

## Katrin Baumert, Ming Liu

# Zeitreihenanalyse



### Einführung

Die Zeitreihenanalyse ist die mathematisch-statistische Untersuchung von Beobachtungswerten in zeitlicher Reihenfolge. Zeitreihen werden bereits seit Jahrtausenden beobachtet und registriert; ihre statistische Theorie gewann jedoch erst im 20. Jahrhundert an Bedeutung. Mit zunehmender Verbreitung der Marktwirtschaft und dem Aufkommen neuer Möglichkeiten zur Datenverwaltung steigert sich das Verlangen nach aussagekräftigen Analysen der beobachteten Werte.

Gegenwärtig werden die Zeitreihen der wichtigsten Konjunkturindikatoren der Bundesrepublik Deutschland nach zwei Verfahren – Census X-12-ARIMA des U.S. Bureau of the Census und dem Berliner Verfahren BV4.1 des Statistischen Bundesamtes – analysiert. Bisher wurde im Amt für Statistik Berlin-Brandenburg das Berliner Verfahren BV4.1 ausschließlich bei der Untersuchung der wichtigsten konjunkturellen Kennziffern verwendet. Aufgrund der international sehr verbreiteten Anwendung des ARIMA-X-12-Verfahrens erschien ein Test und der Vergleich mit dem BV4.1 angebracht. Die Vor- und Nachteile beider Methoden der Zeitreihenanalyse wurden unter den Bedingungen der Berlin-Brandenburger Konjunkturdaten ermittelt. Beispielhaft wird die Untersuchung anhand der brandenburgischen Arbeitslosenzahl erläutert.

Anwendungsmöglichkeiten der Zeitreihenanalyse finden sich in den unterschiedlichsten Bereichen wieder. So nutzt die Ökonomie die Ergebnisse zur Beurteilung wichtiger Konjunkturindikatoren (zum Beispiel: Zahl der Arbeitslosen, Anteil der Exporte). In naturwissenschaftlichen Bereichen, wie der Physik, werden Zeitreihen genutzt, um Aufschluss beispielsweise über Luft-Temperatur-Verhältnisse zu geben. Auch in der Demografie oder der Marktforschung wird die Zeitreihenanalyse zur Gewinnung von Erkenntnissen eingesetzt.

Das Feststellen von charakteristischen Strukturen ist eines der Ziele der Zeitreihenanalyse. Auftretende Periodizität oder zyklische Schwankungen können Auskunft über den weiteren Verlauf der Reihe geben. Weiterhin liefert sie Informationen über die Entwicklungslinie in der Vergangenheit und den Zustand am Reihende. Die Ergebnisse der Analyse bieten ein Hilfsmittel zur aktuellen Konjunktüreinschätzung und können – mit Einschränkungen – zur Vorhersage der Entwicklung genutzt werden.

### Das Zeitreihenmodell

Beobachtungswerte sind kontinuierlich aufeinanderfolgende Daten oder eine Menge von Werten, die zu diskreten Zeitpunkten beobachtet werden. Eine Zeitreihe – deren Werte von Irregularitäten überlagert sind – ist die mögliche Realisation eines stochastischen Prozesses. In der Theorie der Zeitrei-

hen wird davon ausgegangen, dass sich historische Beobachtungswerte aus dem Zusammenwirken verschiedener Komponenten ergeben, die in Abhängigkeit zur Zeit ( $t$ ) stehen. Zur Erleichterung der Analysen wird die Zeitreihe daher in diese Komponenten zerlegt. Beim klassischen Komponentenmodell wird zwischen folgenden imaginären Komponenten unterschieden:

- Trendkomponente:  $T_t$
- Konjunkturkomponente:  $K_t$
- Saisonkomponente:  $S_t$
- Restkomponente:  $R_t$

Die Trendkomponente umfasst langfristig wirkende Ursachen, wie die Veränderung des Lebensstandards aufgrund von technischem Fortschritt. Entsprechen die Beobachtungswerte einem Analysezeitraum, der nicht mehrere Jahrzehnte umfasst, verläuft die Trendkomponente in der Regel linear. Zyklische Schwankungen werden in der Konjunkturkomponente erfasst. Sie verläuft wellenförmig und zeigt beispielsweise den Wandel von Konsumgewohnheiten durch politische Maßnahmen. Aufgrund definitorischer Abgrenzungsprobleme werden Trend- und Konjunkturkomponente zusammengefasst. Die Trend-Konjunktur-Komponente oder auch glatte Komponente gibt die mittel- bis langfristige Grundtendenz einer Reihe wieder. Sie verläuft glatt und ohne unterjährige Schwankungen. Die glatte Komponente spielt eine wichtige Rolle bei Analysen der konjunkturellen Entwicklungen zum Beispiel zu den Arbeitslosen, im Tourismus oder im Baugewerbe. Der jahreszeitliche Wechsel wird durch die Saisonkomponente abgebildet. Sie zeigt eine jährliche Periodizität mit ungefähr gleichbleibendem wellenförmigem Muster auf, welches den Temperaturverhältnissen der Luft entspricht. Die Restkomponente oder auch irreguläre Komponente ist bedingt durch irreguläre Schwankungen. Diese umfassen alle anderen Schwankungen, die die Zeitreihe kurzfristig beeinflussen. Es wird angenommen, dass die Werte der Restkomponente vergleichsweise klein sind und um Null schwanken. Die Komponente unterliegt keiner fassbaren kausalen Gesetzmäßigkeit. In ihr äußern sich zum Beispiel Streiks, Methodenwechsel und Anderes. Bei zeitraumbezogenen Daten ist das Auftreten von kalenderbedingten Schwankungen möglich. Je nach Art der Zeitreihe beeinflussen verschiedene Wochen- und Feiertagskonstellationen die Beobachtungswerte. Die Kalenderkomponente kann zur verbesserten Trennung der Einflüsse von der Restkomponente abgegrenzt werden.

Die Komponenten einer Zeitreihe können unterschiedlich miteinander verknüpft sein.

- Additiv:  $O_t = T_t + K_t + S_t + R_t$
- Multiplikativ:  $O_t = T_t * K_t * S_t * R_t$

Die häufigste Modellannahme, die auch vom Berliner Verfahren (BV4.1) eingesetzt wird, ist das additive Komponentenmodell. Hierbei sind die einzelnen Komponenten der Originalreihe ( $O_t$ ) weitgehend stochastisch unabhängig voneinander, das heißt es werden keine systematischen Zusammenhänge – wie beispielsweise Proportionalität – formuliert. Bei der additiven Verknüpfung werden diese Zusammenhänge nicht ausgeschlossen. Das Analyseverfahren Census X-12-ARIMA bietet darüber hinaus die Möglichkeit, eine multiplikative Verknüpfung zu wählen. Der Ausschlag der Saisonkomponente und die Streuung der Werte nehmen bei wachsendem Trend zu, eine additive Verknüpfung genügt dann der Beschreibung nicht mehr.

### Methoden der Zeitreihenanalyse

Zur Begutachtung der aktuellen wirtschaftlichen Entwicklung werden wichtige Konjunkturindikatoren mithilfe der Zeitreihenanalyse bewertet. Es gibt vielfältige Analysemethoden, die sich sowohl in der Verwendbarkeit der Ergebnisse als auch im Verfahrensaufbau unterscheiden. Im Folgenden wird für die Darstellung der einzelnen Verfahren die monatliche Zahl der Arbeitslosen in Brandenburg vom Januar 1994 bis zum Dezember 2008 verwendet.

### Gleitende Durchschnitte

Eine einfache Methode der Zeitreihenanalyse ist die Methode der gleitenden Durchschnitte. Es handelt sich hierbei um ein Glättungsverfahren zur näherungsweise Ermittlung der Trend-Konjunktur-Komponente. Unter der Voraussetzung, dass die Beobachtungswerte äquidistant sind und periodischen Schwankungen unterliegen, wird die Originalreihe geglättet. Irreguläre Schwankungen werden ausgeschaltet. Bei Monatsreihen wird üblicherweise die Gradzahl entsprechend der Monatsanzahl eines Jahres gewählt. Nichtbeachten der Periodenlänge führt gegebenenfalls bei zu hoher Stützweite zur Ausschaltung der Periodizität, das heißt zu einer zu starken Glättung oder bei zu geringer Stützweite zu einer Anpassung der Trend-Konjunktur-Komponente an die Originalwerte. Der zentrierte gleitende Durchschnitt liefert keine Berechnungen für Werte

am Anfang und Ende der Zeitreihe. Entsprechend der Gradzahl können an diesen Stellen keine Aussagen über die Entwicklung des Trends getroffen werden, der am aktuellen Ende besonders interessant bei der Beurteilung der Konjunkturindikatoren wäre. Verfahren zur gesonderten Berechnung der Randwerte sind eine Möglichkeit, um weitere Erkenntnisse zu erlangen. Eine Alternative sind Methoden der Zeitreihenanalyse mit integrierter Prozedur zur Bestimmung der fehlenden Werte.

### Das Berliner Verfahren

Das Berliner Verfahren (BV) ist ein Komponentenzerglegungsverfahren zur Analyse ökonomischer Monats- und Quartalsdaten. Die Grundlagen wurden Ende der 60er Jahre gemeinsam von der TU Berlin und dem DIW entwickelt. Eine eigenständige Weiterentwicklung BV4.1 wird vom Statistischen Bundesamt zur Analyse der Konjunkturindikatoren eingesetzt. Die Erweiterung enthält vor allem methodische Verbesserungen zur Behandlung von Ausreißern und Kalendereinflüssen. Die Software zum Berliner Verfahren steht als kostenloser Download auf der Homepage des Statistischen Bundesamtes zur Verfügung.

Die Ermittlung der Trend-Konjunktur-Komponente erfolgt unter der Annahme, dass es sich um einen nichtdeterministischen Prozess handelt, der mit Hilfe von Polynomen geringer Ordnung hinreichend genau approximiert werden kann. Während es sich bei der Saisonkomponente um einen schwach stationären Prozess handelt, der durch eine endliche Fourierreihe abgebildet werden kann. BV4.1 besteht aus drei wesentlichen Verfahrensteilen. Der erste Teil befasst sich mit der Identifikation von potenziellen Ausreißern. Darunter werden Beobachtungswerte verstanden, die einzeln und unzusammenhängend auftreten und stark von der restlichen Struktur der Zeitreihe abweichen. Im Allgemeinen sind die Gründe für das Auftreten von diesen Extremwerten unbekannt. Genauere Betrachtungen und Recherchen sind hier nötig. Ein Extremwert kann ein Hinweis auf eine Sprungstelle sein, die beispielsweise mit einem Methodenwechsel einhergeht. Eine Extremwertbereinigung war in den ersten Grundlagen des Verfahrens nicht vorgesehen. Nun erfolgt die Identifikation von potenziellen Ausreißern jedoch mittels der gleitenden Berechnung bedingter Erwartungswerte – für den jeweiligen Folgemonat – und dem Vergleich mit den tatsächlich beobachteten Werten. Überschreitet

Abb. 1 **Monatliche Anzahl der Arbeitslosen und gleitender Durchschnitt 12. Ordnung im Land Brandenburg von 1994 bis 2008**

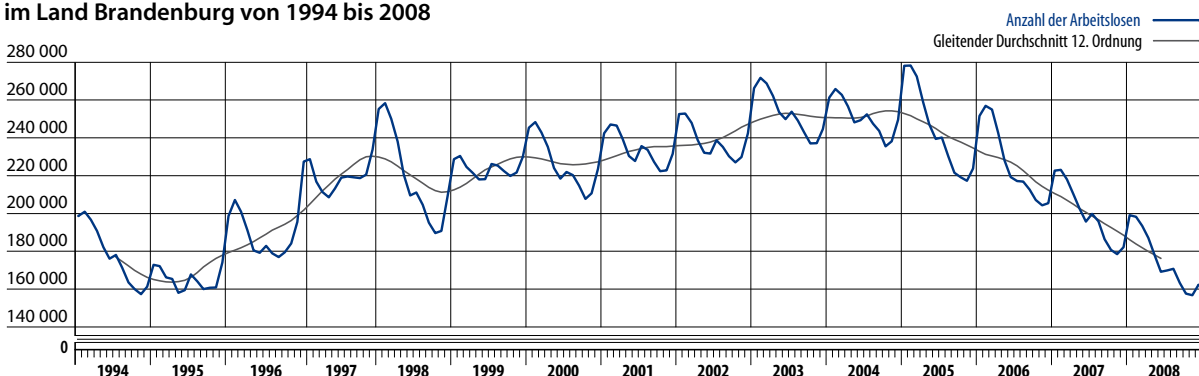


Abb. 2 **Monatliche Anzahl der Arbeitslosen und BV4.1 Trend-Konjunktur-Komponente im Land Brandenburg von 1994 bis 2008**

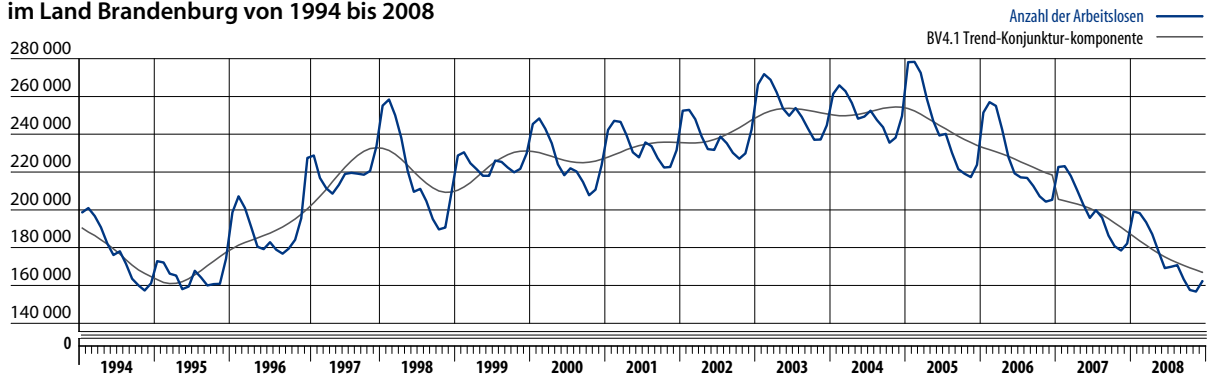
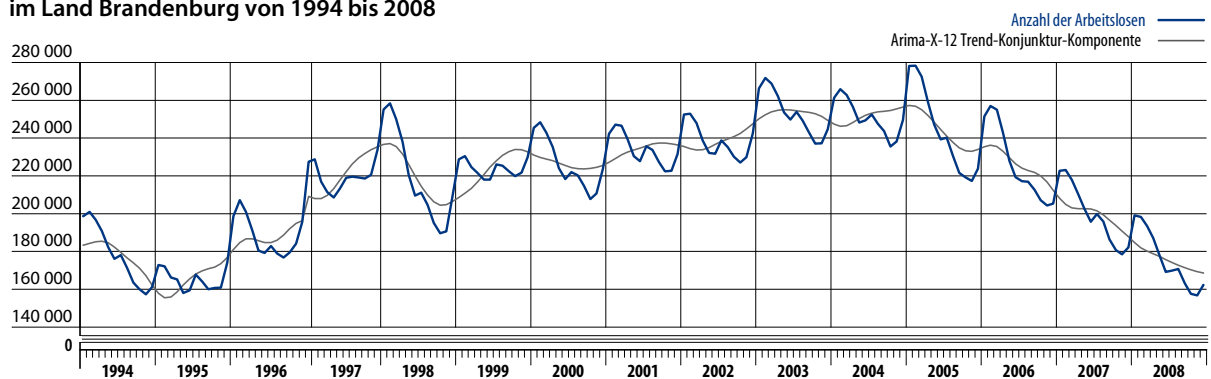


Abb. 3 **Monatliche Anzahl der Arbeitslosen und ARIMA-X-12 Trend-Konjunktur-Komponente im Land Brandenburg von 1994 bis 2008**



der Konfidenzfaktor eine anwenderspezifische Grenze, handelt es sich um einen Ausreißer. Im folgenden Verfahrensteil erfolgt die integrierte Schätzung von Ausreißern, Kalendereinflüssen und gegebenenfalls anwenderbestimmten Einflussgrößen, wie beispielsweise bekannte Sprungstellen. Hierbei wird von einem erweiterten additiven Modell ausgegangen, welches neben Trend-Konjunktur-Komponente, Saison- und Restkomponente auch noch eine Kalender-, eine Ausreißer- und eine anwenderbestimmte Komponente enthält. Die Parameter, die für die eingesetzte Filterprozedur benötigt werden, werden mittels der Methode der kleinsten Quadrate geschätzt. Im dritten Teil des Verfahrens findet die Schätzung der Trend-Konjunktur-Komponente und der Saisonkomponente anhand der zuvor bereinigten Reihe statt.

#### Census X-12-ARIMA

Das Saisonbereinigungsverfahren Census X-12-ARIMA bietet eine weitere Möglichkeit zur Zeitreihenanalyse. Die Grundlagen wurden vom U.S. Bureau of the Census entwickelt. Es handelt es sich um ein international bekanntes und weitverbreitetes Verfahren, welches unter anderem eingesetzt wird von: der Europäischen Zentralbank, der Deutschen Bundesbank und dem Statistischen Bundesamt. Das Statistische Bundesamt veröffentlicht seine Analyseergebnisse teilweise mit beiden Verfahren. Ähnlich wie BV4.1 gliedert sich auch dieses Saisonbereinigungsverfahren in drei Teile. Im ersten erfolgt die Extremwert- und Kalenderbereinigung durch die Modellierung der Originalreihe mit Hilfe von Regressionsmodellen mit ARIMA-Resten (AutoRegressiv-Integrated-Moving-Average). Im Anschluss wird die Saisonbereinigung

auf Basis einer iterativen Anwendung verschiedener gleitender Durchschnitte durchgeführt. Der letzte Teil dieses Verfahrens enthält diagnostische Instrumente zur Überprüfung der Güte der Saison- und Kalenderbereinigung. X-12-ARIMA ermöglicht die Wahl zwischen einer additiven oder multiplikativen Verknüpfung der Komponenten, um auf Besonderheiten der Zeitreihe einzugehen. Außerdem können die Randwerte des Analysezeitraums mithilfe der Berechnung von Prognosewerten stabilisiert werden.

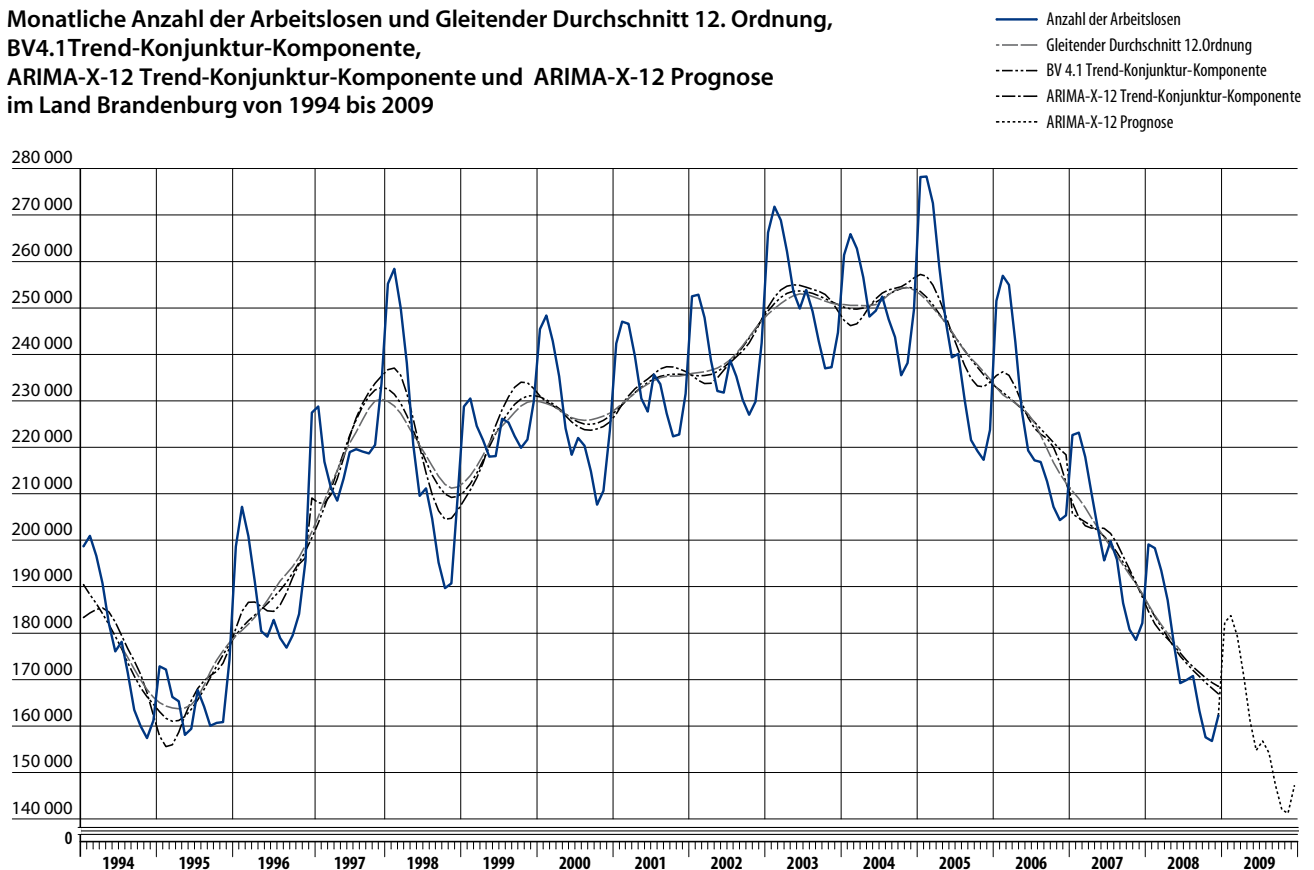
#### Vergleich Census X-12-ARIMA und BV4.1

Ein Vergleich der monatlichen Berliner und Brandenburger Wirtschaftsdaten vom Januar 2000 bis zum Dezember 2008 wurde mit den folgenden Werten durchgeführt:

- Anzahl der Gästeübernachtungen,
- Einfuhr und Ausfuhr in Mill. EUR,
- Preisindex (2005  $\pm$  100) sowie
- Inflationsrate.

Dabei zeigte sich, dass es keine großen Unterschiede zwischen beiden Verfahren bei der Trend-Konjunktur-Komponente gibt. Der lineare Korrelationskoeffizient zwischen den beiden Komponenten lag bei den auf zwei Nachkommastellen gerundeten Werten zwischen 1,00 und 0,98. Selbst wenn man die als problematisch bekannten drei Werte am Ende und Beginn der Trend-Konjunktur-Komponente betrachtet, ergeben sich – mit Ausnahme der brandenburgischen Inflationsrate (bei einem Wert von 0,89) – keine anderen Verhältnisse. Aus mathematischer Sicht sind demzufolge beide Verfahren fast gleichwertig. Bei der praktischen Anwendung ergeben sich jedoch große Unterschiede bei der sicheren

Abb. 4 **Monatliche Anzahl der Arbeitslosen und Gleitender Durchschnitt 12. Ordnung, BV4.1 Trend-Konjunktur-Komponente, ARIMA-X-12 Trend-Konjunktur-Komponente und ARIMA-X-12 Prognose im Land Brandenburg von 1994 bis 2009**



Handhabung, den technischen und personellen Voraussetzungen, den mathematischen Kenntnissen und den Fähigkeiten des Anwenders.

#### Fazit

Der durchgeführte Vergleich der Verfahren zur Zeitreihenanalyse liefert Vor- und Nachteile der verschiedenen Methoden. Zentrierte gleitende Durchschnitte bilden das Fundament der Trendbestimmung und können eingesetzt werden, um die Grundtendenz einer Zeitreihe zu ermitteln. Die fehlenden Werte am Rand des Datenbereichs verringern jedoch die Einsatzmöglichkeiten dieses Verfahrens zur aktuellen Konjunktüreinschätzung.

BV4.1 ermöglicht eine einfache und schnelle Analyse für eine Vielzahl von Zeitreihen. So können in der amtlichen Statistik verschiedene Reihen von Beobachtungswerten auch parallel untersucht werden. Die anwenderfreundliche Benutzeroberfläche gibt verschiedene Wahlmöglichkeiten zur Nutzung von BV4.1, zum Beispiel zum Eingabeformat oder um Angaben zur Kalendereinigung festzulegen. Anwenderbestimmte Regressoren sind hier jedoch nicht notwendig, wodurch weniger subjektive Einflüsse bei der Datenanalyse entstehen. Mithilfe des Verfahrens X-12-ARIMA können Störfaktoren ausgeschaltet und aktuelle Konjunkturbeobachtungen und auch -prognosen erstellt werden. Wobei X-12-ARIMA bei der Erstellung der Analyseergebnisse wesentlich mehr anwenderspezifischen Input als BV4.1 erfordert. Einstellungen zum verwendeten ARIMA-Modell und den Filtern für die Saisonbereinigung müssen hier für jede Zeitreihe neu bestimmt und festgelegt

werden. Das Verfahren Census X-12-ARIMA benötigt für qualitativ hochwertige Ergebnisse eine gut-durchdachte Setzung einer Vielzahl von Parametern, die vom Anwender durchzuführen ist. Dadurch sind die Analyseergebnisse häufig subjektiv und damit schwer vergleichbar, während bei BV4.1 die methodische Suche nach Optimierung bereits in die Filterkonstruktion integriert ist und nicht für jede Zeitreihe wiederholt werden muss. Ein weiterer Vorteil von BV4.1 – im Vergleich zu anderen Verfahren – ist die effiziente Saisonbereinigung, die auch bei rasch veränderlichen Einflüssen eingesetzt werden kann. Durch den klaren Verfahrensaufbau – keine gegenseitige Beeinflussung der Komponenten durch eine eindeutige Trennung – und der Vermeidung freier Parameter liefert BV4.1 eindeutige und übereinstimmende Ergebnisse, die dennoch flexibel und zuverlässig sind. Weiterhin sind bei BV4.1 die Revisionen vorläufiger Analyseergebnisse wesentlich geringer als bei anderen angewandten Verfahren, da für die Analyse sukzessive Berechnungen durchgeführt werden. Die hier vorgestellten Verfahren BV4.1 und X-12-ARIMA beruhen auf unterschiedlichen mathematisch statistischen Verfahren und liefern daher unterschiedliche Ergebnisse. Besonders im Einsatz bei der Zeitreihenanalyse der Berliner und Brandenburger Wirtschaftsdaten und deren Darstellung im Amt für Statistik Berlin-Brandenburg ist das Berliner Verfahren (BV4.1) vorzuziehen.

Frau Katrin Baumert und Herr Ming Liu absolvierten im Rahmen ihrer Ausbildung zum Diplom-Wirtschaftsmathematiker im Jahr 2009 ein Praktikum im Referat Redaktion der Veröffentlichung, Wahlstatistik.