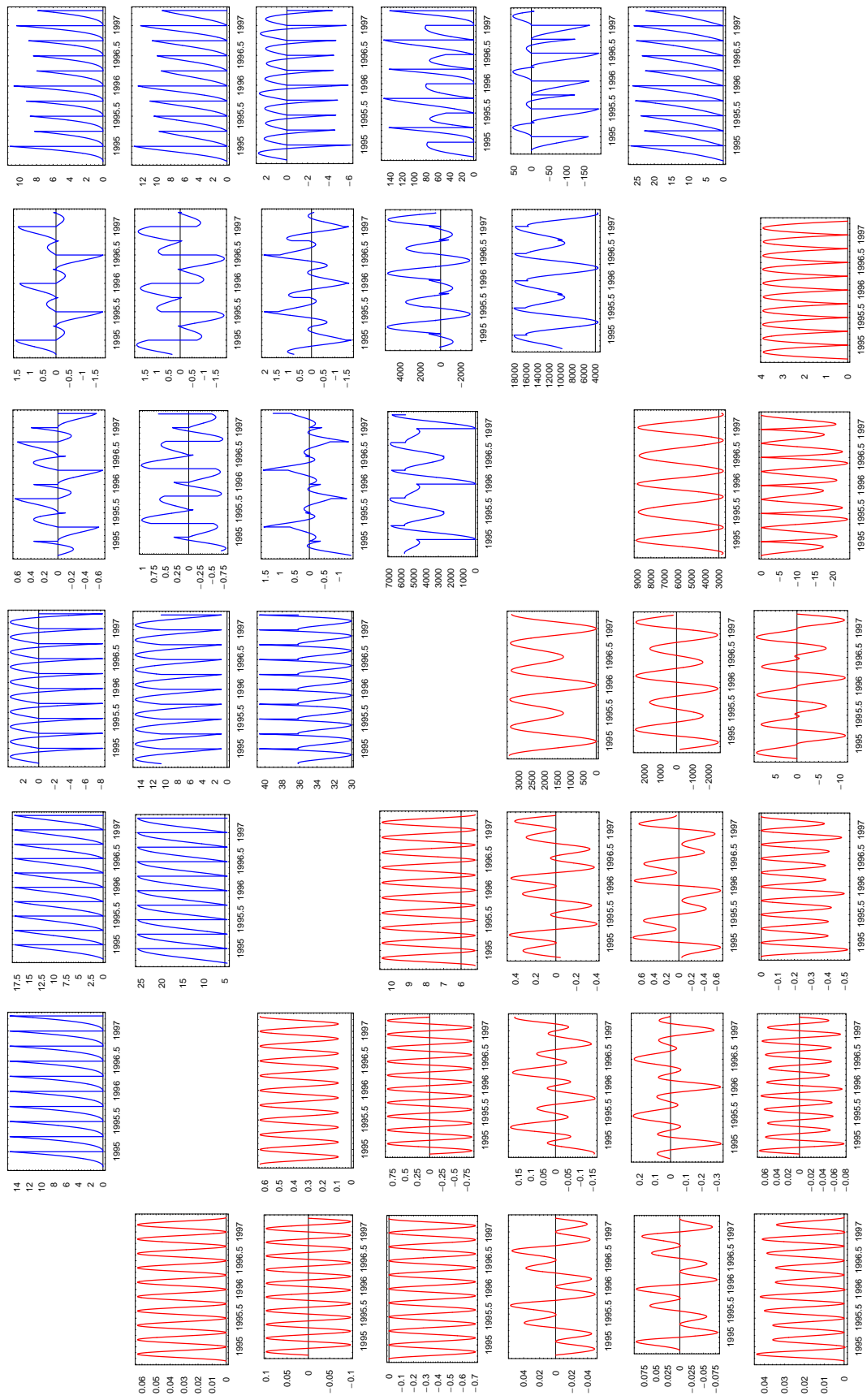


Abbildung 11: EKF-Filterfehler (obere Dreiecksmatrix) und RTS-Glätterfehler (untere Dreiecksmatrix)

erfolgen, die hier nur $\hat{\sigma}^2 = 13.47$ beträgt. Das entspricht einer Reduktion von mehr als 70% gegenüber der Schätzung aus Abschnitt 3. Der Integrationsbereich Q symbolisiert das Quartalsintervall, in dem das BIP aufkumuliert wird.

Zeichnet man die Isoquanten für die Produktionsfunktion, unter Verwendung der mit dem State-Space-Modell geschätzten Parameter, fällt auf, dass die Substitutionselastizität offenbar das Vorzeichen gewechselt hat. Abbildung 12 zeigt die Isoquanten und die Wirkungsfläche der geschätzten Funktion. Dieses Ergebnis mag unerwartet erscheinen, lässt sich aber schlüssig interpretieren.

Formal macht man sich zunächst klar, dass die Produktionsfunktion vom CES-Typ, im Sinne von (4b), mit Skalenelastizität r ist. Ferner gilt offenbar $0 < r < 1 < p$. Die Grenzproduktivität der Faktoren

$$\frac{\partial Y}{\partial x_i} = r\tilde{\theta}_i x_i^{p-1} Y^{1-\frac{p}{r}} \quad (43)$$

für $i = 1, 2$ ist positiv, geht aber für $x_i \rightarrow \infty$ gegen null, wie leicht gezeigt werden kann. Abbildung 13 zeigt den Verlauf der Grenzproduktivität des Faktors Arbeit für konstantes $K = 8000$. Offenbar kommt es in Abhängigkeit des Faktoreinsatzes zunächst zu steigenden und dann fallenden Grenzerträgen, wenn man unterstellt, dass die Produktionssteigerung zu proportionalen Ertragszuwächsen führt. Dieser ertragsgesetzliche³⁵ Verlauf impliziert eine grenztragsmaximale Einsatzmenge eines Faktors, die vom Einsatz des anderen Faktors und den Eigenschaften der Produktionsfunktion³⁶ abhängt. In diesem Zusammenhang wird oft mit einer optimalen (ProKopf-) Kapitalausstattung argumentiert, die nichts anderes darstellt.

Damit ein solcher Verlauf der Grenzproduktivitäts- oder Grenztragskurve für eine Inputmenge $x_{i,j} > 0$ realisiert werden kann, muss die zweite partielle Ableitung

$$\frac{\partial^2 Y}{\partial x_i^2} = r\tilde{\theta}_i x_i^{p-1} Y^{1-\frac{2p}{r}} \left[(r-1)\tilde{\theta}_i x_i^p + (p-1)\tilde{\theta}_j x_j^p \right] \quad (44)$$

für positive Inputs eine Nullstelle aufweisen. Man sieht sofort, dass dazu lediglich der letzte Term in (44) in Frage kommt, woraus sich die Bedingung

$$x_i = - \left[\frac{(r-1)\tilde{\theta}_i}{(p-1)\tilde{\theta}_j} \right]^{\frac{1}{p}} x_j \quad (45)$$

³⁵Vgl. z.B. Woll [36] S. 164 ff.

³⁶Allgemeiner, den Eigenschaften der Technologie, deren effizienter Rand mit Hilfe der Produktionsfunktion abgebildet wird (vgl. Fandel [11] S. 35 ff., S. 51 ff.).

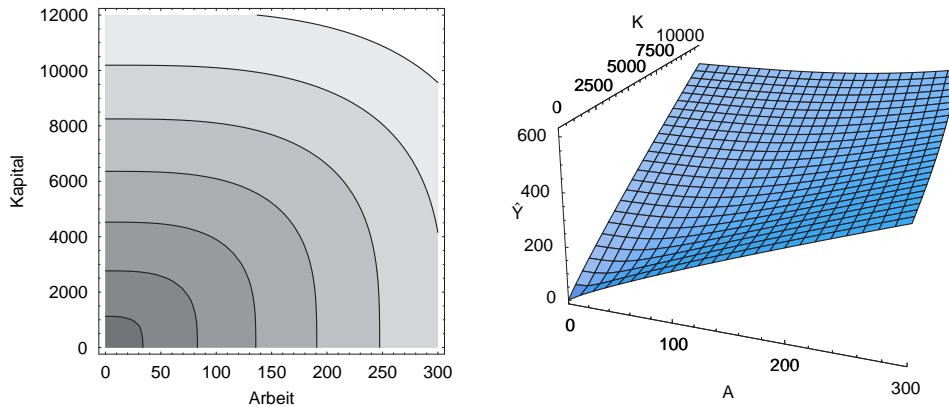


Abbildung 12: Isoquanten und Wirkungsfläche

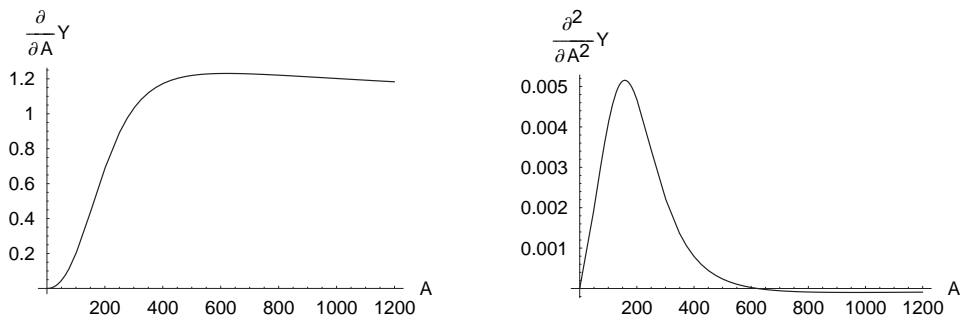


Abbildung 13: Grenzproduktivität und Steigung für $K = 8000$

ableitet. Nun wird klar, dass entweder $r > 1, p < 1$ oder $r < 1, p > 1$ gelten muss, da (45) für $x_{i,j} > 0$ sonst keine Lösung hat. Im Fall $r > 1, p < 1$ gelten offenbar Größensparnisse, da die Grenzproduktivität bei proportionaler Beschäftigungssteigerung aller Faktoren zunimmt, und die Isoquante hat den gewohnt konvexen Verlauf. Diese Konstellation passt gut zu einer betriebswirtschaftlichen Produktionsstruktur. Liegen hingegen sinkende Skalenerträge vor, wie es in der Volkswirtschaft offenbar der Fall ist, müssen die Isoquanten einen konkaven Verlauf aufweisen, da sonst für positive Inputs kein Maximum der partiellen Grenzproduktivität existieren würde. Interessant ist weiterhin, dass die grenzproduktivitätsmaximale Arbeitsbeschäftigung für gegebene Kapitalbeschäftigung aus Abbildung 13 wesentlich höher liegt als die beobachtete. Das weist darauf hin, dass in der Volkswirtschaft eine Unterbeschäftigung oder Unterqualifikation der Arbeitnehmer, oder eine Kombination aus beiden vorliegt.

Anschaulich bedeutet die konkave Krümmung der Isoquante, dass die Produktionsfaktoren in einem eher optionalen Verhältnis³⁷ zueinander stehen, wodurch auch Randlösungen mit einem Produktionsfaktor zulässig werden. Sie sind natürlich nicht realisierbar und resultieren aus der Aggregation aller beobachteten Input-Output-Konstellationen in eine homogene Produktionsfunktion. Idealisiert lassen sich die Randpositionen in Abbildung 12 aber als spezialisierte Ausrichtung der Volkswirtschaft auf Dienstleistung oder Industrie auffassen. Alle dazwischen liegenden Punkte entlang der Isoquante sind Mischformen industrieller Dienstleistungsgesellschaften mit entsprechendem Schwerpunkt. So interpretiert bedeutet die konkave Form, dass eine Spezialisierung auf eine Ausrichtung der Outputerzeugung die Produktivität je eingesetztem Faktor erhöht, was ein völlig schlüssiges und konsistentes Ergebnis darstellt.

Auf der Kostenseite wird im konventionellen Sinn ein optimaler Produktionspunkt dadurch ermittelt, dass der Schnittpunkt zwischen einer konvexen Isoquante und einer Iso-Kostengeraden betrachtet wird. Diese mikroökonomische Vorgehensweise ist im Fall konkaver Isoquanten problematisch, da sie lediglich Randlösungen als optimal identifiziert. Es ist jedoch fraglich, ob volkswirtschaftlich überhaupt so argumentiert werden kann. Geht man im einfachsten Fall davon aus, dass lediglich Unternehmen produzieren und Gewinne erwirtschaften, gelangt man ohne weitere Annahmen zu einer Gleichung für die Unternehmensgewinne (z.B. *Smith* [35] S. 176)

$$\pi = \mathbf{p}'_y \mathbf{y} - \mathbf{p}'_x \mathbf{x}. \quad (46)$$

³⁷ *Frisch* bezeichnet dieses Faktorverhältnis als „alternativ“, benutzt jedoch zur Klassifikation die Änderung der Grenzproduktivität $\frac{\partial^2}{\partial x_i \partial x_j} Y < 0$, nicht die Substitutionselastizität (*Frisch* [13] S. 60).

Das innere Produkt aus Güter- und Güterpreisvektor entspricht der gesamten volkswirtschaftlichen Produktion. Das bedeutet, das Volkseinkommen ist die Summe aus Unternehmensgewinnen und Faktorentgelten, bzw. im untersuchten Modell

$$Y = \pi + w \cdot A + i \cdot K, \quad (47)$$

mit w als Arbeitslohn und i als Kapitalzins. Im langfristigen Gleichgewicht zumindest sind keine Gewinne möglich, da sonst neue Unternehmen in dieselbe Sparte vordringen würden, bis keine Gewinne mehr erzielt werden können (*Woll* [36] S. 181 ff.). Das bedeutet, dass der Output gleich den Kosten ist und zwar unabhängig von der Konstellation der Produktionsfaktoren oder anders formuliert, Isoquante und Iso-Kostenkurve sind kongruent. Ein solcher Verlauf der Iso-Kostenkurve unterstellt die Abhängigkeit der Faktorkosten von ihrer Einsatzmenge im Produktionsprozess, und lässt sich ökonomisch gut interpretieren. Die konkave Iso-Kostenkurve sagt nichts anderes, als dass der Preis für einen Produktionsfaktor, der, infolge vermehrten Einsatzes in der Produktion, auf dem Faktormarkt stärker nachgefragt wird, steigt.

8. Ergebnisse

Wie gezeigt werden konnte, hat die zeitstetige Parameterschätzung Vorteile gegenüber der herkömmlichen Methode. Die Analyse der Residualzeitreihe offenbart eine wesentlich bessere Modellanpassung, durch die von vornherein Verzerrungen minimiert werden. Darüberhinaus kann die Problematik der Stock- und Flowvariablen und die Behandlung fehlender oder irregulärer Messungen korrekt eingebettet werden. Auf diese Vorteile wird von namhaften Ökonometrikern hingewiesen (bspw. *Bergstrom* [4], *Phillips* [26]), die Verfahren konnten sich aber bislang, möglicherweise aufgrund ihrer mathematischen Komplexität, nicht durchsetzen.

Mit Hilfe von Zustandsraummodellen und der sequentiellen Berechnung der Momente mit dem *Kalman*-Filter ist jedoch ein Instrumentarium gegeben, dass in den Ingenieurwissenschaften breite Anwendung findet und eine elegante Lösung der stochastischen Differentialgleichungssysteme ermöglicht. Wie weiterhin gezeigt wurde, beschränken sich die Ergebnisse des Verfahrens nicht auf geringfügig bessere Parameterschätzungen, sondern können im substanzwissenschaftlichen Kontext zu einer völlig anderen Perspektive führen.

Dennoch sollte einschränkend angemerkt werden, dass die abgeleiteten Schätzungen nur die im Rahmen des vorher vereinbarten Modells am besten zu den Beobachtungsdaten passende Manifestation darstellen. Dabei muss das festgelegte Modellgerüst weder vollständig noch richtig sein, noch müssen die geschätzten Parameter den wahren entsprechen. Diesen Einschränkungen sind jedoch alle Modelle unterworfen, wobei der Spezifikationsfehler bei zeitdiskreten Konstrukten bereits bei der Formulierung in Kauf genommen wird.

Literatur

- [1] Ansley, C.F./Kohn, R. (1985):
„*Estimation, filtering and smoothing in State Space Models with incompletely specified initial conditions*“, *The Annals of Statistics*, Vol. 13, No. 4, P. 1286-1316.
- [2] Arnold, L. (1973):
„*Stochastische Differentialgleichungen. Theorie und Anwendung*“, Oldenbourg, München, Wien.
- [3] Arnold, V.I. (2001):
„*Gewöhnliche Differentialgleichungen*“, 2. Auflage, Springer, Berlin, Heidelberg, New York.

- [4] Bergstrom, A.R. (1988):
 „*The History of Continuous-Time Econometric Models*“, *Econometric Theory*, Vol. 4,
 P. 365-383.
- [5] Bergstrom, A.R. (1990):
 „*Continuous Time Econometric Modelling*“, *Oxford University Press, Oxford, New
 York, Toronto*.
- [6] Bronstein, I.N./Semendjajew, K.A./Musiol, G./Mühlig, H. (2001):
 „*Taschenbuch der Mathematik*“, *Harri Deutsch, Thun, Frankfurt/Main*.
- [7] Dennis, J.E./Schnabel, R.B. (1983):
 „*Numerical Methods for Unconstrained Optimization and Nonlinear Equations*“,
Prentice-Hall, New Jersey.
- [8] Dornbusch, R./Fischer, S. (1989):
 „*Makroökonomik*“, 4. Auflage, *Oldenbourg, München, Wien*.
- [9] Fahrmeir, L./Künstler, R./Pigeot, I./Tutz, G. (2003):
 „*Statistik. Der Weg zur Datenanalyse*“, 4. verbesserte Auflage, *Springer, Berlin, Hei-
 delberg, New York*.
- [10] Fahrmeir, L./Hamerle, A./Tutz, G. (1996):
 „*Multivariate statistische Verfahren*“, 2. überarbeitete Auflage, *Walter de Gruyter,
 Berlin, New York*.
- [11] Fandel, G. (1996):
 „*Produktion I - Produktions- und Kostentheorie*“, 5. Auflage, *Springer, Berlin, Hei-
 delberg, New York*.
- [12] Friedrich, H./Lange, C. (1999):
 „*Stochastische Prozesse in Natur und Technik. Modellierung, Simulation, Zuverlässig-
 keit*“, *Harri Deutsch, Thun, Frankfurt/Main*.
- [13] Frisch, R. (1965):
 „*Theory of Production*“, *Dordrecht/Holland*.
- [14] Gard, T.C. (1988):
 „*Introduction to Stochastic Differential Equations*“, *Marcel Dekker, New York*.
- [15] Greene, W.H. (2003):
 „*Econometric Analysis*“, 5. Auflage, *Prentice-Hall Inc., New Jersey*.
- [16] Haken, H. (1990):
 „*Synergetik. Eine Einführung. Nichtgleichgewichts-Phasenübergänge und Selbstorga-
 nisation in Physik, Chemie und Biologie*“, 3. erweiterte Auflage, *Springer, Berlin,
 Heidelberg, New York*.
- [17] Harvey, A.C. (1990):
 „*Forecasting, structural time series models and the Kalman filter*“, *Cambridge Univer-
 sity Press, Cambridge, New York, Port Chester, Melbourne, Sydney*.
- [18] Hodrick, R.J./Prescott, E.C. (1997):
 „*Postwar U.S. Business Cycles: An Empirical Investigation*“, *Journal of Money, Cre-
 dit and Banking*, Vol. 29, No. 1, P. 1-16.
- [19] Jazwinski, A.H. (1970):
 „*Stochastic Processes and Filtering Theory*“, *Academic Press, New York, London*.
- [20] Kalman, R.E. (1960):
 „*A New Approach to Linear Filtering and Prediction Problems*“, *Transactions of the
 ASME-Journal of Basic Engineering*, Band 82, Serie D, S. 35-45.
- [21] Krelle, W. (1969):
 „*Produktionstheorie*“, 2. Auflage, *J.C.B. Mohr (Paul Siebeck), Tübingen*.
- [22] Kuchling, H. (2001):
 „*Taschenbuch der Physik*“, 17. Auflage, *Carl Hanser, München, Wien*.
- [23] Liptser, R.S./Shiryaev, A.N. (2001):
 „*Statistics of Random Processes - II Applications*“, 2. Edition, *Springer, Berlin, Hei-
 delberg, New York*.

- [24] Lütkepohl, H. (1993):
 „*Introduction to Multiple Time Series Analysis*“, 2. Edition, Springer, Berlin, Heidelberg, New York.
- [25] Mardia, K.V./Kent, J.T./Bibby, J.M. (1979):
 „*Multivariate Analysis*“, Academic Press, New York, London.
- [26] Phillips, P.C.B. (1972):
 „*The Structural Estimation of a Stochastic Differential Equation System*“, *Econometrica*, Vol. 40, P. 1021-1041.
- [27] Rauch, H.E./Tung, F./Striebel, C.T. (1965):
 „*Maximum Likelihood Estimates of Linear Dynamic Systems*“, *AIAA Journal*, Band 3, Nr. 8, S. 1445-1450.
- [28] Risken, H. (1989):
 „*The Fokker-Planck Equation. Methods of Solution and Applications*“, 2. Edition, Springer, Berlin, Heidelberg, New York.
- [29] Samuelson, P.A. (1975):
 „*Volkswirtschaftslehre - Band II*“, Bund, Köln.
- [30] Schweppe, F. (1965):
 „*Evaluation of Likelihood Functions for Gaussian Signals*“, *IEEE Transactions on Information Theory*, Band 11, S. 61-70.
- [31] Singer, H. (1992):
 „*Zeitkontinuierliche Dynamische Systeme*“, Campus, Frankfurt/Main, New York.
- [32] Singer, H. (1999):
 „*Finanzmarktökonomie. Zeitstetige Systeme und ihre Anwendung in Ökonometrie und empirischer Kapitalmarktforschung*“, Physica, Heidelberg.
- [33] Singer, H. (2002):
 „*Parameter Estimation of Nonlinear Stochastic Differential Equations: Simulated Maximum Likelihood versus Extended Kalman Filter and Itô-Taylor Expansion*“, *Journal of Computational and Graphical Statistics*, Vol. 11, No. 4, P. 972-995.
- [34] Singer, H. (2004):
 „*Moment equations and Hermite expansion for nonlinear stochastic differential equations with application to stock price models*“, paper presented at the 6th International Conference on Social Science Methodology, Amsterdam.
- [35] Smith, A. (1994):
 „*Einführung in die Volkswirtschaftslehre*“, 2. Auflage, Oldenbourg, München, Wien.
- [36] Woll, A. (1993):
 „*Allgemeine Volkswirtschaftslehre*“, 11. überarbeitete und ergänzte Auflage, Vahlen, München.