

Dr. Martina Rengers

Trendschätzung für die monatlichen Arbeitsmarktzeitreihen des Mikrozensus

Hintergründe und Umsetzung mit dem Berliner Verfahren BV4.1

Mit der Veröffentlichung zum Berichtsmonat März 2011 hat das Statistische Bundesamt Neuerungen bei der monatlichen Arbeitsmarktberichterstattung aus den Ergebnissen des Mikrozensus beziehungsweise der in den Mikrozensus integrierten Arbeitskräfteerhebung vorgenommen. Diese betreffen zum einen die Berechnung der Erwerbslosenquote, zum anderen die erstmalige Herausgabe von Trendschätzergebnissen, bei gleichzeitigem vorläufigem Verzicht auf eine eigenständige Saisonbereinigung der Arbeitsmarktzeitreihen. Eine Saisonbereinigung der Zeitreihen auf Grundlage der Arbeitskräfteerhebung wird allerdings erst nach weiteren methodischen Verbesserungen möglich sein.

Dieser Beitrag erläutert die Hintergründe für die neue Vorgehensweise und legt den Schwerpunkt der Betrachtungen auf die Beschreibung und transparente Darstellung des eingesetzten Trendschätzverfahrens BV4.1. Diese Transparenz ist die Grundvoraussetzung für eine sinnvolle Interpretation der veröffentlichten Trendschätzergebnisse. Dazu werden in einem ersten Kapitel die Neuerungen in der monatlichen Arbeitsmarktberichterstattung dargestellt und die Besonderheiten bei der zeitreihenanalytischen Behandlung der Mikrozensusdaten im Kontext genereller Empfehlungen des Statistischen Amtes der Europäischen Gemeinschaften (Eurostat) und der Europäischen Zentralbank sowie ergänzender spezieller Empfehlungen der Internationalen Arbeitsorganisation ILO erläutert.

Im nachfolgenden zweiten Kapitel wird sodann das Berliner Verfahren BV4.1 detailliert vorgestellt und mit einem konkreten Anwendungsbeispiel veranschaulicht.

Das abschließende dritte Kapitel fasst die wesentlichen Erkenntnisse noch einmal kurz zusammen.

1 Die monatliche Arbeitsmarktstatistik auf Basis des Mikrozensus

1.1 Neuerungen bei der monatlichen Berichterstattung ab Berichtsmonat März 2011

Im Mikrozensus beziehungsweise bei der in den Mikrozensus integrierten Arbeitskräfteerhebung wird die Zahl der Erwerbstätigen und der Erwerbslosen nach dem Labour-Force-Konzept der Internationalen Arbeitsorganisation (ILO-Konzept) erfasst.¹ Eine solche international vergleichbare Arbeitsmarktstatistik ist für Analysten, Forscher und politische Entscheidungsträger von zentraler Bedeutung. Aufgrund des hohen Bedarfs an kurzfristigen und zeitnahen Arbeitsmarktdaten wurde schon vor vielen Jahren der Ruf nach monatlichen Statistiken laut. Seit dem 1. Januar 2005 werden der Mikrozensus und die Arbeitskräfteerhebung als kontinuierliche unterjährige Erhebung durchgeführt, sodass Monatswerte ab dem Jahr 2005 vorliegen. Im gleichen Jahr hat das Statistische Bundesamt mit der Veröffentlichung monatlicher Ergebnisse begonnen, deren Datenbasis jedoch zunächst noch die Telefonerhebung „Arbeitsmarkt in Deutschland“ war, die zu diesem Zeitpunkt schon eine fast zweijährige Pilotphase hinter sich hatte.² Erst im Jahr 2007 wurde die Datenbasis für die monatliche Berichterstattung von der Telefonerhebung auf die Arbeitskräfte-

¹ Nähere Informationen zum Labour-Force-Konzept findet man zum Beispiel bei Rengers, M.: „Das international vereinbarte Labour-Force-Konzept“ in WiSta 12/2004, Seite 1369 ff.

² Siehe Rengers, M./Körner, T./Klass, J.: „Telefonerhebung ‚Arbeitsmarkt in Deutschland‘ 2005 bis 2007“ in WiSta 7/2008, Seite 553 ff.

erhebung umgestellt. In den Jahren 2005 und 2006 genügten die Monatsergebnisse der Arbeitskräfteerhebung nicht den Qualitätsanforderungen. Es hatte sich herausgestellt, dass nach dem Wechsel von einer jährlichen auf eine unterjährige Periodizität eine längere Entwicklungs- und Anpassungsphase benötigt wurde, um qualitativ ausreichende monatliche Schnellauswertungen des unterjährigen Mikrozensus beziehungsweise der unterjährigen Arbeitskräfteerhebung zu erarbeiten. Monatswerte der Arbeitskräfteerhebung werden deshalb erst ab dem Berichtsmonat Januar 2007 veröffentlicht.

Die monatliche Auswertung des unterjährigen Mikrozensus beziehungsweise der unterjährigen Arbeitskräfteerhebung stellt nach wie vor eine erhebliche organisatorisch-technische, erhebungsmethodische und mathematisch-konzeptionelle Herausforderung dar. Im Oktober 2007, als mit der Veröffentlichung monatlicher Ergebnisse der Arbeitskräfteerhebung begonnen wurde, war noch eine Reihe methodischer Fragen zur Erzeugung qualitativ hochwertiger Monatsergebnisse weiter zu untersuchen. Darüber hinaus konnten einige generelle, mit der besonderen Erhebungsform (stichprobenbasierte Haushaltsbefragung) und der schwierigen Operationalisierung des Labour-Force-Konzeptes verbundene Probleme ebenfalls noch nicht zufriedenstellend gelöst werden. Deshalb mussten zunächst einige Übergangslösungen eingesetzt werden; zwei dieser Übergangslösungen konnten mit der Veröffentlichung zum Berichtsmonat März 2011 durch verbesserte Verfahren abgelöst werden.³

So wurde zum einen die monatliche Erwerbslosenquote, definiert als Anteil der Erwerbslosen an den Erwerbspersonen (Summe von Erwerbstätigen und Erwerbslosen), in der Übergangslösung nur teilweise auf Basis der Arbeitskräfteerhebung ermittelt. Lediglich die Zahl der Erwerbslosen wurde dieser Datenquelle entnommen, die Zahl der Erwerbstätigen stammte dagegen aus der Erwerbstätigenrechnung.⁴ In Kauf nehmen musste man bei dieser Vorgehensweise, dass die veröffentlichten Ergebnisse zur Erwerbstätigkeit und Erwerbslosigkeit nicht konsistent waren, da die Zahl der Erwerbstätigen nach den Ergebnissen der Arbeitskräfteerhebung deutlich niedriger war als nach der Erwerbstätigenrechnung. Im Jahr 2004 lag die Erwerbstätigenzahl der Arbeitskräfteerhebung um 2,8 Millionen unter derjenigen der Erwerbstätigenrechnung. Durch methodische Weiterentwicklungen des Mikrozensus beziehungsweise der Arbeitskräfteerhebung seit 2005 hat sich der Abstand zwischen den Ergebnissen der beiden Datenquellen verringert: Bis zum Jahr 2010 konnte die Differenz auf rund 1,5

Millionen Erwerbstätige verkleinert werden. Aufgrund der höheren Kohärenz der Datenquellen erfolgt die Berechnung der Erwerbslosenquote nunmehr ausschließlich auf Basis der Arbeitskräfteerhebung. Dadurch wird nicht nur die Konsistenz der Ergebnisse zur Erwerbstätigkeit und Erwerbslosigkeit gewährleistet, sondern die Berechnung der Erwerbslosenquote allein aus den Ergebnissen der Arbeitskräfteerhebung entspricht jetzt auch dem Standardverfahren des Europäischen Statistischen Systems.

Eine zweite Übergangslösung bestand darin, die saisonbereinigten Erwerbslosenzahlen mithilfe von saisonbereinigten Ergebnissen der registrierten Arbeitslosenzahlen der Bundesagentur für Arbeit zu schätzen. Diese Vorgehensweise wurde gewählt, weil in der ersten Zeit der Veröffentlichung monatlicher Ergebnisse im Jahr 2007 die Zeitreihe der Zahl der Erwerbslosen für die Anwendung eines Saisonbereinigungsverfahrens zu kurz war. Darüber hinaus zeigte diese Zeitreihe methodisch bedingt eine hohe Volatilität. Rein technisch gesehen könnte mit Vollendung des Berichtsjahres 2010 eine eigenständige Saisonbereinigung für die Erwerbslosenzahlen der Arbeitskräfteerhebung vorgenommen werden, da nunmehr eine ausreichend lange Zeitreihe vorliegt. Die Erwerbslosenzeitreihe der Arbeitskräfteerhebung weist (weiterhin) allerdings keine regelmäßigen saisonalen Muster auf, was im Wesentlichen auf die erwähnte methodisch bedingte Volatilität zurückzuführen sein dürfte. Bei einer solch volatilen Ausgangsreihe kann der Einsatz eines Saisonbereinigungsverfahrens zu statistischen Artefakten führen und eine saisonbereinigte Reihe erzeugen, die ebenfalls sehr volatil ist und kaum zu interpretieren wäre. Insbesondere Ökonomen und Konjunkturforscher benutzen unterjährige makroökonomische Zeitreihen wie die der Erwerbslosenzahlen, um ein aktuelles Bild der wirtschaftlichen Lage zu erhalten und gegebenenfalls neue Entwicklungsrichtungen und Wendepunkte möglichst frühzeitig zu erkennen. Eine saisonbereinigte Reihe mit hoher Volatilität ist hierfür kein brauchbares Analyseinstrument. Neben saisonbereinigten Zeitreihen dienen jedoch auch einen Trend beschreibende Zeitreihen als statistisches Hilfsmittel für derartige ökonomische Analysen. Bei einer Schätzung der Trend-Konjunktur-Komponente einer Zeitreihe (Trendschätzung) wird diese nicht nur um regelmäßig wiederkehrende saisonale Schwankungen, sondern auch um irreguläre Effekte bereinigt. Durch das Bereinigen um irreguläre Effekte werden somit zum einen zufallsbedingte, zum anderen beispielsweise aber auch methodisch bedingte Schwankungen entfernt beziehungsweise herausgefiltert. Die zweite Übergangslösung (Schätzung von saisonbereinigten Erwerbslosenzahlen mithilfe der saisonbereinigten registrierten Arbeitslosen) wurde deshalb mit der Veröffentlichung zum Berichtsmonat März 2011 durch die Ergebnisse einer Trendschätzung ersetzt. Die Aussagekraft dieser Ergebnisse ist im Vergleich zu der der Ergebnisse nach dem bisherigen Verfahren deutlich höher, da beim neuen Verfahren ausschließlich Originalwerte der Arbeitskräfteerhebung verwendet werden und nicht auf externe Datenquellen mit definitorisch anders abgegrenzten Zahlen (hier: registrierte Arbeitslose der Bundesagentur für Arbeit) zurückgegriffen werden muss.

Bei der Berechnung der Erwerbslosenquote werden – wie bereits erwähnt – die in den Nenner einfließenden Erwerbs-

³ Der Verfahrenswechsel wurde gemeinsam vom Statistischen Bundesamt und vom Statistischen Amt der Europäischen Gemeinschaften vorbereitet und umgesetzt. Siehe zu den Verfahrensänderungen auch Statistisches Bundesamt (Herausgeber): „Hinweise zur monatlichen Berichterstattung zur Erwerbslosigkeit ab Berichtsmonat März 2011“ unter www.destatis.de im Bereich Arbeitsmarkt (abgerufen am 12. August 2011).

⁴ Auch in der Erwerbstätigenrechnung werden die Erwerbstätigenzahlen nach dem Labour-Force-Konzept ermittelt; dafür werden jedoch *alle* zur Verfügung stehenden erwerbsstatistischen Quellen (derzeit etwa 50) herangezogen. Bei der Mehrheit dieser Datenquellen handelt es sich um Registerangaben für die Grundgesamtheit und nicht um Stichprobenerhebungen, weshalb die Ergebnisse der Erwerbstätigenrechnung eine hohe Datenqualität aufweisen. Für die relativ große Abweichung zwischen den Ergebnissen der Erwerbstätigenrechnung und der Arbeitskräfteerhebung sind mehrere Ursachen verantwortlich. Bezogen auf den Mikrozensus beziehungsweise die Arbeitskräfteerhebung sind hier zum Beispiel Stichprobenfehler, Hochrechnung, ungenaue Bevölkerungszahlen, Antwortausfälle und Messfehler zu nennen. Eine ausführliche Untersuchung findet man bei Körner, T./Puch, K.: „Der Mikrozensus im Kontext anderer Arbeitsmarktstatistiken“ in WiSta 6/2009, Seite 528 ff.

tätigenzahlen nunmehr ebenfalls der Arbeitskräfteerhebung entnommen. Die Monatszeitreihen der Erwerbstätigen und der Erwerbslosenquoten unterliegen deshalb denselben Restriktionen wie die Zeitreihe der Erwerbslosen, sodass auch hier anstelle der Ermittlung saisonbereinigter Ergebnisse bis auf weiteres zunächst nur eine Trendschätzung vorgenommen wird. Die Trendschätzverfahren für die monatliche Erwerbslosen- und die monatliche Erwerbstätigenzeitreihe sind identisch. Beiden Trendschätzungen werden Monatszeitreihen ab 1991 zugrunde gelegt. Da die Arbeitskräfteerhebung erst ab dem Jahr 2005 als kontinuierliche unterjährige Erhebung durchgeführt wurde, wurden die Werte für Januar 1991 bis Dezember 2004 über eine Rückrechnung geschätzt. Die jeweiligen jährlichen Ergebnisse der Arbeitskräfteerhebung und die in den Jahren 2005 bis 2010 beobachteten monatlichen Verläufe sind Basis dieser Rückrechnung. Die Rückrechnungsergebnisse selbst sind ein reines Hilfskonstrukt der Trendschätzung und eignen sich nicht zur Veröffentlichung. Im Gegensatz dazu lassen sich Ergebnisse der Trendschätzung für die Berichtsmonate ab Januar 1991 unter der Bedingung veröffentlichen, dass sie sehr vorsichtig interpretiert werden.

1.2 Besonderheiten bei der zeitreihenanalytischen Behandlung der Mikrozensusdaten

Auf die Bedeutung unterjähriger Arbeitsmarktzahlen für politische Entscheidungsträger und die Wissenschaft wurde bereits hingewiesen. Insbesondere sollen damit konjunkturelle Entwicklungstendenzen und mögliche Wendepunkte der Konjunktur frühzeitig erkannt werden. Aus der Zeitreihenanalyse stehen hierfür grundsätzlich zwei Hilfsmittel zur Verfügung, mit denen die gewünschten Informationen gewonnen werden können: die Trend-Konjunktur-Komponente und die saisonbereinigte Reihe. Bei der Schätzung der Trend-Konjunktur-Komponente, kurz Trend oder auch glatte Komponente genannt, wird die Zeitreihe nicht nur um regelmäßig wiederkehrende saisonale Schwankungen, sondern auch um die irreguläre Komponente bereinigt. Bei der Schätzung saisonbereinigter Werte werden dagegen lediglich Saison- und eventuelle zusätzliche Kalendereffekte ausgeschaltet.

Aus theoretischer Sicht werden konjunkturelle Entwicklungen am besten über die Trend-Konjunktur-Komponente beschrieben. Zur historischen Konjunkturdiagnose, also in retrospektiver Sicht, fällt die Entscheidung zwischen Trend oder saisonbereinigter Reihe deshalb eindeutig zugunsten des Trends aus. Zur Beurteilung der aktuellen oder gar der künftigen Wirtschaftslage geht die Wahl zwischen Trend und saisonbereinigter Reihe zunächst unentschieden aus. Das liegt daran, dass zwischen den verschiedenen Anforderungen, die an ein prospektives Konjunkturanalyseinstrument gestellt werden, Zielkonflikte bestehen, zum Beispiel zwischen dem Wunsch nach einem *klaren Signal* und demjenigen nach *geringen Revisionen* beziehungsweise zwischen der Anforderung nach *schneller Aufdeckung* neuer Entwicklungstendenzen und der Minimierung des *Risikos eines falschen Alarms*.

Am aktuellen Rand einer Zeitreihe ist der Trend zunächst mit größeren Unsicherheiten behaftet, weil die Entwicklung künftiger Werte unbekannt ist. Mit dem monatlichen Hinzufügen neuer Werte besteht deshalb die Gefahr relativ großer Revisionen. Die saisonbereinigte Reihe zeigt am Reihenende dagegen eine größere Stabilität, weil sich Saison- und Kalendereinflüsse im Laufe der Zeit nur sehr allmählich ändern und deshalb recht zuverlässig geschätzt werden können. Das Ausmaß der durch nachfolgende Werte der nächsten Monatsergebnisse ausgelösten Revisionen ist deshalb bei saisonbereinigten Reihen kleiner als bei Trendzeitreihen. Bezüglich des monatlichen Revisionsausmaßes gewinnt zwar die saisonbereinigte Reihe, für die Revisionsdauer gilt dies jedoch nicht, da diese 12 oder sogar 24 Monate betragen kann. Während das Revisionsausmaß in den ersten Monaten vergleichsweise gering ist, kann das Gesamtausmaß der Revisionen nach ein oder zwei Jahren dagegen erheblich sein.

Unabhängig von den Revisionen besteht das Problem des gewünschten klaren Signals bei möglichst schneller Aufdeckung. Trendwerte können zwar prinzipiell ein klares Signal geben, aufgrund der Unsicherheit am aktuellen Rand ist dies ohne das Risiko eines falschen Alarms jedoch nur mit zeitlicher Verzögerung möglich. Saisonbereinigte Reihen weisen eventuell mit geringerer zeitlicher Verzögerung auf mögliche konjunkturelle Wendepunkte hin. Da saisonbereinigte Reihen neben der Trendkomponente allerdings auch die irreguläre Komponente enthalten, stellt das Herausfiltern eines klaren Signals bezüglich der konjunkturellen Entwicklungstendenz aber eine erhebliche Schwierigkeit dar, die zusätzlich bedeutendes Wissen über den ökonomischen Sachverhalt erfordert. Für die „richtige“ Interpretation saisonbereinigter Reihen muss darüber hinaus gleichzeitig auch das mathematisch-methodische Wissen um das eingesetzte – in der Regel sehr komplexe – Saisonbereinigungsverfahren vorhanden und der Zugang zu den Informationen über zeitreihenspezifische Festlegungen konkreter Modellspezifikationen gegeben sein.

Die Frage, ob Trendwerte oder saisonbereinigte Reihen bei der Konjunkturanalyse zum Einsatz kommen sollen, ist ein seit Jahrzehnten viel diskutiertes Thema, zu dem hier nur die wichtigsten Argumente wiedergegeben werden konnten.⁵ Daneben gibt es eine nicht weniger ausgeprägte Debatte um die Wahl des geeigneten methodischen Verfahrens zur Saisonbereinigung beziehungsweise Trendschätzung. Auf europäischer Ebene entstand – angesichts der Vielzahl möglicher Verfahren – das Bedürfnis nach einer harmonisierten Vorgehensweise. Aus diesem Grund und um die Beiträge aus unterjährigen Statistiken der einzelnen Mitgliedstaaten zu EU-Aggregaten vergleichbar zu machen, entwickelten Eurostat und die Europäische Zentralbank die sogenannten *ESS Guidelines on Seasonal Adjustment* (ESS-Guidelines)⁶, also Richtlinien zur Saisonbereinigung. Die Bezeichnung

⁵ Siehe zu dieser Diskussion zum Beispiel auch Eurostat: „Seasonal Adjustment Methods: A Comparison“, Luxemburg 1998, Seite 52 ff., und Deutsche Bundesbank: „Der Übergang vom Saisonbereinigungsverfahren Census X-11 zu Census X-12-ARIMA“, Monatsbericht September 1999, Seite 39 ff., hier: Seite 41 f.

⁶ Eurostat: „ESS Guidelines on Seasonal Adjustment“, Luxemburg 2009. Die Guidelines können unter http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-09-006/EN/KS-RA-09-006-EN.PDF eingesehen werden (abgerufen am 12. August 2011).

lässt bereits erkennen, dass als Standardvorgehensweise eine Saisonbereinigung und nicht etwa eine Trendschätzung gewählt wurde. Methodisch-technisch fokussieren die Guidelines ausschließlich auf zwei vergleichsweise weitverbreitete Ansätze: TRAMO-SEATS und Census X-12-ARIMA. Es handelt sich um die am häufigsten verwendeten Verfahren innerhalb des Europäischen Statistischen Systems (ESS).

Neben der Harmonisierung und dem damit verbundenen Ziel, robustere Zahlen für die EU-Aggregate zu bekommen, sollen die Nutzer und Nutzerinnen eingehend über das verwendete Saisonbereinigungsverfahren informiert werden. Dazu gehört auch eine vollständige Darstellung der im Saisonbereinigungsprozess verwendeten Spezifikationen. Beide Ziele werden explizit in den ESS-Guidelines erwähnt, allerdings wird das letztgenannte Ziel in der Praxis bisher kaum oder zumindest nicht offensiv umgesetzt. Eine Saisonbereinigung soll nach diesen Richtlinien außerdem nur dann erfolgen, wenn es bei einer Zeitreihe eine klare statistische Evidenz und eine ökonomische Interpretation der Saison- und Kalendereffekte gibt. Bei den Vorarbeiten zur Saisonbereinigung sind die üblichen, den Verfahren zugrunde liegenden Stationaritäts- und Homoskedastizitätsannahmen zu überprüfen.⁷ Bei der Saisonbereinigung selbst ist auf eine hohe Qualität zu achten, die über verschiedene Modellierungsansätze und die Prüfung unterschiedlicher Qualitätsmaße kontrolliert werden soll. Da Saisonbereinigungsverfahren sehr komplex sind, wird in den ESS-Guidelines darauf hingewiesen, dass eine Saisonbereinigung einen hohen Zeitaufwand mit entsprechenden technischen und personellen Ressourcen erfordert. Bei kurzen Zeitreihen wird als beste Lösung empfohlen, keine Saisonbereinigung vorzunehmen, wenn die Länge der Zeitreihen weniger als drei Jahre beträgt. Zeitreihen, die drei bis sieben Jahre umfassen, können einer Saisonbereinigung unterzogen werden, wobei zur Stabilisierung der Ergebnisse eine Rückrechnung der Zeitreihe befürwortet wird. Bezüglich der Präsentation und Veröffentlichung der Daten *sollten* als Minimum originäre Zeitreihe, saisonbereinigte Zeitreihe und Zeitreihenidentifikatoren angegeben werden. Zusätzlich *können* Trendwerte, Saisonfaktoren und sonstige Metadaten zur Beschreibung der im Saisonbereinigungsprozess gewählten Optionen veröffentlicht werden. Bei einer Veröffentlichung von Trendwerten empfehlen die ESS-Guidelines aufgrund des Problems des unsicheren Reihenendes als beste Lösung, die jüngsten Werte nicht zu publizieren.

Hinweise zur zeitreihenanalytischen Behandlung monatlicher Arbeitsmarktzeitreihen des Mikrozensus beziehungsweise der Arbeitskräfteerhebung findet man aber nicht nur in den ESS-Guidelines. So gibt auch die Internationale Arbeitsorganisation (International Labour Organization – ILO) in ihren *ILO-Recommendations on Seasonal Adjustment of Employment and Unemployment Data*⁸ ergänzende Empfehlungen speziell zur Saisonbereinigung von Erwerbstätigen- und Erwerbslosenzeitreihen. In den

ILO-Empfehlungen wird zunächst darauf hingewiesen, dass es weder eine einheitliche noch eine perfekte Saisonbereinigung gibt und Saisonbereinigungsergebnisse deshalb immer als Schätzungen betrachtet werden müssen. Der Versuch einer internationalen oder zumindest europaweiten Harmonisierung der Vorgehensweise bei der Saisonbereinigung durch die ESS-Guidelines wird in den ILO-Empfehlungen begrüßt. Die ESS-Guidelines sind allerdings für eine Vielzahl ökonomisch relevanter Zeitreihen entwickelt worden, wobei diese Zeitreihen häufig aus den Volkswirtschaftlichen Gesamtrechnungen hervorgehen. Mit Blick auf die Besonderheiten von Erwerbstätigen- und Erwerbslosenzeitreihen hat die Internationale Arbeitsorganisation die ergänzenden Empfehlungen verfasst und dabei die speziellen Charakteristika dieser umfragebasierten Daten untersucht und berücksichtigt. Umfragebasierte Daten zeigen allein dadurch, dass sie aus einer Stichprobe gewonnen werden, zwangsläufig eine höhere Volatilität als Daten aus Vollerhebungen und Registern – selbst bei hoher Qualität des Stichprobendesigns. Ausgangspunkt der ILO-Empfehlungen sind umfangreiche Untersuchungen von umfragebasierten Erwerbstätigen- und Erwerbslosenzeitreihen aus den weltweit durchgeführten Arbeitskräfteerhebungen (Labour Force Surveys – LFS). Dabei wurde unter anderem festgestellt, dass in der Kalendereinigung zwar keine Handeltage (trading days) berücksichtigt werden müssen, wohl aber wechselnde Feiertage (zum Beispiel Ramadan-Effekt bei der Erwerbstätigenzeitreihe in muslimischen Ländern). Nach ILO-Angaben enthalten Erwerbstätigenzeitreihen aus stichprobenbasierten Umfragedaten darüber hinaus häufig Brüche, wie sie beispielsweise durch Änderungen in der Stichprobenmethodik verursacht werden. Manche Brüche führen sogar dazu, dass die Zeitreihe vor dem Bruch und die nachfolgende als zwei verschiedene Zeitreihen betrachtet werden müssen.

Genau wie die ESS-Guidelines beschränken sich auch die ILO-Empfehlungen bei der Wahl des Saisonbereinigungsverfahrens auf TRAMO-SEATS und Census X-12-ARIMA. Bei ihren Untersuchungen von mehr als Tausend Erwerbstätigen- und Erwerbslosenzeitreihen hat die Internationale Arbeitsorganisation jedoch festgestellt, dass beide Verfahren nicht immer zu ähnlichen Ergebnissen führen. Für den Fall, dass TRAMO-SEATS keine eindeutige Saisonkomponente identifiziert, wird deshalb empfohlen, zusätzlich Census X-12-ARIMA zu verwenden, um die Abwesenheit eines Saisonmusters zu validieren. Das Problem möglicherweise fehlender saisonaler Muster wird indes zusätzlich explizit erwähnt, denn die ILO-Analysen haben gezeigt, dass bei einer Vielzahl von Erwerbstätigenzeitreihen keine eindeutigen Saisonmuster identifiziert werden können. In diesem Fall wird davon abgeraten, eine Saisonbereinigung durchzuführen.⁹

Vor diesen Hintergrundinformationen – Probleme der Trendermittlung und Saisonbereinigung; europaweite und

⁷ Sind Stationaritäts- und Homoskedastizitätsannahme erfüllt, spricht man auch vom sogenannten weißen Rauschen. Siehe dazu die Ausführungen im Abschnitt 2.1.

⁸ Internationale Arbeitsorganisation (ILO, Herausgeber): "ILO Department of Statistics Technical Note: Recommendations on Seasonal Adjustment of Employment and Unemployment Data", September 2010, im Internet unter http://laborsta.ilo.org/sti/DATA_FILES/ILO_SA_GUIDELINES.pdf (abgerufen am 12. August 2011).

⁹ Siehe Internationale Arbeitsorganisation: "Finally, it is worth pointing out the lack of a clearly identifiable seasonal pattern for a non-negligible number of employment time series (e.g. employment in services activities). In such cases, there is no point in adjusting the original data. But this is valuable information to the user and the analyst, since the raw data can be used for short-term analysis. In such cases, it is recommended to attach a note to the raw time series, mentioning the lack of an identifiable seasonal pattern." Fußnote 8, hier: Seite 6.

internationale Empfehlungen; Besonderheiten von Umfragedaten – sind die in Abschnitt 1.1 erwähnten Neuerungen bezüglich der zeitreihenanalytischen Behandlung der monatlichen Arbeitsmarktzeitreihen des Mikrozensus beziehungsweise der Arbeitskräfteerhebung zu beurteilen. Anzumerken ist, dass diese nicht vollständig im Einklang mit den ESS-Guidelines und den ILO-Empfehlungen stehen. Diese Abweichungen erfolgen jedoch nicht willkürlich, sondern sind in den Besonderheiten der unterjährigen Mikrozensusdaten begründet. Diese gehen über die generellen – in den ILO-Empfehlungen erwähnten – Besonderheiten umfragebasierter Daten hinaus. Die monatlichen Auswertungen des unterjährigen Mikrozensus sind mit *zusätzlichen* methodischen Problemen behaftet, die zu einer hohen methodisch bedingten Volatilität der Zeitreihen führen.

Ohne die speziellen Schwierigkeiten (Fragebogendesign, Stichprobendesign, Stichprobenrahmen, Hochrechnung und so weiter) hier im Einzelnen anzusprechen und zu erörtern, sei an dieser Stelle lediglich auf die Schaubilder im Anhang zu diesem Beitrag auf Seite 751 f. verwiesen, die einen optischen Eindruck von der unterschiedlichen Volatilität von monatlichen Arbeitsmarktzeitreihen aus dem Mikrozensus im Vergleich zu denen aus Registerdaten und der Erwerbstätigenrechnung geben.¹⁰

Darüber hinaus ist darauf hinzuweisen, dass es sich bei dieser zeitreihenanalytischen Behandlung der Mikrozensusdaten um eine vorläufige Vorgehensweise handelt. Es wird auch weiterhin daran gearbeitet, die Qualität der monatlichen Ergebnisse aus dem Mikrozensus beziehungsweise der deutschen Arbeitskräfteerhebung zu verbessern. Die Untersuchungen zur Frage, inwieweit Änderungen etwa am Stichprobendesign, am Hochrechnungsverfahren oder an den Datenerhebungsverfahren die methodisch bedingte Volatilität der Zeitreihe verringern können, werden ebenso fortgesetzt wie die Analysen zur Weiterentwicklung oder Ergänzung des eingesetzten Trendschätzverfahrens. Dies beinhaltet auch die Möglichkeit, gegebenenfalls parallel die Ergebnisse einer zweiten Trendschätzung, die auf einem anderen Verfahren beruht, zu veröffentlichen.

Auf eine Saisonbereinigung der Ergebnisse wird aber aus den geschilderten methodischen Gründen im Statistischen Bundesamt und beim Statistischen Amt der Europäischen Gemeinschaften zunächst weiterhin verzichtet.

Bei den monatlichen Zeitreihen des Mikrozensus erfolgt die Trendschätzung auf Basis des *Berliner Verfahrens, Version 4.1 (BV4.1)*. Angesichts der erwähnten Schwierigkeiten von Trendschätzungen und Saisonbereinigungen wird immer wieder betont, dass die *Transparenz* der eingesetzten Verfahren von größter Bedeutung für die Interpretation der Ergebnisse ist. Dies bekräftigen nicht nur die ESS-Guidelines, sondern auch die ILO-Empfehlungen, in denen dieser Aspekt für so wichtig erachtet wird, dass ein eigener Vor-

schlag zur Standardisierung der Dokumentation vorgestellt wird. Um die zurzeit veröffentlichten Trendschätzungen für die monatlichen Arbeitsmarktzeitreihen des Mikrozensus beurteilen zu können, ist es unabdingbar, das verwendete Trendschätzverfahren zu kennen. Das eingesetzte Trendschätzverfahren wird deshalb im Folgenden detailliert vorgestellt.¹¹

2 Trendschätzung mit BV4.1

In der Formelnotation dieses Beitrages wird der Unterstrich verwendet, um Zufallsvariablen, Zufallsvektoren oder Zufallsmatrizen zu kennzeichnen. Realisationen von Zufallsgrößen sind ohne Unterstrich, Vektoren beziehungsweise Matrizen durch fett markierte Klein- und Großbuchstaben dargestellt.

2.1 Das Basismodell von BV4.1 und die Schätzung seiner Parameter

Dem generellen Ansatz zur Zerlegung einer Zeitreihe in ihre Komponenten folgend setzt sich das Basismodell des Berliner Verfahrens, Version 4.1 (BV4.1) additiv aus drei stochastischen Teilprozessen zusammen, welche die Trend-Konjunktur-Schwankungen, die saisonalen Schwankungen und die Zufallsschwankungen einer unterjährigen ökonomischen Zeitreihe beschreiben.

Die systematischen Komponenten Trend-Konjunktur und Saison werden dabei durch bestimmte funktionale Zusammenhänge modelliert und für die sogenannte irreguläre Komponente wird ein Zufallsprozess angenommen, für den weißes Rauschen (white-noise) vorliegt. Ein Zufallsprozess $\{\underline{\varepsilon}_t | t=1, \dots, n\}$ wird als weißes Rauschen bezeichnet, wenn die n Zufallsvariablen $\underline{\varepsilon}_t$ paarweise unkorreliert sind, den Erwartungswert Null und die gleiche konstante Varianz σ_ε^2 haben. Für den $[n \times 1]$ -Vektor der Zufallsvariablen $\underline{\varepsilon}_t$ gilt somit:

$$(1) \quad E(\underline{\varepsilon}_t) = \mathbf{0}_n \text{ und}$$

$$(2) \quad \text{Cov}(\underline{\varepsilon}) = \sigma^2 \cdot \mathbf{I}_n.$$

Bei Monatsreihen wird zur Modellierung der systematischen Trend-Konjunktur-Komponente (im Folgenden kurz als Trendkomponente bezeichnet) eine Polynomfunktion dritten Grades und für die systematische Saisonkomponente eine trigonometrische Funktion sechsten Grades verwendet. Das stochastische Grundmodell nimmt somit folgende Form an:¹²

11 Eine Kurzbeschreibung der Trendschätzung mit BV4.1 findet sich in Rengers, M.: „Monatliche Erwerbslosenstatistik: Trendschätzung mit BV4.1“ im Informationsdienst „Methoden – Verfahren – Entwicklungen“ des Statistischen Bundesamtes, Ausgabe 1/2011, Seite 6 ff. Ausführliche Informationen – nicht nur zur Trendschätzung, sondern zur vollständigen Komponentenzersetzung mit BV4.1 – liefern Speth, H.-T.: „Komponentenzersetzung und Saisonbereinigung ökonomischer Zeitreihen mit dem Verfahren BV4.1“ in der Schriftenreihe „Methodenberichte“ des Statistischen Bundesamtes, Heft 3, Wiesbaden 2004, und Nourney, M.: „Umstellung der Zeitreihenanalyse“ in WiSta 11/1983, Seite 841 ff., im Internet unter www.destatis.de, im Bereich Wissenschaftsforum → Methoden und Verfahren.

12 Die Anwendung des Basismodells erfordert grundsätzlich, die zu analysierenden ökonomischen Zeitreihen von anderen systematischen Einflüssen zu bereinigen, die nicht der Trend-Konjunktur- oder der Saisonkomponente zuzuordnen sind (zum Beispiel kalenderbedingte Einflüsse oder andere systematische Teile der Restkomponente). Darüber hinaus ist es sinnvoll, die Zeitreihe von Ausreißern zu bereinigen. Siehe dazu Speth, H.-T., Fußnote 11, hier: Seite 3 f. und Seite 16 ff.

10 Ausführlichere Informationen über die methodischen Probleme beim unterjährigen Mikrozensus enthält Kapitel 5 beziehungsweise Abschnitt 5.2 „Methodological problems related to the intra-annual results from the German LFS“ in Körner, T./Puch, K.: „Coherence of German Labour Market Statistics. The German Labour Force Survey in Comparison with the Employment Accounts and the Unemployment Register“, Band 19 der Schriftenreihe „Statistik und Wissenschaft“ des Statistischen Bundesamtes, Wiesbaden 2011 (im Erscheinen).

$$\begin{aligned}
 (3) \quad y_t = & \underbrace{\sum_{j=0}^3 \alpha_j \cdot t^j}_{\text{Trend-Konjunktur-Komponente } m_t} \\
 & + \underbrace{\sum_{j=1}^6 (\beta_j \cdot \cos \pi \cdot \frac{j}{6} \cdot t + \gamma_j \sin \pi \cdot \frac{j}{6} \cdot t)}_{\text{Saison-Komponente } s_t} + \varepsilon_t \\
 & \text{Basismodell} \\
 & (t=1, \dots, n)
 \end{aligned}$$

Die Modellierung erfolgt dabei über einen gleitenden Stützbereich n , der in der Regel 27 Monatswerte umfasst. In Matrixschreibweise lautet das Basismodell

$$(4) \quad \underset{n \times 1}{\mathbf{y}} = \underset{n \times 15}{\mathbf{X}} \cdot \underset{n \times 1}{\boldsymbol{\beta}} + \underset{n \times 1}{\boldsymbol{\varepsilon}},$$

wobei \mathbf{X} die Regressormatrix mit den Zeitpunkten $t=1, \dots, n$ des verwendeten Stützbereiches beschreibt, das heißt die Zeitpunkte t einer Zeitreihe werden als Regressorvariablen der Matrix \mathbf{X} verwandt. Bei einer Stützbereichslänge von n Monatswerten werden die Monate immer von 1 bis n durchnummeriert. Die Nummerierung beginnt mit dem zeitlich am weitesten zurückliegenden Monat. In der Notation wird dies im Folgenden teilweise einfach durch $t=1, \dots, n$ oder durch t_1, t_2, \dots, t_n symbolisiert, wobei, $t_1=1, t_2=2, t_3=3$, und so weiter gilt.

Der $[n \times 1]$ -Zufallsvektor \mathbf{y} in Gleichung (4) stellt die Zufallsvariablen y_t ($t=1, \dots, n$) dar, deren Realisationen die beobachteten Zeitreihenwerte abbilden. Bei Einbezug der Oberwelle π in das Basismodell ist zu beachten, dass bei ganzzahligen Beobachtungszeitpunkten die letzte Spalte der Regressormatrix \mathbf{X} nur aus Nullen bestehen würde. Das sin-Glied der letzten Oberwelle wird deshalb nicht miteinbezogen, das heißt die Spalte $[\sin \pi \cdot \frac{6}{6} \cdot t_1 \quad \sin \pi \cdot \frac{6}{6} \cdot t_2 \quad \dots \quad \sin \pi \cdot \frac{6}{6} \cdot t_n]^T$ in \mathbf{X} entfällt, sodass die Regressormatrix die Dimension $[n \times 15]$ annimmt. Entsprechend enthält der Parametervektor $\boldsymbol{\beta}$ ebenfalls nur die 15 Parameter α_j ($j=0, 1, 2, 3$), β_j ($j=1, 2, \dots, 6$) und γ_j ($j=1, 2, \dots, 5$) der Basisfunktionen von Trend- und Saisonkomponente.

Die 15 Parameter α_j, β_j und γ_j der Basisfunktionen von Trend- und Saisonkomponente werden nach dem *Prinzip der gleitenden Schätzung* jeweils für einen Stützbereich der Länge n ermittelt. Als Schätzverfahren wird ein *gewichteter Kleinste-Quadrate-Ansatz (Weighted Least Squares – WLS)* verwendet (siehe dazu auch Exkurs 3 auf Seite 744 f.). Durch die Einfügung der Gewichtung w_t in den Regressionsansatz werden die Zeitreihenwerte in der unmittelbaren Umgebung der Schätzstelle bei der Parameterschätzung stärker gewichtet. Es gilt:

$$(5) \quad w_t = 1 - \frac{|t - t^*|}{D + 1} \quad \text{für } t=1, \dots, n.$$

Hierbei ist t^* der Zeitpunkt innerhalb des Stützbereiches, der das höchste Gewicht $w_t = 1$ erhalten soll, und D die größere Anzahl der Monate zwischen dem Zeitpunkt t^* und den

beiden Endpunkten 1 und n des Stützbereiches. Für den sogenannten mittleren Schätzbereich mit $n=27$ ist $t^*=14$ und $D=13$. Es gilt:

$$(6) \quad w_t = 1 - \frac{|t - 14|}{13 + 1}.$$

Gemäß den Gleichungen (XXII) und (XXIII) des Exkurses 3 lassen sich die Parameter des Basismodells (4) bei der gewichteten Kleinste-Quadrate-Schätzung über

$$(7) \quad \hat{\boldsymbol{\beta}}^{WLS} = \underbrace{[(\mathbf{W} \cdot \mathbf{X})' \cdot (\mathbf{W} \cdot \mathbf{X})]^{-1}}_{[15 \times 15]} \cdot \underbrace{(\mathbf{W} \cdot \mathbf{X})'}_{[15 \times 27]} \cdot \underbrace{\mathbf{W}}_{[27 \times 27]} \cdot \underbrace{\mathbf{y}}_{[27 \times 1]}$$

beziehungsweise

$$(8) \quad \hat{\boldsymbol{\beta}}^{WLS} = (\mathbf{X}^* \cdot \mathbf{X}^*)^{-1} \cdot \mathbf{X}^* \cdot \mathbf{Q} \cdot \mathbf{y}$$

ermitteln. Die Matrix \mathbf{W} in (7) stellt dabei die Gewichtungsbeziehungsweise Transformationsmatrix dar, die als Diagonalmatrix auf der Hauptdiagonalen die Quadratwurzeln der Gewichte w_t aus Gleichung (6) enthält:

$$(9) \quad \mathbf{W}_{27 \times 27} = \begin{bmatrix} \sqrt{w_1} & 0 & \dots & 0 \\ 0 & \sqrt{w_2} & & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{w_n} \end{bmatrix} = \begin{bmatrix} \sqrt{1 - \frac{|1-14|}{13+1}} & 0 & \dots & 0 \\ 0 & \sqrt{1 - \frac{|2-14|}{13+1}} & & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{1 - \frac{|27-14|}{13+1}} \end{bmatrix}.$$

Die Kreuzproduktmatrix $\mathbf{W}' \cdot \mathbf{W}$ wird mit \mathbf{Q} bezeichnet und ist eine Diagonalmatrix mit den Gewichten w_t als Hauptdiagonalelementen. Die Matrix \mathbf{X}^* in Gleichung (8) ist die transformierte Regressormatrix $\mathbf{X}^* = \mathbf{W} \cdot \mathbf{X}$ (siehe dazu Exkurs 3).

Mit dem geschätzten Parametervektor (7) beziehungsweise (8) können nun Schätzwerte \hat{y}_t für das Basismodell (3) ermittelt werden, die die Summe aus geschätzter Trendkomponente \hat{m}_t und geschätzter Saisonkomponente \hat{s}_t darstellen.

In diesem Beitrag interessiert jedoch ausschließlich die Schätzung der Trendkomponente, sodass letztlich nur der in $\hat{\boldsymbol{\beta}}^{WLS}$ enthaltene $[4 \times 1]$ -Teilvektor der α_j -Parameter ($j=0, 1, 2, 3$) der Trendkomponente benötigt wird. $\hat{\boldsymbol{\beta}}^{WLS}$ kann in zwei Teilvektoren $\hat{\boldsymbol{\alpha}}_{11}^{WLS}$ und $\hat{\boldsymbol{\delta}}_{21}^{WLS}$ partitioniert werden,

$$(10) \quad \hat{\boldsymbol{\beta}}^{WLS} = \begin{bmatrix} \hat{\boldsymbol{\alpha}}_{11}^{WLS} \\ \hat{\boldsymbol{\delta}}_{21}^{WLS} \end{bmatrix}, \quad \begin{matrix} [4 \times 1] \\ [11 \times 1] \end{matrix}$$

die einerseits die Parameterschätzer der Trendkomponente ($\hat{\boldsymbol{\alpha}}_{11}^{WLS}$) und andererseits die Parameterschätzer der Saisonkomponente ($\hat{\boldsymbol{\delta}}_{21}^{WLS}$) enthalten.

Zwar wird – wie erwähnt – zur Ermittlung der Trendkomponente nur der Teilvektor $\hat{\boldsymbol{\alpha}}_{11}^{WLS}$ benötigt, die Schätzung der

Parameter von Trend- und Saisonkomponente des Basismodells erfolgt jedoch *simultan*. Im Folgenden wird gezeigt, wie die konkrete Formel zur alleinigen Berechnung der α_j -Parameter ($j=0,1,2,3$) bei dennoch simultaner Schätzung von Trend- und Saisonkomponente aussieht. Dazu wird zunächst auch die Regressormatrix X partitioniert. Es sei:

$$(11) \quad X_{[27 \times 15]} = [X_{11} \quad X_{12}]$$

Die Teilmatrix X_{11} enthält dabei die Regressoren der Trendkomponente,

$$(12) \quad X_{11} = \begin{bmatrix} 1 & t_1 & t_1^2 & t_1^3 \\ 1 & t_2 & t_2^2 & t_2^3 \\ 1 & t_3 & t_3^2 & t_3^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_{27} & t_{27}^2 & t_{27}^3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1^2 & 1^3 \\ 1 & 2 & 2^2 & 2^3 \\ 1 & 3 & 3^2 & 3^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 27 & 27^2 & 27^3 \end{bmatrix},$$

während die Teilmatrix X_{12} von den Regressoren der Saisonkomponente bestimmt wird:

$$(13) \quad X_{12} = \begin{bmatrix} \cos \pi \cdot \frac{1}{6} \cdot t_1 & \cos \pi \cdot \frac{2}{6} \cdot t_1 & \cos \pi \cdot \frac{3}{6} \cdot t_1 \\ \cos \pi \cdot \frac{1}{6} \cdot t_2 & \cos \pi \cdot \frac{2}{6} \cdot t_2 & \cos \pi \cdot \frac{3}{6} \cdot t_2 \\ \vdots & \vdots & \vdots \\ \cos \pi \cdot \frac{1}{6} \cdot t_{27} & \cos \pi \cdot \frac{2}{6} \cdot t_{27} & \cos \pi \cdot \frac{3}{6} \cdot t_{27} \\ \\ \cos \pi \cdot \frac{4}{6} \cdot t_1 & \cos \pi \cdot \frac{5}{6} \cdot t_1 & \cos \pi \cdot \frac{6}{6} \cdot t_1 \\ \cos \pi \cdot \frac{4}{6} \cdot t_2 & \cos \pi \cdot \frac{5}{6} \cdot t_2 & \cos \pi \cdot \frac{6}{6} \cdot t_2 \\ \vdots & \vdots & \vdots \\ \cos \pi \cdot \frac{4}{6} \cdot t_{27} & \cos \pi \cdot \frac{5}{6} \cdot t_{27} & \cos \pi \cdot \frac{6}{6} \cdot t_{27} \\ \\ \sin \pi \cdot \frac{1}{6} \cdot t_1 & \sin \pi \cdot \frac{2}{6} \cdot t_1 & \sin \pi \cdot \frac{3}{6} \cdot t_1 \\ \sin \pi \cdot \frac{1}{6} \cdot t_2 & \sin \pi \cdot \frac{2}{6} \cdot t_2 & \sin \pi \cdot \frac{3}{6} \cdot t_2 \\ \vdots & \vdots & \vdots \\ \sin \pi \cdot \frac{1}{6} \cdot t_{27} & \sin \pi \cdot \frac{2}{6} \cdot t_{27} & \sin \pi \cdot \frac{3}{6} \cdot t_{27} \\ \\ \sin \pi \cdot \frac{4}{6} \cdot t_1 & \sin \pi \cdot \frac{5}{6} \cdot t_1 \\ \sin \pi \cdot \frac{4}{6} \cdot t_2 & \sin \pi \cdot \frac{5}{6} \cdot t_2 \\ \vdots & \vdots \\ \sin \pi \cdot \frac{4}{6} \cdot t_{27} & \sin \pi \cdot \frac{5}{6} \cdot t_{27} \end{bmatrix}$$

Unter Verwendung der Teilmatrizen X_{11} und X_{12} nimmt die transformierte Regressormatrix $X^* = W \cdot X$ folgende Form an:

$$(14) \quad X^* = W \cdot X = \begin{bmatrix} W \cdot X_{11} & W \cdot X_{12} \\ [27 \times 4] & [27 \times 11] \end{bmatrix}$$

Ihre Transponierte lässt sich durch

$$(15) \quad X^{*'} = \begin{bmatrix} X_{11}' \cdot W \\ X_{12}' \cdot W \\ [4 \times 27] \\ [11 \times 27] \end{bmatrix}$$

ausdrücken und die Kreuzproduktmatrix $X^{*'} \cdot X^*$ lautet:

$$(16) \quad X^{*'} \cdot X^* = \begin{bmatrix} X_{11}' \cdot W \cdot W \cdot X_{11} & X_{11}' \cdot W \cdot W \cdot X_{12} \\ X_{12}' \cdot W \cdot W \cdot X_{11} & X_{12}' \cdot W \cdot W \cdot X_{12} \end{bmatrix} = \begin{bmatrix} X_{11}' \cdot Q \cdot X_{11} & X_{11}' \cdot Q \cdot X_{12} \\ X_{12}' \cdot Q \cdot X_{11} & X_{12}' \cdot Q \cdot X_{12} \\ [4 \times 4] & [4 \times 11] \\ [11 \times 4] & [11 \times 11] \end{bmatrix}$$

Für die Inverse der Kreuzproduktmatrix erhält man nach den Rechenregeln der Matrizenrechnung:¹³

$$(17) \quad (X^{*'} \cdot X^*)^{-1} = \begin{bmatrix} B^{-1} & C \\ D & E \end{bmatrix},$$

mit

$$B = X_{11}' \cdot Q \cdot X_{11} - (X_{11}' \cdot Q \cdot X_{12}) \cdot [X_{12}' \cdot Q \cdot X_{12}]^{-1} \cdot (X_{12}' \cdot Q \cdot X_{11}),$$

$$C = -B^{-1} \cdot (X_{11}' \cdot Q \cdot X_{12}) \cdot [X_{12}' \cdot Q \cdot X_{12}]^{-1},$$

$$D = -[X_{12}' \cdot Q \cdot X_{12}]^{-1} \cdot (X_{12}' \cdot Q \cdot X_{11}) \cdot B^{-1},$$

$$E = [X_{12}' \cdot Q \cdot X_{12}]^{-1} + [X_{12}' \cdot Q \cdot X_{12}]^{-1} \cdot X_{12}' \cdot Q \cdot X_{11} \cdot B^{-1} \cdot (X_{11}' \cdot Q \cdot X_{12}) \cdot [X_{12}' \cdot Q \cdot X_{12}]^{-1}$$

Die Formel (7) beziehungsweise (8) zur Berechnung des gesamten Parametervektors (Parameterschätzung für Trend- und Saisonkomponente bei simultaner Schätzung)

$$(18) \quad \hat{\beta}^{WLS} = (X^{*'} \cdot X^*)^{-1} \cdot X' \cdot Q \cdot y$$

lässt sich unter Verwendung der Gleichungen (10), (17) und (11) wie folgt partitionieren:

$$(19) \quad \begin{bmatrix} \hat{\alpha}_{11}^{WLS} \\ \hat{\delta}_{21}^{WLS} \\ [4 \times 1] \\ [11 \times 1] \end{bmatrix} = \begin{bmatrix} B^{-1} & C \\ D & E \end{bmatrix} \cdot \begin{bmatrix} X_{11}' \cdot Q \\ X_{12}' \cdot Q \\ [4 \times 27] \\ [11 \times 27] \end{bmatrix} \cdot y$$

Damit erhält man für den Teilvektor der gesuchten Trendparameter

$$(20) \quad \hat{\alpha}_{11}^{WLS} = \begin{bmatrix} B^{-1} \cdot (X_{11}' \cdot Q) + C \cdot (X_{12}' \cdot Q) \end{bmatrix} \cdot y$$

BV4.1-Trendschätzparameter (simultane Trend- und Saisonschätzung)

Gleichung (20) zeigt, dass zwischen den Parameterschätzern der Trend-Konjunkturkomponente ($\alpha_j, j=0,1,2,3$) und den Parameterschätzern der Saisonkomponente ($\beta_j, j=1, \dots, 6; \gamma_j, j=1, \dots, 5$) keine statistische Unabhängigkeit besteht, denn der WLS-Schätzer eines einfachen Trendmodells einer Polynomfunktion dritten Grades, $y_t = \alpha_0 + \alpha_1 \cdot t + \alpha_2 \cdot t^2 + \alpha_3 \cdot t^3 + u_t$ ($t=1, \dots, n$), hat folgendes Aussehen:

¹³ Siehe dazu zum Beispiel das ausführliche Kapitel zur Matrixalgebra in Rengers, M.: „Varianzanalyse – Ursachen und Folgen ungleicher Zellenbesetzungen und ihre Behandlung über verschiedene Lösungsansätze“, Aachen 2004, Seite 53 ff. (insbesondere Seite 67 ff.).

(21) $\hat{\alpha}^{WLS} = [X_{11}' \cdot Q \cdot X_{11}]^{-1} \cdot X_{11}' \cdot Q \cdot y$
Trendschätzparameter
(alleinige, isolierte Trendschätzung)

Schätzung der Parameter von Trend- und Saisonkomponente des Basismodells (3). Trotz dieser statistischen Abhängigkeit der Parameterschätzung von Trend- und Saisonkomponente besteht zwischen beiden Komponenten kein streng systematischer Zusammenhang im Sinne einer Proportionalität, da ein additives Zusammenwirken der Komponenten angenommen wird und nicht etwa ein multiplikatives.

Die Parameterschätzer eines einfachen Trend-Konjunkturmodells führen also zu anderen Ergebnissen als die kombinierte

Exkurs 1

Das allgemeine lineare Modell ALM und gewisse Formen von Nicht-Linearitäten

Das stochastische Grundmodell

(Ia) $y_i = \beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \dots + \beta_j \cdot x_{ij} + \dots + \beta_{h-1} \cdot x_{i,h-1} + \varepsilon_i, \quad i = 1, \dots, n$

ist aus der linearen multiplen Regressionsanalyse bekannt und ein Spezialfall des allgemeinen linearen Modells. Im sogenannten Modell I des allgemeinen linearen Modells werden die unabhängigen exogenen Variablen X_{ij} als nichtzufällig beziehungsweise deterministisch und die h unbekannt Parameter $\beta_0, \dots, \beta_{h-1}$ als fest und somit ebenfalls nichtzufällig angenommen. Die unbeobachtbaren Stör- oder Fehlervariablen $\varepsilon_i, i = 1, \dots, n$, werden dagegen als Zufallsvariablen aufgefasst. Aufgrund der Zufälligkeit der ε_i sind die beobachtbaren Werte y_i ebenfalls als Realisationen von Zufallsvariablen zu verstehen.

In Matrixschreibweise lässt sich (Ia) über

(Ib)
$$\underset{n \times 1}{y} = \underset{n \times h}{X} \cdot \underset{h \times 1}{\beta} + \underset{n \times 1}{\varepsilon}$$

ausdrücken, worin y ein n -dimensionaler Zufallsvektor der abhängigen, endogenen y -Variablen, X eine $[n \times h]$ -Matrix der unabhängigen, exogenen x -Variablen (Prädiktoren) und β der Vektor der h Parameter ist. Die Prädiktormatrix X enthält zusätzlich die exogene Variable x_{0i} , die für alle $i = 1, \dots, n$ gleich 1 gesetzt wird (Scheinvariable zur Modellierung eines Modells mit Konstante, Interzept).

Dem Modell (Ia) beziehungsweise (Ib) liegt also die Vorstellung zugrunde, dass sich die beobachtbaren Werte der abhängigen, exogenen Variablen y_i additiv aus einer *systematischen* Komponente $\beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \dots + \beta_j \cdot x_{ij} + \dots + \beta_{h-1} \cdot x_{i,h-1}$ und einer *unsystematischen* oder *zufälligen* Komponente $\varepsilon_i, i = 1, \dots, n$ zusammensetzen. Für die zufällige Komponente wird davon ausgegangen, dass der Erwartungswert der Störvariablen null beträgt, das Mittel der Fehler bei unendlicher Stichprobenziehung folglich null wird:

(IIa) $E(\varepsilon_i) = 0, \quad i = 1, \dots, n,$

beziehungsweise

(IIb) $E(\underline{\varepsilon}) = \underline{0}.$

Zusätzlich wird von nicht korrelierten Störvariablen ε_i mit konstanter Fehlervarianz σ^2 (Homoskedastizität) ausgegangen:

(IIIa) $Cov(\varepsilon_i, \varepsilon_{i^*}) = 0, \quad i, i^* = 1, \dots, n, \quad i \neq i^*,$

(IVa) $Var(\varepsilon_i) = \sigma^2, \quad i = 1, \dots, n.$

Die Annahmen (IIIa) und (IVa) lassen sich in Matrixdarstellung zu

(V) $Cov(\underline{\varepsilon}) = \sigma^2 \cdot I_n$

vereinfachen, wobei I_n eine Einheitsmatrix der Dimension $[n \times n]$ ist. Die Kovarianzmatrix der Störvariablen ist also eine Diagonalmatrix mit konstanter Fehlervarianz σ^2 auf der Hauptdiagonalen. Alle Störvariablen stammen somit annahmegemäß aus derselben Population mit der Varianz σ^2 . Außerhalb der Hauptdiagonalen von $\sigma^2 \cdot I_n$ sind die Matrixelemente null, die Störvariablen also unkorreliert.

Das Modell (Ia) beziehungsweise (Ib) ist *linear in den Parametern*; alle Parameter erscheinen in der ersten Potenz. Es ist ebenfalls *linear in den unabhängigen Variablen*. Durch geeignete nichtlineare Transformationen lassen sich auch gewisse Formen von Nichtlinearitäten wie zum Beispiel das Quadrat oder der Logarithmus einer beobachteten Größe als erklärende Variablen einsetzen und über ein allgemeines lineares Modell abbilden. Angenommen, die systematische Komponente zur Erklärung der abhängigen Variablen wäre ein Polynom vom Grad 3 der unabhängigen, nichtzufälligen x -Variablen, dann verändert sich das obige Modell (Ia) zu:

(VI) $y_i = \beta_0 + \beta_1 \cdot x_i + \beta_2 \cdot x_i^2 + \beta_3 \cdot x_i^3 + \varepsilon_i, \quad \forall i = 1, \dots, n.$

Modell (VI) ist weiterhin linear in den Parametern, aber nicht mehr linear in den unabhängigen Variablen. Durch die einfache Transformation

(VII) $z_{i1} = x_i, \quad i = 1, \dots, n,$

$z_{i2} = x_i^2, \quad i = 1, \dots, n,$

$z_{i3} = x_i^3, \quad i = 1, \dots, n,$

lässt sich (VI) jedoch wieder in ein Modell überführen, das linear in den transformierten unabhängigen Variablen z_{i1}, z_{i2}, z_{i3} ist.

Exkurs 2

Gewöhnliche Kleinste-Quadrate-Schätzung (OLS = ordinary least squares) der Parameter des ALM¹

Zur Schätzung der unbekanntem β -Parameter des Modells (Ia) beziehungsweise (Ib) (siehe Exkurs 1) wird die Methode der kleinsten Quadrate (Ordinary Least Squares – OLS) herangezogen.

Unter den im Exkurs 1 genannten Voraussetzungen (IIb) und (V) und der numerisch-technischen Annahme, dass die Matrix \mathbf{X} vollen Rang² besitzt, sind die mit der Methode der kleinsten Quadrate geschätzten Parameter erwartungstreu und haben von allen linearen Schätzern die kleinste Varianz (best linear unbiased estimators = BLUE).

Die Parameterschätzung erfolgt bei der gewöhnlichen Kleinste-Quadrate-Schätzung über die Lösung des folgenden Minimierungsproblems:

$$(VIII) \quad \sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \text{Min!},$$

wobei \hat{y}_t der mit den zu schätzenden Parametern $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{h-1}$ ermittelte Wert der systematischen Komponente ist:

$$(IX) \quad \hat{y}_i = f_i(\hat{\beta}_0, \dots, \hat{\beta}_{h-1}) = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_{1i} + \hat{\beta}_2 \cdot x_{2i} + \dots + \hat{\beta}_{h-1} \cdot x_{h-1,i}.$$

Der Minimierungsansatz (VIII) kann entsprechend über

$$(X) \quad \sum_{i=1}^n [y_t - f_i(\hat{\beta}_0, \dots, \hat{\beta}_{h-1})]^2 \rightarrow \text{Min!}$$

beziehungsweise

$$(XI) \quad \sum_{i=1}^n (y_t - \hat{\beta}_0 - \hat{\beta}_1 \cdot x_{i1} - \hat{\beta}_2 \cdot x_{i2} - \dots - \hat{\beta}_{h-1} \cdot x_{i,h-1})^2 \rightarrow \text{Min!}$$

formuliert werden. Die $x_{i1}, x_{i2}, \dots, x_{i,h-1}$ sind in diesem Minimierungsproblem (XI) konstante Größen, während $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{h-1}$ die Veränderlichen sind.

Um das Minimierungsproblem zu lösen, müssen die partiellen Ableitungen nach den Veränderlichen gleich Null gesetzt werden:

$$(XII) \quad \begin{array}{l} \frac{\partial f(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{h-1})}{\partial \hat{\beta}_0} = 0 \\ \vdots \\ \frac{\partial f(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{h-1})}{\partial \hat{\beta}_{h-1}} = 0. \end{array}$$

Zur Bestimmung der Parameterschätzer $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{h-1}$ ist sodann das aus diesem Vorgang resultierende lineare Gleichungssystem, das aus h sogenannten Normalgleichungen besteht, zu lösen.

Es lässt sich zeigen, dass dieses Normalgleichungssystem in Matrixschreibweise durch

$$(XIII) \quad (\mathbf{X}' \cdot \mathbf{X}) \cdot \hat{\boldsymbol{\beta}} = \mathbf{X}' \cdot \mathbf{y}$$

dargestellt werden kann.

Durch Prämultiplikation beider Seiten von (XIII) mit der Inversen von $(\mathbf{X}' \cdot \mathbf{X})$, also mit $(\mathbf{X}' \cdot \mathbf{X})^{-1}$, erhält man die gesuchte Lösung des Minimierungsproblems:

$$(XIV) \quad \hat{\boldsymbol{\beta}}^{\text{OLS}} = (\mathbf{X}' \cdot \mathbf{X})^{-1} \cdot \mathbf{X}' \cdot \mathbf{y}.$$

Die Inverse $(\mathbf{X}' \cdot \mathbf{X})^{-1}$ existiert jedoch nur dann, wenn die $[h \times h]$ -Matrix $(\mathbf{X}' \cdot \mathbf{X})$ vollen Rang besitzt. Es muss

$$\text{rg}(\mathbf{X}' \cdot \mathbf{X}) = h$$

gelten.

Eine notwendige (allerdings nicht hinreichende) Bedingung für die Existenz dieser Inversen ist, dass die Anzahl n der Beobachtungen die Anzahl h der exogenen Variablen (einschließlich der Scheinvariablen) übersteigt, denn der Rang der Matrix $(\mathbf{X}' \cdot \mathbf{X})$ stimmt mit dem Rang von \mathbf{X} überein:

$$\text{rg}(\mathbf{X}' \cdot \mathbf{X}) = \text{rg}(\mathbf{X}) = h.$$

1 Ausführliche Informationen zum allgemeinen linearen Modell und zur Schätzung seiner Parameter finden sich in Rengers, M.: „Varianzanalyse – Ursachen und Folgen ungleicher Zellenbesetzungen und ihre Behandlung über verschiedene Lösungsansätze“, Aachen 2004, Seite 86 ff., oder in anderen einschlägigen Lehrbüchern.

2 Eine $[n \times h]$ -Matrix \mathbf{X} besitzt vollen Rang, wenn ihr Rang dem Minimum aus Spalten- und Zeilenanzahl entspricht: $\text{rg}(\mathbf{X}) = \min\{n, h\}$. Ist die Spaltenanzahl h kleiner als die Zeilenanzahl n ($h < n$), dann gibt der Rang die Anzahl der linear unabhängigen Spaltenvektoren an. Generell kann der Rang einer Matrix von 0 (für jede Nullmatrix) bis h (für eine Matrix mit vollem Rang) betragen. Übersteigt andererseits die Spaltenanzahl h die Zeilenanzahl n ($h > n$), dann gibt der Rang der Matrix \mathbf{X} die Anzahl der linear unabhängigen Zeilen an.

Exkurs 3

Gewichtete Kleinste-Quadrate-Schätzung (WLS = weighted least squares) der Parameter des ALM

Die Parameter des Modells (Ia) beziehungsweise (Ib) aus Exkurs 1 können auch über einen gewogenen Kleinste-Quadrate-Ansatz geschätzt werden. Der gewogenen Kleinste-Quadrate-Schätzung WLS liegt dabei der Minimierungsansatz

$$(XV) \quad \sum_{i=1}^n w_i \cdot (y_i - \hat{y}_i)^2 \rightarrow \text{Min!}$$

beziehungsweise

$$(XVI) \quad \sum_{i=1}^n w_i \cdot [y_i - f_i(\hat{\beta}_0, \dots, \hat{\beta}_{h-1})]^2 \rightarrow \text{Min!}$$

zugrunde, wobei w_i unterschiedliche Gewichte sind, die dazu führen, dass die quadrierten Abweichungen $(y_i - \hat{y}_i)^2$ im Gegensatz zum gewöhnlichen Kleinste-Quadrate-Ansatz OLS (siehe Exkurs 2) nicht für alle $i = 1, \dots, n$ Beobachtungen gleich gewichtet werden. Mit den Gewichten w_i kann der Einfluss einzelner Abweichungsquadrate $(y_i - \hat{y}_i)^2$ auf die gesamte Parameterschätzung reduziert oder erhöht werden.

Wie die nachfolgenden Ausführungen zeigen, kann der gewöhnliche Kleinste-Quadrate-Schätzer OLS als Lösung dieses gewichteten Minimierungsproblems herangezogen werden, wenn das ursprüngliche Modell $y_i = \beta_0 + \beta_1 \cdot x_{1i} + \beta_2 \cdot x_{2i} + \dots + \beta_{h-1} \cdot x_{h-1,i} + \varepsilon_i$ in geeigneter Weise transformiert wird.¹

Eine Multiplikation des gesamten Modells (Ia) mit der Quadratwurzel der Gewichte w_i führt zu transformierten Werten der Form:

$$(XVII) \quad y_i^* = \sqrt{w_i} \cdot y_i$$

beziehungsweise

$$(XVIII) \quad x_{ji}^* = \sqrt{w_i} \cdot x_{ji} \quad (\text{für alle } j=1, \dots, h-1 \text{ exogenen Variablen}).$$

Das transformierte Modell lautet:

$$\begin{aligned} y_i^* &= \beta_0 \cdot \sqrt{w_i} + \beta_1 \cdot x_{1i}^* + \dots + \beta_{h-1} \cdot x_{h-1,i}^* + \varepsilon_i^* \\ &= \beta_0 \cdot \sqrt{w_i} + \beta_1 \cdot \sqrt{w_i} \cdot x_{1i} + \dots + \beta_{h-1} \cdot \sqrt{w_i} \cdot x_{h-1,i} + \sqrt{w_i} \cdot \varepsilon_i \\ &= \sqrt{w_i} \cdot (\beta_0 + \beta_1 \cdot x_{1i} + \dots + \beta_{h-1} \cdot x_{h-1,i} + \varepsilon_i), \end{aligned}$$

und seine Schätzwerte \hat{y}_i^* sind entsprechend über

$$\begin{aligned} \hat{y}_i^* &= \hat{\beta}_0 \cdot \sqrt{w_i} + \hat{\beta}_1 \cdot x_{1i}^* + \dots + \hat{\beta}_{h-1} \cdot x_{h-1,i}^* \\ &= \hat{\beta}_0 \cdot \sqrt{w_i} + \hat{\beta}_1 \cdot \sqrt{w_i} \cdot x_{1i} + \dots + \hat{\beta}_{h-1} \cdot \sqrt{w_i} \cdot x_{h-1,i} \\ &= \sqrt{w_i} \cdot (\hat{\beta}_0 + \hat{\beta}_1 \cdot x_{1i} + \dots + \hat{\beta}_{h-1} \cdot x_{h-1,i}) \\ &= \sqrt{w_i} \cdot f(\hat{\beta}_0, \dots, \hat{\beta}_{h-1}) \end{aligned}$$

zu ermitteln, also:

$$(XIX) \quad \hat{y}_i^* = \sqrt{w_i} \cdot f(\hat{\beta}_0, \dots, \hat{\beta}_{h-1}) = \sqrt{w_i} \cdot \hat{y}_i.$$

Das Minimierungsproblem der gewichteten Kleinste-Quadrate-Schätzung WLS des Modells (Ia) entspricht dem Minimierungsproblem der gewöhnlichen Kleinste-Quadrate-Schätzung OLS des transformierten Modells (XVII). Es gilt:

$$(XX) \quad \underbrace{\sum_{i=1}^n w_i \cdot (y_i - \hat{y}_i)^2 \rightarrow \text{Min!}}_{\substack{\text{WLS-Minimierungsansatz} \\ \text{des} \\ \text{ursprünglichen Modells}}} = \underbrace{\sum_{i=1}^n (y_i^* - \hat{y}_i^*)^2 \rightarrow \text{Min!}}_{\substack{\text{OLS-Minimierungsansatz} \\ \text{des} \\ \text{transformierten Modells}}}.$$

Herleitung:

$$\begin{aligned} \sum_{i=1}^n w_i [y_i - \hat{y}_i]^2 &= \sum_{i=1}^n w_i [y_i^2 - 2 \cdot y_i \cdot \hat{y}_i + \hat{y}_i^2] \quad \langle \text{nach der 2. binomischen Formel} \rangle \\ &= \sum_{i=1}^n [w_i \cdot y_i^2 - 2 \cdot \sqrt{w_i} \cdot y_i \cdot \sqrt{w_i} \cdot \hat{y}_i + w_i \cdot \hat{y}_i^2] \quad \langle \text{nach Ausmultiplikation} \rangle \\ &= \sum_{i=1}^n [\sqrt{w_i} \cdot y_i - \sqrt{w_i} \cdot \hat{y}_i]^2 \quad \langle \text{nach Rückanwendung der binomischen Formel} \rangle \\ &= \sum_{i=1}^n (y_i^* - \hat{y}_i^*)^2 \quad \langle \text{nach Anwendung von (XVII) und (XIX)} \rangle \end{aligned}$$

In der Matrixdarstellung bedeutet die oben erwähnte Transformation des Modells (Ia) nichts anderes als die Prämultiplikation des Modells $y = X \cdot \beta + \varepsilon$ mit einer Gewichtungs- beziehungsweise Transformationsmatrix W , bei der es sich um eine Diagonalmatrix handelt, deren Hauptdiagonalelemente die Quadratwurzeln der Gewichte w_i sind:

¹ Siehe dazu auch Carroll, R.J./Ruppert, D.: "Transformation and Weighting in Regression", London, New York 1988, Seite 11.

noch: Exkurs 3

$$(XXI) \quad \mathbf{W} = \begin{bmatrix} \sqrt{w_1} & 0 & \dots & 0 \\ 0 & \sqrt{w_2} & & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{w_n} \end{bmatrix}$$

Das transformierte Modell lautet in Matrixschreibweise:

$$\mathbf{W} \cdot \mathbf{y} = \mathbf{W} \cdot \mathbf{X} \cdot \boldsymbol{\beta} + \mathbf{W} \cdot \mathbf{u}$$

beziehungsweise

$$\mathbf{y}^* = \mathbf{X}^* \cdot \boldsymbol{\beta} + \mathbf{u}^*$$

Die Parameter des transformierten Modells können mit dem gewöhnlichen Kleinste-Quadrate-Ansatz OLS geschätzt werden und sind gleichzeitig die gewichteten Kleinste-Quadrate-Schätzer (im Folgenden mit $\hat{\boldsymbol{\beta}}^{WLS}$ gekennzeichnet) des ursprünglichen (nicht transformierten) Modells. Es gilt:

$$\hat{\boldsymbol{\beta}}^{OLS} = (\mathbf{X}^* \cdot \mathbf{X}^*)^{-1} \cdot \mathbf{X}^* \cdot \mathbf{y}^* \quad (\text{OLS-Schätzer des transformierten Modells})$$

und nach Retransformation von $\mathbf{X}^* = \mathbf{W} \cdot \mathbf{X}$ und $\mathbf{y}^* = \mathbf{W} \cdot \mathbf{y}$

$$(XXII) \quad \hat{\boldsymbol{\beta}}^{WLS} = [(\mathbf{W} \cdot \mathbf{X})' \cdot (\mathbf{W} \cdot \mathbf{X})]^{-1} \cdot (\mathbf{W} \cdot \mathbf{X})' \cdot \mathbf{W} \cdot \mathbf{y} \quad (\text{WLS-Schätzer des ursprünglichen Modells})$$

Es sei

$$\mathbf{W}' \cdot \mathbf{W} = \mathbf{Q} = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & w_n \end{bmatrix}$$

dann lässt sich (XXII) wie folgt vereinfachen:

$$(XXIII) \quad \hat{\boldsymbol{\beta}}^{WLS} = [\underbrace{\mathbf{X}' \cdot \mathbf{W}' \cdot \mathbf{W} \cdot \mathbf{X}}_{\mathbf{Q}}]^{-1} \cdot \underbrace{\mathbf{X}' \cdot \mathbf{W}' \cdot \mathbf{W} \cdot \mathbf{y}}_{\mathbf{Q} \cdot \mathbf{y}}$$

$$= [\mathbf{X}' \cdot \mathbf{Q} \cdot \mathbf{X}]^{-1} \cdot \mathbf{X}' \cdot \mathbf{Q} \cdot \mathbf{y}$$

2.2 Die geschätzten Parameter und die zugehörigen Filter zur Trendschätzung

Wie bereits erwähnt verwendet BV4.1 für die Trendschätzung in der Regel einen Stützbereich von 27 Monaten. Bei der Wahl der Stützbereichslänge war entscheidend, dass diese einerseits lang genug ist, um den Trendverlauf nicht von individuellen Unregelmäßigkeiten beeinflussen zu lassen, aber andererseits gleichzeitig kurz genug, um auch noch kürzere konjunkturelle Entwicklungen (zum Beispiel in zweijährigen Zyklen) abzubilden. Für einen gesamten Stützbereich U_1 im Zeitreihenintervall $U_1 = [t_1, \dots, t_k]$ der Länge $n = 27$ ließe sich nun unter Verwendung der WLS-Trendparameterschätzer aus (20) über

$$(22) \quad \hat{\mathbf{m}}_{27 \times 1} = \mathbf{X}_{11} \cdot \hat{\boldsymbol{\alpha}}_{4 \times 1}^{WLS}$$

eine Trendfunktion mit 27 Trendwerten schätzen. Nach dem Prinzip der gleitenden Schätzung kann für einen unmittelbar nachfolgenden zweiten Stützbereich U_2 , der ebenfalls eine Länge von 27 Monaten hat, erneut eine Parameterschätzung gemäß (20) vorgenommen werden. Dieser zweite Stützbereich U_2 deckt das Zeitintervall $U_2 = [t_2, \dots, t_k]$ ab. Auch hier ließen sich sodann über (22) wieder 27 Trendschätzwerte für dieses zweite Intervall ermitteln. Bei einer Zeitreihe mit der Gesamtlänge N kann dieses Prinzip – am Anfang der Reihe beginnend, Monat für Monat zu aktuelleren Daten hin versetzt – bis zum Stütz- beziehungsweise Schätzbereich $U_{N-27+1} = [t_{N-27+1}, \dots, t_N]$ bei jeweils gleicher Stützbereichslänge

von $n = 27$ Monaten fortgesetzt werden. Für ein und denselben Zeitpunkt t^* würde man auf diese Weise, entsprechend der Anzahl der Stützbereiche U_l , $l = 1, \dots, N - 27 + 1$, die den Zeitpunkt t^* überdecken, *mehrere* Trendschätzwerte \hat{m}_{t^*} erhalten.

BV4.1 verwendet jedoch nicht alle 27 Trendschätzwerte eines Stützbereiches U_l . Stattdessen wird von jeder gleitenden Trendschätzung mit Stützbereichslänge 27 nur *ein einziger* Trendschätzwert benutzt. Es handelt sich dabei um denjenigen Trendschätzwert, der genau in der Mitte des gesamten Stützbereiches liegt, das heißt von den in Gleichung (22) dargestellten 27 Trendschätzwerten eines Stützbereiches wird lediglich der zentrale 14. Wert verwendet.¹⁴ Dieser zentrale Trendschätzwert eines Stützbereiches (14. Wert) wird somit über den WLS-Trendparameterschätzer aus (20) und den 14. Zeilenvektor der Teilmatrix \mathbf{X}_{11} der Regressoren der Trendkomponente ermittelt. Formal lässt sich die Reduzierung der 27 Trendschätzwerte aus (22) zu einem einzigen zentralen Trendschätzwert eines Stützbereiches über

$$(23) \quad \hat{m}_{t=14} = \mathbf{x}_{(14)}' \cdot \hat{\boldsymbol{\alpha}}_{11}^{WLS}$$

**zentraler Trendschätzwert:
Berechnung über Parameterschätzer**

14 Zur Begründung dieser Vorgehensweise siehe Speth, H.-T., Fußnote 11, hier: Seite 6: „Ergebnisse der Spektralanalyse zeigen [...], dass die Schätzwerte \hat{m}_t und \hat{s}_t nur dann keine zeitlichen Vor- oder Nachläufe – sogenannte Phasenverschiebungen – gegenüber den entsprechenden Komponenten in den Zeitreihen aufweisen, wenn sie auf Schätzbereichen beruhen, bei denen der Zeitpunkt t exakt in der Mitte liegt.“

ausdrücken, wenn für X_{11} die folgende Partitionierung gilt:

$$(24) \quad X_{11} = \begin{matrix} \begin{matrix} x_{(1)}' \\ \vdots \\ x_{(14)}' \\ \vdots \\ x_{(27)}' \end{matrix} \\ [27 \times 4] \end{matrix} = \begin{bmatrix} 1 & 1 & 1^2 & 1^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 14 & 14^2 & 14^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 27 & 27^2 & 27^3 \end{bmatrix}$$

Die Berechnung des zentralen Trendschätzwertes $\hat{m}_{t=14}$ aus Gleichung (23) lässt sich in Verbindung mit (20) als Funktion beziehungsweise Linearkombination der Zeitreihenwerte y_t darstellen und als *Filtergewicht* oder kurz *Filter* deuten:¹⁵

(25)

$$\hat{m}_{t=14} = x_{(14)}' \cdot \hat{\alpha}_{11}^{WLS} = \begin{bmatrix} 1 & 14 & 14^2 & 14^3 \end{bmatrix} \cdot \underbrace{\left[B^{-1} \cdot (X_{11}' \cdot Q) + C \cdot (X_{12}' \cdot Q) \right]}_{\text{Filtergewichtungsvektor } f_{14}'} \cdot y$$

mit

$$B = X_{11}' \cdot Q \cdot X_{11} - (X_{11}' \cdot Q \cdot X_{12}) \cdot [X_{12}' \cdot Q \cdot X_{12}]^{-1} \cdot (X_{12}' \cdot Q \cdot X_{11})$$

$$C = -B^{-1} \cdot (X_{11}' \cdot Q \cdot X_{12}) \cdot [X_{12}' \cdot Q \cdot X_{12}]^{-1},$$

also:

$$\hat{m}_{t=14} = f_{14}' \cdot y$$

[1x27] [27x1]

**zentraler Trendschätzwert:
Berechnung über Filtergewichtungsvektor**

Während die geschätzten α_j -Parameter der Trendkomponente von den zugrunde liegenden Zeitreihenwerten y_t abhängig sind, sich also der Vektor $\hat{\alpha}_{11}^{WLS}$ mit jeder neuen Zeitreihe beziehungsweise jedem neuen Stützbereich verändert, gilt dies nicht für den sogenannten *Filtergewichtungsvektor* f_{14}' der Gleichung (25). Dieser Filtergewichtungsvektor ist unabhängig von wechselnden Zeitreihenwerten y_t und bleibt für sämtliche gleitenden Stützbereiche immer unverändert, solange Länge des Stützbereiches, Basisfunktion, Gewichtung der Abweichungsquadrate und Schätzstelle gleich bleiben, das heißt die Filter sind *zeitinvariant*.

Der für die Berechnung des zentralen Trendschätzwertes (Stützbereich $n=27$; Schätzstelle $t=14$) benötigte Filtergewichtungsvektor f_{14}' wird auch *symmetrischer Mittelfilter* genannt.

Im Folgenden wird zur Veranschaulichung gezeigt, welche konkreten Werte einzelne Elemente des Filtergewichtungsvektors f_{14}' aus Gleichung (25) annehmen. Aus allen bisherigen Angaben und Beschreibungen können sie jedoch auch selbstständig nachgebaut werden.

Da die Matrizen X_{11} und X_{12} aus (12) und (13) bereits bekannt sind, lässt sich das Aussehen der Produkte $X_{11}' \cdot Q$ beziehungsweise

$X_{12}' \cdot Q$ recht einfach ermitteln, denn es handelt sich bei Q um die Diagonalmatrix der 27 Gewichte w_i .

Allgemein formuliert führt die Postmultiplikation einer beliebigen $[p \times q]$ -Matrix A mit einer $[q \times q]$ -Diagonalmatrix D , deren Diagonalelemente d_1, d_2, \dots, d_q sind, dazu, dass die Elemente der ersten Spalte der Matrix A mit d_1 multipliziert, die Elemente der zweiten Spalte der Matrix A mit d_2 multipliziert (und so weiter) und schließlich die Elemente der letzten Spalte q der Matrix A mit d_q multipliziert werden. Für die Matrizen $X_{11}' \cdot Q \cdot X_{11}$ und $X_{12}' \cdot Q \cdot X_{11}$ gilt:

(26)

$$\underbrace{X_{11}' \cdot Q \cdot X_{11}}_{[4 \times 4]} = \begin{bmatrix} \sum_{i=1}^n w_i \cdot 1 & \sum_{i=1}^n w_i \cdot t_i & \sum_{i=1}^n w_i \cdot t_i^2 & \sum_{i=1}^n w_i \cdot t_i^3 \\ \sum_{i=1}^n w_i \cdot t_i & \sum_{i=1}^n w_i \cdot t_i^2 & \sum_{i=1}^n w_i \cdot t_i^3 & \sum_{i=1}^n w_i \cdot t_i^4 \\ \sum_{i=1}^n w_i \cdot t_i^2 & \sum_{i=1}^n w_i \cdot t_i^3 & \sum_{i=1}^n w_i \cdot t_i^4 & \sum_{i=1}^n w_i \cdot t_i^5 \\ \sum_{i=1}^n w_i \cdot t_i^3 & \sum_{i=1}^n w_i \cdot t_i^4 & \sum_{i=1}^n w_i \cdot t_i^5 & \sum_{i=1}^n w_i \cdot t_i^6 \end{bmatrix}$$

$$= \begin{bmatrix} 14 & 196 & 3.199 & 57.526 \\ 196 & 3.199 & 57.526 & 1.108.303 \\ 3.199 & 57.526 & 1.108.303 & 22.492.666 \\ 57.526 & 1.108.303 & 22.492.666 & 475.351.579 \end{bmatrix} \text{ und}$$

(27)

$$\underbrace{X_{12}' \cdot Q \cdot X_{11}}_{[11 \times 4]}$$

$$= \begin{bmatrix} 0,13329 & 6,60252 & 153,89481 & 4.229,33384 \\ -0,10714 & -0,32143 & 2,53571 & 213,53571 \\ -0,14286 & -2,00000 & -41,71429 & -968,00000 \\ -0,03571 & 0,03571 & 6,03571 & 237,89286 \\ 0,00957 & 0,54034 & 11,24805 & 256,38044 \\ = & 0,00000 & 0,00000 & -7,00000 & -294,00000 \\ 0,23086 & 0,49744 & -39,72325 & -2.067,71818 \\ 0,18558 & 3,27852 & 71,81825 & 1.720,23575 \\ 0,00000 & -0,14286 & -4,00000 & -125,28571 \\ -0,06186 & -1,17532 & -24,18685 & -554,19440 \\ -0,01658 & 0,00256 & 6,79468 & 281,93246 \end{bmatrix}$$

Die $[11 \times 11]$ -Inverse $[X_{12}' \cdot Q \cdot X_{12}]^{-1}$ ist eine symmetrische Matrix mit den in Formel (28) auf der Seite 748 dargestellten Elementen.

Die $[4 \times 4]$ -Inverse B^{-1} lautet:

(29)

$$B^{-1} = \begin{bmatrix} 5,55917 & -1,39595 & 0,10160 & -0,00223 \\ -1,39595 & 0,38808 & -0,02988 & 0,00068 \\ 0,10160 & -0,02988 & 0,00239 & -0,00006 \\ -0,00223 & 0,00068 & -0,00006 & 0,00000 \end{bmatrix}$$

¹⁵ Solche Linearkombinationen von Zeitreihenwerten y_t werden im Zusammenhang mit der Komponentenzerlegung auch als *lineare Filter* bezeichnet.

Tabelle 1 BV4.1-Filtergewichte zur Schätzung der Trend-Konjunktur-Komponente bei Monatsreihen für die zeitlich aktuellsten 14 Reihenwerte bei 30 Monatswerten: symmetrischer Mittelfilter und asymmetrische Randfilter

Teil 1: Filter M_t (Schätzbereich, Schätzstelle, Polynomgrad) beziehungsweise M_t (Schätzbereich, Schätzstelle, Stelle mit maximalem Gewicht, Polynomgrad); bei zwei Filtern: gewogenes Mittel der Filter							
	f_1	f_2	f_3	f_4	f_5	f_6	f_7
i	$M_t(27, 27, 3)$ $M_t(20, 20, 20, 1)$	$M_{t-1}(26, 35, 3)$ $M_{t-1}(21, 20, 21, 1)$	$M_{t-2}(25, 23, 3)$ $M_{t-2}(22, 20, 22, 1)$	$M_{t-3}(25, 22, 3)$ $M_{t-3}(23, 20, 23, 1)$	$M_{t-4}(25, 21, 3)$ $M_{t-4}(24, 20, 24, 1)$	$M_{t-5}(25, 20, 3)$ $M_{t-5}(25, 20, 25, 1)$	$M_{t-6}(26, 20, 3)$
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0
4	-0,01649	0	0	0	0	0	0
5	-0,02865	-0,01665	0	0	0	0	0,01612
6	-0,04086	-0,03038	-0,01284	0,00026	0,01121	0,01881	0,02417
7	0,01106	0,00311	0,00154	0,01061	0,01739	0,02159	0,01350
8	0,02333	0,01192	0,00568	0,00998	0,01326	0,01431	0,00143
9	0,03471	0,02083	0,00842	0,00684	0,00463	0,00124	-0,01306
10	0,04337	0,02446	0,01011	0,00183	-0,00632	-0,01425	-0,02739
11	0,04063	0,02491	0,00968	-0,00456	-0,00632	-0,02929	-0,03929
12	0,03225	0,02062	0,00616	-0,01192	-0,00632	-0,04132	-0,04666
13	0,01655	0,01015	-0,00140	-0,01992	-0,00632	-0,04801	-0,04351
14	-0,00808	-0,00792	-0,01393	-0,02826	-0,03982	-0,04337	-0,03198
15	-0,04318	-0,03494	-0,03231	-0,03667	-0,03529	-0,02930	-0,01143
16	-0,08462	-0,07228	-0,05743	-0,04190	-0,02449	-0,00512	0,01832
17	-0,14075	-0,11555	-0,08636	-0,04599	-0,00707	0,02934	0,04949
18	-0,21161	-0,16627	-0,11614	-0,04888	0,01215	0,06490	0,09169
19	0,07227	0,08022	0,08179	0,07272	0,06594	0,06175	0,06984
20	0,06000	0,07141	0,07765	0,07336	0,07007	0,06902	0,08191
21	0,04862	0,06251	0,07491	0,07650	0,07870	0,08209	0,09639
22	0,03996	0,05887	0,07323	0,08151	0,08966	0,09758	0,11073
23	0,04270	0,05843	0,07365	0,08789	0,10111	0,11262	0,12262
24	0,05108	0,06271	0,07717	0,09525	0,11147	0,12466	0,12999
25	0,06678	0,07319	0,08474	0,10325	0,11927	0,13134	0,12684
26	0,09142	0,09125	0,09726	0,11159	0,12316	0,12671	0,11531
27	0,12652	0,11828	0,11564	0,12000	0,11863	0,11263	0,09476
28	0,18444	0,15561	0,14077	0,12523	0,10782	0,08845	0,06501
29	0,25273	0,21553	0,16969	0,12933	0,09040	0,05400	0,01773
30	0,33580	0,27998	0,21231	0,13196	0,05997	-0,00037	-0,03252

Teil 2: Filter M_t (Schätzbereich, Schätzstelle, Polynomgrad)							
	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}
i	$M_{t-7}(28, 21, 3)$	$M_{t-8}(29, 21, 3)$	$M_{t-9}(30, 21, 3)$	$M_{t-10}(30, 20, 3)$	$M_{t-11}(29, 18, 3)$	$M_{t-12}(28, 16, 3)$	$M_{t-13}(27, 14, 3)$
1	0	0	0,00506	0,00290	0	0	0
2	0	0,00715	0,00621	0,00315	-0,00010	0	0
3	0,00953	0,00942	0,00526	0,00214	-0,00092	-0,00427	0
4	0,01317	0,00901	0,00360	0,00094	-0,00172	-0,00677	-0,01059
5	0,01360	0,00770	0,00234	0,00032	-0,00200	-0,00776	-0,01609
6	0,01299	0,00687	0,00231	0,00087	-0,00147	-0,00757	-0,01549
7	-0,00444	-0,01235	-0,01681	-0,01672	-0,01487	-0,01162	-0,01175
8	-0,01555	-0,02128	-0,02303	-0,02085	-0,01376	-0,00675	-0,00196
9	-0,02585	-0,02854	-0,02716	-0,01953	-0,00996	0,00074	0,01088
10	-0,03395	-0,03309	-0,02500	-0,01501	-0,00324	0,01070	0,02577
11	-0,03861	-0,03039	-0,01898	-0,00702	0,00647	0,02284	0,04167
12	-0,03491	-0,02289	-0,00886	0,00450	0,01905	0,03671	0,05757
13	-0,02537	-0,01032	0,00323	0,01812	0,03424	0,05178	0,07245
14	-0,00965	0,00429	0,02085	0,03604	0,05162	0,06733	0,08530
15	0,00818	0,02570	0,04320	0,05738	0,07097	0,08534	0,09509
16	0,03417	0,05291	0,06927	0,08107	0,09110	0,10126	0,10941
17	0,06708	0,08463	0,09784	0,10589	0,11071	0,11392	0,11551
18	0,10532	0,11936	0,12757	0,13049	0,12842	0,12216	0,10941
19	0,08778	0,09569	0,10014	0,10006	0,09820	0,09495	0,09509
20	0,09889	0,10462	0,10636	0,10418	0,09709	0,09009	0,08530
21	0,10918	0,11188	0,11049	0,10287	0,09329	0,08259	0,07245
22	0,11729	0,11643	0,10833	0,09834	0,08657	0,07263	0,05757
23	0,12194	0,11372	0,10231	0,09035	0,07687	0,06050	0,04167
24	0,11824	0,10623	0,09219	0,07883	0,06428	0,04662	0,02577
25	0,10870	0,09365	0,07505	0,06231	0,04910	0,03156	0,01088
26	0,09298	0,07189	0,05627	0,04414	0,03181	0,01600	-0,00196
27	0,06562	0,04821	0,03488	0,02381	0,01329	0,00226	-0,01175
28	0,03599	0,02142	0,01047	0,00133	-0,00605	-0,01115	-0,01549
29	0,00265	-0,00900	-0,01685	-0,02288	-0,02538	-0,02283	-0,01609
30	-0,03497	-0,04290	-0,04654	-0,04803	-0,04361	-0,03126	-0,01059

$$(28) \begin{matrix} 0,14406 & 0,00007 & 0,00137 & 0,00130 & 0,00040 & -0,00006 & -0,00183 & -0,00227 & -0,00124 & 0,00011 & 0,00055 \\ & 0,14336 & -0,00137 & 0,00111 & 0,00130 & 0,00032 & 0,00227 & 0,00058 & -0,00214 & -0,00180 & -0,00010 \\ & & 0,14311 & -0,00136 & 0,00135 & 0,00135 & 0,00238 & 0,00237 & 0,00000 & -0,00235 & -0,00234 \\ & & & 0,14337 & 0,00006 & 0,00103 & 0,00011 & 0,00180 & 0,00214 & -0,00055 & -0,00224 \\ & & & & 0,14409 & -0,00129 & -0,00055 & -0,00010 & 0,00124 & 0,00224 & 0,00179 \\ & & & & & 0,07153 & -0,00010 & -0,00055 & 0,00000 & 0,00179 & 0,00224 \\ & & & & & & 0,14195 & -0,00270 & 0,00071 & 0,00142 & 0,00024 \\ & & & & & & & 0,14269 & -0,00124 & 0,00097 & 0,00141 \\ & & & & & & & & 0,14297 & -0,00124 & 0,00071 \\ & & & & & & & & & 0,14273 & -0,00265 \\ & & & & & & & & & & 0,14202 \end{matrix}$$

Der Filtergewichtungsvektor f_{14}' der Gleichung (25) konkretisiert sich somit zu:

$$(30) \quad f_{14}' = \begin{bmatrix} -0,01059 & -0,01609 & -0,01549 & -0,01175 \\ -0,00196 & 0,01088 & 0,02577 & 0,04167 \\ 0,05757 & 0,07245 & 0,08530 & 0,09509 \\ 0,10941 & 0,11551 & 0,10941 & 0,09509 \\ 0,08530 & 0,07245 & 0,05757 & 0,04167 \\ 0,02577 & 0,01088 & -0,00196 & -0,01175 \\ -0,01549 & -0,01609 & -0,01059 & \end{bmatrix}$$

*symmetrischer
Mittelfilter*

Das bisher vorgestellte Grundprinzip der gleitenden Schätzung mit zentraler Schätzstelle kann nicht weiter fortgesetzt werden, sobald der Stützbereich das Ende der Zeitreihe erreicht hat. Da der Stützbereich bei dem symmetrischen Mittelfilter eine Länge von $n=27$ Zeitreihenwerten hat, müssen dem für einen bestimmten Monat zu schätzenden Trendwert zeitlich noch 13 weitere Monatsergebnisse folgen. Trendschätzwerte für die letzten 13 Monate am aktuellen Rand einer Zeitreihe können daher nicht über den symmetrischen Mittelfilter geschätzt werden. Stattdessen kommen hier sogenannte *asymmetrische Randfilter* zum Einsatz.¹⁶

Zur Ermittlung der asymmetrischen Randfilter kommt weiterhin das Basismodell zum Einsatz, es verändern sich jedoch die Länge des Stützbereiches und die Schätzstelle, die nun nicht mehr in der Mitte des Stützbereiches liegt. Für die letzten sechs Zeitpunkte wird darüber hinaus ergänzend zur Trendkomponente des Basismodells eine weitere Modellierung der Trendkomponente vorgenommen. Der Trendschätzwert wird dann über ein gewogenes Mittel der Einzelschätzwerte beider Modellvarianten bestimmt. Während die Trendkomponente im Basismodell des ersten Einzelschätzwertes weiterhin ein Polynom dritten Grades ist, wird die Trendkomponente bei der Berechnung des zweiten

Einzelschätzwertes durch eine Gerade modelliert. Diese Vorgehensweise wurde gewählt, um die Stabilität der vorläufigen Schätzergebnisse am aktuellen Rand zu erhöhen, sodass der Revisionsbedarf – im Vergleich zur ausschließlichen Modellierung der Trendkomponente über ein Polynom dritten Grades – durch neuere Schätzungen nach Hinzutreten weiterer Beobachtungswerte am aktuellen Rand verringert werden konnte.¹⁷

Tabelle 1 auf Seite 747 zeigt die Werte der vorläufigen 13 asymmetrischen Randfiltervektoren f_1 bis f_{13} und des endgültigen symmetrischen Mittelfilters f_{14} . Dabei liegen 30 Zeitreihenwerte vor, die den Index i bekommen, beginnend mit dem am längsten zurückliegenden Wert. Mithilfe dieser Filtergewichte lässt sich bei Zeitreihen, für die keine Extremwerte identifiziert werden, einfach durch Linearkombination mit den Ursprungswerten eine Trend-Konjunktur-Komponente schätzen. Nach 13 Monaten verändert sich der geschätzte Trendwert endgültig nicht mehr. Der Revisionsbedarf ist jedoch in den ersten Monaten am höchsten und wird immer geringer.

2.3 Ein Beispiel zur Berechnung der Trend-Konjunktur-Komponente

In Tabelle 2 sind beispielhaft 30 Zeitreihenwerte der *erwerbslosen Frauen im Alter von 25 bis 74 Jahren* für die Monate von Januar 2009 bis Juni 2011 dargestellt.

Tabelle 2 Ursprungswerte y_t der erwerbslosen Frauen im Alter von 25 bis 74 Jahren

	2009	2010	2011
Januar	1 211 610	1 201 294	1 039 577
Februar	1 312 723	1 202 657	1 008 979
März	1 185 598	1 112 101	965 730
April	1 175 165	1 059 255	926 410
Mai	1 128 510	1 083 201	908 887
Juni	1 169 307	1 014 391	951 246
Juli	1 191 249	989 586	–
August	1 226 839	955 557	–
September	1 113 627	1 005 996	–
Oktober	1 106 628	1 046 145	–
November	1 138 614	1 006 257	–
Dezember	1 129 088	1 023 829	–

¹⁶ Das Problem, dass symmetrische Mittelfilter nicht angewendet werden können, besteht natürlich für beide Enden der Zeitreihe, sodass auch die Trendschätzwerte der ersten 13 Monate einer Zeitreihe mit asymmetrischen Randfiltern ermittelt werden müssen. Die hier dargestellten asymmetrischen Randfilter für das Reihenende sind jedoch analog auch auf den Reihenanfang – allerdings spiegelverkehrt – anzuwenden.

¹⁷ Siehe Nourney, M., Fußnote 11, hier: Seite 843 ff., und Speth, H.-T., Fußnote 11, hier: Seite 10 ff.

In Verbindung mit den Filtergewichten der Tabelle 1 lassen sich somit für die 14 aktuellsten Werte die nachfolgend gezeigten Trend-Konjunktur-Komponenten \hat{m}_t als Linearkombination aus Filtergewicht und entsprechendem Zeitreihenursprungswert ermitteln.¹⁸

Am aktuellen Rand wird zunächst der asymmetrische Filtervektor f_1' (siehe Tabelle 1) eingesetzt:¹⁹

$$\begin{aligned} \hat{m}_{\text{Jun}2011} &= \mathbf{f}_1' \cdot \mathbf{y} \\ &= 0 \cdot y_{\text{Jan}2009} + 0 \cdot y_{\text{Feb}2009} + 0 \cdot y_{\text{Mrz}2009} \\ &\quad + (-0,01649) \cdot y_{\text{Apr}2009} + (-0,02865) \cdot y_{\text{Mai}2009} \\ &\quad + (-0,04086) \cdot y_{\text{Jun}2009} + 0,01106 \cdot y_{\text{Jul}2009} \\ &\quad + 0,02333 \cdot y_{\text{Aug}2009} + 0,03471 \cdot y_{\text{Sep}2009} \\ &\quad + 0,04337 \cdot y_{\text{Okt}2009} + 0,04063 \cdot y_{\text{Nov}2009} \\ &\quad + 0,03225 \cdot y_{\text{Dez}2009} + 0,01655 \cdot y_{\text{Jan}2010} \\ &\quad + (-0,00808) \cdot y_{\text{Feb}2010} + (-0,04318) \cdot y_{\text{Mrz}2010} \\ &\quad + (-0,08462) \cdot y_{\text{Apr}2010} + (-0,14075) \cdot y_{\text{Mai}2010} \\ &\quad + (-0,21161) \cdot y_{\text{Jun}2010} + 0,07227 \cdot y_{\text{Jul}2010} \\ &\quad + 0,06000 \cdot y_{\text{Aug}2010} + 0,04862 \cdot y_{\text{Sep}2010} \\ &\quad + 0,03996 \cdot y_{\text{Okt}2010} + 0,04270 \cdot y_{\text{Nov}2010} \\ &\quad + 0,05108 \cdot y_{\text{Dez}2010} + 0,06678 \cdot y_{\text{Jan}2011} \\ &\quad + 0,09142 \cdot y_{\text{Feb}2011} + 0,12652 \cdot y_{\text{Mrz}2011} \\ &\quad + (0,18444) \cdot y_{\text{Apr}2011} + 0,25273 \cdot y_{\text{Mai}2011} \\ &\quad + (0,33580) \cdot y_{\text{Jun}2011} \\ &= 936\,000 \end{aligned}$$

Hierbei ist \mathbf{y} ein $[30 \times 1]$ -Spaltenvektor der aufsteigend nach Aktualität sortierten Zeitreihenwerte. Da BV4.1 nicht mit Prognosen arbeitet, entspricht der letzte Wert der Zeitreihe, der für die Trendschätzung am aktuellen Rand eingesetzt wird, gleichzeitig dem aktuellsten tatsächlichen Wert. In diesem Beispiel ist dies die Zahl der erwerbslosen Frauen im Berichtsmonat Juni 2011. Dieser Ursprungswert fließt in die Schätzung des Trendwertes für Juni 2011 mit einem Gewicht von 0,33580 ein. Zusätzlich berücksichtigt werden die 26 vorangegangenen Monatswerte, wobei der am weitesten in der Vergangenheit zurückliegende Wert ein Gewicht von lediglich -0,01649 erhält. Sobald der nächste Monatswert vorliegt, hier der für Juli 2011, erfolgt die Schätzung des Trendwertes für Juni 2011 über den Filtervektor f_2' in Kombination mit den letzten 30 aktuellsten Beobachtungswerten. Dann erhält der Ursprungswert des Juni 2011 nur noch ein Gewicht in Höhe von 0,21553, während gleichzeitig der aktuellste Wert mit einer Gewichtung von 0,27998 einfließt.

¹⁸ Es sei an dieser Stelle erneut darauf hingewiesen, dass die Trendschätzwerte bei BV4.1 nur dann eine Linearkombination der originären Zeitreihenwerte des Stützbereiches sind, wenn zuvor keine Extremwerte identifiziert wurden. Liegt ein Extremwert vor, dann wird der originäre Wert durch einen geschätzten, um den Extremwert bereinigten Wert ersetzt.

¹⁹ Alle Berechnungen wurden mit Excel durchgeführt, wobei sämtliche Nachkommastellen der Filtergewichte berücksichtigt wurden. Das Nachrechnen mit den hier ausgewiesenen Zahlen führt entsprechend zu leicht abweichenden Ergebnissen. Die Trendschätzergebnisse sind auf Tausend gerundet.

Im obigen Beispiel, bei dem die Zeitreihe mit dem Monat Juni 2011 endet, wird entsprechend der Trendschätzwert für den Monat Mai 2011 über diesen Filtervektor f_2' ermittelt:

$$\begin{aligned} \hat{m}_{\text{Mai}2011} &= \mathbf{f}_2' \cdot \mathbf{y} \\ &= 0 \cdot y_{\text{Jan}2009} + 0 \cdot y_{\text{Feb}2009} + 0 \cdot y_{\text{Mrz}2009} \\ &\quad + 0 \cdot y_{\text{Apr}2009} + (-0,01665) \cdot y_{\text{Mai}2009} \\ &\quad + (-0,03038) \cdot y_{\text{Jun}2009} + 0,00311 \cdot y_{\text{Jul}2009} \\ &\quad + 0,01192 \cdot y_{\text{Aug}2009} + 0,02083 \cdot y_{\text{Sep}2009} \\ &\quad + 0,02446 \cdot y_{\text{Okt}2009} + 0,02491 \cdot y_{\text{Nov}2009} \\ &\quad + 0,02062 \cdot y_{\text{Dez}2009} + 0,01015 \cdot y_{\text{Jan}2010} \\ &\quad + (-0,00792) \cdot y_{\text{Feb}2010} + (-0,03494) \cdot y_{\text{Mrz}2010} \\ &\quad + (-0,07228) \cdot y_{\text{Apr}2010} + (-0,11555) \cdot y_{\text{Mai}2010} \\ &\quad + (-0,16627) \cdot y_{\text{Jun}2010} + 0,08022 \cdot y_{\text{Jul}2010} \\ &\quad + 0,07141 \cdot y_{\text{Aug}2010} + 0,06251 \cdot y_{\text{Sep}2010} \\ &\quad + 0,05887 \cdot y_{\text{Okt}2010} + 0,05843 \cdot y_{\text{Nov}2010} \\ &\quad + 0,06271 \cdot y_{\text{Dez}2010} + 0,07319 \cdot y_{\text{Jan}2011} \\ &\quad + 0,09125 \cdot y_{\text{Feb}2011} + 0,11828 \cdot y_{\text{Mrz}2011} \\ &\quad + 0,15561 \cdot y_{\text{Apr}2011} + 0,21553 \cdot y_{\text{Mai}2011} \\ &\quad + 0,27998 \cdot y_{\text{Jun}2011} \\ &= 943\,000 \end{aligned}$$

Das System lässt sich so fortsetzen, bis schließlich nach 13 Monaten ein endgültiger Trendschätzwert feststeht, der nicht mehr revidiert wird. Erst mit dem Wert für Juni 2011 kann ein endgültiger Trendschätzwert für Mai 2010 über den symmetrischen Filtervektor f_{14}' berechnet werden:

$$\begin{aligned} \hat{m}_{\text{Mai}2010} &= \mathbf{f}_{14}' \cdot \mathbf{y} \\ &= 0 \cdot y_{\text{Jan}2009} + 0 \cdot y_{\text{Feb}2009} + 0 \cdot y_{\text{Mrz}2009} \\ &\quad + (-0,01059) \cdot y_{\text{Apr}2009} + (-0,01609) \cdot y_{\text{Mai}2009} \\ &\quad + (-0,01549) \cdot y_{\text{Jun}2009} + (-0,01175) \cdot y_{\text{Jul}2009} \\ &\quad + (-0,00196) \cdot y_{\text{Aug}2009} + 0,01088 \cdot y_{\text{Sep}2009} \\ &\quad + 0,02577 \cdot y_{\text{Okt}2009} + 0,04167 \cdot y_{\text{Nov}2009} \\ &\quad + 0,05757 \cdot y_{\text{Dez}2009} + 0,07245 \cdot y_{\text{Jan}2010} \\ &\quad + 0,08530 \cdot y_{\text{Feb}2010} + 0,09509 \cdot y_{\text{Mrz}2010} \\ &\quad + 0,10941 \cdot y_{\text{Apr}2010} + 0,11551 \cdot y_{\text{Mai}2010} \\ &\quad + 0,10941 \cdot y_{\text{Jun}2010} + 0,09509 \cdot y_{\text{Jul}2010} \\ &\quad + 0,08530 \cdot y_{\text{Aug}2010} + 0,07245 \cdot y_{\text{Sep}2010} \\ &\quad + 0,05757 \cdot y_{\text{Okt}2010} + 0,04167 \cdot y_{\text{Nov}2010} \\ &\quad + 0,02577 \cdot y_{\text{Dez}2010} + 0,01088 \cdot y_{\text{Jan}2011} \\ &\quad + (-0,00196) \cdot y_{\text{Feb}2011} + (-0,01175) \cdot y_{\text{Mrz}2011} \\ &\quad + (-0,01549) \cdot y_{\text{Apr}2011} + (-0,01609) \cdot y_{\text{Mai}2011} \\ &\quad + (-0,01059) \cdot y_{\text{Jun}2011} \\ &= 1\,071\,000 \end{aligned}$$

3 Zusammenfassung

Die monatliche Auswertung des seit 2005 unterjährig durchgeführten Mikrozensus warf und wirft immer noch erhebliche Schwierigkeiten auf, die über die generellen Probleme einer stichprobenbasierten Haushaltsbefragung hinausgehen. Viele Probleme organisatorisch-technischer und/oder mathematisch-konzeptioneller Art ließen sich in den ersten beiden Jahren nach Einführung der kontinuierlichen Erhebung lösen, sodass Monatswerte aus dem Mikrozensus beziehungsweise der Arbeitskräfteerhebung ab dem Berichtsmontat Januar 2007 veröffentlicht werden können. Dennoch zeigen die monatlichen Zeitreihen des Mikrozensus weiterhin eine methodisch bedingte hohe Volatilität.

Mit Vollendung des Berichtsjahres 2010 liegen monatliche Arbeitsmarktzeitreihen mit einer Länge von vier Jahren vor, die rein technisch gesehen nunmehr einem Standardsaisonbereinigungsverfahren unterzogen werden könnten. Aufgrund der methodisch bedingten hohen Volatilität der Ausgangszeitreihen besteht bei einer solchen Saisonbereinigung allerdings die Gefahr, statistische Artefakte zu produzieren und eine saisonbereinigte Reihe zu erzeugen, die ebenfalls sehr volatil ist und kaum zu interpretieren wäre. Saisonbereinigte Reihen mit hoher Volatilität sind als konjunkturelles Analyseinstrument kaum geeignet, sodass hier als zeitreihenanalytisches Instrument eher eine Trendschätzung eingesetzt werden sollte. Im Gegensatz zur Saisonbereinigung werden bei einer Trendschätzung nicht nur regelmäßig wiederkehrende saisonale Schwankungen, sondern auch irreguläre, zufallsbedingte, aber auch methodisch bedingte Effekte herausgefiltert. Das Statistische Bundesamt hat mit der Veröffentlichung von monatlichen Arbeitsmarktdaten des Mikrozensus zum Berichtsmontat März 2011 entschieden, eine solche Trendschätzung einzusetzen und deren Ergebnisse neben den Werten der Originalreihe zu veröffentlichen. Auf eine Veröffentlichung saisonbereinigter Ergebnisse wird dagegen vorerst verzichtet.

Da die Frage, ob eine Trendschätzung oder eine Saisonbereinigung die bessere Wahl für den Konjunkturforscher darstellt, seit Jahrzehnten kontrovers diskutiert wird, kann es gegen diese Vorgehensweise Bedenken geben. In den von Eurostat und der Europäischen Zentralbank entwickelten ESS Guidelines on seasonal adjustment²⁰ wird als Standardvorgehensweise die Saisonbereinigung empfohlen. Gemäß diesen Richtlinien sollte dabei eines der international am meisten verwendeten Verfahren, also TRAMO-SEATS oder Census X-12-ARIMA, zum Einsatz kommen.

Während die ESS-Guidelines für eine Vielzahl ökonomischer Zeitreihen entwickelt wurden, wobei diese Zeitreihen häufig aus den Volkswirtschaftlichen Gesamtrechnungen hervorgehen, enthalten die ILO-Recommendations on Seasonal Adjustment of Employment and Unemployment Data²¹ ergänzende Empfehlungen speziell zur Saisonbereinigung von umfragebasierten Erwerbstätigen- und Erwerbslosenzeitreihen. Zwar stützen sich die ILO-Empfehlungen generell auf die ESS-Guidelines und auch hier wird die Saisonberei-

nigung einer Trendschätzung vorgezogen, bei der zeitreihenanalytischen Behandlung von umfragebasierten Daten sind jedoch laut Internationaler Arbeitsorganisation zusätzliche Aspekte zu berücksichtigen. So weisen die von der Internationalen Arbeitsorganisation untersuchten Erwerbstätigen- und Erwerbslosenzeitreihen spezielle Charakteristika umfragebasierter Daten auf, das heißt sie haben allein dadurch, dass sie aus einer Stichprobe gewonnen werden, zwangsläufig eine höhere Volatilität als Daten aus Vollerhebungen und Registern. Darüber hinaus hat die Internationale Arbeitsorganisation bei einer Vielzahl von Erwerbstätigenzeitreihen keine eindeutigen Saisonmuster identifizieren können und in solchen Fällen von der Durchführung einer Saisonbereinigung abgeraten.

Die Neuerungen bei der zeitreihenanalytischen Behandlung der monatlichen Arbeitsmarktzeitreihen des Mikrozensus stehen nicht vollständig im Einklang mit den ESS-Guidelines und den ILO-Empfehlungen. Diese Abweichung ist jedoch in den Besonderheiten der unterjährigen Mikrozensusdaten begründet, die über die generellen Charakteristika umfragebasierter Daten hinausgehen. Monatsauswertungen des Mikrozensus sind derzeit noch mit *zusätzlichen* methodischen Problemen behaftet, die zu einer hohen methodisch bedingten Volatilität der Zeitreihen führen. Es wird jedoch fortwährend untersucht, inwieweit Änderungen etwa am Stichprobendesign, am Hochrechnungsverfahren oder an den Datenerhebungsverfahren die methodisch bedingte Volatilität der Zeitreihe verringern können, damit neben Trendschätzergebnissen zum frühestmöglichen Zeitpunkt auch zuverlässige saisonbereinigte Ergebnisse veröffentlicht werden können. In der Zwischenzeit werden darüber hinaus auch Analysen zu Weiterentwicklung oder Ergänzung des eingesetzten Trendschätzverfahrens durchgeführt, um gegebenenfalls zusätzlich Ergebnisse eines zweiten Trendschätzverfahrens zu veröffentlichen.

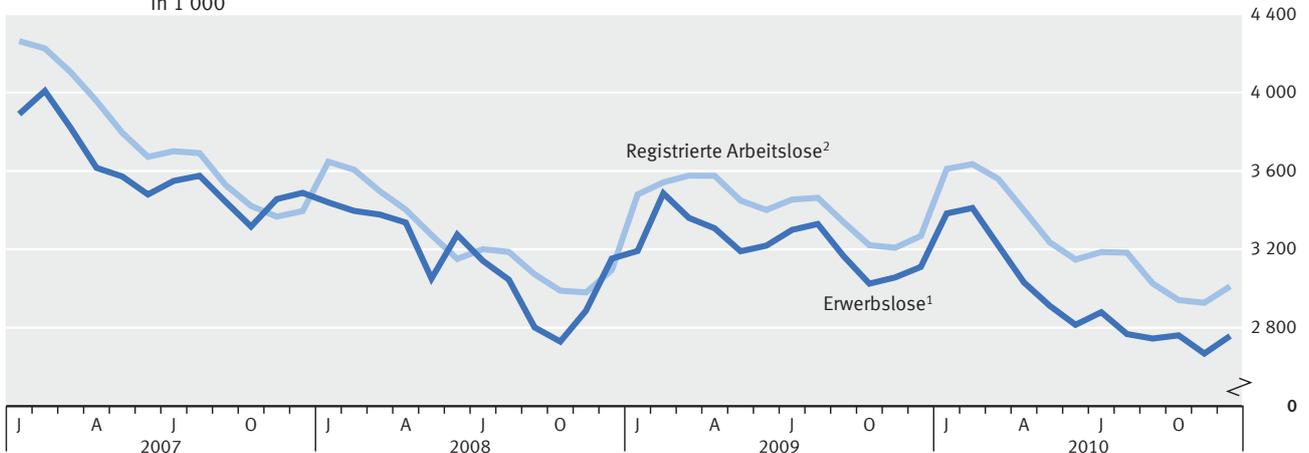
Derzeit erfolgt die Trendschätzung bei den monatlichen Zeitreihen des Mikrozensus auf Basis des Berliner Verfahrens, Version 4.1 (BV4.1). Dieses Verfahren hat unter anderem den Vorteil, dass es eine deutlich geringere Komplexität hat als zum Beispiel das Trendschätzverfahren von Census X-12-ARIMA. Somit kann eine Transparenz geschaffen werden, die für die Interpretation der Ergebnisse der Trendschätzung von herausragender Bedeutung ist. Transparenz und Nachvollziehbarkeit des eingesetzten Verfahrens sind eine Grundvoraussetzung, um die zurzeit veröffentlichten Trendschätzungen für die monatlichen Arbeitsmarktzeitreihen des Mikrozensus beurteilen zu können. Darüber hinaus gilt es, auch das allgemeine Bewusstsein für die generelle Unsicherheit solcher Trendschätzungen zu schärfen. [u](#)

²⁰ Siehe Fußnote 6.

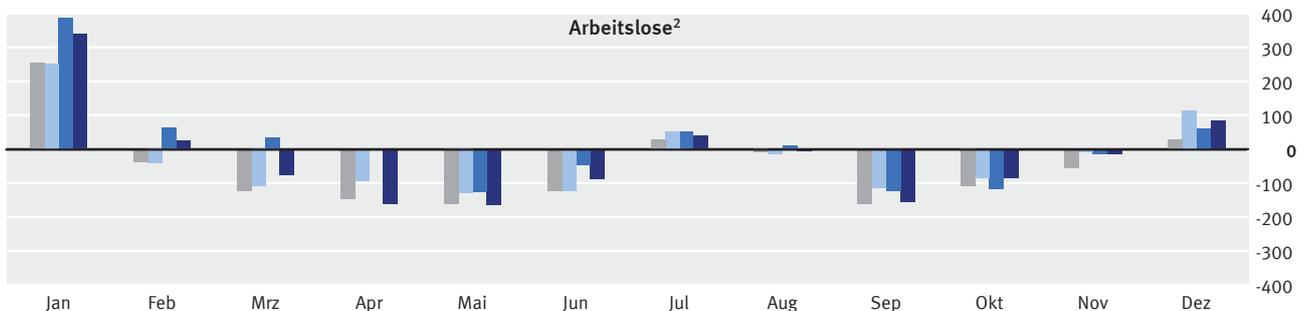
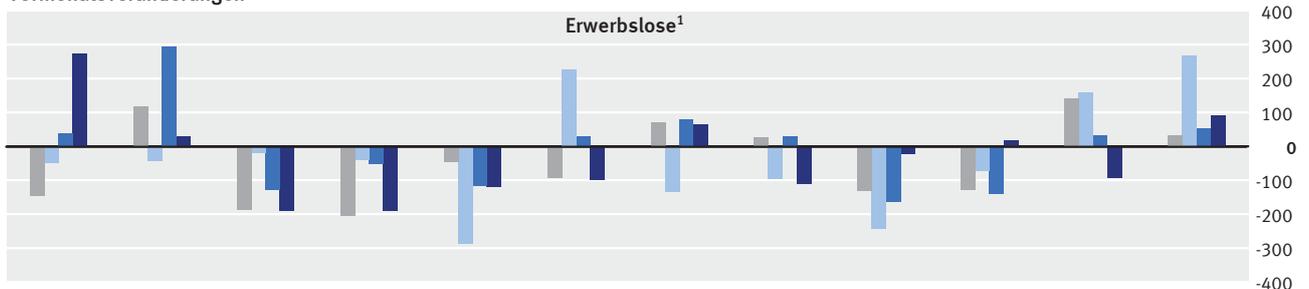
²¹ Siehe Fußnote 8.

Anhang: Monatliche Arbeitsmarktzeitreihen von Umfragedaten im Vergleich zu Registerdaten und Erwerbstätigenrechnung

Schaubild 1: Monatsreihen der Erwerbslosen¹ und registrierten Arbeitslosen²
in 1 000



Vormonatsveränderungen

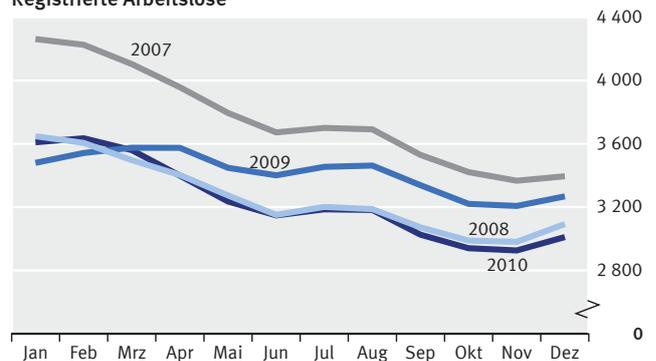


■ 2007 ■ 2008 ■ 2009 ■ 2010

Erwerbslose¹

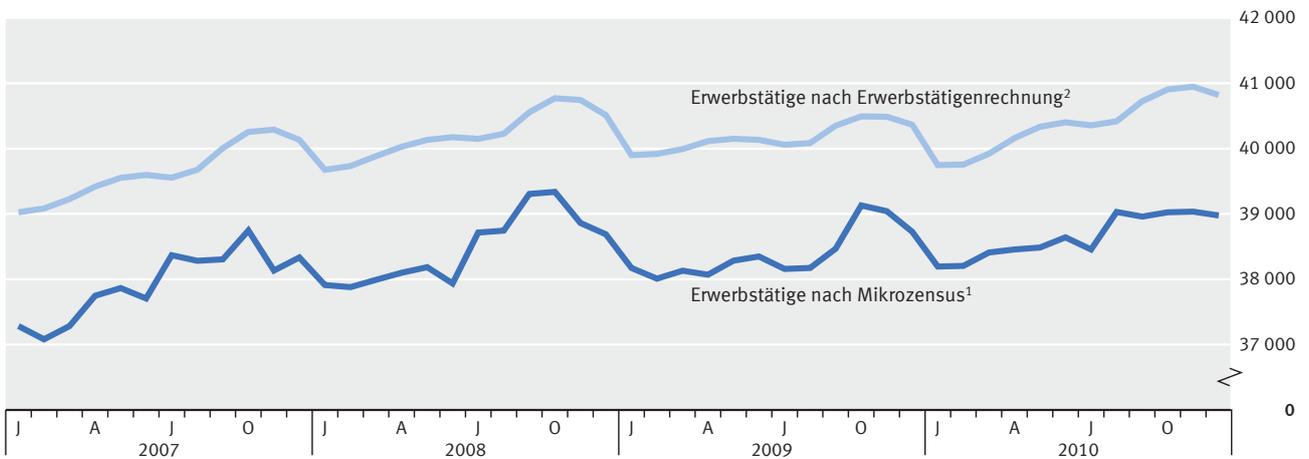


Registrierte Arbeitslose²

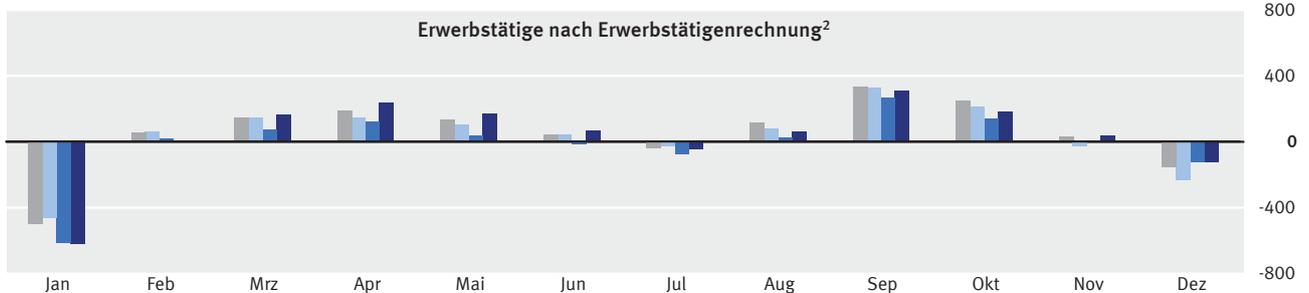
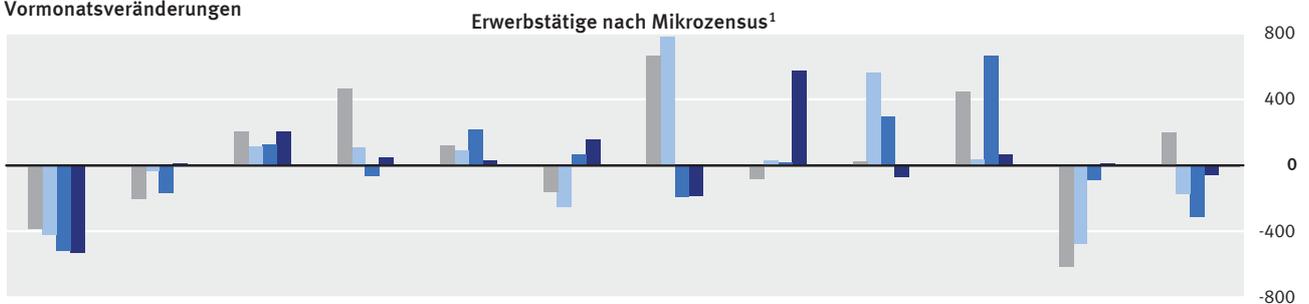


1 Erwerbslose auf Basis des Mikrozensus; Personen im Alter von 15 bis 74 Jahren in Privathaushalten.
2 Registrierte Arbeitslose der Bundesagentur für Arbeit.

Schaubild 2: Monatsreihen der Erwerbstätigen nach Mikrozensus¹ und Erwerbstätigenrechnung²
in 1 000

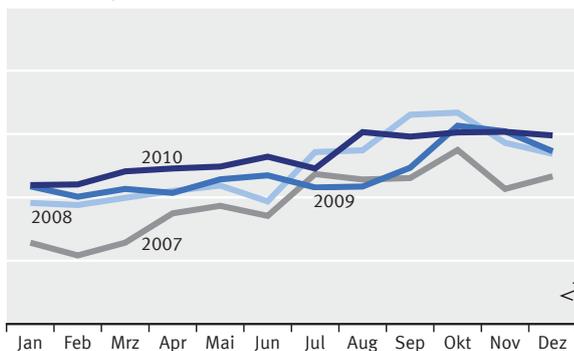


Vormonatsveränderungen

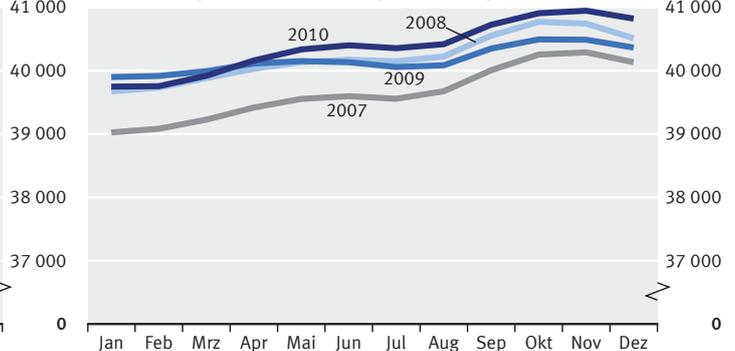


■ 2007 ■ 2008 ■ 2009 ■ 2010

Erwerbstätige nach Mikrozensus¹



Erwerbstätige nach Erwerbstätigenrechnung²



1 Erwerbstätige im Alter von 15 bis 74 Jahren in Privathaushalten, ohne Wehr- und Zivildienstleistende.

2 Erwerbstätige mit Wohnort in Deutschland, einschließlich Personen in Gemeinschaftshaushalten, einschließlich Wehr- und Zivildienstleistender.