Analyse von Zeitreihen mit EViews

Diese Übung zeigt anhand einer (nur einer!!) Zeitreihe, wie man wichtige Methoden der Zeitreihenanalyse mit EViews leicht durchzuführen kann.

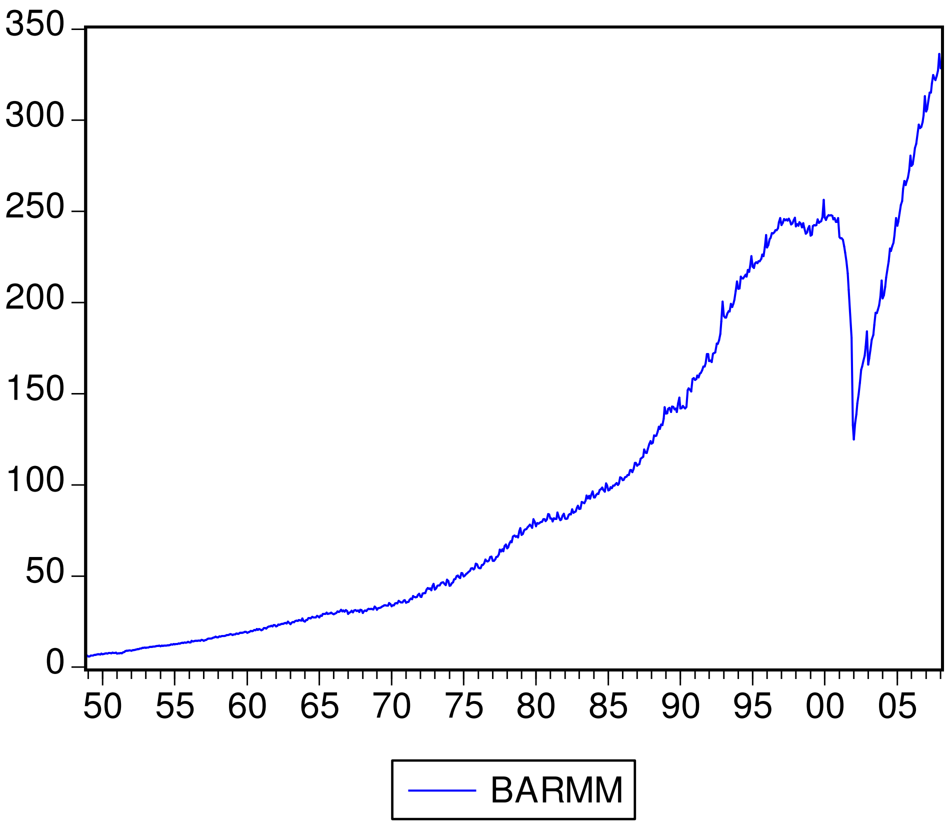
A. Eine lange Zeitreihe mit Strukturbruch

1. Graphische Darstellung

Bei der folgenden Zeitreihe handelt es sich um den Bargeldumlauf (in Mrd. DM) in Deutschland von Dez. 1948 bis Jan. 2008 (insgesamt 710 Monatsdaten). Ab Januar 1999 stellen die Zahlen den Beitrag Deutschlands zum Euro-Bargeldumlauf dar. Die Angaben sind im Interesse der Vergleichbarkeit über längere Zeiträume von € zu DM Angaben nach dem offiziellen Kurs 1 € = 1,95583 DM umgerechnet worden. Zur Illustration:

	Dez. 98	Jan. 99	Febr. 99	März 99
€	-	120,997	121,297	123,651
DM	241,961	236,650	237.236	241,840

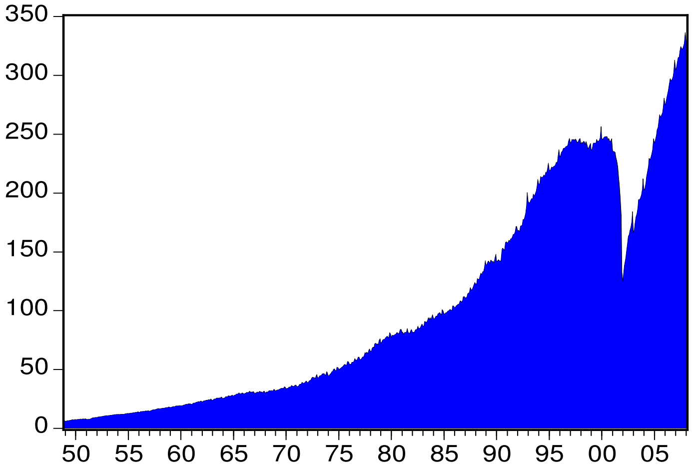
Ein Problem mit den Daten ist jedoch durch die außergewöhnliche Situation einige Zeit vor Einführung des Euros als Bargeld im Januar 2002 entstanden, als der Bargeldumlauf in Erwartung der Währungsumstellung erheblich zurückgefahren wurde. Das zeigt sich auch deutlich an den von EViews angebotenen graphischen Darstellungen der Zeitreihe



Die Zeitreihe heißt "Barmm" (= Bargeldumlauf DM monatlich). Der Einbruch Ende 2001 wird auch an den Zahlen, die dem obigen Bild zugrunde liegen, deutlich

Monat	2001-10	01-11	01-12	02-01	02-02	02-03	02-04
Mrd. DM	195,4	180,7	132,9	124,8	133,6	138,7	144,7
Mrd. €	99,90	92,41	67,97	83,83	68,32	70,90	74,01

Der Bargeldumlauf ist mithin von Oktober 2001 bis Jan. 2002 von 195,4 Mrd. DM[1] auf 124,8 Mrd. DM, also um 36% gesunken.



Neben der obigen "Line" Graphik bietet EViews auch das nebenstehende "Area"-Diagramm mit "bars", d.h. Blöcken der Zeitreihe barmm (Legende ausgeblendet) als eine graphische Darstellung an. Auch hier sieht man deutlich den durch die Währungsumstellung 2001 bis 2004 entstandenen Einbruch in der Zeitreihe.

Es empfiehlt sich - wegen des Strukturbruchs - die Zeitreihe durch Veränderung der Stichprobe[2] in zwei Intervalle zu zerlegen (Teil B dieses Papiers – unten Seite 4) und/oder durch frequency conversion von Monatsdaten zu Quartals- oder Jahresdaten überzugehen (Teil C, Seite 6).

2. Komponentenzerlegung, Trendberechnung

Auf die Demonstration von Verfahren der Saisonbereinigung und die Betrachtung von ARIMA-Modellen wird hier aus Platzgründen verzichtet. Diese Verfahren sind auch nicht Gegenstand der Vorlesung. Die einzige Methode aus diesem Bereich, die im Folgenden dargestellt wird ist der

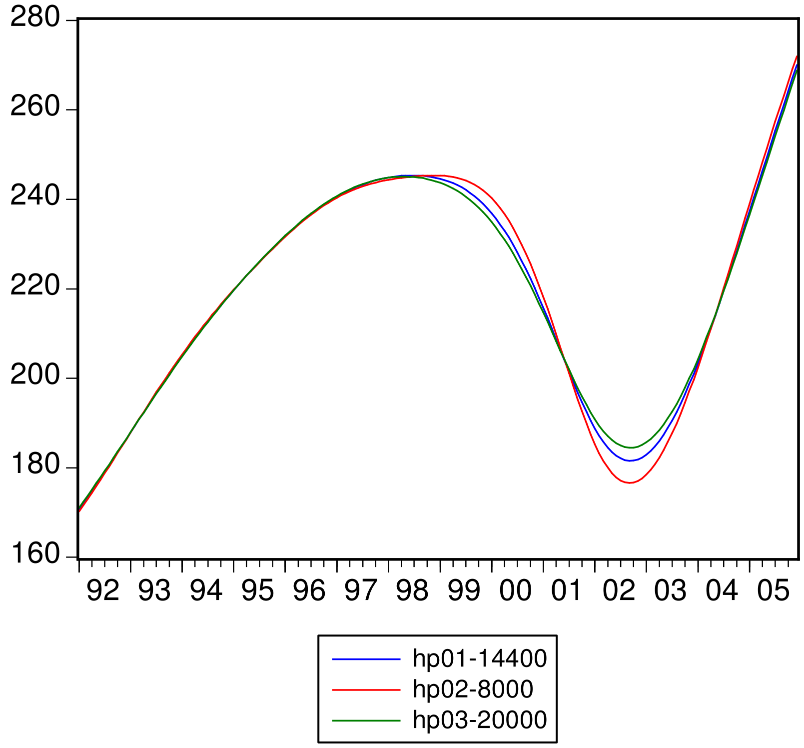
Hodrick Prescott (HP-) Filter mit verschiedenem Lamda

Damit wird eine sehr beliebte Methode der Trendberechnung angewendet. Ähnlich wie bei gleitenden Mittelwerten liegt auch dieser Betrachtung ein Filter zugrunde, der die glatte Komponente[3] ($y_{g,t}$) aus den Ursprungswerten (y_t) bestimmt, indem man den folgenden Ausdruck $\sum_{t=1}(y_t - y_{g,t})^2 + \lambda\sum_{t=2}[(y_{g,t+1} - y_{g,t}) - (y_{g,t} - y_{g,t-1})]^2$ bezüglich $y_{g,t}$ minimiert.

Der erste Summand ist der quadrierte Zyklus (Abweichung der Ursprungswerte von der glatten Komponente), der zweite $\lambda\sum(\Delta y_{g,t+1} - \Delta y_{g,t})^2 = \lambda\sum(\Delta^2 y_{g,t+1})^2$ bewirkt, ob $y_{g,t}$ mehr oder weniger glatt verläuft ("glatt" heißt, dass sich aufeinander folgende Differenzen Δ wenig unterscheiden).[4] Das wird mit dem einzustellenden Parameter λ (Voreinstellung 14.400 bei Monatsdaten[5] wie im vorliegenden Fall) "gesteuert". Wie die folgende Graphik zeigt, führt ein kleines λ (bei hp02 ist λ = 8.000)[6] zu einer guten Anpassung (die Schwankungen der Ursprungswerte werden mit dem Trend mitgezeichnet) und entsprechend geringen Glätte und ein großes λ (bei hp03 ist λ = 20.000) bewirkt, dass die glatte Komponente die Schwingungen der Daten "mitmacht" und die Reihe der $y_{g,t}$ somit weniger glatt verläuft. Die Unterschiede zwischen den HP-Filtern hp01 bis hp03 sind gering und eigentlich nur im kritischen Be-

[1] Der Spitzenwert war 1999/12 mit 256,4 Mrd. DM.
[2] Nicht zu verwechseln mit der Prozedur **resampling**.
[3] Darunter versteht man die Summe von Trend und Konjunkturkomponente.
[4] Bei Linearität (im höchsten Maße "glatt") beispielsweise ist Δ (und damit auch die zweiten Differenzen Δ^2) = 0.
[5] Die Voreinstellung von λ ist abhängig von der Frequenz der Zeitreihe. EViews verwendet λ = 100 bei Jahres, 1600 bei Quartals und 14400 bei Monatsdaten.
[6] Wie die Namen hp01,…,hp03 zeigen ist es auch bei solchen Berechnungen – wie immer bei EViews - wichtig, dass die errechneten Variablen benannt werden. Nur wenn sie einen Namen haben kann man mit ihnen weiter rechnen oder sie mit anderen Variablen graphisch darstellen.

reich des Einbruchs vor und nach der Währungsumstellung sichtbar zu machen. Das Bild enthält deshalb auch nur den Ausschnitt[7] von 1992 bis 2005.



Wie man sieht, tritt mit größerem λ eine stärkere Glättung ein (der Einbruch wird nicht voll nachgezeichnet).

Man kann neben der hier abgebildeten Zeitreihe des Filters (in drei Varianten), also der errechneten glatten Komponente) auch der Zyklus (y_t - $y_{g,t}$) graphisch darstellen, wenn man $y_{g,t}$ benennt (etwa mit hp01 usw.) und die Reihe barmm - hp01 berechnet und benannt hat.

3. Test auf Stationarität (Einheitswurzeltest) für den ganzen Zeitraum 1949 bis 2008

Einzelheiten zu der Bedeutung und Methode der Unit-Root-Tests vgl. meinen Download Nr. 5.

a) Level von Barmm (= y)

Die hier gewählten Optionen für den " Augmented Dickey-Fuller Test" (ADF test) sind: Level, (with) intercept (= c)[8], der "augmented" Teil bis zum lag 4. Berechnet wird also (wenn y_t die Variable "barmm" ist) $\Delta y_t = \mu + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \delta_3 \Delta y_{t-3} + \delta_4 \Delta y_{t-4} + u_t$.
Das Rechenergebnis wird von EViews wie folgt dargestellt:

Null Hypothesis: BARMM has a unit root; Exogenous: Constant

ADF Test Statistic	1.204364	1% Critical Value*	-3.4422
		5% Critical Value	-2.8660
		10% Critical Value	-2.5691

*MacKinnon critical values for rejection of hypothesis of a unit root.

Das Ergebnis bedeutet, dass H_0 **nicht** verworfen wird, oder anders gesagt: y hat eine unit root = einen stochastischen Trend = ist nicht-stationär, was bei Betrachtung der Graphik auch nicht überrascht.

Weitere Angaben von EViews: Augmented Dickey-Fuller Test Equation; Dependent Variable: D(BARMM); Method: Least Squares; Date: ** Time: ** **Sample(adjusted): 1949:05 2008:01** Included observations: 705 after adjusting endpoints: Die folgenden Angaben zur Schätzgleichung sind für die Testentscheidung nicht sehr relevant. Gezeigt wird jetzt die Gleichung, die geschätzten Koeffizienten und ihre Signifikanz (dieser Teil des "estimation output" wird in den folgenden Anwendungen nicht mehr gezeigt):

Variable		Coefficient	Std. Error	t-Statistic	Prob.
BARMM(-1)	(γ)	0.001594	0.001324	1.204364	0.2289
D(BARMM(-1))	(δ_1)	0.190381	0.038212	4.982177	0.0000
D(BARMM(-2))	(δ_2)	-0.049792	0.038815	-1.282810	0.2000

[7] Durch Umstellung des samples kann man in der Graphik oder auch bei bestimmten Berechnungen (z.B. Schätzung einer Regressionsfunktion) leicht einen Ausschnitt aus der gesamten Zeitreihe konstruieren.
[8] Deshalb im equation output die Angabe "Exogenous: Constant".

D(BARMM(-3))	(δ_3)	0.135715	0.038804	3.497432	0.0005
D(BARMM(-4))	(δ_4)	0.056334	0.038430	1.465871	0.1431
C	(μ)	0.141939	0.174261	0.814520	0.4156

Es folgen die üblichen Angaben zur Güte der Schätzung, die hier nur zum Teil wiedergegeben werden

R-squared	0.066386	Mean dependent var	0.456589
Adjusted R-squared	0.059708	S.D. dependent var	3.150251
S.E. of regression	3.054756	Akaike info criterion	5.079751
Sum squared resid	6522.743	Schwarz criterion	5.118544
Log likelihood	-1784.612	F-statistic	9.940652
Durbin-Watson stat	1.993329	Prob(F-statistic)	0.000000

Man könnte jetzt auch andere Optionen wählen, etwa den ganzen Test nicht für das level (y)[9] der Zeitreihe, sondern für die first (oder second) differences durchführen oder auch einen (deterministischen, linearen) Trend einbauen (vgl. unten Abschn. B). Im Unterschied zu dieser Option ist im Folgenden zunächst mit exponential smoothing ein Trend bestimmt worden (EViews fügt automatisch beim Namen ein "sm" an) und der Unit Root Test ist durchgeführt worden für die Differenzen (Abweichungen von y_t von diesem Trend), also für die konstruierte Zeitreihe Diff = Barmm – Barmmsm (nicht zu verwechseln mit D(BARMM) was die ersten Differenzen $\Delta y_t = y_t - y_{t-1}$ sind).

b) Abweichungen vom (nichtlinearen) Trend: Unit Root Test mit der Trendabweichung (= Variable Diff)

ADF Test Statistic	-10.39133	1% Critical Value*	-2.5686*
		5% Critical Value	-1.9399
		10% Critical Value	-1.6159

* MacKinnon (1996) one-sided p-values (for rejection of hypothesis of a unit root).

Wie man jetzt sieht ist H_0 (Nichtstationarität) zu verwerfen weil -10,39 < -2,57 (der Test ist stets einseitig); die Zeitreihe der Differenzen hat keine Einheitswurzel (keinen stochastischen Trend), sie hat ist somit stationär. Die folgenden Angaben, die das Programm auswirft, sind nützlich, um zu dokumentieren, was hier berechnet wurde:

Augmented Dickey-Fuller Test Equation…., Dependent Variable: D(DIFF); Sample(adjusted): 1949:05 2008:01; es folgt die Liste der Variablen[10] DIFF(-1), D(DIFF(-1)) bis D(DIFF(-4)); es wurde also die Option "level, none, number of lags 4"[11] gewählt.

B. Analyse von zwei Teilintervallen der gegebenen langen Zeitreihe

1. Erzeugung zweier zusätzlicher Seiten des workfiles mit speziellen samples

Man kann mit copy und paste die volle Zeitreihe auf zwei weitere Seiten "Intervall 1" und "Intervall 2" kopieren und kann dort nach Doppelklick auf die "Sample Zeile" im Workfile Directory (man erhält damit ein Menü in dem man das sample umstellen kann) neue Teilintervalle definieren.[12] Hier wurde definiert (auch als Namen für zusätzliche Seiten im workfile)

Intervall 1: 1948:12 2000:10 (geschrieben 1948m12 2000m01) und
Intervall 2: 2005:01 2008:01.

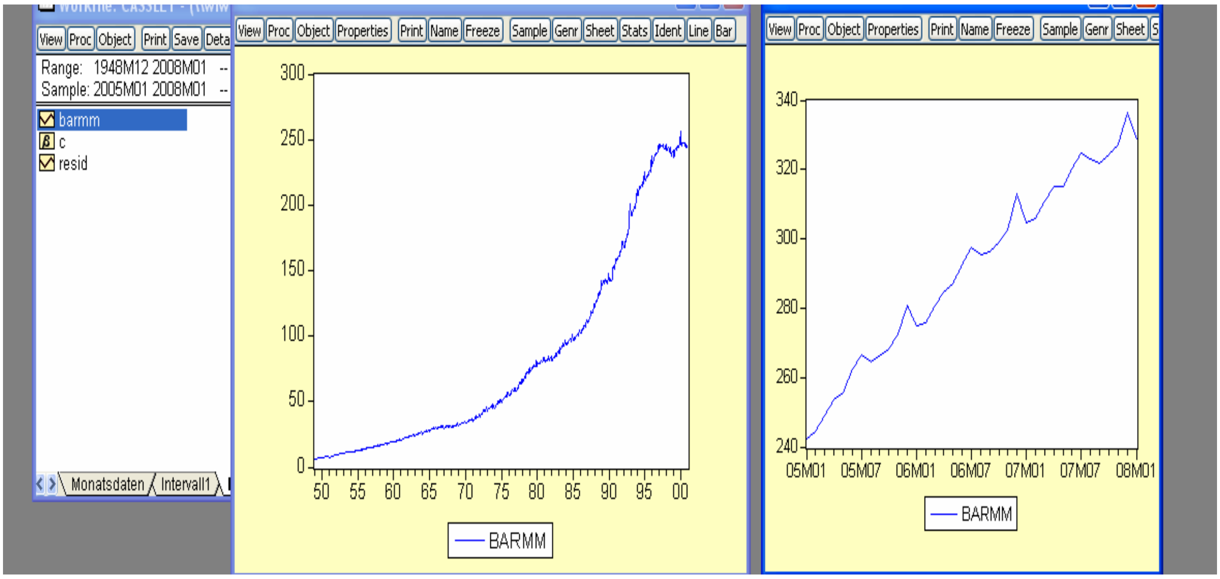
[9] Die oben angegebene Schätzgleichung $\Delta y_t = \mu + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \ldots + \delta_4 \Delta y_{t-4} + u_t$. sollte nicht den Eindruck entstehen lassen, hier sei der Test für die "first differences" durchgeführt worden.

[10] d.h. die rechte Seite der Gleichung jeweils mit den üblichen hier nicht wiedergegeben den Angaben Coefficient, Std. Error, t-Statistic und Prob.

[11] Die dep. var. (abhängige Variable) ist D(Diff), keine Konstante c und kein Trend und lags von 1 bis 4.

[12] Auf den entsprechenden beiden Seiten erscheint ein Objekt barmma (selbst gewählter Name) mit einem Doppelpfeilen $\boxed{\leftrightarrow}$. Das ist anders als die (blau markierte) Zickzacklinie von "barmm" keine Zeitreihe, sondern nur eine Festlegung des Stichprobenintervalls, weil dieses von der "Range" (also dem zeitlichen Bereich), der Daten im workfile abweicht.

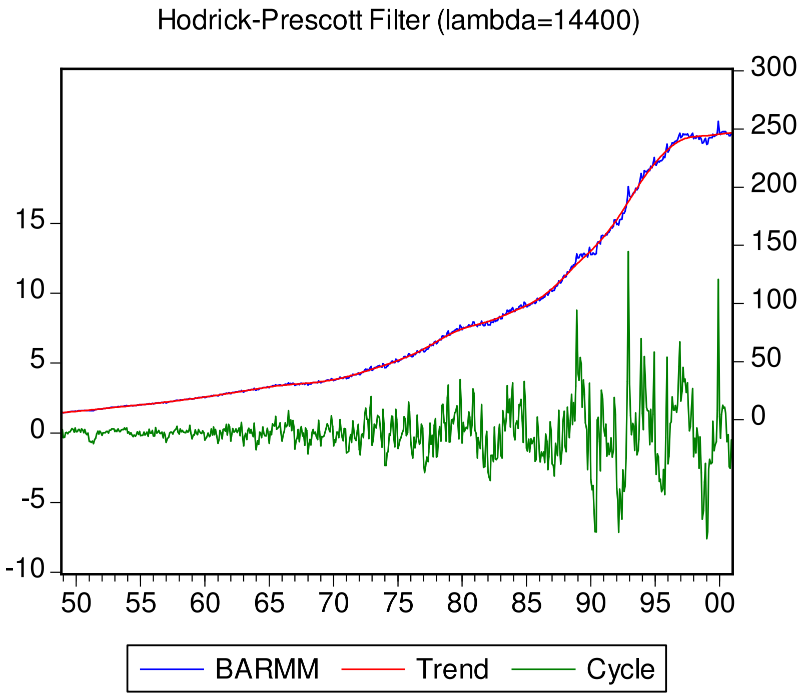
Die kompletten vier Jahre 2001 bis 2004 entfallen damit als außergewöhnliche Jahre wegen der Währungsumstellung.
Die Zeitreihe **barmm** behält jeweils ihren Namen, ist aber auf den beiden Seiten "Intervall1" und "Intervall2" jeweils eine kürzere Reihe (wie auch das nachfolgende "screen shot" Bild zeigt) als die gleichnamige Reihe auf der ursprünglichen Seite "Monatsdaten".



Mit den jeweiligen Zeitreihen barmm kann nun weiter gerechnet werden. Wenn die entsprechende Zeitreihe geöffnet ist kann man mit **view** verschiedene beschreibende Statistiken , das Korrelogramm (ACF und PACF), den Einheitswurzeltest (für level, 1st und 2nd differences) oder den BDS Test abrufen, entsprechend mit **proc** Gleichungen berechnen, oder Saisonbereinigung und Trendberechnungen durchführen. Im Folgenden sieht man den HP-Filter (eine Trendberechnung) für das erste (1948/12 bis 2000/12) und das zweite (2005/01 bis 2008/01) Teilintervall.

2. HP-Filter für die beiden Teilintervalle

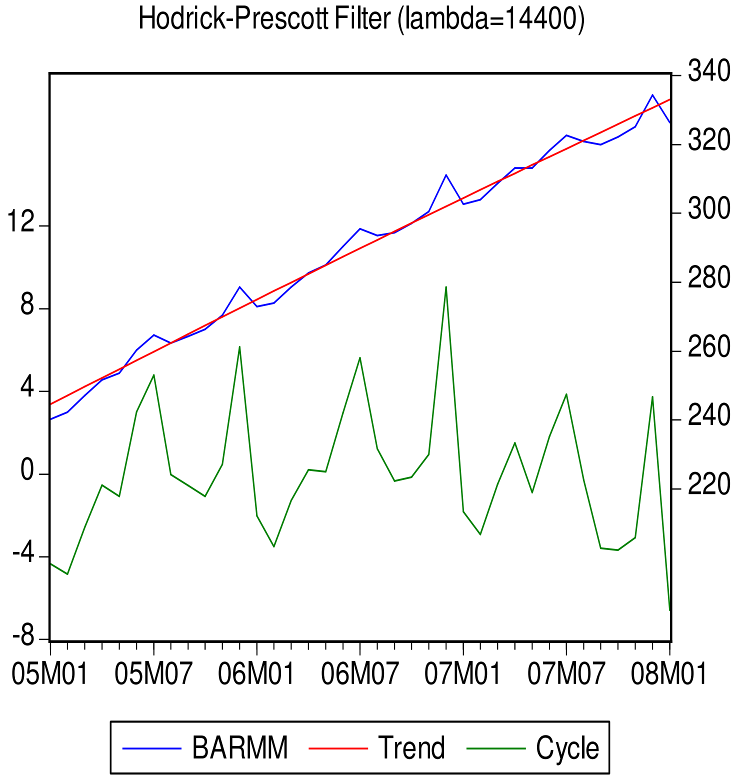
Für 1948/12 – 2000/12 gilt



Wie man an dem Bild sieht ist der Zyklus (grün) - anders als im Zeitraum 2005 bis 2008 (vgl. nächste Seite) - keineswegs regelmäßig (z.B. eine regelmäßige Saisonkomponente wiedergebend). Vielmehr hat sich ab Beginn der 90er Jahre, also schon lange vor der Umstellung der DM auf den € die Volatilität des Bargeldumlaufs in Deutschland stark erhöht.

Die Einführung der DM in den neuen Bundesländern hat offenbar keinen so beträchtlichen Bruch erzeugt wie die Einführung des € als Bargeld.

Für 2005/01 – 2008/01 erhält man



Wie man sieht ist der Charakter der beiden Zeitreihen sehr unterschiedlich. Für das zweite Teilintervall erhält man praktisch einen linearen Trend (rote Linie) und einen ziemlich regelmäßigen Zyklus, während für den ersten Zeitraum ein sehr viel unregelmäßiger Zyklus bestimmt wurde und auch der Trend (rote Linie) keineswegs linear verläuft. Die Berechnungen für den Trend bleiben im workfile auch unter den Namen hptrend01, hptrend02 usw. erhalten so dass mit ihnen weiter gerechnet werden kann.

3. Unit-Root Test für ein Teilintervall

Auch für das kürzere aktuellere Intervall (05/01 bis 08/01) kann man für die Zeitreihe barmm Einheitswurzeltest durchführen. Die *auf der nächsten Seite* wiedergegebenen Tests betreffen jedoch nicht dieses Teilintervall, sondern das sample 92/01 bis 05/12 wie im Abschn. A2 (Bild auf S. 3 oben). Man kann an der Tabelle gut erkennen, welche Optionen für die Spezifizierung gewählt worden sind und wie die Testentscheidung (über H_0) zu fällen ist.

C. Jahresdaten

Die Umformung der als Daten vorgegebenen Monatsdaten zu Jahresdaten ist nicht ganz einfach und kann auf eine im Folgenden beschriebene Art erfolgen

Hinweise zur frequency conversion: In einem workfile (wf) kann nur mit Daten ***einer*** Frequenz gerechnet werden (z.B. nur mit Monatsdaten), will man die Daten in eine andere Frequenz transformieren (etwa von Monatsdaten zu Quartalsdaten) und auch mit diesen Daten weiterrechnen, ***muss ein neuer wf gebildet werden***. Der folgende screen shot (*auf der nächsten Seite*) zeigt, wie man dabei vorgehen kann.[13]

Man sollte beide workfiles öffnen und zunächst im Ausgangs-wf (hier "Cassel 1") eine group bilden, bzw. die bereits gebildete **group** anklicken (das Folgende funktioniert nur, wenn eine Gruppe geöffnet ist) dann **View/Dated Data Table** und man erhält dann die Tabelle mit den Monats- und Jahreswerten, die im screen shot – *nächste Seite unterer Teil* - zu sehen ist).

[13] Der ursprüngliche workfile hieß Cassel 1, der neue Cassel 2 (die Namensgebung kommt daher, dass von mir die hier im Vordergrund stehende Variable "Bargeldumlauf" vor allem mit dem Ziel einer empirischen Arbeit in Zusammenarbeit mit Herrn Prof. Dr. Cassel untersucht wurde).

Zu Abschn. B Teil 3

Zwei Unit-Root Tests für das Teilintervall 92 – 05
(sample: 1992M01 2005M12, Included observations: 168)

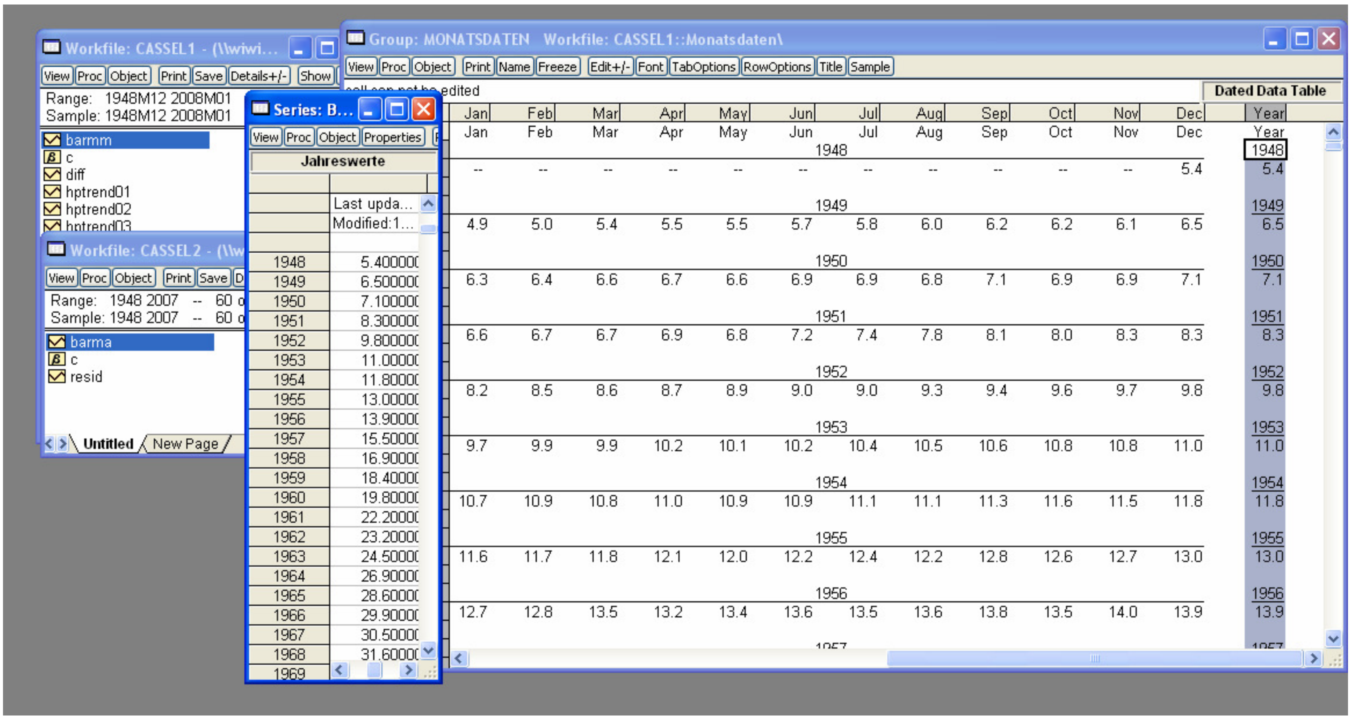
gewählte Optionen (und wie es im output angegeben wird)	1st differences, Kon-stan-te kein Trend, kein Lag Exogenous: Constant Lag Length: 0	2nd differences, Konstan-te und Trend, Lag 3* Exogenous: Constant, Lin-ear Trend, Lag Length: 3
Dependent Variable	D(BARMM,2)	D(BARMM,3)
Null Hypothesis	D(BARMM) has a unit root	D(BARMM,2) has a unit root
ADF[14] test statistic (prob value)	-9.567657 (0.0000)	-10.54399 (0.0000)
Test critical values: 1% level (5% level)**	-3.469451 (-2.878618)	-4.013608 (-3.436795)
Entscheidung	stationär (H_0 ablehnen)	stationär (H_0 ablehnen)
Liste der Regressoren (und deren prob values)	D(BARMM(-1)), C (0.0000, 0.2680)	D(BARMM(-1),2), D(BARMM(-1),3), D(BARMM(-2),3), D(BARMM(-3),3), C @TREND(1992M01)*** (nur C und Trend sind nicht signifikant)
F statistic (prob value)	91.54005 (0.000000)	94.24882
Adjusted R-squared; Durbin-Watson stat	0.351557; 2.001687	0.736279; 1.993342

* diese Betrachtung ist an sich überflüssig (und nur zu Demonstrationszwecken durchgeführt worden), denn wenn die ersten Differenzen stationär sind, dann sind es auch die zweiten. Die komplizierte Schätzgleichung bei 2nd differences lautet $\Delta^3 y_t = \gamma\Delta^2 y_{t-1} + \delta_1\Delta^3 y_{t-1} + \delta_2\Delta^3 y_{t-2} + \delta_3\Delta^3 y_{t-3} + c + \beta t + u_t$. Die Koeffizienten c und β sind Konstante und Trend. Die zweiten Differenzen sind Differenzen der ersten Differenzen also $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = y_t - y_{t-1} - (y_{t-1} - y_{t-2})$ mit $\Delta = \Delta^1$. Die dritten sind definiert als $\Delta^3 y_t = \Delta^2 y_t - \Delta^2 y_{t-1}$ usw.

** EViews gibt auch den kritische Wert beim 10% level an (er ist für unsere Zwecke hier jedoch uninteressant und wird deshalb auch nicht wiedergegeben)

*** das ist die Variable t beginnend mit t = 1 für Jan. 1992 und der hierfür angegebenen Koeffizient ist β.

Screen shot der "Dated Data Table"

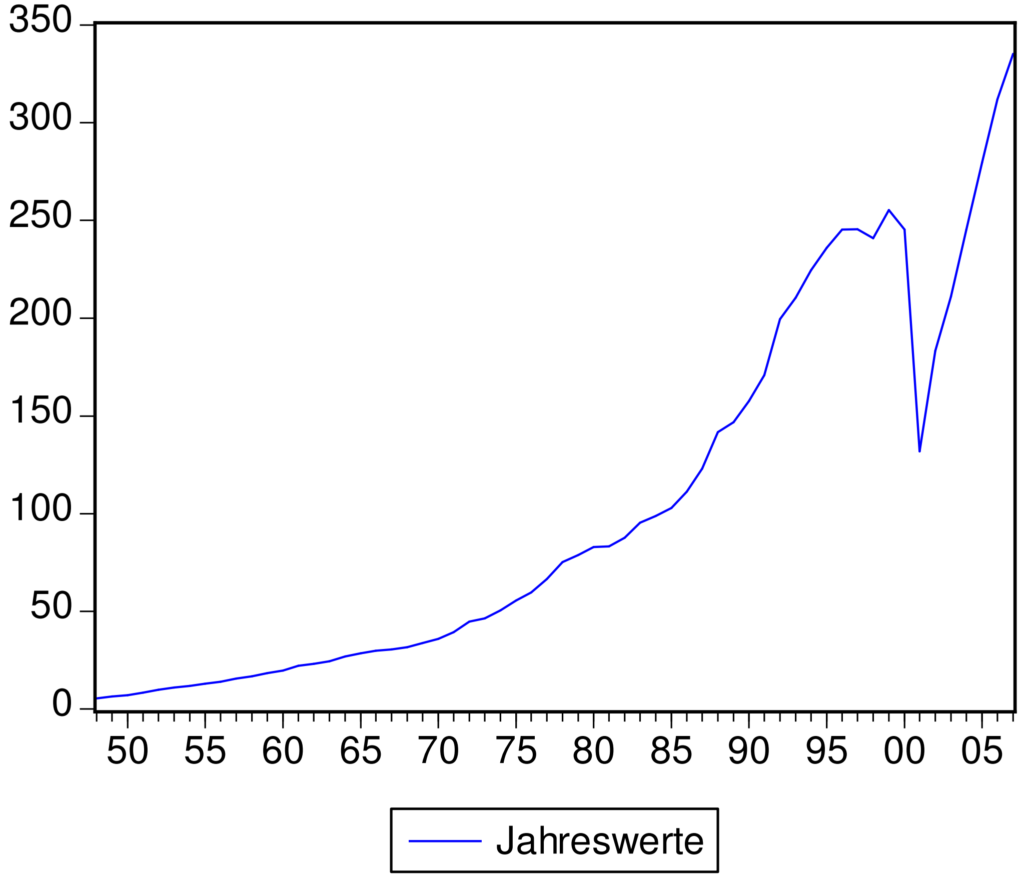


[14] Augmented Dickey-Fuller

Bei dieser Tabelle erhält man oben buttons für neue Befehle, insbes. ***TabOptions*** (mit diesen Table-Options kann man die spezielle frequency conversion Methode wählen [hier ist der Jahreswert stets als der Wert des 12.ten Monats gewählt worden; man könnte z.B. auch als Jahreswert einen Durchschnitt der Monatswerte definieren])[15] und ***RowOptions*** (betrifft das Format der Zahlendarstellung).

Die von EViews zu Jahresdaten transformierten Daten (oben im Bild grau markiert) kann man mit copy und paste in den neuen wf ziehen und die Reihe dann anders benennen (hier barma wegen annual; diese Reihe existiert somit neben der ursprünglichen Reihe barmm [m wegen monthly]). Die schmale Tabelle (spread sheet Ansicht) "Jahreswerte" enthält die Daten der im neuen workfile aufgenommenen Variable "barma" (der Zahlenvergleich mit der markierten Spalte in der Dated Data Table kann zur Kontrolle nützlich sein).

Es ist bemerkenswert, dass auch die Zeitreihe der Jahresdaten, nicht nur die der Monatsdaten stark geprägt ist von dem Bruch (oder besser "Einbruch") durch die Einführung des €-Bargelds und Abschaffung der DM.



Durch Änderung der Legende ist in diesem Bild die Bezeichnung "Jahreswerte" statt "barma" eingefügt worden.

Die Umstellung der Monatsdaten zu Jahresdaten hat den Vorteil, dass sich leichter weitere Reihen finden lassen, mit denen der Bargeldumlauf korreliert werden kann, weil viele volkswirtschaftliche Daten nur mit geringerer Frequenz (Jahresdaten statt Monatsdaten) vorliegen. Dies gilt insbesondere für Daten der Volkswirtschaftlichen Gesamtrechnung (VGR).

[15] Man kann im dated data table einstellen: first row = native frequency (hier Monatwerte, also die Frequenz wie im ursprünglichen wf [hier: Cassel 1]) und second = row annual (das entspricht der Darstellung oben auf S. 7) oder auch umgekehrt, also in der first row die berechneten Jahreswerte und in der zweiten die als Daten vorgegebenen Monatswerte darstellen lassen.