

# Eingereichte Beiträge zum BMWK-Beteiligungsprozess zur Potenzialschätzung

Anlage zum Gutachten

Potenzialschätzung und Produktionslücken der Bundesregierung – Darstellung und Vergleich von Vorschlägen zur Anpassung der Bestimmungsmethode<sup>1</sup>

## Vorbemerkungen

Im Zuge des Beteiligungsprozesses gab es eine Vielzahl von Einreichungen. Sieben Einreichungen (1-7) wurden von den Einreichenden zur Evaluation im Rahmen des Gutachtens des IfW Kiel vorgesehen. Teilweise wurden kam es im Zuge des Bearbeitungsprozesses des Gutachtens zu einem Austausch zwischen dem IfW Kiel und den Einreichenden, in dem die Vorschläge für die Echtzeitberechnungen präzisiert bzw. modifiziert, weshalb teilweise Darstellungen im Gutachten von denen in den ursprünglichen Einreichungen, die hier dokumentiert sind, abweichen können.

1. Christian Breuer: Arbeitslosigkeit als natürlicher Konjunkturindikator
2. Kai Carstensen, Felix Kießner, Thies Rossian: Faktormodell zur Bestimmung der TFP
3. Dezernat Zukunft: Vorschläge zur Projektion des Arbeitspotenzials und der Produktionsfunktion
4. Marcell Göttert und Timo Wollmershäuser: Umfragegestützte strukturelle Finanzierungssalden
5. Thorsten Hild: Vorschlag zur Bestimmung des Produktionspotenzials im Rahmen der grundgesetzlich verankerten Schuldenregel
6. Carsten-Patrick Meier: Modellierung der NAWRU im D\*-Modell von Kiel Economics
7. Christian Proaño: Für eine ökonomischere Fundierung der Trend-TFP in der Ermittlung des Produktionspotenzials<sup>2</sup>

Zudem gab es im Zuge des Beteiligungsprozesses weitere Beiträge, die die Potenzialschätzung thematisiert haben (8-12), die von den Einreichenden aber nicht für eine Berücksichtigung im Rahmen der empirischen Evaluation im Gutachten vorgesehen waren. Diese Beiträge sind hier ebenfalls angefügt. Eine weitere Einreichung von Heimberger und Schütz wird hier nicht aufgeführt, da diese die Frage der Budgetsemielastizität und nicht die der Potenzialschätzung bearbeitet.

8. Tino Berger und Christian Ochsner: Robust Real-Time Estimates of the German Output Gap based on a Multivariate Trend-Cycle Decomposition
9. Deutsche Bundesbank: Zur Potenzialschätzung im Rahmen der Schuldenregel des Bundes – Ein Beitrag zum Beteiligungsprozess
10. Veronika Grimm: Konjunkturbereinigungsverfahren: Verbesserung der Schätzung, aber keine Abschwächung der Fiskalregeln durch die Hintertür
11. Christian Ochsner, Lars Other und Christopher Zuber: Der SVR-Ansatz zur Bestimmung des Produktionspotenzials
12. Torsten Schmidt: Forschungs- und Entwicklungsaktivitäten in der Potenzialschätzung berücksichtigen

Die Einreichungen finden sich in der Reihung 1-12 auf den folgenden Seiten.

---

<sup>1</sup> Gutachten des Kiel Institut für Weltwirtschaft im Auftrag des BMWK, Mai 2023.

<sup>2</sup> In einer zuvor eingereichten Version wurde zudem die Nutzung der CES-Produktionsfunktion diskutiert, die im Wirtschaftsdienstartikel nicht mehr enthalten ist.

# Arbeitslosigkeit als natürlicher Konjunkturindikator

Christian Breuer

15. September 2022

Update: 8. Juni 2023

## Abstract

### Einleitung

Die Schätzungen von Produktionspotenzial bzw. Produktionslücken als nicht direkt beobachtbare Indikatoren der ökonomischen Analyse sind schon seit Jahrzehnten der Kritik ausgesetzt. Für die Analyse und Bewertung der Ausrichtung der Finanzpolitik wies Blanchard (1990) darauf hin, dass auch direkt beobachtbare Konjunkturindikatoren, insbesondere die Arbeitslosigkeit, als natürlicher Konjunkturindikator und Proxy für die Produktionslücke verwendet werden können.

Die Umsetzung der sogenannten „Blanchard Methode“ bzw. der „Blanchard Fiscal Indicator“ (BFI) blieb in der wissenschaftlichen Literatur bisher jedoch umstritten (u.a. Breuer, 2019). In diesem Papier soll die Idee von Blanchard (1990) aufgegriffen werden, arbeitsmarktorientierte und beobachtbare Variablen als Größen für die (Unter-) Auslastung der Volkswirtschaft zur Nutzung für die Ermittlung des konjunkturbereinigten Budgetsaldo heranzuziehen.

### Ansatz

In diesem Ansatz wird die Arbeitslosenquote als natürlicher Konjunkturindikator benutzt. Anders als in früheren Arbeiten zur Analyse der Wirkungen fiskalpolitischer Maßnahmen (bspw. Alesina und Perotti (1995) und folgende) wird jedoch nicht die Reaktion der Staatsfinanzen direkt in Abhängigkeit von der Arbeitslosigkeit geschätzt.<sup>1</sup> Stattdessen wird zunächst der zyklische Zusammenhang zwischen

---

<sup>1</sup> Das aktuelle Verfahren der Schuldenbremse in Deutschland bzw. der Fiskalregeln in Europa sieht insbesondere eine Trennung zwischen automatischen und diskretionären Budgeteffekten vor, die dieses Verfahren nicht leisten

Bruttoinlandsprodukt und Arbeitslosigkeit (Okun-Zusammenhang) geschätzt, mit deren Hilfe unter Verwendung der etablierten und geschätzten Budgetelastizitäten der zyklische Anteil des Staatsdefizits berechnet werden kann. Die Methode hat den Vorteil, dass erstens, auf einen beobachtbaren Referenzindikator zurückgegriffen wird, der kaum Revisionen ausgesetzt ist, nämlich Arbeitslosigkeit, zweitens, lediglich ein theoretisch einfacher und plausibler Okun-Zusammenhang zweier Größen (Arbeitslosigkeit und Bruttoinlandsprodukt) zu schätzen ist, und drittens, der weitere Instrumentenkasten zur Berechnung der strukturellen Budgetsalden, insbesondere die Budgetelastizität, verwendet bzw. beibehalten werden kann.

Zur Schätzung des Zusammenhangs zwischen BIP und Arbeitslosigkeit kommen verschiedene Verfahren in Betracht (Schätzung in Niveaus, ersten Differenzen, inkl. oder exkl. Lags, trendbereinigt oder in Wachstumsraten). Hier soll zunächst das Niveau des zyklische Teil der Arbeitslosigkeit  $X_t$  erklärt werden mit bisherigen Schätzungen des zyklischen Teils des BIP  $Y_t$ :

$$X_t = a + bY_t + u_t \quad (1)$$

Dabei entspricht  $X_t$  zunächst der Differenz zwischen beobachteter Arbeitslosigkeit und NAWRU und  $Y_t$  der Differenz zwischen beobachtetem BIP und geschätztem Potenzial.<sup>2</sup>

Die Schätzung verwendet auf beiden Seiten der Gleichung geschätzte Größen, was nicht der Idee entspricht, auf beobachtbare Größen zurückzugreifen. Um dies zu vermeiden, wird in einem zweiten Verfahren die Veränderungsrate der Arbeitslosigkeit auf erste Differenzen des Bruttoinlandsprodukts regressiert. Zudem wird in diesem dynamischen Ansatz ein Verzögerungsterm verwendet, da der Zusammenhang in ersten Differenzen nicht notwendigerweise im ersten Jahr wirken muss, die Schätzung jedoch für Abweichungen in Produktionslücken über mehrere Jahre hinweg geeignet sein soll.

$$X_t = a + bY_t + cX_{t-1} + u_t \quad (2)$$

Der mit den beiden Verfahren geschätzte Okun-Zusammenhang wird in Tabelle 1 dargestellt. Um weitgehend unabhängig von geschätzten Größen zu arbeiten, wird im Folgenden der unter (2) genannte Ansatz verfolgt. Wir rechnen insofern unter Berücksichtigung der sich durch den Verzögerungsterm ergebenden langfristigen Dynamik mit einer langfristigen Okun-Elastizität von -0,59 bzw. einem inversen Okun-Zusammenhang von  $1 / -0,59 = -1,695$ .<sup>3</sup>

---

konnte. Für weitere Kritik an der praktischen Umsetzung der „Blanchard Methode“ in der Literatur in Folge von Alesina und Perotti (1995), siehe u.a. Breuer (2019).

<sup>2</sup> Der Koeffizient von Interesse  $b$  beträgt in dieser ersten Schätzung -0,32. Demnach erhöht sich die Arbeitslosigkeit bei einem Rückgang des zyklischen BIP um 1 Prozentpunkt um 0,32 Prozentpunkte.

<sup>3</sup> Es ist zu beachten, dass die geschätzten Elastizitäten abhängig von den gewählten Größen, dem Sample, der Frist (kontemporär, inkl. Verzögerungen, lange Frist) sein muss. Ähnlich wie bei den gewählten Größen der Budget-Elastizitäten ist hier eine Auswahl sinnvoller Indikatoren auf Basis von Annahmen geboten. In dieser kurzen Notiz kann zwingenderweise nur eine sehr kleine Auswahl besprochen werden.

Ganz rechts in der Tabelle befindet sich der mit den beiden Methoden geschätzte Kehrwert des langfristigen Okun-Zusammenhangs. Eine Erhöhung der Arbeitslosigkeit um einen Prozentpunkt würde im geschätzten Sample in etwa einer Reduktion des BIP zwischen 1.7 und 3.1 Prozent entsprechen.

**Tabelle 1: Okun-Zusammenhang lt. Gleichung (1) und (2)**

Gleichung	Okun-Elastizität	Lag	Langfristelastizität (b/(1-c))	Kehrwert
(1)	-0,324	-	-0,324	-3,086
(2)	-0,236	0,600	-0,590	-1,695

## Ergebnisse

Wir verwenden die unter Berücksichtigung von Verzögerungstermen in Gleichung (2) geschätzte Okun-Beziehung zur Berechnung der Produktionslücken, indem wir die zyklische Komponente des BIP (OG) mit Hilfe einer Schätzung oder Definition des zyklischen Teils der Arbeitslosigkeit (UG) und dem Kehrwert der Okun-Elastizität  $e$  berechnen.

$$OG = e * UG \quad (3)$$

Für die Berechnung des zyklischen Teils der Arbeitslosigkeit kommen verschiedene Verfahren in Betracht. Fontanari et al. (2019) verwenden zum einen a) die von Okun verwendete Definition der Vollbeschäftigung bei einer Arbeitslosigkeit in Höhe von 4%. Am aktuellen Rand dürfte Deutschland hiervon nicht weit entfernt sein, obwohl zur Berechnung der Unterbeschäftigung sowie der strukturellen Arbeitslosigkeit/ Unterbeschäftigung wiederum verschiedene Definitionen / Annahmen getroffen werden können.

Weiterhin könnte die ebenfalls von Fontanari et al. verwendete Definition b) genutzt werden, nach der Vollbeschäftigung bei der niedrigsten im Sample beobachteten Arbeitslosigkeit angenommen werden kann.

In Frage kommt ebenfalls, die strukturelle Arbeitslosigkeit c) bei einem Durchschnittswert anzusetzen, wobei wiederum zu definieren wäre, für welches Sample der Durchschnitt berechnet wird.

Eine weitere Möglichkeit bestünde darin, d) die mit etablierten Methoden geschätzte NAWRU als Schätzung der strukturellen Arbeitslosigkeit zu verwenden. Ein Nachteil dieser Methode wäre, dass man weiterhin auf den etablierten Methoden der Potenzialschätzung aufbaut und zumindest der Teil der Revisionsproblematik ungelöst bliebe, der sich auf die Revision der NAWRU bezieht. Dieser Nachteil könnte auch ein Vorteil sein, wenn die Probleme der aktuellen Methode bspw. weniger in der

Berechnung der NAWRU gesehen würden (bzw. wenn hier Verbesserungspotenzial gesehen würde), sondern in anderen Aspekten, wie etwa der Berechnung bzw. Revision der TFP.

Wir verwenden im Folgenden die Methoden a) und d), nehmen also an, dass die strukturelle Arbeitslosigkeit bei ca. 4% liegt, oder in etwa der NAWRU entspricht. Zudem nutzen wir neben der Arbeitslosenquote auch ein Maß für die zyklische Unterbeschäftigung unter Verwendung von längeren Zeitreihen zur Kurzarbeit sowie der durchschnittlichen Arbeitszeitreduktion der Kurzarbeitenden.<sup>4</sup>

Nach Abzug der jeweiligen Definitionen der strukturellen Arbeitslosigkeit bzw. Vollbeschäftigung von der Gesamtzahl der Arbeitslosigkeit bzw. Unterbeschäftigung errechnet sich das jeweils als konjunkturell angesehene Maß der Arbeitslosigkeit bzw. Unterbeschäftigung. Multipliziert mit dem Reziprok der Okun-Elastizität erhalten wir die errechnete Konjunkturkomponente des BIP, mit dem wiederum in Verbindung mit dem bekannten Instrumentenkasten der Budgetelastizitäten die zyklischen Anteile der Staatseinnahmen und –ausgaben (bzw. die Konjunkturkomponente) berechnet werden können.

Abbildung 1 zeigt die auf Basis der hier beschriebenen Methodik berechneten Produktionslücken. Methode 1 greift dabei auf die Arbeitslosigkeit zurück, während Methode 2 zudem die Kurzarbeitenden sowie die durchschnittliche Arbeitszeitreduktion der Kurzarbeitenden verwendet. Methode UB („Unterbeschäftigung“) recurriert dabei auf die Abweichung von Vollbeschäftigung (hier: 4%), während die Methode mit dem Index „Z“ sich hier auf die zyklische Arbeitslosigkeit im Sinne der Abweichung von der NAWRU bezieht. Beide Methoden sind keinesfalls alternativlos. Denkbar wären ebenfalls andere Methoden auf Basis von Gewichten, Durchschnittswerten oder andere Berechnungen struktureller Arbeitslosigkeit (NAWRU), die möglichst die Schwächen der bisherigen Methodik (Revisionsanfälligkeit) vermeiden.

## Diskussion

Welcher Indikator zu verwenden ist, sollte in Abhängigkeit von den verfolgten Zielen gesehen werden. Geht es darum, die Revisionsanfälligkeit zu minimieren (oder gar zu eliminieren), würde sich die Nutzung von Modellvariante a) unter Setzung eines Maßes für Vollbeschäftigung (bspw. 4%, wie hier verwendet) anbieten. Wird das Ziel für die Vollbeschäftigung nicht erreicht, wäre eine Diskussion darüber denkbar, ob der Zielkonflikt (wenn Vollbeschäftigung nicht erreicht werden kann) im Rahmen der Fiskalregeln gelöst werden sollte, oder mit anderen Methoden zu beheben wäre. Dabei kann es sich je nach Analyse und politischer Entscheidung um Nachfrage- oder Angebots- bzw. Arbeitsmarktpolitik

---

<sup>4</sup> Als Quellen verwenden wir für den Zeitraum ab 1991 „Der Monatsbericht zum Arbeits- und Ausbildungsmarkt in Deutschland, August 2022, Seite 59“ sowie „histat-Datenkompilation online: Erwerbstätigkeitsstatistik in der Bundesrepublik Deutschland. Ausgewählte Daten von 1950 bis 2000“ von Sensch, Jürgen, für den Zeitraum vor 1991.

handeln. Nicht notwendigerweise müssen diese wirtschaftspolitischen Ziele innerhalb der Schuldenbremse gelöst oder überhaupt adressiert werden. Die aktuellen Verfahren, versuchen, diese Problematik in die Schätzung der NAWRU zu verlagern, die aber aktuell angesichts der Revisionen mit einer stabilen Bestimmung der strukturellen Arbeitslosigkeit im Euroraum überfordert zu sein scheint. Die Annahme, dass die mit den bisherigen Verfahren geschätzte NAWRU der strukturellen Arbeitslosigkeit entspricht, wäre eine Alternative (d), mit der jedoch die Revisionsproblematik (insofern sie sich auf die NAWRU bezieht) nicht vollständig beseitigt werden kann. Die Revisionsproblematik ist selbstverständlich kein Wert an sich, sondern deshalb ein Problem, da mit prozyklischen Revisionen eine Unterschätzung der Produktionslücke und damit auch prozyklische (tendenziell destabilisierende) Politik verbunden wäre. Mit dieser Modellvariante d) einhergehen würde (wie im bisherigen Verfahren), dass Staaten mit steigender bzw. hoher (persistenter bzw. struktureller) Arbeitslosigkeit höhere Sparanstrengungen vornehmen müssen. Ob diese prozyklische fiskalische Kontraktion ein geeignetes und gesundes Mittel ist, um Arbeitslosigkeit im 21. Jahrhundert abzubauen oder ob diese so eher verfestigt würde, darüber lässt sich je nach Theorie und Situation streiten. Um die Schätzungen der NAWRU in der EU zu verbessern, könnte aber auch eine Diskussion darüber wertvoll sein, inwiefern gerade in der Europäischen Währungsunion die nationalen Schätzungen der NAWRU (und des Produktionspotenzials) und ihre Revisionen miteinander korrespondieren.

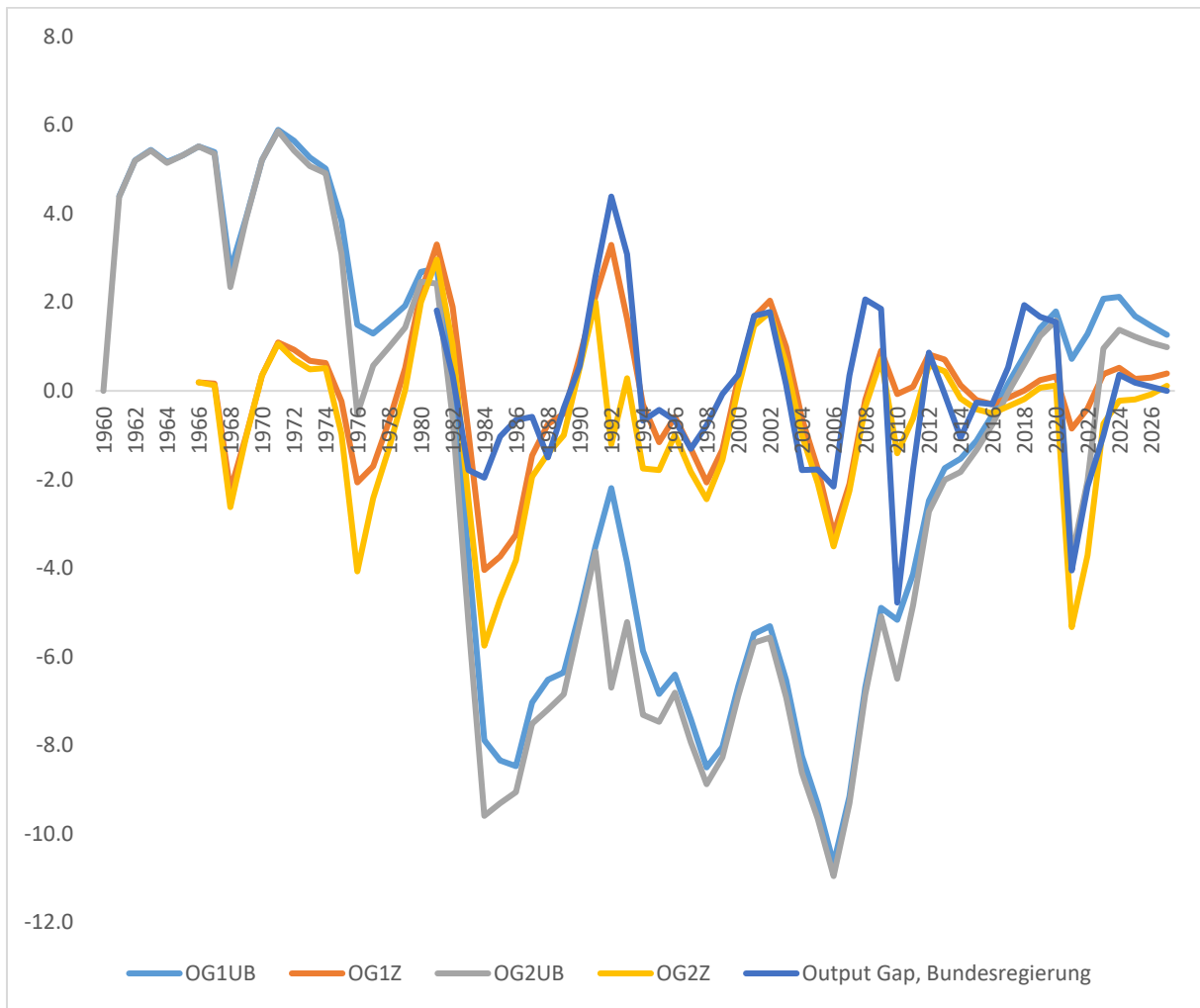
Die Unsicherheit über das Produktionspotenzial (insbesondere in nominaler Rechnung) ist erheblich. Dabei spielt neben der Unsicherheit über die Entwicklung des realen BIP auch die Unsicherheit über die Entwicklung der Deflatoren in den verschiedenen Regionen des (nicht optimalen) Währungsraums für die jeweils auf nationaler Ebene stattfindenden Schätzungen eine Rolle. So sind die Schätzungen zur strukturellen Arbeitslosigkeit sowie des Produktionspotenzials in den Mitgliedsländern in einer Währungsunion aus theoretischen Erwägungen miteinander verbunden.<sup>5</sup> Eine hohe strukturelle Arbeitslosigkeit (bspw. wegen realer Aufwertung und hohen Deflatoren) dürfte mit geringer Wettbewerbsfähigkeit und Leistungsbilanzdefiziten verbunden sein und vice versa. Die Frage, ob Zielverfehlungen in Bezug auf Preise/Wettbewerbsfähigkeit/ Leistungsbilanz bzw. Arbeitslosigkeit (die sich nach einigen Jahren an dieser Stelle auf die Schätzungen zur NAWRU und ihre Revision auswirken) mit Hilfe von Fiskalregeln gelöst werden sollen, ist nicht einfach zu beantworten. Denkbar wäre, Fragen der Wettbewerbsfähigkeit außerhalb des fiskalischen Rahmens zu lösen. So werden Pathologien bzgl. Leistungsbilanz als auch bzgl. Arbeitslosigkeit bereits im „Verfahren bei makroökonomischen Ungleichgewichten“ adressiert. Ob eine deutliche Zielabweichung der Wettbewerbsfähigkeit oder der Arbeitslosigkeit den Verschuldungsspielraum der Fiskalregeln eher erhöhen oder senken sollte (wegen

---

<sup>5</sup> Bspw. wird durch die starke Verflechtung des Währungsraums und durch Mobilität von Produktionsfaktoren bei Wanderungsbewegungen das Potenzial eines Landes auf Kosten eines anderen Landes erhöht. Dieser Effekt ist ein Kanal, der prozyklische Revisionen der Schätzungen des Produktionspotenzials auslösen dürfte. Bspw. führen prozyklischen Wanderungsbewegungen, wie sie sich aus Gravity Modellen in der Migrationsforschung ergeben und wie sie in der Eurokrise beobachtet wurden, zu erheblichen prozyklischen Revisionen – bei den Produktionsfaktoren selbst, die sich systematisch asymmetrisch über den Währungsraum verteilen.

„struktureller Arbeitslosigkeit“), lässt sich insofern nicht a priori beantworten, sondern unterliegt einer politischen Zielvorstellung und wäre insofern eine politische, weniger eine technische Frage, obgleich die bisherige Technik hier eine eindeutige Antwort zu geben scheint. Nach den gravierenden Erfahrungen mit prozyklischen fiskalpolitischen Maßnahmen während der Eurokrise und ihren ökonomischen, sozialen und politischen Folgen wäre es m.E. nicht sinnvoll, diese grundlegenden Entscheidungen und Zielkonflikte den technischen Details der Ausgestaltung zu überlassen. Wäre die Vermeidung von prozyklischer Politik ein relevantes Ziel, dann wären prozyklisch revisionsanfällige Indikatoren grds. zu vermeiden (Breuer, 2021). Dies spräche für den hier vorgestellten Indikator. Wäre antizyklische Politik (zu einem ggf. politisch zu definierenden Grad) ein Ziel, wäre ggf. das Instrumentarium zur Schätzung der automatischen Stabilisatoren (und ihrer Elastizitäten) für den Gebrauch im Rahmen von Fiskalregeln anzupassen. Fraglich wäre aber weiterhin, an welchem Indikator sich die Ausrichtung der Fiskalpolitik orientieren sollte. In diesem Beitrag wurde ein Ansatz auf Basis Arbeitsmarktindikatoren für Deutschland vorgestellt, der geeignet ist, die Revisionsproblematik zu minimieren, insbesondere wenn arbeitsmarktpolitische ad hoc Maßnahmen (bspw. Kurzarbeitergeld) adäquat berücksichtigt werden.

**Abb. 1: Output Gaps 1960-2026: vier Methoden nach Blanchard-Okun sowie EU Methode**



Quelle: eigene Berechnungen.



## Literatur

Alesina, A. and R. Perotti (1995), Fiscal Expansions and Adjustments in OECD Countries, *Economic Policy* 10, 205-248.

Blanchard, O. (1990), Suggestions for a new set of fiscal indicators, OECD Working Paper, No. 79.

Breuer, C. (2019), Expansionary Austerity and Reverse Causality: A Critique of the Conventional Approach, Institute for New Economic Thinking Working Paper Series No. 98.

Breuer, C. (2021), Structural Indicators and the Fiscal Uncertainty Principle, *Intereconomics*, 56, 182–183.

Fontanari, C., Palumbo, A., and Salvatori, C. (2019), Potential Output in Theory and Practice: A Revision and Update of Okun's Original Method, Institute for New Economic Thinking Working Paper No. 93

Okun, A. (1962), Potential GNP: Its measurement and significance, Proceedings of the American Statistical Association, Business and Economic Statistics Section, ASA, Washington, 98-104

# Estimation of the TFP Gap for the Largest Five EMU Countries

*Kai Carstensen, Felix Kießner, Thies Rossian*

## **Impressum:**

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email [office@cesifo.de](mailto:office@cesifo.de)

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: [www.SSRN.com](http://www.SSRN.com)
- from the RePEc website: [www.RePEc.org](http://www.RePEc.org)
- from the CESifo website: <https://www.cesifo.org/en/wp>

# Estimation of the TFP Gap for the Largest Five EMU Countries

## Abstract

In this paper we augment the Bayesian unobserved components model of the EU Commission to estimate the cyclical component of total factor productivity (TFP gap) with a factor structure to include a wide array of business cycle indicators. We demonstrate that this model extension considerably stabilizes the estimate of the of the TFP gap. Specifically, consider the usual autumn forecast of the EU Commission in October of a year  $T$ . For the last two “in-sample” years  $T - 2$  and  $T - 1$ , and for the “now-cast” year  $T$ , the year-to-year revisions can be reduced by up to 30 percent. Improvements for the two “out-of-sample” years  $T + 1$  and  $T + 2$  also considered relevant by the EU Commission are quantitatively smaller (up to 10 percent) but still relevant. The results do vary across countries but are qualitatively robust with respect to different indicator sets, model specifications or vintages considered.

JEL-Codes: C320, E370.

Keywords: trend-cycle decomposition, unobserved components model, factor model, Bayesian estimation, total factor productivity, EU Commission.

*Kai Carstensen\**  
*University of Kiel / Germany*  
*carstensen@stat-econ.uni-kiel.de*

*Felix Kießner*  
*University of Kiel / Germany*  
*kiessner@stat-econ.uni-kiel.de*

*Thies Rossian*  
*University of Kiel / Germany*  
*thies.rossian@gmx.de*

\*corresponding author

January 15, 2023

We thank Jens Boysen-Hogrefe, Alessandro Rossi, Werner Roeger, Anna Thum-Thysen, Francesca Maria D’Auria and the participants of the meeting of the Output Gap Working Group in Bucharest on 17th May 2019 for helpful discussion and comments. Financial support by the German Federal Ministry for Economics and Energy for the research project 34/17 (IC4 - 80 14 38/034) is gratefully acknowledged.

# 1 Introduction

As part of the European Semester, the EU Commission annually assesses the fiscal plans of its member countries. A cornerstone of the procedure is to calculate the structural government budgets, i.e., budgets adjusted for the business cycle situation. To this end, a reliable estimate of the output gap is needed that exhibits only small revisions in order to keep fiscal (and political) over- and under-reactions as small as possible.

The EU Commission estimates potential output from a Cobb-Douglas production function determining output as a function of labour, capital and total factor productivity (TFP), see Havik et al. (2014). Assuming the capital stock is largely non-cyclical, the output gap is determined the cyclical components of labour and TFP where the latter is quantitatively more important, see for a discussion Doornik and Zuber (2019), which is why we concentrate on it.

In this paper we augment the Bayesian unobserved components model of the EU Commission, henceforth referred to as the “EU model”, to estimate the cyclical component of TFP (TFP gap) with a factor structure to include a wide array of business cycle indicators. Thereby, we intend to stabilize the estimate of the TFP gap. We demonstrate that this model extension considerably improves the nowcast of the TFP gap. To this end, we mimic the usual autumn forecast of the EU Commission that is executed in October of a year  $T$  and thus based on incomplete knowledge of this year’s macroeconomic situation. For the last two fully known in-sample years  $T - 2$  and  $T - 1$ , the year-to-year revisions can be reduced by up to 30 percent. For the partly unknown nowcast year  $T$ , the reduction in the revision size is still around 20 percent. Improvements for the two fully unknown out-of-sample years  $T + 1$  and  $T + 2$ , which are also considered relevant by the EU Commission, are quantitatively smaller (up to 10 percent) but still relevant.

A strength of our approach is that it does not change the general setup of the EU model. Both the basic model structure and the Bayesian estimation method remain unaltered. Technically, we simply add a block of factor model equations to improve the stability of the TFP gap estimates. Thereby, it should be easier to convince the EU Commission and the member states to adopt the extension, while they might be much more hesitant to radically change the modeling philosophy.<sup>1</sup>

Our paper relates to the large literature on output gap estimation. One strand of this literature shows that it is beneficial in terms of estimation precision and economic interpretability to extract the output gap by means of multivariate unob-

---

<sup>1</sup>In fact, we were invited to present our approach at the meeting of the Output Gap Working Group in Bucharest on 17th May 2019 and have been in contact to the EU Commission since then. Our impression has generally been that both member states and the EU Commission are highly interested in modifications of the current EU model that reduce the revisions of the TFP gap.

served component (UC) models applied to GDP plus a low-dimensional vector of business cycle indicators such as inflation and unemployment, see Kuttner (1994), Planas et al. (2008), Basistha and Startz (2008), and Fleischman and Roberts (2011) for the US, Valle e Azevedo et al. (2006) and Jarociński and Lenza (2018) for the euro area, Melolinna and Toth (2019) for the UK, and Blagrove et al. (2015) for a large set of countries.<sup>2</sup>

Another strand of the literature extracts cyclical factors from high-dimensional panels of indicator variables. Aastveit and Trovik (2014) apply a dynamic factor model to a set of 54 US indicators in order to obtain GDP nowcasts that reduce the end-of-sample problem of the HP filter documented by Orphanides and van Norden (2002); Mise et al. (2005). Weiske (2018) smoothes the first principal component of a set of 37 (stationary) business cycle indicators and shows that it is a reliable real-time output gap estimator for the four largest euro area countries. Barigozzi and Luciani (2019) extract the common cycle from a set of 103 (nonstationary) US variables and show that it is a competitive output gap estimator exhibiting only moderate revisions over time.

This paper integrates these two strands by including a factor structure in an otherwise standard UC model of the output gap as used by the EU Commission. Specifically, we assume that TFP and our panel of business cycle indicators share the same common cycle. We estimate this cycle in a joint model which has the advantage that extraction of the common cycle is supervised in the sense that it needs to account for both the cyclical variation in the indicator panel and in TFP.

The paper is structured as follows. Section 2 introduces the EU model and our extensions of it. In Section 3 discusses the priors, Section 4 presents the data sets used, and Section 5 describes the estimation procedure. Section 6 presents the results of a recursive revision analysis for our preferred specifications, while Section 7 conducts a sensitivity analysis. Section 8 discusses the plausibility of the estimated trend and cycle components and Section 9 concludes.

## 2 Trend-cycle models

In this section we first describe the UC model currently used by the EU Commission to estimate the TFP gap. Subsequently, we introduce and discuss our extensions.

---

<sup>2</sup>Another way to increase the information set is to use geographically disaggregate data such as state-level GDP for the US output gap (González-Astudillo, 2019b) or country-level data for the euro area output gap (González-Astudillo, 2019a; Huber et al., 2020). We do not follow this approach as we are interested in output gap estimates for individual member states of the euro area. However, we include aggregate euro area indicators in our factor approach for each country.

## 2.1 The EU model

The EU Commission currently uses an unobserved components model to estimate the TFP gap. Log TFP,  $y_t$ , is decomposed linearly into trend,  $p_t$ , and cycle,  $c_t$ , by means of the observation equation

$$y_t = p_t + c_t. \quad (1)$$

The trend is assumed to follow a random walk with AR(1) slope,  $\mu_t$ ,

$$\Delta p_t = \mu_{t-1} \quad (2)$$

$$\mu_t = \omega(1 - \rho) + \rho\mu_{t-1} + a_{\mu t}, \quad a_{\mu t} \stackrel{\text{iid}}{\sim} N(0, V_\mu), \quad (3)$$

where  $|\rho| < 1$  to ensure stationarity of  $\mu_t$ . The first observation  $\mu_1$  is assumed to be drawn from its unconditional distribution (we use the same assumption for all stationary AR processes discussed below). The cycle is modeled as a mean zero AR(2) process

$$c_t = 2A \cos\left(\frac{2\pi}{\tau}\right)c_{t-1} - A^2c_{t-2} + a_{ct}, \quad a_{ct} \stackrel{\text{iid}}{\sim} N(0, V_c), \quad (4)$$

where the parameterization with periodicity  $\tau > 0$  and amplitude  $0 < A < 1$  ensures stationarity and cyclical behavior.

To reduce trend revisions at the sample end, Havik et al. (2014) relate the TFP cycle to the Capacity-Utilization-and-Business-Sentiment (CUBS) indicator,  $u_t$ , a weighted average of survey-based capacity utilization in the manufacturing sector and business sentiment indicators in the construction and services sectors that are derived from the Business Tendency Surveys of the EU Commission.<sup>3</sup> This gives rise to the second observation equation

$$u_t = \mu_u + \beta c_t + e_{ut}. \quad (5)$$

To allow for somewhat persistent deviations between CUBS and the cycle, the error is modeled as a mean-zero AR(1) process:

$$e_{ut} = \delta_u e_{u,t-1} + a_{ut}, \quad a_{ut} \stackrel{\text{iid}}{\sim} N(0, V_u), \quad (6)$$

where  $|\delta_u| < 1$  to ensure stationarity.

Altogether, the EU model consists of the observation equations (1) and (5), and the state equations (2), (3), (4), and (6). It is cast in state space form and estimated with Bayesian methods using a Metropolis-within-Gibbs sampler. The priors are country-specific and discussed below.

---

<sup>3</sup>The reliability of capacity utilization as an indicator of the output gap and TFP gap is documented by Graff and Sturm (2010), D'Auria et al. (2010), Planas et al. (2013), and Turner et al. (2016).

## 2.2 Extension 1: Two-step factor model

Our first extension of the EU model simply adds an externally estimated business cycle factor,  $f_t$ , by means of another observation equation that links it to the TFP cycle:

$$f_t = \alpha c_t + e_{ft}, \quad (7)$$

where we neglect an intercept because the factor will be standardized and the cycle has a mean of zero by construction.

We again allow for persistence in the error, hence we model it as the AR(1) process

$$e_{ft} = \delta_f e_{f,t-1} + a_{ft} \quad a_{ft} \stackrel{\text{iid}}{\sim} N(0, V_f), \quad (8)$$

where  $|\delta_f| < 1$  to ensure stationarity.

Furthermore, we make the following restriction regarding the smoothness of the trend in (2)

$$|\Delta_2 p_t| \leq \zeta_p, \quad (9)$$

where  $\zeta_p$  is a threshold derived as the minimal smoothness of the trend growth rate of all countries and vintages of an HP-filter estimate with  $\lambda = 6.25$ . We introduce this restriction to keep the trend from taking up business cycle fluctuations.

Estimation proceeds in two steps. First, we estimate the factor from a set of business cycle indicators described below. To this end, we either use principle components (PCA) or a Bayesian factor model (Bay) presented in Appendix A.1 and A.2, respectively. Then we feed the factor in the extended model which consists of the observation equations (1), (5), and (7), and the state equations (2), (3), (4), (6), (8) and (9). Depending on which factor we use, we denote the model by 2-Step-PCA or 2-Step-Bay. We estimate it with Bayesian methods.

## 2.3 Extension 2: One-step factor model

In our second extension of the EU model, we integrate factor estimation and trend-cycle decomposition into one model which we estimate in a single step. This one-step approach has the advantage that parameter estimation and factor extraction is “supervised” in the sense that both the co-movement of the factor and the cyclical component of TFP is taken into account.

To facilitate one-step estimation, we add a set of  $k$  measurement equations to the EU model. Defining the  $k \times 1$  vector of observed yearly standardized business cycle indicators,  $x_t$ , we assume the factor structure

$$x_t = \Lambda f_t + e_{xt}, \quad (10)$$



where  $\Lambda = (\lambda_1, \dots, \lambda_k)'$  is a  $k \times 1$  vector. The errors follow the diagonal VAR(1) process

$$e_{xt} = \Gamma e_{x,t-1} + a_{xt} \quad a_{xt} \stackrel{\text{iid}}{\sim} N(0, V_x), \quad (11)$$

where  $V_x$  is a diagonal  $k \times k$  covariance matrix with coefficients  $v_{ii}$ ,  $i = 1, \dots, k$ , on the main diagonal, and  $\Gamma$  is a diagonal  $k \times k$  matrix with coefficients  $|\Gamma_{ii}| < 1$ ,  $i = 1, \dots, k$ , on the main diagonal to ensure stationarity. Altogether, this model, henceforth denoted by 1-Step, consists of the observation equations (1), (5), and (10), and the state equations (2), (3), (4), (6), (7), (8), (9) and (11). We estimate it with Bayesian methods.

## 2.4 Hodrick-Prescott (HP) filters

We compare the EU model and its extensions to the HP filter both with standard settings and with estimated signal-to-noise ratio. The standard HP filter is applied with two parameterizations. While a usual choice of the smoothing parameter for annual data is  $\lambda = 100$ , Ravn and Uhlig (2002) recommend  $\lambda = 6.25$ .

To estimate the signal-to-noise ratio, we consider the state space form of the HP filter (Harvey and Trimbur, 2008):

$$y_t = \tilde{p}_t + \tilde{c}_t \quad \tilde{c}_t \stackrel{\text{iid}}{\sim} N(0, \tilde{\sigma}_c^2), \quad (12)$$

$$\Delta \tilde{p}_t = \tilde{\mu}_{t-1} \quad (13)$$

$$\tilde{\mu}_t = \tilde{\mu}_{t-1} + \tilde{\varepsilon}_t, \quad \tilde{\varepsilon}_t \stackrel{\text{iid}}{\sim} N(0, \tilde{\sigma}_\varepsilon^2). \quad (14)$$

While the underlying assumption of an iid cycle  $\tilde{c}_t$  is certainly unrealistic, it is routinely made in all papers using the HP filter. We estimate the two variances  $\tilde{\sigma}_c^2$  and  $\tilde{\sigma}_\varepsilon^2$ , and thus implicitly the signal-to-noise ratio  $\tilde{\lambda}^{-1} = \tilde{\sigma}_\varepsilon^2 / \tilde{\sigma}_c^2$ , with Bayesian methods. We denote this estimated HP model by HP-EST.

## 3 Priors

We estimate the models with Bayesian methods, using a Metropolis-within-Gibbs algorithm, for the five largest EU countries Germany (DE), France (FR), Italy (IT), Spain (ES), and the Netherlands (NL). To replicate the results of the EU Commission, we use the EU model with the priors proposed by Havik et al. (2014). The priors for  $\omega$  and  $\rho$  of the slope equation (3) are independently normal, the prior for the variance  $V_\mu$  is independently inverse Gamma. A natural conjugate prior does not exist because we draw the first observation from its unconditional distribution. The priors for  $A$  and  $\tau$  of the cycle equation (4) are independently

Beta, the prior for the variance  $V_c$  is independently inverse Gamma. For the parameters  $\mu_u$ ,  $\beta$ , and  $V_u$  of the CUBS equations (5) and (6) we assume a natural conjugate Normal-inverse Gamma (NiG) prior implying marginal  $t$  distributions for  $\mu_u$  and  $\beta$ . The persistence parameter  $\delta_u$  is a priorily independently normal. Details of the priors is shown in Table 1. The parameterization (including truncation) for the coefficients of the EU model,  $\omega$ ,  $\rho$ ,  $V_\mu$ ,  $\tau$ ,  $A$ ,  $V_c$ ,  $\beta$ ,  $\mu_u$ ,  $\delta_u$ , and  $V_u$ , are taken from Havik et al. (2014).

The two-step factor models add the observation equation (7) and the AR(1) error equation (8) which altogether feature three additional parameters. The prior for  $\delta_f$  is truncated normal with mean zero, fairly large standard deviation of 1, and stationarity-preserving truncation bounds. The prior for  $\alpha$  and  $V_f$  is NiG. The prior parameters are empirical in the sense that they are taken from a preliminary regression of the factor,  $f_t$ , on an initial estimate of the cycle obtained by applying the HP filter to log TFP. Specifically, the prior means of  $\alpha$  and  $V_f$  are set to their OLS estimates, the prior standard deviations to four times their OLS standard errors.

The one-step factor model adds  $k$  observable business cycle indicators and treats the factor,  $f_t$ , as an unobservable state variable. To achieve identification, we set  $\alpha = 1$ , while the inverse Gamma prior  $V_f$  is parameterized empirically as described for the two-step factor model. Concerning the factor equations (10), the diagonal structure of  $\Gamma$  and  $V_x$  allows to specify equationwise priors for  $\lambda_i$ ,  $v_{ii}$ , and  $\Gamma_{ii}$ . For the persistence parameters  $\Gamma_{ii}$  we use a truncated normal prior with mean zero, fairly large standard deviation of 1, and stationarity-preserving truncation bounds. For  $\lambda_i$  and  $v_{ii}$ , we apply an empirical NiG prior. It is specified by regressing an initial estimate of the factor, obtained by applying the HP filter to log TFP, on each business cycle indicator. Specifically, the prior means of  $\lambda_i$  and  $v_{ii}$  are set to their OLS estimates, the prior standard deviations to four times their OLS standard errors.

Table 1: Priors

Parameter	marginal pdf	mean	s.d.	range
<i>Parameters of the EU model</i>				
$\omega$	$N$	.015	.01	(0,.03)
$\rho$	$N$	.8	.24	(0,.99)
$V_\mu (\times 10^{-6})$	$iG$	2.4	2.4	(0, $\infty$ )
$\tau$	$B$	8.0	3.5	(2,32)
$A$	$B$	.42	.17	(0,1)
$V_c (\times 10^{-4})$	$iG$	3.0	3.0	(0, $\infty$ )
$\beta$	$t$	1.4	.7	(0,5)
$\mu_u$	$t$	0	.03	(-1,1)
$\delta_u$	$N$	0	.4	(0,.99)
$V_u (\times 10^{-3})$	$iG$	4.5	4.5	(0, $\infty$ )
<i>Additional parameters of the 2-step factor model</i>				
$\alpha$	$t$	empirical		$(-\infty, \infty)$
$\delta_f$	$N$	0	1	(-.99,.99)
$V_f$	$iG$	empirical		(0, $\infty$ )
<i>Additional parameters of the 1-step factor model, <math>i = 1, \dots, k</math></i>				
$\lambda_{ii}$	$t$	empirical		$(-\infty, \infty)$
$\Gamma_{ii}$	$N$	0	1	(-.99,.99)
$v_{ii}$	$iG$	empirical		(0, $\infty$ )

Notes:  $N$  denotes the normal density,  $iG$  the inverse Gamma density,  $B$  the four-parameter Beta density, and  $t$  the three-parameter  $t$  density. Note that the prior mean and standard deviation of  $V_\mu$ ,  $V_c$ , and  $V_u$  presented in Havik et al. (2014) are partly incorrect. We use the values implemented in the actual software used by the EU Commission which is available at the 'Output Gap' interest group of the CIRCABC website of the EU Commission. In addition, the values for some countries deviate from the ones in the Table above. Specifically, prior mean and standard deviation of  $V_\mu$  for France are .3 and .2, for Spain .5 and .5, and for the Netherlands .4 and .4. The prior mean and standard deviation of  $V_c$  for Spain are 2.5 and 2. The prior mean and standard deviation of  $V_u$  for France are 2.5 and 2.5, for Italy 3 and 3, for Spain 1 and .4, and for the Netherlands 1 and 1. Finally, to achieve identification, the prior of  $\alpha$  has mean 1 and standard deviation zero in the one-step factor model.

## 4 Data

We run the models with annual data from 1980 to 2021. Since we perform a recursive revision analysis that mimics the autumn projections of the EU Commission, we take the real-time nature of the data into account as far as possible.

Log TFP is calculated as

$$\log(TFP) = \log(GDP) - 0.65 \log(L) - 0.35 \log(K),$$

where  $L$  and  $K$  are labor and capital, respectively. It is provided by the EU Commission as vintage data that reflect the information known at the end of October each year. Since TFP data for year  $T$  are not fully available in the autumn, the EU Commission uses its economic projections to fill up the missing quarters. It even provides TFP forecasts of the years  $T+1$  and  $T+2$  to alleviate the usual end-point problem of gap estimates. We follow this approach and integrate these forecasts into our estimation approach.

The CUBS indicator is also published by the EU Commission and is available in vintages from 2009 onwards when it was introduced. We construct pseudo real time vintages for earlier years.

The business cycle factor of each country is constructed from a set of 42 indicators that are similar across countries and mostly available since 1980. They consist of three parts: domestic survey indicators, domestic hard indicators, and international indicators. We use 12 domestic survey indicators, all of which are taken from the harmonized EU surveys and provided by the EU Commission. They include, inter alia, capacity utilization, economic sentiment, appreciation of new orders, and production expectations. The domestic hard indicators cover the areas production and sales (8 indicators), labor market (3 indicators), prices (3 indicators), income and consumption (4 indicators), and finance (3 indicators). The international indicators cover the euro area (5 indicators), and the US (4 indicators). For details see Appendix B.

We convert all indicators to annual frequency before we make them stationary by essentially applying the transformations recommended by Stock and Watson (2002) for US data. While we only have revised data, we take the usual publication lags into account. This entails that at the sample end we include only observations that are known at the end of October of that year. While it would be optimal to have real-time vintages, only a few of our indicators are revised markedly at the monthly or quarterly frequency and these revisions typically average out to a large extent once they are aggregated to the annual frequency. In addition, Marcellino and Musso (2011) document that data revisions are a minor source of real-time output gap revisions for the euro area and Aastveit and Trovik (2014) show that using a factor model can strongly reduce the sensitivity of output gap estimates to data revisions. Hence, a quasi real-time approach to our indicators appears sufficient to compare output gap revisions across different models.

To study the relevance of the indicators, we run the factor models with three different sets: only the domestic survey indicators (henceforth denoted by SUR), all domestic indicators (DOM), and all indicators (BIG).

To account for the possibility that an indicator leads or lags the TFP cycle, we proceed as follows. We estimate a preliminary cycle by HP filtering, shift each indicator up to four quarters forward and backwards, and compute the correlation between the preliminary cycle and the shifted indicator. We include in the indicator panel the lead or lag which results in the largest absolute correlation. The indicator sets based on the optimally time-shifted indicators are labeled SURo, DOMo, and BIGo, respectively, while the indicator sets without time shift are denoted as SUR, DOM, and BIG.

## 5 Estimation

In the following we describe estimation of the one-step factor model because it encompasses both the EU model and the two-step factor model. We follow the Bayesian approach of Havik et al. (2014) but apply it to our extended model.

### 5.1 Likelihood

Since the one-step factor model is linear and the variables are normally distributed, it is straightforward to write it in state space form and derive the likelihood function.

### 5.2 Posterior simulation

The structure of the model suggests to apply a Metropolis-within-Gibbs simulation of the posterior distribution. We briefly comment on each Gibbs step.

*Sampling the unobserved components.* To sample trend, cycle, factor, and the missing observations in the CUBS indicator and the business cycle indicators conditional on the parameters, we directly draw from the conditionally normal posterior distribution imposing a diffuse initial condition for the trend.

*Sampling  $\omega$ ,  $\rho$ ,  $V_\mu$ .* Conditional on the trend,  $p_t$ , these are the coefficients of a stationary AR(1) model with the first observation drawn from its unconditional distribution. The posterior of  $\omega$  given  $\rho$  and  $V_\mu$  is truncated normal and the posterior of  $V_\mu$  given  $\rho$  and  $\omega$  is inverse Gamma, while the posterior of  $\rho$  given  $\omega$  and  $V_\mu$  needs to be sampled by means of a Metropolis-Hastings (MH) step. The proposal distribution is a truncated normal distribution as suggested by Havik et al. (2014).

*Sampling  $A$ ,  $\tau$ ,  $V_c$ .* Conditional on the cycle,  $c_t$ , these are the coefficients of a stationary AR(2) model with the first two observations drawn from their unconditional distribution. Due to the specific parameterization, only the posterior of  $V_c$  given  $A$  and  $\tau$  is analytical (inverse Gamma). To sample  $A$  we apply a MH

step with the asymptotic normal distribution of the maximum likelihood estimator truncated to the unit interval as proposal pdf. To sample  $\tau$  we again apply a MH step with the asymptotic normal distribution of a crude maximum likelihood estimator truncated to the prespecified range as proposal pdf.

*Sampling  $\mu_u, \beta, V_u$ .* Conditional on the CUBS indicator  $u_t$ , the cycle  $c_t$ , and the AR(1) parameter  $\delta_u$ , these are coefficients of a regression model with known autoregressive error structure. Given the NiG prior, the posterior is again NiG from which we can sample easily.

*Sampling  $\delta_u$ .* Conditional on the residuals of the CUBS equation obtained in the previous step, this is the coefficient of an AR(1) model with known variance, where the first observation is drawn from its unconditional distribution. To sample  $\delta_u$  we apply a MH step with the proposal pdf being the truncated normal posterior of an AR(1) model that conditions on the first observation.

*Sampling  $\Lambda$  and  $V_x$ .* By construction of the model, we can sample the elements  $\lambda_i$  and  $v_{ii}$  equationwise. Conditional on the factor and the AR(1) parameters  $\Gamma_{ii}$ , these are coefficients of a regression model with known autoregressive error structure. Given the NiG prior, the posterior is again NiG from which we can sample easily.

*Sampling  $\Gamma$ .* Due to the diagonal structure of  $\Gamma$ , we can sample each element  $\Gamma_{ii}$  separately. Conditional on the residuals of the indicator equations obtained in the previous step,  $\Gamma_{ii}$  is the coefficient of an AR(1) model with known variance, where the first observation is drawn from its unconditional distribution. We apply a MH step with the proposal pdf being the truncated normal posterior of an AR(1) model that conditions on the first observation.

*Sampling  $\alpha$  and  $V_f$ .* Conditional on the factor  $f_t$ , the cycle  $c_t$ , and the AR(1) parameter  $\delta_f$ , these are coefficients of a regression model with known autoregressive error structure. Given the NiG prior, the posterior is again NiG from which we can sample easily.

*Sampling  $\delta_f$ .* Conditional on the residuals of the factor equation obtained in the previous step, this is the coefficient of an AR(1) model with known variance, where the first observation is drawn from its unconditional distribution. To sample  $\delta_f$  we apply a MH step with the proposal pdf being the truncated normal posterior of an AR(1) model that conditions on the first observation.

We take 10,000 draws from the Gibbs sampler of which we discard the first 2,000 as burn-in sample. Convergence is checked by a diagnostic test that compares the means of the first 10 percent and the last 40 percent of the retained draws (Geweke, 1992).

## 6 Recursive revision analysis

We conduct a recursive revision analysis of the vintage years  $T = 2005, \dots, 2021$  to quantify the year-on-year cycle revisions of the various estimators presented above. We first describe the setup and then report the results.

### 6.1 Setup

To closely mimic the situation of the EU Commission which estimates the cycle in the autumn of each year, we use only information that is available by the end of October in each of those years. In particular, all steps of the estimation procedures, such as indicator selection, factor estimation, and empirical prior specification, are repeated recursively.

Havik et al. (2014) include in each year  $T$  the EU Commission's TFP forecast of years  $T + 1$  and  $T + 2$  as if they were observations in order to alleviate the end-point problem of any filter. While it renders all results dependent on the quality of those forecasts, we nevertheless follow this approach to be as close as possible to the EU Commission. This allows us to study how changes in the estimation methodology affect the cycle revisions keeping all other modeling choices constant.

We measure the year-on-year revisions of the cycle by the root mean-squared error (RMSE). It is calculated as an average over each vintage year  $T = 2005, \dots, 2021$  of the horizons  $T + h$ , where  $h = -2, \dots, 2$ . Let  $c_{h,T}^{(C,M)}$  denote the TFP cycle estimate (in percentage points) of model  $M$  for country  $C$  and horizon  $h$  of vintage  $T$ . Then the revision RMSE for horizon  $h$  of model  $M$  and country  $C$  is defined as

$$R_h^{(C,M)} = \sqrt{\frac{1}{16} \sum_{T=2005}^{2021} \left( c_{h,T}^{(C,M)} - c_{h-1,T+1}^{(C,M)} \right)^2} / s_{2021}^{(C,M)}, \quad (15)$$

where we normalize by the standard deviation,  $s_{2021}^{(C,M)}$ , of the cycle estimated in the last vintage year. This normalization ensures that we do not prefer specifications that yield cycles of small amplitudes.

Below we report, for each country, the relative revision of model  $M$  compared to the EU model,

$$R_{h,rel}^{(C,M)} = R_h^{(C,M)} / R_h^{(C,EU \text{ model})}.$$

Since the number of vintages analyzed is only 17 and thus quite small, we also report the average relative revision of model  $M$  across all five countries,

$$R_{h,rel}^{(M)} = \frac{1}{5} \sum_C R_{h,rel}^{(C,M)}.$$

Table 2: Revisions to next vintage, average over all five countries

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
EU model	-	$\delta_u = 0$	0.39	0.49	0.57	0.98	0.92
	-		1.17	1.13	1.06	1.01	<b>0.99</b>
HP-EST	-		<b>0.99</b>	1.19	1.40	1.02	0.69
HP	-	$\lambda = 6.25$	<b>0.82</b>	1.04	1.28	<b>0.96</b>	<b>0.64</b>
	-	$\lambda = 100$	1.14	1.18	1.09	0.76	0.66
2-Step-Bay	SUR <sub>o</sub>	$\delta_u = 0$	<b>0.83</b>	<b>0.84</b>	<b>0.90</b>	1.04	1.06
	DOM <sub>o</sub>	$\delta_u = 0$	0.81	0.81	0.82	<b>0.96</b>	0.90
	BIG <sub>o</sub>	$\delta_u = 0$	0.80	0.79	0.81	1.00	0.94
2-Step-PCA	SUR <sub>o</sub>	$\delta_u = 0$	0.76	0.74	0.88	1.10	1.16
	DOM <sub>o</sub>	$\delta_u = 0$	0.78	0.76	<b>0.79</b>	1.05	1.15
	BIG <sub>o</sub>	$\delta_u = 0$	0.80	0.75	0.83	1.13	1.24
1-Step	SUR <sub>o</sub>	$\delta_u = 0, \Gamma = 0$	<b>0.69</b>	0.72	0.84	1.00	0.88
	DOM <sub>o</sub>	$\delta_u = 0, \Gamma = 0$	<b>0.69</b>	<b>0.71</b>	0.82	1.02	0.86
	BIG <sub>o</sub>	$\delta_u = 0, \Gamma = 0$	0.70	0.73	0.84	1.01	0.83

Notes: The first row shows the revision,  $R_h^{(C,M)}$ , of the EU model averaged across countries. The remaining rows show the revision  $R_{h,rel}^{(M)}$  relative to the EU model. Improvements over the EU model are marked by shaded cells, the best model is highlighted by a bold number.

## 6.2 Aggregate results of the revision analysis

The results of the revision analysis averaged over all countries are shown in Table 2. We treat the EU model with the parameter restriction  $\delta_u = 0$  as baseline because it is used by the EU Commission for all five countries except France. The restriction precludes persistent deviations between the CUBS indicator and the TFP gap thus giving large weight to the indicator. For this baseline model we report the average normalized revision of the TFP gap in percentage points in row 1. For the backcast of  $T - 2$  and  $T - 1$ , the average revision size of the EU model is 39 percent and 49 percent of the standard deviation of the cycle, while it is 57 percent for the nowcast of  $T$ . Not surprisingly, the revision size jumps upward for the out-of-sample forecasts  $T + 1$  and  $T + 2$ , exhibiting values of almost 100 percent of the cycle.

The results for the EU model without a restriction on  $\delta_u$  are shown for completeness in row 2 because this is the specification preferred by the EU Commission for France. Here and for all other models we display the average relative revision  $R_{h,rel}^{(M)}$  compared to the baseline model. Relative revisions that are larger than 1 indicate that the baseline EU model exhibits, on average, smaller revisions. For example, the average revision of the unrestricted EU model at the sample end,  $h = 0$ , is 6 percent larger than the average revision of the baseline EU model. It turns out that it does not pay off to relax the restriction  $\delta_u = 0$  although for the out-of-sample forecast horizons the EU model without parameter restriction is on par.



The HP model with estimated signal-to-noise ratio (HP-EST) is reported in row 3. It does not outperform the baseline model except for the two-year ahead forecast  $h = 2$ . The classical HP filter with  $\lambda = 6.25$  and  $\lambda = 100$  shown in rows 4 and 5 again lead – with one exception – to much larger revisions than the baseline model within the sample, i.e., for forecast horizons  $h = -2$  to  $h = 0$ . However, it works surprisingly well for the out-of-sample periods  $h = 1$  and  $h = 2$  given that the end-of-sample problem of the HP filter is well-documented in the literature.

Let us now turn to our first extension of the EU model, the two-step factor model, where the factor is estimated separately – either with a Bayesian approach or by PCA – and then is fed into the trend-cycle model. To be as comparable as possible to our baseline EU model, we impose the restriction  $\delta_u = 0$ . It turns out that the two-step factor model improves markedly over the baseline EU model no matter how the factor is estimated and which indicator set is used. It reduces the average revisions by 20-25 percent for the backcasts ( $h = -2$  and  $h = -1$ ) and by 10-20 percent for the nowcast ( $h = 0$ ), while it is comparable to the baseline EU model for the forecasts ( $h = 1$  and  $h = 2$ ).

Our second extension of the EU model is the one-step factor model which estimates the factor jointly with the trend-cycle decomposition. We again impose the restriction  $\delta_u = 0$  as in the baseline EU model. In a similar vein, we restrict  $\Gamma = 0$  and thus exclude persistent deviations between the factor and the indicators.<sup>4</sup> We find that this model is superior to both the baseline EU model and the two-step model, no matter which indicator set is used. It reduces the average revisions by around 30 percent for the backcasts ( $h = -2$  and  $h = -1$ ), almost 20 percent for the nowcast ( $h = 0$ ), and around 15 percent for the two-step ahead forecast ( $h = 2$ ), while it is on par with the baseline EU model for the one-step ahead forecast ( $h = 1$ ).

We note that the differences between the three indicator sets are fairly small. We thus prefer the survey indicator set, SURo, because the data are easy to obtain and not revised by construction. This implies that a fully-fledged real-time analysis would deliver the same results. In contrast, the revision size using the indicator sets DOMo and BIGo might increase somewhat with truly real-time data even though probably not much as discussed above.

### 6.3 Country by country results of the revision analysis

Country by country results are shown in Tables 3 to 7. In the following we will discuss, for each country, the most important differences to the results averaged over all five countries presented above.

**Germany.** The baseline EU model works comparably to the five-country

---

<sup>4</sup>We discuss the effects of relaxing these restriction in Section 7.1.

Table 3: Revisions to next vintage, Germany

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
EU model	-	$\delta_u = 0$	0.34	0.40	0.47	0.79	0.61
	-		1.06	1.08	1.05	0.90	0.93
HP-EST	-		0.82	0.85	1.03	0.73	0.54
HP	-	$\lambda = 6.25$	0.70	0.80	1.06	0.84	0.55
	-	$\lambda = 100$	1.02	1.07	1.03	0.65	0.59
2-Step-Bay	SURo	$\delta_u = 0$	0.94	0.99	0.93	1.01	1.00
	DOMo	$\delta_u = 0$	0.94	0.94	0.89	1.02	0.88
	BIGo	$\delta_u = 0$	0.94	0.91	0.87	1.02	0.89
2-Step-PCA	SURo	$\delta_u = 0$	0.90	0.77	0.68	1.15	1.39
	DOMo	$\delta_u = 0$	0.85	0.79	0.68	1.13	1.35
	BIGo	$\delta_u = 0$	0.91	0.80	0.70	1.14	1.38
1-Step	SURo	$\delta_u = 0, \Gamma = 0$	0.93	1.03	1.02	0.91	0.83
	DOMo	$\delta_u = 0, \Gamma = 0$	0.85	0.85	0.88	1.03	0.92
	BIGo	$\delta_u = 0, \Gamma = 0$	0.83	0.83	0.88	0.99	0.88

Notes: The first row shows the average revision,  $R_h^{(C,M)}$ , of the EU model. The remaining rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the EU model. Improvements over the EU model are marked by shaded cells, the best model is highlighted by a bold number.

average for the backcasts and the nowcast, while it exhibits much smaller revisions for the forecasts (see row 1 of Table 3). For example, the two-year ahead forecast revision is only 61 percent of the cycle standard deviation in Germany, while it is 92 percent in the five-country average shown in Table 2. Another notable result is that the HP filter generally outperforms the baseline EU model by far, reducing, e.g., two-year ahead forecast revision to almost 50 percent.

The two-step factor models behave similarly to the five-country average: It clearly reduces the revisions for the backcasts and nowcast, while it is comparable to the baseline EU model for the forecasts. Extracting the factor by PCA from the survey indicators set (SURo) is a particularly well suited to nowcast the TFP gap with revisions being reduced by almost one third.

The one-step factor model also improves over the baseline EU model but slightly less markedly than in the five-country average. This may reflect the extremely good performance of the baseline EU model for Germany. While the full indicator set (BIGo) is preferable, using the survey indicators set (SURo) is nevertheless a good choice, particularly with respect to forecasting the TFP gap. At the two-year horizon ( $h = 2$ ), the revisions are reduced by 17 percent.

**France.** For France, the EU Commission uses the EU model without the restriction  $\delta_u = 0$ , i.e., they allow for a persistent error term in the CUBS equation. We find that this specification (row 2 in Table 4), performs slightly worse across all forecast horizons compared to the EU baseline model which we use as baseline. As in the case of Germany, the HP filter again outperforms the baseline EU model

Table 4: Revisions to next vintage, France

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
EU model	-	$\delta_u = 0$	0.48	0.76	0.74	0.94	0.90
	-		1.11	1.08	1.03	1.01	1.02
HP-EST	-		0.82	0.75	0.77	0.72	0.54
HP	-	$\lambda = 6.25$	0.84	0.80	0.79	0.72	0.55
	-	$\lambda = 100$	0.86	0.80	0.69	<b>0.50</b>	<b>0.50</b>
2-Step-Bay	SURo	$\delta_u = 0$	0.76	0.68	0.59	0.81	0.89
	DOMo	$\delta_u = 0$	0.76	0.64	0.54	0.72	0.79
	BIGo	$\delta_u = 0$	0.77	0.65	0.57	0.86	0.97
2-Step-PCA	SURo	$\delta_u = 0$	0.81	0.76	0.76	1.02	1.22
	DOMo	$\delta_u = 0$	0.80	0.71	0.65	0.88	1.15
	BIGo	$\delta_u = 0$	0.90	0.76	0.82	1.20	1.54
1-Step	SURo	$\delta_u = 0, \Gamma = 0$	<b>0.65</b>	<b>0.56</b>	0.50	0.80	0.75
	DOMo	$\delta_u = 0, \Gamma = 0$	0.66	0.57	<b>0.47</b>	0.80	0.66
	BIGo	$\delta_u = 0, \Gamma = 0$	0.68	0.58	0.49	0.80	0.65

Notes: The first row shows the average revision,  $R_h^{(C,M)}$ , of the EU model. The remaining rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the EU model. Improvements over the EU model are marked by shaded cells, the best model is highlighted by a bold number.

by far. In particular, the HP filter with smoothing parameter  $\lambda = 100$  cuts the revisions sizes by 50 percent for forecast horizons  $h = 1$  and  $h = 2$ .

Of the two-step factor models, the one using Bayesian techniques to extract the factor in the first step is particularly well suited. It strongly reduces the revision sizes, no matter which indicator set is used.

The one-step factor model dominates all competitors for the backcast and the nowcast, reducing the revision sizes by up to 53 percent. It is also very well-suited for forecasting the TFP gap. While the best indicator set for forecasting is the full set (BIGo), the survey set (SURo) follows close behind and is optimal for backcasting.

**Italy.** For Italy, the EU Commission uses the baseline EU model. We find that relaxing the restriction  $\delta_u = 0$  yields considerably smaller revision sizes (row 2 in Table 5). Again, the HP filter improves of the baseline EU model, however this time only for forecasting but not for backcasting and nowcasting the TFP gap.

Of the two-step factor models, again the one using Bayesian techniques to extract the factor in the first step is preferable, especially if combined with the survey indicator set (SURo). It strongly reduces the revision sizes for backcasting and nowcasting the TFP gap but also outperforms the baseline model when forecasting the TFP gap.

The one-step factor model is very well suited for backcasts and nowcasts, for which it reduces the revision sizes by 25-40 percent over the baseline EU model. However, it does not outperform the benchmark in terms of forecast revisions, no

Table 5: Revisions to next vintage, Italy

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
EU model	-	$\delta_u = 0$	0.28	0.39	0.40	0.96	1.03
	-		0.98	0.86	0.76	0.87	0.85
HP-EST	-		0.95	1.28	1.52	0.84	0.64
HP	-	$\lambda = 6.25$	1.02	1.25	1.51	0.79	<b>0.60</b>
	-	$\lambda = 100$	1.57	1.36	1.25	<b>0.71</b>	0.68
2-Step-Bay	SURo	$\delta_u = 0$	0.76	0.69	0.62	0.90	0.79
	DOMo	$\delta_u = 0$	0.69	0.67	0.60	0.99	0.87
	BIGo	$\delta_u = 0$	<b>0.66</b>	0.63	<b>0.57</b>	1.09	0.97
2-Step-PCA	SURo	$\delta_u = 0$	0.69	0.61	0.67	1.01	0.93
	DOMo	$\delta_u = 0$	0.67	0.57	0.67	1.20	1.17
	BIGo	$\delta_u = 0$	<b>0.66</b>	<b>0.53</b>	0.64	1.31	1.30
1-Step	SURo	$\delta_u = 0, \Gamma = 0$	0.70	0.60	0.65	1.12	1.03
	DOMo	$\delta_u = 0, \Gamma = 0$	0.75	0.61	0.62	1.20	1.10
	BIGo	$\delta_u = 0, \Gamma = 0$	0.75	0.65	0.62	1.14	1.03

Notes: The first row shows the average revision,  $R_h^{(C,M)}$ , of the EU model. The remaining rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the EU model. Improvements over the EU model are marked by shaded cells, the best model is highlighted by a bold number.

matter which indicator set is used. Nevertheless, the survey set (SURo) is still a reasonable choice as it leads to forecast revisions that are comparable to those of the baseline EU model.

**Spain.** For Spain it is impossible to beat the baseline EU model when it comes to nowcasting or forecasting (see Table 6). In particular, both the HP filters and the two-step factor models are clearly outperformed at almost all horizons and for all indicator sets.

The one-step factor model that uses the survey indicator set (SURo) reduces the revision size of the TFP backcasts considerably over the EU baseline model but exhibits larger revisions for  $h \geq 1$ . This is surprising because it stands in contrast to the results for the other countries. We argue below that the revision stability of the baseline EU model comes at a cost: It reflects the difficulty to estimate a cycle for Spain that appears sensible at all.

Table 6: Revisions to next vintage, Spain

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
EU model	-	$\delta_u = 0$	0.32	0.34	<b>0.42</b>	<b>0.70</b>	<b>0.60</b>
	-		1.66	1.62	1.47	1.33	1.25
HP-EST	-		1.50	1.90	2.79	2.17	1.38
HP	-	$\lambda = 6.25$	0.89	1.50	2.22	1.78	1.16
	-	$\lambda = 100$	1.41	1.65	1.71	1.40	1.20
2-Step-Bay	SURo	$\delta_u = 0$	0.95	1.06	1.49	1.70	1.86
	DOMo	$\delta_u = 0$	1.01	1.09	1.35	1.48	1.44
	BIGo	$\delta_u = 0$	0.96	1.04	1.32	1.43	1.35
2-Step-PCA	SURo	$\delta_u = 0$	0.64	0.75	1.35	1.48	1.41
	DOMo	$\delta_u = 0$	0.98	1.00	1.24	1.43	1.44
	BIGo	$\delta_u = 0$	0.89	0.95	1.24	1.38	1.33
1-Step	SURo	$\delta_u = 0, \Gamma = 0$	<b>0.54</b>	<b>0.74</b>	1.33	1.33	1.13
	DOMo	$\delta_u = 0, \Gamma = 0$	0.60	0.82	1.41	1.40	1.08
	BIGo	$\delta_u = 0, \Gamma = 0$	0.62	0.86	1.48	1.43	1.06

Notes: The first row shows the average revision,  $R_h^{(C,M)}$ , of the EU model. The remaining rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the EU model. Improvements over the EU model are marked by shaded cells, the best model is highlighted by a bold number.

**Netherlands.** The baseline EU model exhibits particularly large forecast revision of up to 150 percent of the cycle standard deviation for  $h = 1$ . The HP filter can reduce these revisions considerably, particularly if the smoothing parameter  $\lambda = 100$  is chosen.

The two-step factor models strongly outperform the baseline EU model at all horizons, no matter which indicator set is used and how the factor is extracted from it. The advantage is particularly pronounced for the Bayesian factor estimation which yields reductions of the revision sizes between 25 and almost 50 percent.

The one-step factor model is comparable to the two-step factor models. It is best for backcasting and nowcasting the TFP gap but also well-suited for forecasting. Using the survey indicator set (SURo) is once again a reasonable choice as it yields excellent back- and nowcasts and also very good forecasts.

Table 7: Revisions to next vintage, Netherlands

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
EU model	-	$\delta_u = 0$	0.51	0.58	0.84	1.50	1.44
	-		1.03	1.02	0.99	0.94	0.88
HP-EST	-		0.84	1.16	0.87	0.63	0.36
HP	-	$\lambda = 6.25$	0.65	0.87	0.83	0.65	<b>0.34</b>
	-	$\lambda = 100$	0.85	1.01	0.77	<b>0.53</b>	<b>0.34</b>
2-Step-Bay	SUR <sub>o</sub>	$\delta_u = 0$	0.74	0.80	0.87	0.79	0.75
	DOM <sub>o</sub>	$\delta_u = 0$	0.65	0.73	0.74	0.60	0.53
	BIG <sub>o</sub>	$\delta_u = 0$	0.65	0.74	0.73	0.60	0.52
2-Step-PCA	SUR <sub>o</sub>	$\delta_u = 0$	0.75	0.83	0.92	0.84	0.87
	DOM <sub>o</sub>	$\delta_u = 0$	0.62	0.71	0.73	0.62	0.63
	BIG <sub>o</sub>	$\delta_u = 0$	0.62	0.72	0.74	0.63	0.65
1-Step	SUR <sub>o</sub>	$\delta_u = 0, \Gamma = 0$	0.62	<b>0.68</b>	<b>0.68</b>	0.82	0.65
	DOM <sub>o</sub>	$\delta_u = 0, \Gamma = 0$	<b>0.59</b>	0.69	0.70	0.68	0.55
	BIG <sub>o</sub>	$\delta_u = 0, \Gamma = 0$	0.61	0.72	0.71	0.68	0.52

Notes: The first row shows the average revision,  $R_h^{(C,M)}$ , of the EU model. The remaining rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the EU model. Improvements over the EU model are marked by shaded cells, the best model is highlighted by a bold number.

## 7 Sensitivity analysis

In this Section, we discuss how the results of both the two-step and the one-step factor models presented above depend on the specification choices made. We discuss various parameter restrictions (or relaxations thereof) in Section 7.1, analyze the effect of using the indicators without shifting them forward or backward in Section 7.2, assess whether preselecting only a few indicators improves the results in Section 7.3, and exclude the Covid-19 period from the sample in Section 7.4. We present and discuss the main results in the following and present more detailed results in the Tables in Appendix C.

### 7.1 Parameter restrictions

While we generally maintain the parameterization of the baseline EU model which is the core also of the extended models, we experiment with restrictions of the AR(1) parameters.  $\delta_u$ ,  $\delta_f$ , and  $\Gamma$ . Detailed results can be found in Tables 13-18 in Appendix C.1.

#### 7.1.1 The two-step factor models

Let us first consider the two-step factor models. Their baseline specification reported above imposes  $\delta_u = 0$  and thus precludes persistent deviations of the cycle

from the CUBS indicator. To check whether this restriction is too strong, we relax it and estimate completely unrestricted two-step factor models. It turns out that they exhibit larger revision sizes than our baseline specification for all countries, forecast horizons and indicator sets (except the one-year ahead forecast for Germany). Therefore, imposing  $\delta_u = 0$  and thereby mimicking the preferred specification of the EU Commission clearly pays off.

We also analyze whether a more restrictive model can reduce the revision sizes. Specifically, we add to the baseline specification the restriction that  $\delta_f = 0$  in the factor equation (8). It entails that the deviations of the factor from the cycle is non-persistent and thus appears quite strong because the factor is extracted from the indicators in the first step in an unsupervised manner, i.e., without any relationship to the TFP. Nevertheless, the results indicate that this restriction reduces the revision sizes of the backcasts and nowcasts of the TFP gap both for the five-country average and for (most of) the individual countries, while the revision sizes of the forecasts mostly increase. These results hold — with few exceptions — for both the Bayesian and the PCA approach and for all three indicator sets.

Overall, our takeaway from these sensitivity checks is that the baseline restriction  $\delta_u = 0$  is clearly advantageous, while specifications with the additional restriction  $\delta_f = 0$  of the factor equation lead to mixed results that do not uniformly dominate the baseline specification.

### 7.1.2 The one-step factor models

The baseline specification of the one-step factor model imposes both  $\delta_u = 0$  and  $\Gamma = 0$ . The latter restrictions ensures that the deviations of the individual indicators from the factor is non-persistent and appears to be strong because of the diversity of the indicators. The results indicate that relaxing one of the restrictions or both of them jointly by and large does not reduce the revision sizes.

As a final exercise, we add to the baseline specification the restriction that  $\delta_f = 0$  in the factor equation (8). The results suggest that this extra restriction leads to smaller revisions than the baseline specification for the backcasts and, for some countries, also for the nowcasts of the TFP gap. However, forecasting performance is worse, partly by a large margin. Allowing for a persistent error term in the factor equation, as in our baseline, seems to balance the in-sample and out-of-sample fit well.

We conclude that our baseline specification of the one-step factor model dominates alternative specifications in most cases.

## 7.2 Using unshifted indicators

Furthermore we check whether shifting the indicators forward or backward to maximize their correlation with the cycle as applied in our baseline specification is sensible, since it might lead to in-sample gains at the cost of a deteriorated out-of sample performance due to overfitting. However, the results shown in Tables 19 to 24 in Appendix C.2 suggest that this is not an issue. In fact, using the indicators without a time shift and then applying the one-step or two-step factor models in the baseline specification lower the revision sizes of the backcasts and nowcasts somewhat but worsen the performance of the forecasts.

We conclude that time-shifting the indicators as in our baseline specification is beneficial with respect to forecasting the TFP gap but can be omitted especially for backcasting. We nevertheless prefer the baseline specification because it already strongly outperforms the baseline EU model in terms of backcasting but is less far ahead in terms of forecasting.

## 7.3 Preselecting indicators

We also considered preselecting indicators as for example suggested by Boivin and Ng (2006), Bai and Ng (2008), Schumacher and Breitung (2008) or Fuentes et al. (2015). To this end, we followed Bai and Ng (2008) and Carstensen et al. (2020) and applied, at each forecast origin, the elastic net of Zou and Hastie (2005) to select the 5, 10, 15 and 20 most relevant indicators. However, we found strong evidence that this approach does not improve revision performance.<sup>5</sup>

The result is not surprising as we have chosen all our indicators because they have typically be proved useful for business cycle analysis and thus are relevant cyclical indicators. Now, repeatedly applying the elastic net for each forecast origin leads to changes in the indicators selected which in turn transmits to larger revision sizes, while the potential benefit of the elastic net to exclude largely irrelevant indicators does not apply in our case.

## 7.4 Excluding the Covid-19 period

Finally, the results of the revision analysis might also depend on the sample considered, especially in case of huge shocks that potentially induce a structural break. In fact, our revision sample discussed above includes the beginning of the COVID-19 pandemic in 2020 which led to unprecedented slumps in output throughout the EU. Since the European Commission still needs to estimate the TFP gaps of the member states, we check how the models reacted to this event. To this end, we repeat the revision analysis but this time choose 2019 as the last vintage year. We

---

<sup>5</sup>Detailed results are available upon request.



use our baseline setup discussed in section 6.2 with the only exception that we now normalize our revision measure (15) by the standard deviation of the estimated TFP cycle as of 2019,  $s_{2019}^{(C,M)}$ .<sup>6</sup>

The result for the five-country average are shown in Table 25 in Appendix C.3. The baseline EU model (row 1) exhibits a revision size for the backcasts and the nowcast of the TFP gap that is about the same compared to the complete vintage analysis discussed in section 6.2. Not surprisingly, however, revisions are smaller for both out-of-sample forecast horizons. The relative performance of the HP filters deteriorates. Nevertheless, it yields smaller revisions than the baseline EU model for two-step forecasts of the TFP gap.

The advantage of both the two-step and the one-step factor models over the baseline EU model with respect to backcasting and nowcasting remains qualitatively but diminishes in size, regardless which indicator set is used. However, they are outperformed in terms of forecasting one year ahead, while especially the one-step factor model is roughly on par with the baseline EU model.

The revision results for the individual countries is displayed in Tables 26 to 30. They reveal that the one-step factor model still outperforms the baseline EU model with respect to backcasting and nowcasting the TFP gap for Germany, France, Italy, and the Netherlands. The exception is Spain which now — in contrast to the full vintage sample — rather favors the baseline EU model. If it comes to forecasting, the one-step factor model yields smaller revisions than the baseline EU model for Germany, France, and the Netherlands, while it is slightly worse for Italy and severely worse for Spain. In fact, once we exclude Spain, the one-step factor model still outperforms the baseline EU model on average over Germany, France, Italy, and the Netherlands.

Overall, comparing the vintage samples with and without the COVID-19 period suggests that the factor models are more flexible than the baseline EU model to accommodate a large shock which is why they increase their advantage in the full sample. However, Spain appears to be an outlier as, e.g., the one-step factor model exhibits more than 100 percent larger revisions than the baseline EU model. We suspect that this has to do with the specific business cycle history of Spain as we discuss in the next section.

---

<sup>6</sup>Note however that the standard deviation of the cycle estimate of 2019 is very similar to the one of 2021 in almost all specifications. Therefore, using  $s_{2021}^{(C,M)}$  instead renders our results qualitatively unchanged.

## 8 Plausibility of the estimated trend and cycle components

So far we measured and discussed only the revision RMSE as an indicator of model performance. While cycle and trend are unknown, some properties of trend and cycle may nevertheless be known from external sources, for example recession periods, and estimation results should be consistent over time, i.e., over different vintages. In the following we will discuss and compare our factor model to the outcomes of the EU baseline model in these regards.

### 8.1 Decomposition of trend and cycle

One key difference of all factor models and HP filters compared to the baseline EU model is that the former tend to predict a more optimistic state of the TFP cycle at the last 5 to 10 years of a sample, including the two-year ahead TFP gap forecasts. Conversely, the trend growth rate of the baseline EU model tends to be higher. Consequently, the baseline EU model attributes low TFP growth at the sample end more likely to the cycle than to the trend, while the factor models' estimate of the trend reacts faster.

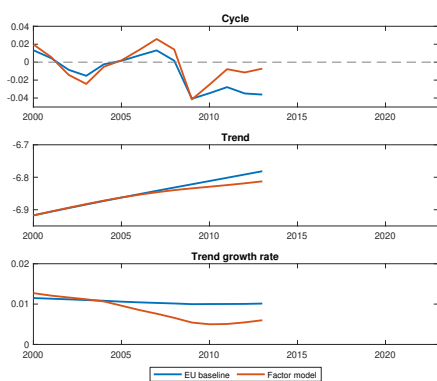
In Figure (1) we illustrate this behavior by comparing a one-step factor model (SURO, baseline specification with  $\delta_u = 0$  and  $\Gamma = 0$ , left panels) and the HP-filter with  $\lambda = 100$  (right panels) to the EU-model for the vintages 2011 (upper panels) and 2021 (lower panels) for the Netherlands as an example. In each subfigure we report the estimated cycle (to facilitate the comparison, the cycles of the factor model and HP-filter are rescaled to have the same variance as the baseline EU model), the trend as well as the trend growth rate from 2000 to the respective sample end including the two-year ahead forecasts. The horizontal dashed grey line corresponds to a TFP gap of zero.

For vintage 2011 we observe a large discrepancy between the cycle estimates of the baseline EU model (blue line) and the factor model or HP-filter (red line), especially for the years 2010-2013. The baseline EU model predicts a substantially negative cycle, while the other two models predict a lower trend growth rate. Therefore, trend estimates of the competing models are strikingly different.

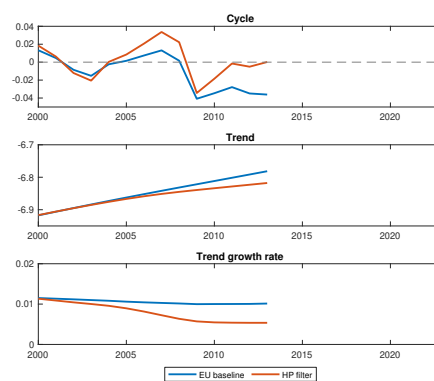
By comparing the results to the 2021 vintage shown in the lower panels, we find that the trend growth rate estimated by the baseline EU model eventually also drops considerably. In fact, in the 2021 vintage, it is almost identical to the trend growth rates of the factor model and the HP filter. Since this slowdown of the trend growth rate is pervasive across the EU, it leads to relatively large revisions of the baseline EU model, particularly for France, Italy, and the Netherlands. In fact, the factor models signal this development — which is probably a consequence

of the Great Recession — already in 2011, while the baseline EU model capture it much later, beginning with vintage 2015.

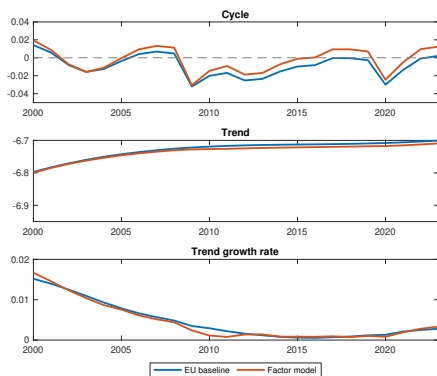
Our interpretation is that the CUBS indicator alone does not sufficiently strongly indicate in early vintages that it is not the cycle but the trend that needs to react. In contrast, our factor models include more information and attribute a larger weight to it so that they react much earlier. For the period after the Great Recession described here, this would have had strong policy implications. If the EU Commission would have found a lower trend growth and a less negative output gap, government would perhaps urged much more towards structural policies than to business cycle stabilization.



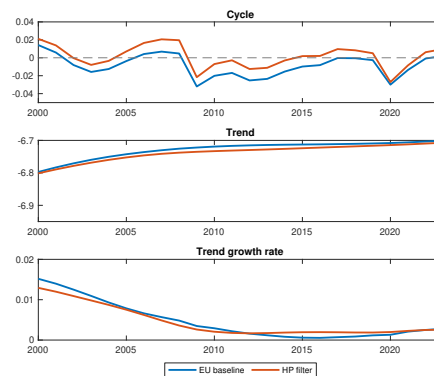
(a) SURo, vintage 2011



(b)  $\lambda = 100$ , vintage 2011



(c) SURo, vintage 2021



(d)  $\lambda = 100$ , vintage 2021

Figure 1: Cycle and trend estimates of factor model (left) and HP filter (right) compared to EU approach for the Netherlands

## 8.2 Cycle length

In Figure (2) we plot the estimated cycle of each country applying both the EU baseline model (blue) and our one-step factor model (SU<sub>Ro</sub>, baseline specification with  $\delta_u = 0$  and  $\Gamma = 0$ ) using the information available in 2021. The horizontal dashed grey line corresponds to a TFP gap of zero, and the shaded areas indicate recessions as defined by the OECD.<sup>7</sup>

For Germany both approaches lead to a fairly similar cycle even if again the TFP gap of the factor model is slightly more positive towards the sample end. Much larger differences at the sample end occur in the case of France, the Netherlands as well as Italy. For the latter we note that the periodicity of the cycle using the EU commission approach is larger (12 years on average) compared to our factor extension (9 years on average). Especially after 2008, the EU model always predicts a negative cycle.

The most striking difference however can be observed for Spain for which the estimated cycles differ substantially. The periodicity of the EU baseline is much larger (15 years) than that of the factor model (10 years on average). In particular, the EU model predicts a positive TFP gap for all twenty years from 1985 to 2004 and a consistently negative TFP gap thereafter. This roughly coincides with the long housing boom and subsequent bust in Spain. Our interpretation is that the CUBS indicator did not strongly enough signal that the bust was to a relevant extent structural — after all, many resources of the construction sector became unproductive — as opposed to only cyclical. Again using a broader information base appears to deliver a more realistic picture of the business cycle stance.

---

<sup>7</sup>FRED data, derived from the OECD Composite Leading Indicators: Reference Turning Points and Component Series.

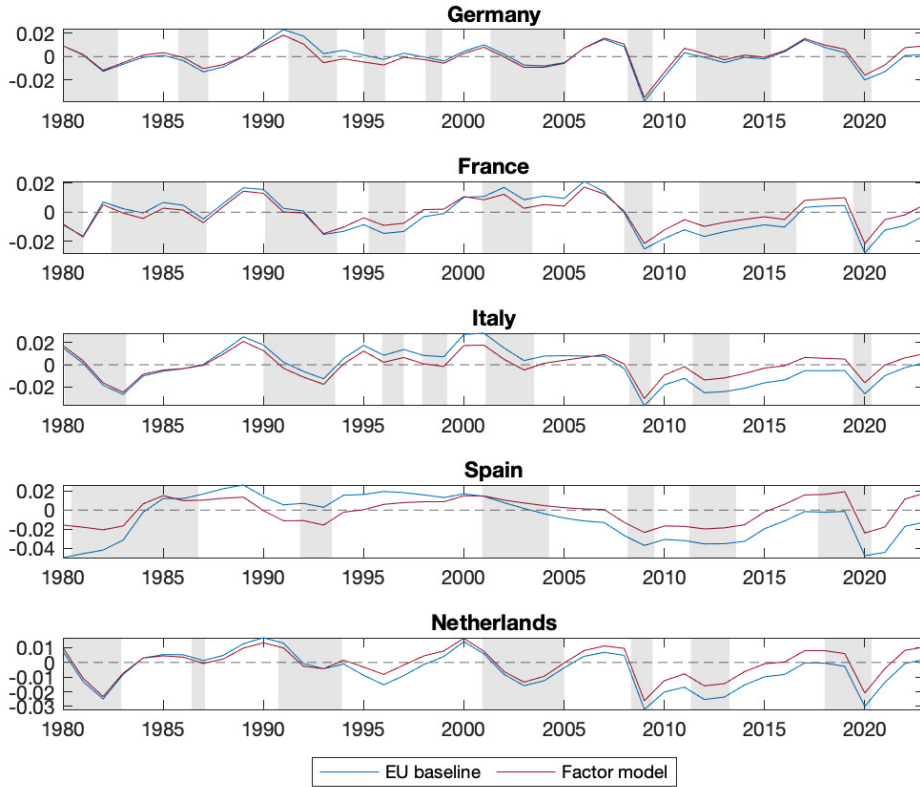


Figure 2: Estimated cycles for vintage 2021

## 9 Conclusion

In this paper we showed that estimating and forecasting the TFP gap of the largest five EU countries — and presumably also for the smaller countries — can benefit considerably from including a set of business cycle indicators via a factor model extension in the otherwise unchanged unobserved components model of the EU Commission. Based on a vintage of data as of October of a year  $T$ , the advantage is particularly pronounced for the last two in-sample years  $T - 2$  and  $T - 1$ , for which the year-to-year revisions can be reduced by up to 30 percent, and for the nowcast year  $T$ , for which the reduction is around 20 percent. Improvements for the two out-of-sample years  $T + 1$  and  $T + 2$  are also feasible but quantitatively smaller (around 10 percent) and dependent on the country specifics.

We also argued that the EU model produces implausible results for Spain. Especially for this country, its estimate of the trend growth rate is very stable over vintages leading to small revisions and thereby outperforming our factor models.

However, its cycle length is implausibly long indicating that Spain went through only one upswing and one recession since 1985. A factor constructed from a set of business cycle indicators indicates that this is not a complete description of the Spanish business cycle history which is why the factor models yield much smaller cycle lengths.

In a sensitivity analysis we showed that the results of the baseline specification of our factor models yield very good results compared to relevant alternatives. Overall, the results do not depend overly on the specification of the factor model or the indicator set chosen — it appears sufficient to include “enough” relevant information.

It remains to be analyzed which role the prior choices made by the EU commission play. In an attempt to make our factor models as comparable as possible to the baseline EU model, we have refrained from optimizing on this margin.

## References

- Aastveit, K. A. and Trovik, T. (2014). Estimating the output gap in real time: A factor model approach. *The Quarterly Review of Economics and Finance*, 54(2):180–193.
- Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317.
- Barigozzi, M. and Luciani, M. (2019). Measuring the Output Gap Using Large Datasets. Technical report, Available at SSRN.
- Basistha, A. and Startz, R. (2008). Measuring the NAIRU with reduced uncertainty: a multiple-indicator common-cycle approach. *The Review of Economics and Statistics*, 90(4):805–811.
- Blagrove, P., Garcia-Saltos, M. R., Laxton, M. D., and Zhang, F. (2015). *A simple multivariate filter for estimating potential output*. Number 15-79. International Monetary Fund.
- Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132:169–194.
- Carstensen, K., Heinrich, M., Reif, M., and Wolters, M. H. (2020). Predicting Ordinary and Severe Recessions with a Three-State Markov-Switching Dynamic Factor Model — An Application to the German Business Cycle. *International Journal of Forecasting*, forthcoming.

- D’Auria, F., Denis, C., Havik, K., Morrow, K. M., Planas, C., Raciborski, R., Röger, W., and Rossi, A. (2010). The production function methodology for calculating potential growth rates and output gaps. Technical Report Economic Papers 420, European Economy.
- Dovern, J. and Zuber, C. (2019). Revisions of potential output estimates in the EU after the Great Recession. *CESifo Working Paper*, (7681).
- Fleischman, C. and Roberts, J. (2011). From many series, one cycle: improved estimates of the business cycle from a multivariate unobserved components model. Finance and Economics Discussion Series 2011-46, Board of Governors of the Federal Reserve System (U.S.).
- Fuentes, J., Poncela, P., and Rodríguez, J. (2015). Sparse partial least squares in time series for macroeconomic forecasting. *Journal of Applied Econometrics*, 30(4):576–595.
- Geweke, J. (1992). *Bayesian statistics*, volume 4, chapter Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, pages 169–193. Oxford University Press.
- González-Astudillo, M. (2019a). An output gap measure for the euro area: Exploiting country-level and cross-sectional data heterogeneity. *European Economic Review*, 120(C).
- González-Astudillo, M. (2019b). Estimating the U.S. output gap with state-level data. *Journal of Applied Econometrics*, 34(5):795–810.
- Graff, M. and Sturm, J.-E. (2010). The information content of capacity utilisation rates for output gap estimates. KOF Working papers 10-269, KOF Swiss Economic Institute, ETH Zurich.
- Harvey, A. and Trimbur, T. (2008). Trend estimation and the Hodrick-Prescott filter. *Journal of the Japanese Statistical Society*, 380(1):41–49.
- Havik, K., Morrow, K. M., Orlandi, F., Planas, C., Raciborski, R., Röger, W., Rossi, A., Thum-Thysen, A., and Vandermeulen, V. (2014). The Production Function Methodology for Calculating Potential Growth Rates & Output Gaps. Technical report, EU Commission Economic Papers 535.
- Huber, F., Pfarrhofer, M., and Piribauer, P. (2020). A multi-country dynamic factor model with stochastic volatility for euro area business cycle analysis. *Journal of Forecasting*, n/a(n/a).

- Jarociński, M. and Lenza, M. (2018). An inflation-predicting measure of the output gap in the euro area. *Journal of Money, Credit and Banking*, 50(6):1189–1224.
- Kuttner, K. N. (1994). Estimating potential output as a latent variable. *Journal of Business & Economic Statistics*, 12(3):361–368.
- Marcellino, M. and Musso, A. (2011). The reliability of real-time estimates of the euro area output gap. *Economic Modelling*, 28(4):1842 – 1856.
- Melolinna, M. and Toth, M. (2019). Output gaps, inflation and financial cycles in the uk. *Empirical Economics*, 56(3):1039–1070.
- Mise, E., Kim, T.-H., and Newbold, P. (2005). On suboptimality of the hodrick–prescott filter at time series endpoints. *Journal of Macroeconomics*, 27(1):53 – 67.
- Orphanides, A. and van Norden, S. (2002). The unreliability of output-gap estimates in real time. *The Review of Economics and Statistics*, 84(4):569–583.
- Planas, C., Roeger, W., and Rossi, A. (2013). The information content of capacity utilization for detrending total factor productivity. *Journal of Economic Dynamics and Control*, 37(3):577 – 590.
- Planas, C., Rossi, A., and Fiorentini, G. (2008). Bayesian analysis of the output gap. *Journal of Business & Economic Statistics*, 26:18–32.
- Ravn, M. O. and Uhlig, H. (2002). On adjusting the Hodrick-Prescott filter for the frequency of observations. *Review of Economics and Statistics*, 84(2):371–380.
- Schumacher, C. and Breitung, J. (2008). Real-time forecasting of German GDP based on a large factor model with monthly and quarterly data. *International Journal of Forecasting*, 24(3):386–398.
- Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.
- Turner, D., Cavalleri, M. C., Guillemette, Y., Kopoin, A., Ollivaud, P., and Rusticelli, E. (2016). An investigation into improving the real-time reliability of OECD output gap estimates. OECD Economics Department Working Papers 1294, OECD Publishing.
- Valle e Azevedo, J., Koopman, S. J., and Rua, A. (2006). Tracking the business cycle of the euro area: A multivariate model-based bandpass filter. *Journal of Business & Economic Statistics*, 24:278–290.



Weiske, S. (2018). Indicator-based estimates of the output gap in the euro area. Technical Report Working Paper No. 12/2018, Sachverständigenrat zur Begutachtung der Gesamtwirtschaftlichen Entwicklung.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.

# A Estimation of factor models

## A.1 Factor model estimated by principal components

To estimate a single business cycle factor from a set of  $k$  standardized indicators by principal component analysis, we apply the EM algorithm of Stock and Watson (2002) to cope with missing data at the beginning of the sample. It is initialized by inserting zeros where observations are missing in the  $T \times k$  data matrix  $\mathbf{X}$ . In the first step, the factor and its loading vector are estimated by standard principal component analysis. In the second step, the missing observations in  $\mathbf{X}$  are updated by setting them to factor times loading coefficient. We iterate these two steps until convergence to obtain a final estimate of the factor.

## A.2 Factor model estimated with Bayesian techniques

### A.2.1 The model

The purpose of the model is to extract a single factor,  $f_t$ , from a  $k \times 1$  vector of observed business cycle indicators,  $x_t$ , which are standardized. The measurement equation is

$$x_t = \Lambda f_t + v_t, \quad t = 1, \dots, T, \quad v_t \stackrel{\text{iid}}{\sim} N(0, \Sigma_v), \quad (16)$$

where  $\Sigma_v$  is diagonal with diagonal elements  $\sigma_1^2, \dots, \sigma_k^2$ . The factor is assumed to follow an AR(2) process,

$$f_t = 2\varphi \cos(2\pi/\varrho) f_{t-1} - \varphi^2 f_{t-2} + w_t, \quad t = 3, \dots, T, \quad w_t \stackrel{\text{iid}}{\sim} N(0, 1), \quad (17)$$

with the first two observations being drawn from the unconditional distribution. The parameterization follows Havik et al. (2014). Assuming  $0 < \varrho < 1$  and  $2 < \varphi < 32$  ensures stationarity.

Defining  $\Lambda = (\lambda_1, \dots, \lambda_k)'$ , identification of the factor is achieved by setting the first element to one, i.e.,  $\lambda_1 = 1$ .

The parameters of interest are defined as  $\theta = (\lambda_2, \dots, \lambda_k, \sigma_1^2, \dots, \sigma_k^2, \varrho, \varphi)'$ .

### A.2.2 Priors

Conditional on the factor, (16) defines a set of  $k$  independent regression equations of which the first one is special because its mean parameter is restricted to one. To obtain closed-form solutions, we use conjugate priors. For the first regression equation we thus use an inverse Gamma<sup>8</sup> prior,

$$\sigma_1^2 \sim iG(\underline{s}_1^{-2}, \underline{\nu}_1), \quad (18)$$

---

<sup>8</sup>We parameterize the pdf of the inverse Gamma distribution,  $y \sim iG(s^{-2}, \nu)$ , as  $p(y) = c_{iG} y^{-\nu/2-1} \exp(-0.5\nu s^2/y)$ , where  $c_{iG}$  is an integration constant.

and for the remaining equations we use Normal-inverse Gamma priors,

$$\lambda_i | \sigma_i^2 \sim N(\underline{\lambda}_i, \sigma_i^2 \underline{V}), \quad \sigma_i^2 \sim iG(\underline{s}_i^{-2}, \underline{\nu}_i). \quad (19)$$

For the AR(2) parameters we follow Havik et al. (2014) and use

$$\varphi \sim B(\underline{\alpha}_1, \underline{\beta}_1, 0, 1) \quad (20)$$

and

$$\varrho \sim B(\underline{\alpha}_2, \underline{\beta}_2, 2, 32), \quad (21)$$

where  $B(\alpha, \beta, a, b)$  is the Beta distribution with parameters  $\alpha$  and  $\beta$  rescaled to the interval  $(a, b)$ .

The implied prior for  $f_1, f_2$  is

$$[f_1, f_2]' | \varrho, \varphi \sim N(0, \Upsilon(\varrho, \varphi)), \quad (22)$$

where  $\Upsilon(\varrho, \varphi)$  is the unconditional  $2 \times 2$  variance matrix of an AR(2) process and thus a function of  $\varrho$  and  $\varphi$ . For  $f_3, \dots, f_T$  the prior can be written as

$$f_t | f_{t-1}, f_{t-2}, \varrho, \varphi \sim N(2\varrho \cos(2\pi/\varphi) f_{t-1} - \varphi^2 f_{t-2}, 1). \quad (23)$$

### A.2.3 Posterior simulation

The structure of the model suggests to apply a Metropolis-within-Gibbs simulation of the posterior distribution. We briefly comment on each Gibbs step.

*Sampling unobserved components.* To sample factor and the missing observations in the indicators conditional on the parameters, we directly draw from the conditionally normal posterior distribution.

*Sampling  $\varphi$  and  $\varrho$ .* Conditional on the factor,  $f_t$ , these are the coefficients of a stationary AR(2) model with the first two observations drawn from their unconditional distribution. To sample  $\varrho$  we apply a MH step with the asymptotic normal distribution of the maximum likelihood estimator truncated to the unit interval as proposal pdf. To sample  $\varphi$  we again apply a MH step with the asymptotic normal distribution of a crude maximum likelihood estimator truncated to the prespecified range as proposal pdf.

*Sampling  $\lambda_2, \dots, \lambda_k$  and  $\sigma_1^2, \dots, \sigma_k^2$ .* Conditional on everything else the likelihood is the standard likelihood of  $k$  independent regressions. As prior we choose the natural conjugate Normal-Inverse-Gamma prior. We have to make the following case distinction:

- Due to identification we set  $\lambda_1 = 1$  which implies an inverse-Gamma posterior for  $\sigma_1^2$ .
- For  $i = 2, \dots, k$  the posterior of  $\lambda_i, \sigma_i^2$  is also Normal-Inverse-Gamma.

## B Data

Table 8: Indicators for Germany

Indicator	Freq.	Provider	Source	Mnemonic	Tr.	Qu.
<i>Hard indicators</i>						
Total industrial production	m	FRED	OECD MEI	DEUPROINDMISMEI	3	3
Manufacturing production	m	FRED	OECD MEI	DEUPROMANMISMEI	3	3
Investment goods production	m	FRED	OECD MEI	DEUPRMNVG01IXOBSAM	3	3
Construction	m	FRED	OECD MEI	DEUPROCONMISMEI	3	3
Retail sales	m	FRED	OECD MEI	DEUSARTMISMEI	3	3
Passenger car registrations	m	FRED	OECD MEI	DEUSLRTC03IXOBSAM	3	3
Building permits	m	FRED	OECD MEI	DEUODCNPI03MLSAM	3	3
Unemployment rate	m	FRED	OECD MEI	LMUNRRTTDEM156S	2	3
Number of employees	q	FRED	OECD MEI	LFEMTTTTDEQ647S	3	3
Hours worked	q	Eurostat	Eurostat		3	3
CPI	m	FRED	OECD MEI	DEUCPIALLMINMEI	3	3
PPI	m	FRED	OECD MEI	DEUPPDMMINMEI	3	3
House prices	q	FRED	BIS	QDEN628BIS	3	3
GDP	q	FRED	OECD MEI	NAEXKP01DEQ661S	3	2
Income	q	FRED	OECD QNA	DEUCOMPQDSNAQ	3	2
Hourly earnings	q	FRED	OECD MEI	DEUHOUREAQISMEI	3	2
Private consumption	q	FRED	OECD MEI	NAEXKP02DEQ661S	3	2
Interest rate spread	m	FRED	OECD	DEULOCOSIORSTM	1	3
Stock market index	m	FRED	OECD MEI	SPASTT01DEM661N	3	3
REER	m	FRED	BIS	RNDEBIS	3	3
<i>Survey indicators</i>						
Capacity utilization	q	DG EcFin	DG EcFin	INDU.DE.TOT.13.QPS.Q	1	3
New Orders, total	q	DG EcFin	DG EcFin	INDU.DE.TOT.11.BS.Q	1	3
New Orders, Intermediate goods	q	DG EcFin	DG EcFin	INDU.DE.INTM.11.BS.Q	1	3
New Orders, Investment goods	q	DG EcFin	DG EcFin	INDU.DE.INVE.11.BS.Q	1	3
Assessment of order-book levels	m	DG EcFin	DG EcFin	INDU.DE.TOT.2.BS.M	1	3
Production expectations	m	DG EcFin	DG EcFin	INDU.DE.TOT.5.BS.M	1	3
Industrial confidence	m	DG EcFin	DG EcFin	DE.INDU	1	3
Services confidence	m	DG EcFin	DG EcFin	DE.SERV	1	3
Consumers confidence	m	DG EcFin	DG EcFin	DE.CONS	1	3
Retail confidence	m	DG EcFin	DG EcFin	DE.RETA	1	3
Construction confidence	m	DG EcFin	DG EcFin	DE.BUIL	1	3
Economic sentiment	m	DG EcFin	DG EcFin	DE.ESI	1	3
<i>International indicators</i>						
Economic sentiment, Euro Area	m	DG EcFin	DG EcFin	EA.ESI	1	3
Unemployment rate, Euro Area	m	FRED	OECD MEI	LRHUTTTTEZM156S	1	3
Eurocoin	m	CEPR	CEPR	Eurocoin	1	3
European Stock market index	m	FRED	OECD	SPASTT01EZM661N	3	3
Oil price / Brent	m	FRED	IMF	POILBREUSDM	3	3
Employment, US	m	FRED	BLS	PAYEMS	3	3
GDP, US	q	FRED	BEA	GDPC1	3	2
Real personal income, US	m	FRED	BEA	RPI	3	3
Federal Funds Rate, US	m	FRED	Fed	FEDFUNDS	2	3

Notes: m and q denote monthly and quarterly frequency. Tr denotes the stationarity-generating transformations: 1 = level, 2 = difference, 3 = log difference. Qu denotes the last quarter of a year available at the end of October of a year.

Table 9: Indicators for France

Indicator	Freq.	Provider	Source	Mnemonic	Tr.	Qu.
<i>Hard indicators</i>						
Total industrial production	m	FRED	OECD MEI	FRAPROINDMISMEI	3	3
Manufacturing production	m	FRED	OECD MEI	FRAPROMANMISMEI	3	3
Intermediate goods production	m	FRED	OECD MEI	FRAPRMNIG01IXOBSAM	3	3
Investment goods production	m	FRED	OECD MEI	FRAPRMNVG01IXOBSAM	3	3
Construction	m	FRED	OECD MEI	FRAPROCONMISMEI	3	3
Retail sales	m	FRED	OECD MEI	FRASARTMISMEI	3	3
Passenger car registrations	m	FRED	OECD MEI	FRASLRTCRO3IXOBSAM	3	3
Building permits	m	FRED	OECD MEI	FRAPERMITMISMEI	3	3
Unemployment rate	m	FRED	OECD MEI	LRHUTTTTFRM156S	2	3
Number of employees	q	FRED	OECD MEI	LFEMTTTTFRQ647S	3	3
Hours worked	q	Eurostat	Eurostat		3	3
CPI	m	FRED	OECD MEI	FRACPIALLMINMEI	3	3
PPI	m	FRED	OECD MEI	FRAPPDMMINMEI	3	3
House prices	q	FRED	BIS	QFRN628BIS	3	3
GDP	q	FRED	OECD MEI	NAEXKP01FRQ661S	3	2
Income	q	FRED	OECD QNA	FRACOMPQDSNAQ	3	2
Hourly earnings	q	FRED	OECD MEI	LCEAMN01FRQ661S	3	2
Private consumption	q	FRED	OECD MEI	NAEXKP02FRQ661S	3	2
Interest rate spread	m	OECD	OECD		1	3
Stock market index	m	FRED	OECD MEI	SPASTT01FRM661N	3	3
REER	m	FRED	BIS	RNFRBIS	3	3
<i>Survey indicators</i>						
Capacity utilization	q	DG EcFin	DG EcFin	INDU.FR.TOT.13.QPS.Q	1	3
New Orders, total	q	DG EcFin	DG EcFin	INDU.FR.TOT.11.BS.Q	1	3
New Orders, Intermediate goods	q	DG EcFin	DG EcFin	INDU.FR.INTM.11.BS.Q	1	3
New Orders, Investment goods	q	DG EcFin	DG EcFin	INDU.FR.INVE.11.BS.Q	1	3
Assessment of order-book levels	m	DG EcFin	DG EcFin	INDU.FR.TOT.2.BS.M	1	3
Production expectations	m	DG EcFin	DG EcFin	INDU.FR.TOT.5.BS.M	1	3
Industrial confidence	m	DG EcFin	DG EcFin	FR.INDU	1	3
Services confidence	m	DG EcFin	DG EcFin	FR.SERV	1	3
Consumers confidence	m	DG EcFin	DG EcFin	FR.CONF	1	3
Retail confidence	m	DG EcFin	DG EcFin	FR.RETA	1	3
Construction confidence	m	DG EcFin	DG EcFin	FR.BUIL	1	3
Economic sentiment (ESI)	m	DG EcFin	DG EcFin	FR.ESI	1	3

Notes: m and q denote monthly and quarterly frequency. Tr denotes the stationarity-generating transformations: 1 = level, 2 = difference, 3= log difference. Qu denotes the last quarter of a year available at the end of October of a year. The international indicators are the same as shown for Germany and thus not repeated here.

Table 10: Indicators for Italy

Indicator	Freq.	Provider	Source	Mnemonic	Tr.	Qu.
<i>Hard indicators</i>						
Total industrial production	m	FRED	OECD MEI	ITAPROINDMISMEI	3	3
Manufacturing production	m	FRED	OECD MEI	ITAPROMANMISMEI	3	3
Intermediate goods production	m	FRED	OECD MEI	ITAPRMNIG01IXOBSAM	3	3
Investment goods production	m	FRED	OECD MEI	ITAPRMNVG01IXOBSAM	3	3
Construction	m	FRED	OECD MEI	ITAPRCNT001IXOBSAM	3	3
Retail sales	m	FRED	OECD MEI	ITASARTMISMEI	3	3
Passenger car registrations	m	FRED	OECD MEI	ITASLRTC03IXOBSAM	3	3
Building permits	q	ISTAT	ISTAT		3	3
Unemployment rate	m	FRED	OECD MEI	LRHUTTTTITM156S	2	3
Number of employees	q	FRED	OECD MEI	LFEMTTTTITQ647S	3	3
Hours worked	q	Eurostat	Eurostat		3	3
CPI	m	FRED	OECD MEI	ITACPIALLMINMEI	3	3
PPI	m	FRED	OECD MEI	ITAPPDMMINMEI	3	3
House prices	q	FRED	BIS	QITN628BIS	3	3
GDP	q	FRED	OECD MEI	LORSGPORITQ661S	3	2
Income	q	FRED	OECD QNA	ITACOMPQDSNAQ	3	2
Hourly earnings	q	FRED	OECD MEI	LCEAMN01ITQ661S	3	2
Private consumption	q	FRED	OECD MEI	NAEXKP02ITQ661S	3	2
Interest rate spread	m	FRED	OECD		1	3
Stock market index	m	FRED	OECD MEI	SPASTT01ITM661N	3	3
REER	m	FRED	BIS	RNITBIS	3	3
<i>Survey indicators</i>						
Capacity utilization	q	DG EcFin	DG EcFin	INDU.IT.TOT.13.QPS.Q	1	3
New Orders, total	q	DG EcFin	DG EcFin	INDU.IT.TOT.11.BS.Q	1	3
New Orders, Intermediate goods	q	DG EcFin	DG EcFin	INDU.IT.INTM.11.BS.Q	1	3
New Orders, Investment goods	q	DG EcFin	DG EcFin	INDU.IT.INVE.11.BS.Q	1	3
Assessment of order-book levels	m	DG EcFin	DG EcFin	INDU.IT.TOT.2.BS.M	1	3
Production expectations	m	DG EcFin	DG EcFin	INDU.IT.TOT.5.BS.M	1	3
Industrial confidence	m	DG EcFin	DG EcFin	IT.INDU	1	3
Services confidence	m	DG EcFin	DG EcFin	IT.SERV	1	3
Consumers confidence	m	DG EcFin	DG EcFin	IT.CON	1	3
Retail confidence	m	DG EcFin	DG EcFin	IT.RETA	1	3
Construction confidence	m	DG EcFin	DG EcFin	IT.BUIL	1	3
Economic sentiment (ESI)	m	DG EcFin	DG EcFin	IT.ESI	1	3

Notes: m and q denote monthly and quarterly frequency. Tr denotes the stationarity-generating transformations: 1 = level, 2 = difference, 3= log difference. Qu denotes the last quarter of a year available at the end of October of a year. The international indicators are the same as shown for Germany and thus not repeated here.

Table 11: Indicators for Spain

Indicator	Freq.	Provider	Source	Mnemonic	Tr.	Qu.
<i>Hard indicators</i>						
Total industrial production	m	FRED	OECD MEI	ESPPROINDMISMEI	3	3
Manufacturing production	m	FRED	OECD MEI	ESPPROMANMISMEI	3	3
Intermediate goods production	m	FRED	OECD MEI	ESPPRMNIG01IXOBSAM	3	3
Investment goods production	m	FRED	OECD MEI	ESPPRMNVG01IXOBSAM	3	3
Construction	m	FRED	OECD MEI	ESPPROCONMISMEI	3	3
Retail sales	m	FRED	OECD MEI	ESPSARTMISMEI	3	3
Passenger car registrations	m	FRED	OECD MEI	ESPSLRTC03MLSAM	3	3
Building permits	m	FRED	OECD MEI	ESPPERMITMISMEI	3	3
Unemployment rate	m	FRED	OECD MEI	LRHUTTTTESM156S	2	3
Number of employees	q	FRED	OECD MEI	LFEMTTTTESQ647S	3	3
Hours worked	q	Eurostat	Eurostat		3	3
CPI	m	FRED	OECD MEI	ESPCPIALLMINMEI	3	3
PPI	m	FRED	OECD MEI	ESPPDMMINMEI	3	3
House prices	q	FRED	BIS	QESN628BIS	3	3
GDP	q	FRED	OECD MEI	LORSGPORESQ661S	3	2
Income	q	Eurostat	Eurostat		3	2
Hourly earnings	q	FRED	OECD MEI	LCEAMN01ESQ661S	3	2
Private consumption	q	FRED	OECD MEI	NAEXKP02ESQ661S	3	2
Interest rate spread	m	OECD	OECD		1	3
Stock market index	m	FRED	OECD MEI	SPASTT01ESM661N	3	3
REER	m	FRED	BIS	RNESBIS	3	3
<i>Survey indicators</i>						
Capacity utilization	q	DG EcFin	DG EcFin	INDU.ES.TOT.13.QPS.Q	1	3
New Orders, total	q	DG EcFin	DG EcFin	INDU.ES.TOT.11.BS.Q	1	3
New Orders, Intermediate goods	q	DG EcFin	DG EcFin	INDU.ES.INTM.11.BS.Q	1	3
New Orders, Investment goods	q	DG EcFin	DG EcFin	INDU.ES.INVE.11.BS.Q	1	3
Assessment of order-book levels	m	DG EcFin	DG EcFin	INDU.ES.TOT.2.BS.M	1	3
Production expectations	m	DG EcFin	DG EcFin	INDU.ES.TOT.5.BS.M	1	3
Industrial confidence	m	DG EcFin	DG EcFin	ES.INDU	1	3
Services confidence	m	DG EcFin	DG EcFin	ES.SERV	1	3
Consumers confidence	m	DG EcFin	DG EcFin	ES.CONS	1	3
Retail confidence	m	DG EcFin	DG EcFin	ES.RETA	1	3
Construction confidence	m	DG EcFin	DG EcFin	ES.BUIL	1	3
Economic sentiment (ESI)	m	DG EcFin	DG EcFin	ES.ESI	1	3

Notes: m and q denote monthly and quarterly frequency. Tr denotes the stationarity-generating transformations: 1 = level, 2 = difference, 3= log difference. Qu denotes the last quarter of a year available at the end of October of a year. The international indicators are the same as shown for Germany and thus not repeated here.

Table 12: Indicators for Netherlands

Indicator	Freq.	Provider	Source	Mnemonic	Tr.	Qu.
<i>Hard indicators</i>						
Total industrial production	m	FRED	OECD MEI	NLDPROINDMISMEI	3	3
Manufacturing production	m	FRED	OECD MEI	NLDPROMANMISMEI	3	3
Intermediate goods production	m	FRED	OECD MEI	NLDPRMNIG01IXOBSAM	3	3
Investment goods production	m	FRED	OECD MEI	NLDPRMNVG01IXOBSAM	3	3
Construction	m	FRED	OECD MEI	NLDPROCONMISMEI	3	3
Retail sales	m	FRED	OECD MEI	NLDSARTMISMEI	3	3
Passenger car registrations	m	FRED	OECD MEI	NLDSLRTCR03MLSAM	3	3
Building permits	m	FRED	OECD MEI	NLDPERMITMISMEI	3	3
Unemployment rate	m	FRED	OECD MEI	LRHUTTTTNLM156S	2	3
Number of employees	q	FRED	OECD MEI	LFEMTTTTNLQ647S	3	3
Hours worked	q	Eurostat	Eurostat		3	3
CPI	m	FRED	OECD MEI	NLDCPIALLMINMEI	3	3
PPI	m	FRED	OECD MEI	NLDPPDMMINMEI	3	3
House prices	q	FRED	BIS	QNLN628BIS	3	3
GDP	q	FRED	OECD MEI	LORSGPORNQ661S	3	2
Income	q	Eurostat	Eurostat		3	2
Hourly earnings	q	FRED	OECD MEI	LCEAMN01NLQ661S	3	2
Private consumption	q	FRED	OECD MEI	NAEXKP02NLQ661S	3	2
Interest rate spread	m	OECD	OECD		1	3
Stock market index	m	FRED	OECD MEI	SPASTT01NLM661N	3	3
REER	m	FRED	BIS	RNNLBIS	3	3
<i>Survey indicators</i>						
Capacity utilization	q	DG EcFin	DG EcFin	INDU.NL.TOT.13.QPS.Q	1	3
New Orders, total	q	DG EcFin	DG EcFin	INDU.NL.TOT.11.BS.Q	1	3
New Orders, Intermediate goods	q	DG EcFin	DG EcFin	INDU.NL.INTM.11.BS.Q	1	3
New Orders, Investment goods	q	DG EcFin	DG EcFin	INDU.NL.INVE.11.BS.Q	1	3
Assessment of order-book levels	m	DG EcFin	DG EcFin	INDU.NL.TOT.2.BS.M	1	3
Production expectations	m	DG EcFin	DG EcFin	INDU.NL.TOT.5.BS.M	1	3
Industrial confidence	m	DG EcFin	DG EcFin	NL.INDU	1	3
Services confidence	m	DG EcFin	DG EcFin	NL.SERV	1	3
Consumers confidence	m	DG EcFin	DG EcFin	NL.CON	1	3
Retail confidence	m	DG EcFin	DG EcFin	NL.RETA	1	3
Construction confidence	m	DG EcFin	DG EcFin	NL.BUIL	1	3
Economic sentiment (ESI)	m	DG EcFin	DG EcFin	NL.ESI	1	3

Notes: m and q denote monthly and quarterly frequency. Tr denotes the stationarity-generating transformations: 1 = level, 2 = difference, 3= log difference. Qu denotes the last quarter of a year available at the end of October of a year. The international indicators are the same as shown for Germany and thus not repeated here.



## C Results of the sensitivity analysis

### C.1 Alternative parameter restrictions

Table 13: Factor models with alternative parameter restrictions, revisions to next vintage, average over all five countries

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
2-Step-Bay	SURo		0.95	0.96	1.02	1.08	1.12
	DOMo		0.89	0.88	0.90	1.00	0.97
	BIGo		0.86	0.85	0.88	1.04	1.01
	SURo	$\delta_u = 0, \delta_f = 0$	<b>0.79</b>	<b>0.80</b>	<b>0.88</b>	1.06	1.10
	DOMo	$\delta_u = 0, \delta_f = 0$	<b>0.74</b>	<b>0.73</b>	<b>0.78</b>	1.01	<b>0.93</b>
	BIGo	$\delta_u = 0, \delta_f = 0$	<b>0.73</b>	<b>0.72</b>	<b>0.79</b>	1.05	1.01
2-Step-PCA	SURo		0.81	0.80	0.94	1.11	1.19
	DOMo		0.84	0.82	0.86	1.09	1.19
	BIGo		0.84	0.81	0.89	1.15	1.28
	SURo	$\delta_u = 0, \delta_f = 0$	<b>0.72</b>	<b>0.70</b>	<b>0.85</b>	1.12	1.17
	DOMo	$\delta_u = 0, \delta_f = 0$	<b>0.72</b>	<b>0.67</b>	<b>0.74</b>	1.09	1.15
	BIGo	$\delta_u = 0, \delta_f = 0$	<b>0.75</b>	<b>0.68</b>	<b>0.81</b>	1.17	1.27
1-Step	SURo		0.75	0.80	0.90	1.00	0.90
	DOMo		0.75	0.79	0.90	1.08	0.90
	BIGo		0.76	0.80	0.91	1.05	0.87
	SURo	$\delta_u = 0$	0.70	0.74	0.84	<b>0.99</b>	0.88
	DOMo	$\delta_u = 0$	0.69	0.72	0.83	1.05	0.89
	BIGo	$\delta_u = 0$	0.71	0.73	0.85	1.02	0.85
	SURo	$\Gamma = 0$	0.73	0.79	0.90	1.00	0.89
	DOMo	$\Gamma = 0$	0.75	0.78	0.89	1.05	0.88
	BIGo	$\Gamma = 0$	0.75	0.80	0.91	1.03	0.85
	SURo	$\delta_u = 0, \Gamma = 0, \delta_f = 0$	<b>0.66</b>	<b>0.70</b>	0.85	1.01	0.88
	DOMo	$\delta_u = 0, \Gamma = 0, \delta_f = 0$	<b>0.65</b>	<b>0.66</b>	0.83	1.10	0.93
	BIGo	$\delta_u = 0, \Gamma = 0, \delta_f = 0$	<b>0.66</b>	<b>0.68</b>	0.84	1.08	0.89

Notes: All rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the baseline EU model. Improvements over this model are marked by shaded cells. Bold numbers indicate an improvement over the corresponding factor model with the same indicator set but with the baseline restrictions  $\delta_u = 0$  and  $\Gamma = 0$ , where the latter only applies to the one-step model.

Table 14: Factor models with alternative parameter restrictions, revisions to next vintage, Germany

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
2-Step-Bay	SURo		0.99	1.05	0.98	<b>0.96</b>	<b>0.98</b>
	DOMo		0.99	0.98	0.92	1.00	0.88
	BIGo		0.97	0.95	0.89	<b>1.01</b>	0.89
	SURo	$\delta_u = 0, \delta_f = 0$	0.95	<b>0.98</b>	<b>0.92</b>	1.07	1.18
	DOMo	$\delta_u = 0, \delta_f = 0$	<b>0.93</b>	<b>0.90</b>	<b>0.83</b>	1.09	<b>0.98</b>
	BIGo	$\delta_u = 0, \delta_f = 0$	<b>0.93</b>	<b>0.87</b>	<b>0.81</b>	1.09	<b>0.99</b>
2-Step-PCA	SURo		0.93	0.80	0.71	1.13	1.38
	DOMo		0.88	0.82	0.71	<b>1.12</b>	1.35
	BIGo		0.93	0.82	0.73	<b>1.12</b>	<b>1.36</b>
	SURo	$\delta_u = 0, \delta_f = 0$	0.90	<b>0.75</b>	<b>0.62</b>	1.19	1.44
	DOMo	$\delta_u = 0, \delta_f = 0$	<b>0.83</b>	<b>0.75</b>	<b>0.62</b>	1.18	1.41
	BIGo	$\delta_u = 0, \delta_f = 0$	<b>0.89</b>	<b>0.76</b>	<b>0.65</b>	1.20	1.45
1-Step	SURo		1.03	1.12	1.08	<b>0.85</b>	0.86
	DOMo		0.88	0.92	0.93	<b>1.02</b>	0.93
	BIGo		0.91	0.91	0.94	<b>0.95</b>	0.89
	SURo	$\delta_u = 0$	0.99	1.07	1.04	<b>0.89</b>	0.87
	DOMo	$\delta_u = 0$	<b>0.84</b>	0.88	0.91	1.04	0.94
	BIGo	$\delta_u = 0$	0.86	0.87	0.90	<b>0.97</b>	0.88
	SURo	$\Gamma = 0$	0.97	1.08	1.07	<b>0.88</b>	<b>0.82</b>
	DOMo	$\Gamma = 0$	0.90	0.90	0.92	<b>1.01</b>	0.92
	BIGo	$\Gamma = 0$	0.85	0.86	0.91	<b>0.97</b>	0.88
	SURo	$\delta_u = 0, \Gamma = 0, \delta_f = 0$	<b>0.92</b>	1.05	1.07	<b>0.88</b>	<b>0.82</b>
	DOMo	$\delta_u = 0, \Gamma = 0, \delta_f = 0$	<b>0.84</b>	<b>0.84</b>	<b>0.87</b>	1.07	0.93
	BIGo	$\delta_u = 0, \Gamma = 0, \delta_f = 0$	<b>0.80</b>	<b>0.80</b>	<b>0.86</b>	1.01	0.88

Notes: All rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the baseline EU model. Improvements over this model are marked by shaded cells. Bold numbers indicate an improvement over the corresponding factor model with the same indicator set but with the baseline restrictions  $\delta_u = 0$  and  $\Gamma = 0$ , where the latter only applies to the one-step model.

Table 15: Factor models with alternative parameter restrictions, revisions to next vintage, France

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
2-Step-Bay	SURo		0.88	0.76	0.70	0.87	1.07
	DOMo		0.88	0.72	0.62	0.74	0.93
	BIGo		0.90	0.75	0.67	0.91	1.12
	SURo	$\delta_u = 0, \delta_f = 0$	<b>0.73</b>	<b>0.61</b>	<b>0.51</b>	0.87	0.88
	DOMo	$\delta_u = 0, \delta_f = 0$	<b>0.72</b>	<b>0.54</b>	<b>0.46</b>	0.83	0.79
2-Step-PCA	BIGo	$\delta_u = 0, \delta_f = 0$	0.80	<b>0.60</b>	<b>0.55</b>	0.97	1.11
	SURo		0.84	0.80	0.80	1.04	1.31
	DOMo		0.83	0.78	0.70	0.94	1.22
	BIGo		0.94	0.81	0.89	1.22	1.62
	SURo	$\delta_u = 0, \delta_f = 0$	0.84	<b>0.72</b>	<b>0.72</b>	1.06	1.24
1-Step	DOMo	$\delta_u = 0, \delta_f = 0$	<b>0.79</b>	<b>0.62</b>	<b>0.54</b>	0.93	<b>1.08</b>
	BIGo	$\delta_u = 0, \delta_f = 0$	0.98	<b>0.69</b>	<b>0.77</b>	1.25	1.59
	SURo		<b>0.62</b>	0.61	0.53	0.80	0.77
	DOMo		0.69	0.63	0.52	0.81	0.68
	BIGo		0.70	0.62	0.52	0.80	0.66
1-Step	SURo	$\delta_u = 0$	<b>0.62</b>	0.56	<b>0.48</b>	<b>0.78</b>	<b>0.73</b>
	DOMo	$\delta_u = 0$	0.67	0.59	0.48	0.81	0.68
	BIGo	$\delta_u = 0$	<b>0.67</b>	0.58	0.49	0.80	0.66
	SURo	$\Gamma = 0$	<b>0.64</b>	0.60	0.53	0.81	0.78
	DOMo	$\Gamma = 0$	0.69	0.62	0.51	0.82	0.66
	BIGo	$\Gamma = 0$	0.70	0.62	0.53	0.80	0.65
	SURo	$\delta_u = 0, \Gamma = 0, \delta_f = 0$	<b>0.62</b>	<b>0.54</b>	<b>0.46</b>	0.80	<b>0.70</b>
	DOMo	$\delta_u = 0, \Gamma = 0, \delta_f = 0$	0.68	0.58	0.48	0.86	0.71
	BIGo	$\delta_u = 0, \Gamma = 0, \delta_f = 0$	0.69	0.58	<b>0.48</b>	0.85	0.70

Notes: All rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the baseline EU model. Improvements over this model are marked by shaded cells. Bold numbers indicate an improvement over the corresponding factor model with the same indicator set but with the baseline restrictions  $\delta_u = 0$  and  $\Gamma = 0$ , where the latter only applies to the one-step model.

Table 16: Factor models with alternative parameter restrictions, revisions to next vintage, Italy

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
2-Step-Bay	SURo		0.93	0.84	0.80	0.91	<b>0.77</b>
	DOMo		0.76	0.76	0.71	1.01	0.87
	BIGo		0.71	0.69	0.64	1.13	0.99
	SURo	$\delta_u = 0, \delta_f = 0$	<b>0.73</b>	<b>0.67</b>	<b>0.65</b>	0.94	0.82
	DOMo	$\delta_u = 0, \delta_f = 0$	0.69	<b>0.65</b>	0.60	1.02	0.89
	BIGo	$\delta_u = 0, \delta_f = 0$	0.66	<b>0.62</b>	0.59	1.13	1.02
2-Step-PCA	SURo		0.77	0.68	0.76	<b>1.00</b>	<b>0.91</b>
	DOMo		0.72	0.63	0.73	1.23	1.19
	BIGo		0.70	0.58	0.69	1.33	1.33
	SURo	$\delta_u = 0, \delta_f = 0$	<b>0.67</b>	0.61	0.69	1.02	<b>0.93</b>
	DOMo	$\delta_u = 0, \delta_f = 0$	0.67	<b>0.56</b>	0.67	1.23	1.21
	BIGo	$\delta_u = 0, \delta_f = 0$	0.68	0.54	0.66	1.32	1.33
1-Step	SURo		0.76	0.64	0.72	1.14	1.04
	DOMo		0.83	0.69	0.74	1.25	1.12
	BIGo		0.84	0.73	0.75	1.19	1.05
	SURo	$\delta_u = 0$	0.70	0.61	0.66	1.14	1.05
	DOMo	$\delta_u = 0$	0.75	0.61	0.65	1.22	1.11
	BIGo	$\delta_u = 0$	0.76	<b>0.64</b>	0.64	1.16	1.05
	SURo	$\Gamma = 0$	0.75	0.63	0.72	1.14	1.04
	DOMo	$\Gamma = 0$	0.83	0.69	0.72	1.24	1.11
	BIGo	$\Gamma = 0$	0.84	0.74	0.74	1.17	1.03
	SURo	$\delta_u = 0, \Gamma = 0, \delta_f = 0$	<b>0.68</b>	<b>0.59</b>	0.67	1.15	1.06
	DOMo	$\delta_u = 0, \Gamma = 0, \delta_f = 0$	<b>0.73</b>	0.61	0.67	1.30	1.18
	BIGo	$\delta_u = 0, \Gamma = 0, \delta_f = 0$	0.75	0.65	0.67	1.24	1.10

Notes: All rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the baseline EU model. Improvements over this model are marked by shaded cells. Bold numbers indicate an improvement over the corresponding factor model with the same indicator set but with the baseline restrictions  $\delta_u = 0$  and  $\Gamma = 0$ , where the latter only applies to the one-step model.

Table 17: Factor models with alternative parameter restrictions, revisions to next vintage, Spain

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
2-Step-Bay	SURo		1.16	1.26	1.71	1.87	2.05
	DOMo		1.15	1.18	1.50	1.62	1.63
	BIGo		1.08	1.12	1.46	1.56	1.51
	SURo	$\delta_u = 0, \delta_f = 0$	<b>0.80</b>	<b>0.95</b>	<b>1.44</b>	<b>1.61</b>	<b>1.79</b>
	DOMo	$\delta_u = 0, \delta_f = 0$	<b>0.86</b>	<b>1.00</b>	1.37	<b>1.45</b>	<b>1.39</b>
2-Step-PCA	BIGo	$\delta_u = 0, \delta_f = 0$	<b>0.76</b>	<b>0.92</b>	1.34	1.43	<b>1.30</b>
	SURo		0.75	0.85	1.46	1.52	1.44
	DOMo		1.15	1.16	1.42	1.55	1.55
	BIGo		1.02	1.07	1.39	1.47	1.42
	SURo	$\delta_u = 0, \delta_f = 0$	<b>0.50</b>	<b>0.65</b>	1.36	<b>1.46</b>	<b>1.37</b>
1-Step	DOMo	$\delta_u = 0, \delta_f = 0$	<b>0.87</b>	<b>0.93</b>	1.26	<b>1.42</b>	<b>1.37</b>
	BIGo	$\delta_u = 0, \delta_f = 0$	0.70	0.82	1.28	1.40	1.29
	SURo		0.68	0.87	1.44	1.41	1.21
	DOMo		0.74	0.97	1.58	1.53	1.19
	BIGo		0.72	0.96	1.63	1.57	1.18
	SURo	$\delta_u = 0$	0.55	<b>0.73</b>	<b>1.30</b>	1.35	1.15
	DOMo	$\delta_u = 0$	0.60	0.83	1.43	1.44	1.13
	BIGo	$\delta_u = 0$	<b>0.61</b>	<b>0.84</b>	1.49	1.47	1.11
	SURo	$\Gamma = 0$	0.67	0.88	1.48	1.41	1.20
	DOMo	$\Gamma = 0$	0.70	0.94	1.55	1.49	1.13
BIGo	$\Gamma = 0$	0.74	0.99	1.64	1.53	1.15	
SURo	$\delta_u = 0, \Gamma = 0, \delta_f = 0$	<b>0.51</b>	<b>0.70</b>	1.38	1.35	1.14	
DOMo	$\delta_u = 0, \Gamma = 0, \delta_f = 0$	<b>0.51</b>	<b>0.74</b>	1.50	1.48	1.16	
BIGo	$\delta_u = 0, \Gamma = 0, \delta_f = 0$	<b>0.56</b>	<b>0.79</b>	1.57	1.52	1.16	

Notes: All rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the baseline EU model. Improvements over this model are marked by shaded cells. Bold numbers indicate an improvement over the corresponding factor model with the same indicator set but with the baseline restrictions  $\delta_u = 0$  and  $\Gamma = 0$ , where the latter only applies to the one-step model.

Table 18: Factor models with alternative parameter restrictions, revisions to next vintage, Netherlands

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
2-Step-Bay	SURo		0.78	0.89	0.92	0.79	0.75
	DOMo		0.65	0.75	0.77	0.62	0.54
	BIGo		0.65	0.76	0.76	0.61	0.54
	SURo	$\delta_u = 0, \delta_f = 0$	0.74	<b>0.78</b>	0.90	0.83	0.81
	DOMo	$\delta_u = 0, \delta_f = 0$	<b>0.49</b>	<b>0.55</b>	<b>0.65</b>	0.66	0.61
	BIGo	$\delta_u = 0, \delta_f = 0$	<b>0.50</b>	<b>0.61</b>	<b>0.67</b>	0.65	0.61
2-Step-PCA	SURo		0.76	0.88	0.95	0.84	0.89
	DOMo		0.62	0.73	0.76	0.62	0.64
	BIGo		0.63	0.76	0.77	0.63	0.65
	SURo	$\delta_u = 0, \delta_f = 0$	<b>0.71</b>	<b>0.77</b>	<b>0.88</b>	0.86	0.88
	DOMo	$\delta_u = 0, \delta_f = 0$	<b>0.46</b>	<b>0.51</b>	<b>0.62</b>	0.68	0.70
	BIGo	$\delta_u = 0, \delta_f = 0$	<b>0.52</b>	<b>0.60</b>	<b>0.67</b>	0.68	0.69
1-Step	SURo		0.68	0.78	0.73	0.82	<b>0.64</b>
	DOMo		0.63	0.73	0.71	0.80	0.59
	BIGo		0.64	0.78	0.72	0.76	0.55
	SURo	$\delta_u = 0$	0.65	0.73	0.71	<b>0.78</b>	<b>0.61</b>
	DOMo	$\delta_u = 0$	0.60	0.70	0.70	0.72	0.57
	BIGo	$\delta_u = 0$	0.64	0.74	0.71	0.70	0.53
	SURo	$\Gamma = 0$	0.63	0.76	0.72	<b>0.77</b>	<b>0.63</b>
	DOMo	$\Gamma = 0$	0.61	0.74	0.73	0.69	0.56
	BIGo	$\Gamma = 0$	0.61	0.77	0.73	<b>0.67</b>	0.54
	SURo	$\delta_u = 0, \delta_f = 0, \Gamma = 0$	<b>0.57</b>	<b>0.63</b>	<b>0.67</b>	0.86	0.69
	DOMo	$\delta_u = 0, \delta_f = 0, \Gamma = 0$	<b>0.47</b>	<b>0.53</b>	<b>0.62</b>	0.81	0.65
	BIGo	$\delta_u = 0, \delta_f = 0, \Gamma = 0$	<b>0.49</b>	<b>0.58</b>	<b>0.60</b>	0.80	0.60

Notes: All rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the baseline EU model. Improvements over this model are marked by shaded cells. Bold numbers indicate an improvement over the corresponding factor model with the same indicator set but with the baseline restrictions  $\delta_u = 0$  and  $\Gamma = 0$ , where the latter only applies to the one-step model.

## C.2 Contemporaneous indicator sets without time shift

Table 19: Including indicator sets without time shift, revisions to next vintage, average over all five countries

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
2-Step-Bay	SUR	$\delta_u = 0$	0.87	0.86	<b>0.89</b>	<b>1.01</b>	<b>0.98</b>
	DOM	$\delta_u = 0$	0.90	0.86	0.85	1.02	0.96
	BIG	$\delta_u = 0$	0.89	0.87	0.85	1.02	0.96
2-Step-PCA	SUR	$\delta_u = 0$	0.84	0.80	0.90	1.13	<b>1.15</b>
	DOM	$\delta_u = 0$	0.88	0.83	0.95	1.23	1.27
	BIG	$\delta_u = 0$	0.90	0.85	0.97	1.24	1.27
1-Step	SUR	$\delta_u = 0, \Gamma = 0$	<b>0.68</b>	<b>0.67</b>	<b>0.79</b>	1.03	0.98
	DOM	$\delta_u = 0, \Gamma = 0$	<b>0.68</b>	<b>0.67</b>	<b>0.79</b>	1.07	0.99
	BIG	$\delta_u = 0, \Gamma = 0$	<b>0.69</b>	<b>0.70</b>	<b>0.80</b>	1.03	0.93

Notes: All rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the baseline EU model. Improvements over this model are marked by shaded cells. Bold numbers indicate an improvement over the corresponding factor model with the same indicator set but with the baseline restrictions  $\delta_u = 0$  and  $\Gamma = 0$ , where the latter only applies to the one-step model.

Table 20: Including indicator sets without time shift, revisions to next vintage, Germany

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
2-Step-Bay	SUR	$\delta_u = 0$	0.98	1.00	1.02	1.02	<b>0.99</b>
	DOM	$\delta_u = 0$	<b>0.89</b>	<b>0.92</b>	0.95	1.05	0.92
	BIG	$\delta_u = 0$	<b>0.90</b>	0.93	0.97	1.05	0.95
2-Step-PCA	SUR	$\delta_u = 0$	0.96	0.94	0.95	<b>1.11</b>	<b>1.18</b>
	DOM	$\delta_u = 0$	0.90	0.88	0.91	1.25	1.42
	BIG	$\delta_u = 0$	0.93	0.91	0.93	1.28	1.47
1-Step	SUR	$\delta_u = 0, \Gamma = 0$	<b>0.81</b>	<b>0.76</b>	<b>0.79</b>	1.04	1.12
	DOM	$\delta_u = 0, \Gamma = 0$	<b>0.74</b>	<b>0.71</b>	<b>0.77</b>	1.11	1.24
	BIG	$\delta_u = 0, \Gamma = 0$	<b>0.74</b>	<b>0.71</b>	<b>0.78</b>	1.07	1.17

Notes: All rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the baseline EU model. Improvements over this model are marked by shaded cells. Bold numbers indicate an improvement over the corresponding factor model with the same indicator set but with the baseline restrictions  $\delta_u = 0$  and  $\Gamma = 0$ , where the latter only applies to the one-step model.

Table 21: Including indicator sets without time shift, revisions to next vintage, France

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
2-Step-Bay	SUR	$\delta_u = 0$	0.78	0.69	<b>0.58</b>	<b>0.78</b>	<b>0.83</b>
	DOM	$\delta_u = 0$	0.91	0.75	0.59	0.83	0.92
	BIG	$\delta_u = 0$	0.92	0.77	0.59	0.87	0.97
2-Step-PCA	SUR	$\delta_u = 0$	0.95	0.83	<b>0.75</b>	<b>1.01</b>	<b>1.07</b>
	DOM	$\delta_u = 0$	1.04	0.87	0.74	1.11	1.19
	BIG	$\delta_u = 0$	1.12	0.94	0.89	<b>1.17</b>	<b>1.17</b>
1-Step	SUR	$\delta_u = 0, \Gamma = 0$	<b>0.63</b>	0.57	0.50	0.82	0.79
	DOM	$\delta_u = 0, \Gamma = 0$	<b>0.65</b>	0.57	<b>0.45</b>	0.83	0.73
	BIG	$\delta_u = 0, \Gamma = 0$	<b>0.65</b>	0.58	<b>0.45</b>	0.80	0.69

Notes: All rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the baseline EU model. Improvements over this model are marked by shaded cells. Bold numbers indicate an improvement over the corresponding factor model with the same indicator set but with the baseline restrictions  $\delta_u = 0$  and  $\Gamma = 0$ , where the latter only applies to the one-step model.

Table 22: Including indicator sets without time shift, revisions to next vintage, Italy

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
2-Step-Bay	SUR	$\delta_u = 0$	0.77	<b>0.65</b>	<b>0.57</b>	0.93	0.84
	DOM	$\delta_u = 0$	0.73	<b>0.60</b>	<b>0.55</b>	1.04	0.91
	BIG	$\delta_u = 0$	0.74	<b>0.62</b>	<b>0.54</b>	<b>1.01</b>	<b>0.88</b>
2-Step-PCA	SUR	$\delta_u = 0$	0.72	<b>0.56</b>	0.76	1.24	1.21
	DOM	$\delta_u = 0$	0.73	0.57	0.99	1.45	1.44
	BIG	$\delta_u = 0$	0.75	0.58	0.91	1.39	1.37
1-Step	SUR	$\delta_u = 0, \Gamma = 0$	0.76	<b>0.58</b>	0.71	1.19	1.13
	DOM	$\delta_u = 0, \Gamma = 0$	0.76	<b>0.55</b>	0.65	1.26	1.19
	BIG	$\delta_u = 0, \Gamma = 0$	0.75	<b>0.59</b>	<b>0.61</b>	1.20	1.12

Notes: All rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the baseline EU model. Improvements over this model are marked by shaded cells. Bold numbers indicate an improvement over the corresponding factor model with the same indicator set but with the baseline restrictions  $\delta_u = 0$  and  $\Gamma = 0$ , where the latter only applies to the one-step model.



Table 23: Including indicator sets without time shift, revisions to next vintage, Spain

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
2-Step-Bay	SUR	$\delta_u = 0$	1.09	1.20	<b>1.43</b>	<b>1.53</b>	<b>1.56</b>
	DOM	$\delta_u = 0$	1.28	1.31	1.40	<b>1.45</b>	1.48
	BIG	$\delta_u = 0$	1.19	1.26	1.42	1.45	1.46
2-Step-PCA	SUR	$\delta_u = 0$	<b>0.83</b>	<b>0.85</b>	<b>1.17</b>	<b>1.38</b>	<b>1.40</b>
	DOM	$\delta_u = 0$	1.04	1.11	1.33	1.46	1.49
	BIG	$\delta_u = 0$	<b>0.96</b>	1.02	1.31	1.46	1.47
1-Step	SUR	$\delta_u = 0, \Gamma = 0$	<b>0.56</b>	<b>0.73</b>	<b>1.24</b>	<b>1.26</b>	1.14
	DOM	$\delta_u = 0, \Gamma = 0$	<b>0.63</b>	<b>0.84</b>	<b>1.37</b>	<b>1.35</b>	1.11
	BIG	$\delta_u = 0, \Gamma = 0$	<b>0.65</b>	<b>0.88</b>	<b>1.41</b>	<b>1.36</b>	1.07

Notes: All rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the baseline EU model. Improvements over this model are marked by shaded cells. Bold numbers indicate an improvement over the corresponding factor model with the same indicator set but with the baseline restrictions  $\delta_u = 0$  and  $\Gamma = 0$ , where the latter only applies to the one-step model.

Table 24: Including indicator sets without time shift, revisions to next vintage, Netherlands

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
2-Step-Bay	SUR	$\delta_u = 0$	<b>0.72</b>	<b>0.78</b>	<b>0.83</b>	0.81	<b>0.69</b>
	DOM	$\delta_u = 0$	0.68	0.74	0.75	0.73	0.57
	BIG	$\delta_u = 0$	0.69	0.77	0.73	0.70	0.56
2-Step-PCA	SUR	$\delta_u = 0$	<b>0.76</b>	<b>0.80</b>	<b>0.87</b>	0.93	0.88
	DOM	$\delta_u = 0$	0.69	0.74	0.78	0.88	0.82
	BIG	$\delta_u = 0$	0.75	0.82	0.82	0.90	0.85
1-Step	SUR	$\delta_u = 0, \Gamma = 0$	<b>0.63</b>	<b>0.69</b>	<b>0.72</b>	0.84	0.72
	DOM	$\delta_u = 0, \Gamma = 0$	<b>0.61</b>	<b>0.69</b>	<b>0.73</b>	0.79	0.66
	BIG	$\delta_u = 0, \Gamma = 0$	<b>0.64</b>	<b>0.74</b>	<b>0.73</b>	<b>0.73</b>	0.61

Notes: All rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the baseline EU model. Improvements over this model are marked by shaded cells. Bold numbers indicate an improvement over the corresponding factor model with the same indicator set but with the baseline restrictions  $\delta_u = 0$  and  $\Gamma = 0$ , where the latter only applies to the one-step model.

### C.3 Vintages 2005-2019 excluding the COVID-19 years

Table 25: Vintage sample without COVID-19 years, revisions to next vintage, average over all five countries

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
EU model	-	$\delta_u = 0$	0.39	0.51	0.58	0.80	0.78
	-		1.12	1.08	1.02	1.03	0.95
HP-EST	-		1.01	1.23	1.28	1.21	<b>0.80</b>
HP	-	$\lambda = 6.25$	1.00	1.17	1.15	1.10	0.82
	-	$\lambda = 100$	1.18	1.19	1.01	<b>0.87</b>	0.93
2-Step-Bay	SURo	$\delta_u = 0$	0.90	0.91	0.92	1.23	1.25
	DOMo	$\delta_u = 0$	0.89	0.89	0.79	1.01	1.06
	BIGo	$\delta_u = 0$	0.89	0.89	0.77	1.02	1.05
2-Step-PCA	SURo	$\delta_u = 0$	0.82	0.81	0.89	1.37	1.44
	DOMo	$\delta_u = 0$	0.86	0.83	<b>0.73</b>	1.19	1.42
	BIGo	$\delta_u = 0$	0.89	0.85	<b>0.73</b>	1.24	1.47
1-Step	SURo	$\delta_u = 0, \Gamma = 0$	<b>0.75</b>	<b>0.79</b>	0.80	1.18	1.08
	DOMo	$\delta_u = 0, \Gamma = 0$	0.78	0.80	0.76	1.17	1.08
	BIGo	$\delta_u = 0, \Gamma = 0$	0.78	0.82	0.77	1.17	1.05

Notes: The first row shows the average revision,  $R_h^{(C,M)}$ , of the EU model. The remaining rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the EU model. Improvements over the EU model are marked by shaded cells, the best model is highlighted by a bold number.

Table 26: Vintage sample without COVID-19 years, revisions to next vintage, Germany

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
EU model	-	$\delta_u = 0$	0.34	0.43	0.49	0.70	0.50
	-		1.05	1.07	1.03	0.86	0.82
HP-EST	-		0.82	0.83	0.95	0.76	0.69
HP	-	$\lambda = 6.25$	<b>0.69</b>	0.78	0.96	0.86	0.67
	-	$\lambda = 100$	0.92	0.99	0.93	<b>0.65</b>	0.67
2-Step-Bay	SURo	$\delta_u = 0$	0.89	0.97	0.88	1.03	0.98
	DOMo	$\delta_u = 0$	0.90	0.92	0.85	1.06	1.00
	BIGo	$\delta_u = 0$	0.92	0.91	0.82	1.08	0.98
2-Step-PCA	SURo	$\delta_u = 0$	0.82	<b>0.74</b>	<b>0.59</b>	1.19	1.52
	DOMo	$\delta_u = 0$	0.79	0.77	0.60	1.18	1.55
	BIGo	$\delta_u = 0$	0.88	0.79	0.62	1.21	1.58
1-Step	SURo	$\delta_u = 0, \Gamma = 0$	0.93	1.04	1.03	0.84	<b>0.64</b>
	DOMo	$\delta_u = 0, \Gamma = 0$	0.82	0.86	0.87	1.02	0.87
	BIGo	$\delta_u = 0, \Gamma = 0$	0.80	0.84	0.87	0.98	0.82

Notes: The first row shows the average revision,  $R_h^{(C,M)}$ , of the EU model. The remaining rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the EU model. Improvements over the EU model are marked by shaded cells, the best model is highlighted by a bold number.

Table 27: Vintage sample without COVID-19 years, revisions to next vintage, France

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
EU model	-	$\delta_u = 0$	0.52	0.83	0.81	0.83	0.89
	-		1.07	1.06	1.01	1.01	1.02
HP-EST	-		0.97	0.91	0.59	0.74	0.69
HP	-	$\lambda = 6.25$	0.93	0.90	0.62	0.67	0.65
	-	$\lambda = 100$	0.79	0.81	0.56	<b>0.31</b>	<b>0.53</b>
2-Step-Bay	SURo	$\delta_u = 0$	0.78	0.71	0.58	0.73	0.93
	DOMo	$\delta_u = 0$	0.79	0.66	0.51	0.57	0.85
	BIGo	$\delta_u = 0$	0.80	0.68	0.55	0.83	1.08
2-Step-PCA	SURo	$\delta_u = 0$	0.83	0.78	0.77	1.03	1.30
	DOMo	$\delta_u = 0$	0.81	0.72	0.63	0.78	1.24
	BIGo	$\delta_u = 0$	0.91	0.77	0.82	1.26	1.68
1-Step	SURo	$\delta_u = 0, \Gamma = 0$	<b>0.70</b>	<b>0.62</b>	0.46	0.80	0.86
	DOMo	$\delta_u = 0, \Gamma = 0$	0.72	0.64	<b>0.40</b>	0.74	0.76
	BIGo	$\delta_u = 0, \Gamma = 0$	0.75	0.65	<b>0.40</b>	0.72	0.75

Notes: The first row shows the average revision,  $R_h^{(C,M)}$ , of the EU model. The remaining rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the EU model. Improvements over the EU model are marked by shaded cells, the best model is highlighted by a bold number.

Table 28: Vintage sample without COVID-19 years, revisions to next vintage, Italy

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
EU model	-	$\delta_u = 0$	0.27	0.39	0.41	0.68	0.69
	-		1.02	0.83	0.72	0.82	0.76
HP-EST	-		1.04	1.16	1.02	1.14	0.93
HP	-	$\lambda = 6.25$	1.09	1.09	1.01	1.02	0.83
	-	$\lambda = 100$	1.40	0.99	0.76	0.89	0.90
2-Step-Bay	SURo	$\delta_u = 0$	0.74	0.57	0.47	1.02	0.89
	DOMo	$\delta_u = 0$	<b>0.67</b>	0.58	0.43	1.17	1.07
	BIGo	$\delta_u = 0$	<b>0.67</b>	0.60	0.42	1.11	0.98
2-Step-PCA	SURo	$\delta_u = 0$	0.70	0.53	0.57	1.22	1.16
	DOMo	$\delta_u = 0$	<b>0.67</b>	<b>0.52</b>	0.57	1.41	1.46
	BIGo	$\delta_u = 0$	0.68	<b>0.52</b>	<b>0.41</b>	1.31	1.32
1-Step	SURo	$\delta_u = 0, \Gamma = 0$	0.74	0.55	0.49	1.27	1.18
	DOMo	$\delta_u = 0, \Gamma = 0$	0.79	0.59	0.44	1.26	1.13
	BIGo	$\delta_u = 0, \Gamma = 0$	0.79	0.61	0.42	1.22	1.08

Notes: The first row shows the average revision,  $R_h^{(C,M)}$ , of the EU model. The remaining rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the EU model. Improvements over the EU model are marked by shaded cells, the best model is highlighted by a bold number.

Table 29: Vintage sample without COVID-19 years, revisions to next vintage, Spain

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
EU model	-	$\delta_u = 0$	0.28	<b>0.29</b>	<b>0.25</b>	<b>0.27</b>	<b>0.36</b>
	-		1.52	1.46	1.41	1.49	1.25
HP-EST	-		1.24	1.90	2.90	2.69	1.23
HP	-	$\lambda = 6.25$	1.62	2.17	2.38	2.29	1.53
	-	$\lambda = 100$	2.01	2.15	2.08	1.96	2.20
2-Step-Bay	SURo	$\delta_u = 0$	1.36	1.51	1.82	2.57	2.72
	DOMo	$\delta_u = 0$	1.44	1.54	1.40	1.63	1.83
	BIGo	$\delta_u = 0$	1.40	1.49	1.32	1.47	1.70
2-Step-PCA	SURo	$\delta_u = 0$	1.00	1.13	1.57	2.53	2.37
	DOMo	$\delta_u = 0$	1.40	1.40	1.09	1.92	2.21
	BIGo	$\delta_u = 0$	1.34	1.38	1.05	1.80	2.13
1-Step	SURo	$\delta_u = 0, \Gamma = 0$	<b>0.77</b>	1.02	1.34	2.18	2.10
	DOMo	$\delta_u = 0, \Gamma = 0$	0.94	1.20	1.35	2.12	2.07
	BIGo	$\delta_u = 0, \Gamma = 0$	0.94	1.25	1.41	2.22	2.09

Notes: The first row shows the average revision,  $R_h^{(C,M)}$ , of the EU model. The remaining rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the EU model. Improvements over the EU model are marked by shaded cells, the best model is highlighted by a bold number.

Table 30: Vintage sample without COVID-19 years, revisions to next vintage, Netherlands

Model	Indicator set	Parameter restrictions	Relative revision $R_{h,rel}^{(C,M)}$				
			$h = -2$	$h = -1$	$h = 0$	$h = 1$	$h = 2$
EU model	-	$\delta_u = 0$	0.55	0.63	0.92	1.50	1.45
	-		0.96	0.98	0.95	0.96	0.89
HP-EST	-		0.98	1.36	0.94	0.74	0.45
HP	-	$\lambda = 6.25$	0.68	0.93	0.80	0.68	0.40
	-	$\lambda = 100$	0.78	1.00	0.72	<b>0.53</b>	<b>0.33</b>
2-Step-Bay	SURo	$\delta_u = 0$	0.72	0.80	0.86	0.80	0.75
	DOMo	$\delta_u = 0$	0.66	0.76	0.75	0.61	0.53
	BIGo	$\delta_u = 0$	0.66	0.77	0.74	0.59	0.52
2-Step-PCA	SURo	$\delta_u = 0$	0.75	0.86	0.93	0.86	0.87
	DOMo	$\delta_u = 0$	0.65	0.75	0.76	0.65	0.64
	BIGo	$\delta_u = 0$	0.64	0.77	0.77	0.64	0.65
1-Step	SURo	$\delta_u = 0, \Gamma = 0$	<b>0.62</b>	<b>0.70</b>	<b>0.69</b>	0.83	0.64
	DOMo	$\delta_u = 0, \Gamma = 0$	<b>0.62</b>	0.73	0.73	0.69	0.55
	BIGo	$\delta_u = 0, \Gamma = 0$	0.64	0.77	0.74	0.69	0.52

Notes: The first row shows the average revision,  $R_h^{(C,M)}$ , of the EU model. The remaining rows show the revision  $R_{h,rel}^{(C,M)}$  relative to the EU model. Improvements over the EU model are marked by shaded cells, the best model is highlighted by a bold number.

## Vorschlag zur Anpassung des Verfahrens zur Bestimmung des Produktionspotenzials

Im Folgenden wird skizziert, wie sowohl die juristische Kritik am Verfahren zur Bestimmung des Produktionspotenzials als auch die ökonomischen Erkenntnisse der letzten zehn Jahre im Rahmen der Schuldenbremse und der EU-CAM adressiert werden könnten. Die Vorschläge betreffen die Projektion des Arbeitspotenzials und die Elastizität des Faktors Arbeit (nach EU-CAM-Definition des Produktionspotenzials die Variablen  $L$  und  $\alpha$ :  $Y = L^\alpha * K^{1-\alpha} * TFP$ ).

### 1) Verfahren zur Projektion des Arbeitspotenzials

Das heutige Verfahren zur Bestimmung des Produktionspotenzials ist verfassungsrechtlich problematisch, da die Inputs zur Potenzialschätzung nicht allein auf ökonomischer Grundlage bestimmbar sind (u.a. ist das Potenzial nicht unabhängig von der Fiskalpolitik, siehe [Krebs 2017](#), [Fatás 2018](#), [Lechthaler 2021](#), [Ilzetki 2021](#), [Ohler, DVBl. 2009, S. 1265 \(1267\)](#)). Daher sollte künftig der Bundestag das Potenzial soweit möglich auf Basis tatsächlicher Politik setzen und begründen. Insb. die Inputfaktoren des Arbeitspotenzials Erwerbstätige, Partizipationsrate und Arbeitsstunden (nach EU-CAM Definition  $L = POPW * PARTS * (1 - NAWRU) * HOURST$ ) lassen sich direkt mit Politik in Verbindung bringen. Die Kopplung fiskalischen Spielraums an potenzialausweitende Politik ist heute zentral für eine nachhaltige Finanzpolitik ([BMF 2020](#), [Sigl-Glückner 2021](#)).

Mögliche Verfahren zur Bestimmung der Inputs des Arbeitspotenzials:

- **Einzelfallverfahren:** Projektion der Inputvariablen auf Basis einzelner Reformvorhaben und empirischer Evidenz/Expertenwissen. Ein solches Verfahren brächte zwar eine hohe Nachvollziehbarkeit mit sich, wäre aber aufgrund geringer Standardisierung sehr komplex.
- **Empirischer Anker:** Für Inputvariablen ohne klares, mit konkreten Reformvorhaben unterlegtes Ziel: Projektion des Potenzials auf Basis des vergangenen Trends. Für Inputvariablen, deren Potenzial auf Basis empirischer Erkenntnisse oder aktueller Politik nicht dem vergangenen Trend entspricht: Festsetzung des Potenzials bei einem empirisch begründbaren Ankerwert, der einem erreichbaren und mit der aktuellen und geplanten Politik kongruenten Zielzustand entspricht.
- **Empirischer Anker + gesamtwirtschaftliches Tempolimit:** In einer Situation nachfrageseitig getriebener Inflation sollte verhindert werden, dass ein per Anker definiertes Potenzial die Nachfrage übermäßig befeuert. Anstatt Potenziale direkt mit Ankerwerten gleichzusetzen, könnte das Potenzial z.B. linear über fünf Jahre zum Anker ansteigen.
- **Empirischer Anker + vorhabenspezifischer Sicherheitsgurt:** Sollten Preissteigerungen angebotsseitig von Einzelpreissteigerungen getrieben sein, würde ein gesamtwirtschaftliches Tempolimit keinen wirksamen Schutz vor Inflation bieten. Stattdessen sollte jedes Vorhaben in der Gesetzesfolgenanalyse auf preisliche Auswirkungen analysiert werden.

Ein ankerbasiertes Verfahren würde Komplexität und Revisionsanfälligkeit reduzieren, da statistische Trendzerlegungen entfielen.

Alternativ könnte auch ein modellgestützter Ansatz -wie bereits zur Schätzung der Auswirkungen der Aufbau- und Resilienzpläne auf das Potenzial zum Einsatz kam -z.B. [BMF 2021](#)- genutzt werden, dieser ist aber für Nicht-Experten schwer nachvollziehbar.

## 2) Empirische Updates

Die Methodik der Potenzialschätzung sollte mit empirischen Erkenntnissen in Einklang gebracht werden (§ 5 (4) G 115).

Das betrifft zum einen die heute bei 0,65 festgesetzte Produktionselastizität des Faktors Arbeit ( $Y = L^\alpha * K^{1-\alpha} * TFP$ ). Diese liegt für Deutschland konstant bei 0,7 und sollte entsprechend angepasst werden. Die Totale Faktorproduktivität (TFP) hängt von der Arbeitselastizität ab und müsste daher neu geschätzt werden.

Weiter kann die im Arbeitspotenzial enthaltene Schätzung der NAWRU ( $L = POPW * PARTS * (1 - NAWRU) * HOURST$ ) die niedrigstmögliche Arbeitslosigkeit ohne sich beschleunigende Inflation nicht zuverlässig abbilden. Das liegt zum einen an mit Unsicherheit behafteten Annahmen ([EZB 2021](#)), zum anderen an der Beeinflussung der NAWRU durch konjunkturelle Faktoren ([Heimberger 2017](#)). Als erster möglicher Proxy zur Schätzung der niedrigstmöglichen Arbeitslosigkeit könnte die um den Anteil der Langzeitarbeitslosen bereinigte Arbeitslosenquote dienen ([Schuster \(2021\)](#)), da dauerhaft erfolglose Suchbemühungen unter Vollbeschäftigung nicht existieren. Der Proxy sollte auf Basis aktueller Forschung und Diskussionen zwischen Experten und Politik weiterentwickelt werden.

## 3) Vergleich zur EU-CAM

Die obigen Vorschläge modifizieren lediglich die Berechnung der Inputs des Arbeitspotenzials und der Arbeitselastizität im Rahmen der EU-CAM, sodass sie auch auf europäischer Ebene eingebracht werden könnten ([Bundesregierung 2022](#)). Bereits heute unterscheiden sich die von der Bundesregierung und der EU-KOM genutzten Inputs zur Potenzialschätzung. Eine geänderte Potenzialschätzung macht die Neuberechnung der (heute von der EU-KOM geschätzten) Budgetsemielastizität erforderlich ([Mourre 2014](#)).

## 4) Quantitative Auswirkungen

Variable			Effekt auf das Produktionspotenzial 2024* (in Mrd. EUR, im Vgl. zu BReg 04.2022)	
			Empirischer Anker	Empirischer Anker + gesamtwirtschaftliches Tempolimit
A	TFP	keine Änderung, da nicht interpretierbar		
L	Erwerbsbevölkerung	Bevölkerungsprojektion, aktuelle Entwicklungen, Politikmaßnahmen	+ 83	+ 38
	Partizipationsrate	eine der vorgestellten Varianten (s. Schuster 2021)		
	Arbeitsstunden	eine der vorgestellten Varianten (s. Schuster 2021)		
	NAWRU	eine der vorgestellten Varianten (s. Schuster 2021)		
K	Bruttoanlagevermögen	keine Änderung		
a	Produktionselastizität Arbeit	Aktualisierung auf Basis der deutschen Lohnquote	+ 8 (inkl. TFP nötiger Neuberechnung)	

\*Vergleich für 2024 anstelle 2023, da in weiter in der Zukunft liegenden Jahren der Unterschied zwischen BReg Schätzung und EU-CAM zurückgeht und so besser mit beiden verglichen werden kann.

# Umfragegestützte strukturelle Finanzierungssalden

von Marcell Göttert\* und Timo Wollmershäuser†

*Seit vielen Jahren wird über die Ausgestaltung der Fiskalregeln in der EU und Deutschland diskutiert. Die Kontroverse dreht sich dabei unter anderem um die Revisionsanfälligkeit der Konjunkturkomponente des staatlichen Finanzierungssaldos, die zur Bestimmung des strukturellen Finanzierungssaldos geschätzt werden muss. Im vorliegenden Beitrag vergleichen wir eine neue umfragegestützte Methode, die die konjunkturelle Lage anhand der Kapazitätsauslastung misst, mit der bisher im Rahmen der EU-Haushaltsüberwachung verwendeten Produktionsfunktionsmethode. Wir zeigen, dass die Revisionen der Konjunkturkomponente bei der umfragegestützten Methode deutlich geringer und nicht systematisch sind. Somit liefert die Kapazitätsauslastung bereits frühzeitig ein verlässliches und insbesondere unverzerrtes Signal über die aktuelle konjunkturelle Lage und kann damit auch beim aktuellen Vorschlag der EU-Kommission bzgl. der Fiskalregeln zur Anwendung kommen.*

*Abstract: The design of fiscal rules in the EU and Germany has been the subject of debate for many years. Among other things, the controversy revolves around the susceptibility to revision of the cyclical component of the fiscal balance, which has to be estimated to determine the structural fiscal balance. In this paper, we compare a new survey-based method that measures the cyclical situation using capacity utilization with the production function method used so far in EU budgetary surveillance. We show that the revisions to the cyclical component are significantly smaller and non-systematic for the survey-based method. Capacity utilization thus provides a reliable and, above all, unbiased signal about the current economic situation at an early stage and can therefore also be used in the EU Commission's current proposal on fiscal rules.*

*JEL-Codes: E32, E62, H62, H68*

---

\* Agenda Austria, Wien, [marcell.goettert@agenda-austria.at](mailto:marcell.goettert@agenda-austria.at).

† ifo Institut für Wirtschaftsforschung, München, [wollmershaeuser@ifo.de](mailto:wollmershaeuser@ifo.de).

## Einleitung

Seit vielen Jahren wird über die Ausgestaltung der Haushaltsüberwachung in der Europäischen Union (EU) und Deutschland diskutiert. Zwar sind derzeit wichtige Fiskalregeln aufgrund der Corona- und Energiekrise ausgesetzt. Ob sie allerdings in ihrer ursprünglichen Form ab dem kommenden Jahr wieder in Kraft treten, ist unklar. Denn die Europäische Kommission hat mit ihren Vorschlägen für ein neues Regelwerk der Debatte neuen Schwung verliehen. Die Kontroverse bei den bisherigen Fiskalregeln dreht sich unter anderem um ihr Herzstück, den strukturellen staatlichen Finanzierungssaldo, und die mit seiner Schätzung verbundenen hohen und systematischen Revisionen (Ademmer et al. 2019). Egal wie die Debatte endet, ist schon jetzt klar, dass es weiterhin einer konjunkturellen Einschätzung am aktuellen Rand bedürfen wird.

Nach dem im Jahr 2013 in Kraft getretenen Europäischen Fiskalpakt darf das strukturelle Finanzierungsdefizit nicht größer als 0,5% in Relation zum Bruttoinlandsprodukt ausfallen.<sup>1</sup> Darüber hinaus spielen die strukturellen Finanzierungssalden bei beiden Komponenten des Stabilitäts- und Wachstumspakts (SWP) eine bedeutende Rolle. Beim präventiven Arm sind Grenzwerte für die strukturellen Finanzierungssalden definiert. Beim korrektiven Arm werden Maßnahmen zur Beseitigung eines übermäßigen Haushaltsdefizits empfohlen, bei denen ein Ausgabepfad mit einem strukturellen Budgetpfad kombiniert werden (ECOFIN 2017; Europäische Kommission 2018). Darüber hinaus haben viele Staaten in der EU nationale Fiskalregeln, in denen strukturelle Finanzierungssalden eine zentrale Rolle spielen. So erlaubt die Schuldenbremse in Deutschland eine strukturelle Neuverschuldung, die 0,35% in Relation zum Bruttoinlandsprodukt nicht übersteigt. Und in Österreich darf das strukturelle Finanzierungsdefizit des Gesamtstaats nicht höher als 0,45% des Bruttoinlandsprodukts ausfallen.<sup>2</sup>

## Strukturelle Finanzierungssalden nach der EU-Methode

Grundsätzlich ist die Ausrichtung der Fiskalregeln an strukturellen Finanzierungssalden sinnvoll, da sie die Verschuldungsmöglichkeiten eines Staates unabhängig von der konjunkturellen Lage und einmaligen Ausgabe- oder Einnahmeschwankungen festlegen (Feld et al. 2021). Dadurch soll gewährleistet werden, dass die mit kurzfristigen Schwankungen einhergehenden Verbesserungen und Verschlechterung des Finanzierungssaldos den staatlichen Handlungsspielraum nicht erhöhen oder beschränken. Daher wird zur Berechnung der strukturellen Finanzierungssalden ( $X$ ) der staatliche Finanzierungssaldo ( $B$ ) um konjunkturbedingte ( $Y$ ) und einmalige ( $T$ ) Änderungen der Einnahmen und Ausgaben bereinigt:

$$(1) \quad X = B - Y - T.$$

Die Konjunkturkomponente ergibt sich aus dem Produkt der Budgetsemielastizität ( $\varepsilon_{B,OG}$ ) mit der Produktionslücke („Output Gap“) ( $OG$ ):

$$(2) \quad Y = \varepsilon_{B,OG} OG,$$

wobei die Budgetsemielastizität die Reaktivität des staatlichen Finanzierungssaldos auf eine Veränderung der Produktionslücke misst.

Mit dem strukturellen Finanzierungssaldo versucht man also einen staatlichen Haushalt darzustellen, der sich in einem „normalen“ Jahr ohne konjunkturbedingte ( $Y = 0$ ) oder einmalige ( $T = 0$ ) Mehr- oder Minderausgaben ergeben würde. Im Kern drehen sich die meisten Kontroversen im

---

<sup>1</sup> Falls der staatliche Schuldenstand signifikant unter 60% in Relation zum Bruttoinlandsprodukt liegt und die langfristige Stabilität der öffentlichen Finanzen gesichert ist, darf das strukturelle Defizit auch bis zu 1% in Relation zum Bruttoinlandsprodukt betragen.

<sup>2</sup> Hiervon entfallen 0,35 Prozentpunkte auf den Bund und die restlichen 0,1 Prozentpunkte auf Länder sowie Gemeinden.



Zusammenhang mit den strukturellen Finanzierungssalden um die Bestimmung der konjunkturellen Normalsituation. Formal ist diese Situation gegeben, wenn die Produktionslücke ( $OG$ ) gleich Null ist, wobei diese als prozentuale Abweichung des preisbereinigten Bruttoinlandsprodukts ( $BIP$ ) vom Produktionspotenzial ( $POT$ ) definiert ist:

$$(3) \quad OG = \frac{(BIP - POT)}{POT} * 100\% .$$

Da das Produktionspotenzial im Gegensatz zum Bruttoinlandsprodukt eine unbeobachtbare Größe ist, hängt seine Bestimmung und damit die der Produktionslücke sowohl von modelltheoretischen Vorstellungen als auch von empirischen Schätzverfahren ab. Die für die Fiskalregeln der EU gemeinsam vereinbarte Methode („Commonly Agreed Methodology“, kurz EU-CAM) zur Schätzung des Produktionspotenzials basiert dabei auf einer Cobb-Douglas-Produktionsfunktion (Blondeau et al. 2021).<sup>3</sup>

Ademmer et al. (2019) zeigen, dass die Schätzungen der Produktionslücke der EU-Kommission und anderer Institutionen, die auf dem Produktionsfunktionsansatz basieren, umfangreichen Revisionen unterliegen.<sup>4</sup> Eine wesentliche Ursache hierfür ist, dass in die Schätzung auch Prognosen der Variablen der Produktionsfunktion einfließen. Dabei führen Prognosefehler und -anpassungen nicht nur zu Revisionen zukünftiger Produktionslücken (sog. ex-ante Revisionen), sondern auch zu nachträglichen Revisionen der Schätzungen bereits abgelaufener Jahre (sog. ex-post Revisionen). Die damit einhergehende Unzuverlässigkeit der bestehenden Fiskalregeln ist für die Europäische Kommission ein wesentliches Motiv, ein neues Regelwerk vorzuschlagen: „The experience before and during crisis confirmed the difficulties associated with designing policy recommendations on the basis of unobservable indicators that are subject to frequent revisions (such as the ‚output gap‘ and the ‚structural balance‘)“ (Europäische Kommission, 2022, S. 4). Jedoch benötigt auch der aktuelle Vorschlag der EU-Kommission eine Einschätzung der konjunkturellen Lage am aktuellen Rand.

Abbildung 1 veranschaulicht die durchschnittlichen Revisionen der Konjunkturkomponente des strukturellen Finanzierungssaldos für Deutschland, die sich auf Basis der Schätzungen der EU-Kommission im Rahmen der EU-CAM für die Jahre von 2010 bis 2020 ergeben. Dabei wird für unterschiedliche Zeitpunkte („Vintages“)  $\tau$ , zu denen die EU-Kommission die Konjunkturkomponente für das Jahr  $t$  bestimmt hat, folgende Gleichung geschätzt:

$$(4) \quad Y_{t,t+\tau} - Y_{t,t+0,25} = \alpha_{\tau} + u_{\tau,t}.$$

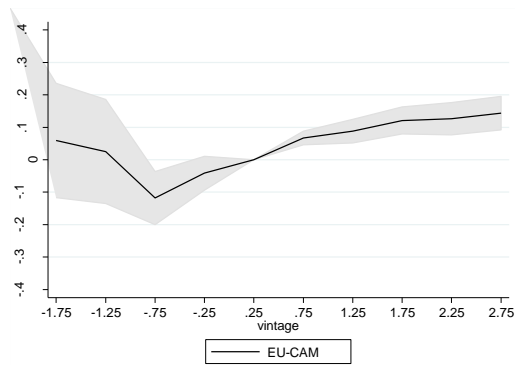
Die Konstante  $\alpha_{\tau}$  misst die durchschnittliche Revision und  $\tau$  den zeitlichen Abstand zum Ende Jahres  $t$  in Bruchteilen eines Jahres. Die erste Konjunkturkomponente für das Jahr  $t = 2010$  wird im Frühjahr des Jahres 2009 ( $t - 1,75$ ) geschätzt; die erste Revision (Neuschätzung) erfolgt ein halbes Jahr später im Herbst 2009 ( $t - 1,25$ ). Berücksichtigte Schätzzeitpunkte ex ante sind  $\tau = \{-1,75; -1,25; -0,75; -0,25\}$ <sup>5</sup>, sowie ex post  $\tau = \{0,25; 0,75; 1,25; 1,75; 2,25; 2,75\}$ . Als Referenz für die Revision wird immer die Konjunkturkomponente der ersten ex-post-Schätzung ( $t + 0,25$ ) verwendet.

<sup>3</sup> Auch bei der deutschen und österreichischen Schuldenbremse wird zur Berechnung der strukturellen Finanzierungssalden die EU-CAM verwendet.

<sup>4</sup> Ähnliche Ansätze verfolgen die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) und der Internationale Währungsfonds (IWF).

<sup>5</sup> Das umfragebasierte Verfahren, das weiter unten vorgestellt wird, greift in der Echtzeitanalyse auf Prognosen der Projektgruppe Gemeinschaftsdiagnose (GD) zurück. Für den Vergleich der Schätzergebnisse der beiden Verfahren werden nur Vintages ab  $t - 1,75$  herangezogen, da die GD-Prognosen erst im Herbst 2016 auf  $t - 2,25$  erweitert wurden und somit für eine aussagekräftige Auswertung nur eine zu geringe Zahl an Beobachtungen zur Verfügung steht.

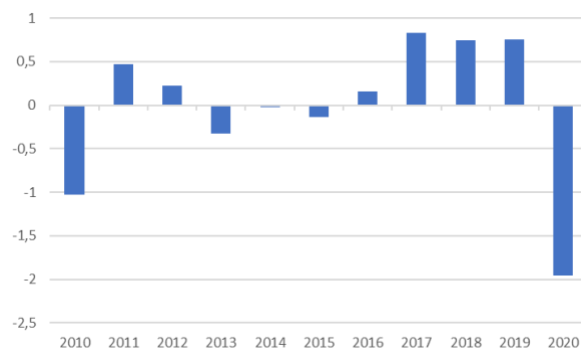
Abbildung 1: Durchschnittliche Revision der Konjunkturkomponente bei der EU-Methode



Anmerkung: Die durchschnittliche Revision der Konjunkturkomponente wird in Prozentpunkten des Produktionspotenzials gemessen. Die grau schraffierte Fläche zeigt das 95%-Konfidenzintervall der Schätzung.

Es zeigt sich zum einen, dass die Streuung der Revisionen der Konjunkturkomponente und damit die Unsicherheit im Hinblick auf Verlässlichkeit der Fiskalregeln durchaus beachtlich ist. Das Ausmaß wird insbesondere durch einen Vergleich mit der durchschnittlichen Schwankung (Standardabweichung) der Konjunkturkomponente im Beobachtungszeitraum (2010 bis 2020) verdeutlicht, die bei 0,8 Prozent des Produktionspotenzials liegt (vgl. Abbildung 2). Dabei ist die Streuung der Revisionen nicht nur ex ante groß, was aufgrund der Prognoseunsicherheit nachvollziehbar wäre, sondern insbesondere auch ex post, nachdem das Fiskaljahr bereits abgeschlossen ist. Zum anderen deuten die Schätzungen auf eine systematische Aufwärtsrevision der Konjunkturkomponente, insbesondere für bereits abgelaufene Jahre hin. Damit wird der strukturelle Finanzierungssaldo und mit ihm die fiskalischen Handlungsspielräume anfangs systematisch überschätzt und im Lauf der Zeit nach unten korrigiert.

Abbildung 2: Konjunkturkomponente des staatlichen Finanzierungssaldos



Anmerkung: Veröffentlichung der EU-Kommission im Frühjahr 2022. Die Konjunkturkomponente wird in Prozent des Produktionspotenzials gemessen.

## Neue umfragegestützte Methode

Um die mit dem Produktionsfunktionsansatz verbundenen Probleme zu vermeiden, wird im vorliegenden Beitrag eine neue Methode zur Berechnung struktureller Finanzierungssalden vorgeschlagen, in der die Konjunkturkomponente durch eine im Rahmen von Unternehmensbefragungen erhobene Kapazitätsauslastung bestimmt wird (vgl. auch Göttert und Wollmershäuser 2021). Bereits Orphanides et al. (2000) haben gezeigt, dass es eine hohe Korrelation zwischen der Produktionslücke und der Kapazitätsauslastung gibt. Dies überrascht wenig, denn „(d)er Potenzial-Output wird in aller Regel als jenes Aktivitätsniveau definiert, das bei einer ‚normalen‘ Auslastung der gesamtwirtschaftlichen Kapazitäten erzielt wird.“ (Deutsche Bundesbank, 2014, S.

14). Wenn die Kapazitätsauslastung über ihrem Normalniveau ist, ist das Bruttoinlandsprodukt größer als das Produktionspotenzial und die Produktionslücke damit positiv. Für eine Kapazitätsauslastung unter dem Normalniveau ist das Bruttoinlandsprodukt kleiner als das Produktionspotenzial und damit die resultierende Produktionslücke negativ.

EU-weit wird die aktuelle Kapazitätsauslastung im Rahmen einer repräsentativen Unternehmensbefragungen einmal im Quartal erhoben. In Deutschland wird diese Befragung, die auf circa 9 000 Meldungen von Unternehmen basiert, vom ifo Institut durchgeführt. Die dabei erfassten Wirtschaftsbereiche decken mit dem Verarbeitenden Gewerbe und den privaten Dienstleistungsbereichen einen Großteil der Bruttowertschöpfung ab und sollten damit die Gesamtwirtschaft gut repräsentieren. Zudem wird bei dieser Befragung implizit die Auslastung beider Produktionsfaktoren, Kapital und Arbeit, berücksichtigt. So argumentiert die Europäische Zentralbank, „dass die Umfrageteilnehmer ihre Beurteilung der Kapazitätsauslastung auf die Nutzung sämtlicher Ressourcen des Betriebs, d. h. sowohl den Kapital- als auch den Arbeitseinsatz, gründen. Diese Annahme basiert auf dem Gesamthalt der Erhebung, in der ausdrücklich nach verschiedenen Produktionshemmnissen wie der Verknappung von Kapital, Arbeitskräften und sonstigen Vorleistungen gefragt wird, wobei davon ausgegangen wird, dass die Teilnehmer bei der Bewertung ihrer Kapazitätsauslastung alle genannten Produktionsfaktoren berücksichtigen.“ (Europäische Zentralbank, 2007, S. 51, Fußnote 2).

#### Bestimmung der Konjunkturkomponente mit Hilfe der Kapazitätsauslastungslücke

Die gesamtwirtschaftliche Kapazitätsauslastung („Capacity Utilization“)  $CU_t$  wird als gewichteter Durchschnitt der Auslastungsgrade im Verarbeitende Gewerbe und in den privaten Dienstleistungsbereichen errechnet. In Analogie zum Indikator der wirtschaftlichen Einschätzung der Europäischen Kommission („Economic Sentiment Indicator“) entfällt dabei ein Gewicht von 4/7 auf das Verarbeitende Gewerbe und die verbleibenden 3/7 auf die Dienstleistungsbereiche (vgl. Europäische Kommission 2021). Die Normalauslastung der Kapazitäten wird im Einklang mit der Vorgehensweise von Marcellino und Musso (2011) als langfristiger Mittelwert  $\overline{CU}$  definiert. Dieser Mittelwert wird zu jedem Schätzzeitpunkt  $t + \tau$  auf Basis aller  $n$  zur Verfügung stehenden Beobachtungen der Kapazitätsauslastung berechnet:

$$(5) \quad \overline{CU}_\tau = \frac{1}{n} \sum_{i=1}^{t+\tau} CU_i.$$

Damit kann es auch rückwirkend zu geringfügigen Änderungen des Mittelwertes kommen, da eine neue Beobachtung sowohl den Summenterm als auch die Anzahl der Beobachtungen  $n$  verändert. Allerdings tendiert der Einfluss einer neuen Beobachtung auf den Mittelwert der Zeitreihe mit wachsender Anzahl an Beobachtungen zunehmend gegen Null. Die konjunkturelle Lage der Volkswirtschaft wird schließlich mit der Kapazitätsauslastungslücke („Capacity Utilization Gap“)  $CUG$  gemessen, die sich in Abhängigkeit vom Schätzzeitpunkt  $t + \tau$  für das Jahr  $t$  aus der Abweichung der in Prozent gemessenen Kapazitätsauslastung  $CU_t$  von ihrem langjährigen Mittelwert  $\overline{CU}_\tau$  berechnet:

$$(6) \quad CUG_{\tau,t} = CU_t - \overline{CU}_\tau.$$

Wenn  $CUG_{\tau,t} > 0$  ist, befindet sich die Volkswirtschaft in einer Überauslastung; wenn  $CUG_{\tau,t} < 0$ , liegt eine Unterauslastung vor.

Abschließend muss die Kapazitätsauslastungslücke  $CUG$  noch in die Konjunkturkomponente  $Y$  des staatlichen Finanzierungssaldos überführt werden. Zu diesem Zweck wird die Vorgehensweise des aktuellen EU-Verfahrens erweitert und eine lineare Abhängigkeit der Konjunkturkomponente von der Kapazitätsauslastungslücke unterstellt:

$$(7) \quad Y = \varepsilon_{B,OG} \varepsilon_{OG,CUG} CUG.$$

Während die Budgetsemielastizität  $\varepsilon_{B,OG}$  aus dem EU-Verfahren übernommen wird, muss die Elastizität der Produktionslücke  $OG$  gegenüber der Kapazitätsauslastungslücke  $CUG$

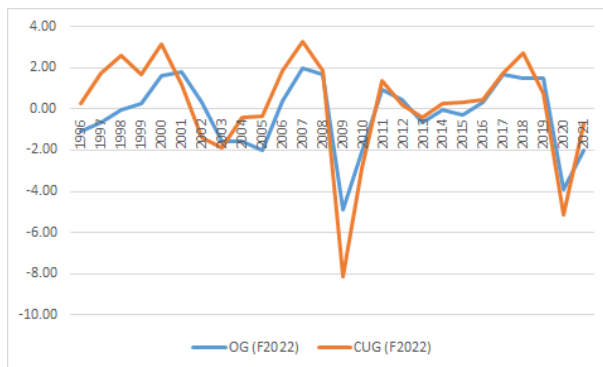
$$(8) \varepsilon_{OG,CUG} = d\ln(OG + 1)/d\ln(CUG + 1)$$

geschätzt werden. Dabei wird der hohe Gleichlauf zwischen beiden Variablen ausgenutzt (vgl. Abbildung 3). Der Schätzzeitraum wurde auf die Jahre zwischen 1996 und 2018 begrenzt, so dass der größte Teil der nachträglichen Revisionen der Schätzungen der Produktionslücke nach der Methode der EU-Kommission bereits stattgefunden haben dürfte. Die Ergebnisse der Schätzung

$$(9) \ln(OG) = \beta \ln(CUG) + u$$

werden in Tabelle 1 zusammengefasst. Der Schätzwert für  $\beta$  wird im Folgenden für die Elastizität  $\varepsilon_{OG,CUG}$  verwendet. Das hohe Bestimmtheitsmaß  $R^2$  verdeutlicht noch einmal die sehr hohe Korrelation zwischen beiden Zeitreihen.

Abbildung 3: Produktions- und Kapazitätsauslastungslücke im Vergleich



Anmerkung: Beide Lücken spiegeln den im Frühjahr 2022 zur Verfügung stehenden Informationsstand wider. Die Produktionslücke wird in Prozent des Produktionspotenzials gemessen, die Kapazitätsauslastungslücke als Abweichung der Kapazitätsauslastung vom Mittelwert in Prozentpunkten.

Tabelle 1: Elastizität der Produktionslücke gegenüber der Kapazitätsauslastungslücke

	$CUG$	Beobachtungen	korrigiertes $R^2$
$OG$ (EU-CAM)	0,55*** (0,07)	23	0,71

Anmerkung: Der Wert in Klammern gibt den Standardfehler an. \*\*\* stehen für einen p-Wert von unter 1 Prozent.

## Prognose der Kapazitätsauslastungslücke

Zur Bewertung der staatlichen Finanzierungssalden im Prognosezeitraum muss die Konjunkturkomponente fortgeschrieben werden. Für die EU-Methode ist ohnehin eine gesamtwirtschaftliche Projektion erforderlich, die u.a. eine Prognose des preisbereinigten Bruttoinlandsprodukts und somit der Produktionslücke umfasst. Bei der umfragegestützten Methode muss in einem zusätzlichen Arbeitsschritt die Kapazitätsauslastungslücke fortgeschrieben werden. Diese Prognose kann sowohl auf Basis von aggregierten Jahresdaten als auch ausgehend von den vierteljährlichen Ursprungswerten der Kapazitätsauslastung durchgeführt werden. Um einen fairen Vergleich der umfragegestützten Methode mit der EU-Methode zu gewährleisten, wird dabei nur auf Daten zurückgegriffen, die zum jeweiligen Zeitpunkt der Erstellung der EU-Projektionen zur Verfügung stehen. So liegt im Frühjahr bereits die Veröffentlichung der Kapazitätsauslastung im ersten Quartal des jeweiligen Jahres vor; im Herbst sind es die ersten drei Quartale.<sup>6</sup>

<sup>6</sup> Die Frage nach der aktuellen Kapazitätsauslastung wird immer im ersten Monat eines Quartals gestellt.

### Prognose mit Jahresdaten (Jahresvariante 1)

Zur Fortschreibung der Kapazitätsauslastung mit Jahresdaten werden zwei Varianten umgesetzt. Bei der ersten Variante werden bereits vorhandene Vierteljahresdaten für das laufende Jahr verworfen. Für die zurückliegenden Jahre werden die Jahreswerte als Durchschnitt der Vierteljahreswerte berechnet. Zur Fortschreibung der Kapazitätsauslastung wird ein einfaches zeitreihenökonomisches Modell spezifiziert

$$(10) \quad CU_t = \alpha + \beta_1 CU_{t-1} + \beta_2 \Delta BIP_t + u_t,$$

in dem als exogene Variable die Jahreswachstumsrate des preisbereinigten Bruttoinlandsprodukts  $\Delta BIP_t$  einfließt. Der Beginn des Schätzzeitraum (1996) wird durch die Verfügbarkeit der Kapazitätsauslastung in den Dienstleistungsbereichen bestimmt, das Ende (2021) durch die zum Zeitpunkt des Erstellens der vorliegenden Studie verfügbaren Realisationen der Kapazitätsauslastung und des preisbereinigten Bruttoinlandsprodukts.

Mit Hilfe der Schätzgleichung und unter Verwendung von Prognosewerten für die Jahreswachstumsrate des preisbereinigten Bruttoinlandsprodukts wird dann im Prognosezeitraum die Kapazitätsauslastung fortgeschrieben:

$$(11) \quad CU_t = \hat{\alpha} + \hat{\beta}_1 CU_{t-1} + \hat{\beta}_2 \Delta BIP_t.$$

Dabei wird auf die Projektionen der Bundesregierung für den Zeitraum von Frühjahr 2011 bis Frühjahr 2022 zurückgegriffen, um Echtzeitdaten der Kapazitätsauslastungslücke einschließlich ihrer Prognose zu erhalten. Abschließend wird die Kapazitätsauslastungslücke unter Verwendung der Gleichungen (5) und (6) konstruiert.

### Prognose mit Jahresdaten (Jahresvariante 2)

Bei der zweiten Variante der Fortschreibungsmethode mit Jahresdaten verwerfen wir bei der Frühjahrsprojektion in Analogie zu Variante 1 die bereits vorhandenen Vierteljahresdaten der Kapazitätsauslastung für das erste Quartal des laufenden Jahres und nutzen zur Prognose der Kapazitätsauslastung (inkl. des laufenden Jahres) das Regressionsmodell (10). Bei der Herbstprojektion greifen wir im Gegensatz dazu auf die drei bereits veröffentlichten Quartale zurück und nutzen ihren Durchschnitt als Schätzwert für das laufende Jahr  $t$ :

$$(12) \quad CU_t = 1/3 (CU_{t,Q1} + CU_{t,Q2} + CU_{t,Q3}).$$

Für die Prognose der darauffolgenden Jahre kommt wiederum das Regressionsmodell (10) zum Einsatz. Wie bei Variante 1 wird auf die Projektionen der Bundesregierung für den Zeitraum von Frühjahr 2011 bis Frühjahr 2022 zurückgegriffen, um Echtzeitdaten der Kapazitätsauslastungslücke einschließlich ihrer Prognose zu erhalten. Abschließend wird die Kapazitätsauslastungslücke unter Verwendung der Gleichungen (5) und (6) konstruiert.

### Prognose mit Vierteljahresdaten (Vierteljahresvariante)

Zur Fortschreibung der Kapazitätsauslastung mit Vierteljahresdaten werden alle zum Zeitpunkt der Schätzung zur Verfügung stehenden Daten verwendet. Dabei wird ein einfaches zeitreihenökonomisches Modell spezifiziert

$$(13) \quad CU_t = \alpha + \beta_1 CU_{t-1} + \beta_2 CU_{t-2} + \beta_3 CU_{t-3} + \beta_4 \Delta BIP_t + \beta_5 \Delta BIP_{t-1} + u_t,$$

in dem als exogene Variable die vierteljährlichen Verlaufsraten (Vorquartalsveränderungsraten) des preisbereinigten Bruttoinlandsprodukts  $\Delta BIP_t$  einfließen. Der Beginn des Schätzzeitraum (1995 Q2) wird durch die Verfügbarkeit der Kapazitätsauslastung in den Dienstleistungsbereichen bestimmt, das Ende (2022 Q1) durch die zum Zeitpunkt des Erstellens der vorliegenden Studie verfügbaren Realisationen der Kapazitätsauslastung und des preisbereinigten Bruttoinlandsprodukts.

Mit Hilfe der Schätzgleichung und unter Verwendung der Prognosewerte für die vierteljährliche Verlaufsrate des preisbereinigten Bruttoinlandsprodukts wird dann im Prognosezeitraum die Kapazitätsauslastung fortgeschrieben:

$$(14) \quad CU_t = \hat{\alpha} + \hat{\beta}_1 CU_{t-1} + \hat{\beta}_2 CU_{t-2} + \hat{\beta}_3 CU_{t-3} + \hat{\beta}_4 \Delta BIP_t + \hat{\beta}_5 \Delta BIP_{t-1}.$$

Da vergangene Prognosen der vierteljährlichen Verlaufsrate der EU-Kommission oder der Bundesregierung nicht verfügbar sind, greifen wir in der vorliegenden Studie auf Prognosen der Projektgruppe Gemeinschaftsdiagnose für den Zeitraum von Frühjahr 2009 bis Frühjahr 2022 zurück.<sup>7</sup> Die Vierteljahreswerte der Kapazitätsauslastung werden abschließend zu Jahresdurchschnitten aggregiert und die Kapazitätsauslastungslücke unter Verwendung der Gleichungen (5) und (6) konstruiert.

### Revision der Konjunkturkomponente bei der umfragegestützten Methode

Abbildung 4 veranschaulicht die durchschnittlichen Revisionen der Konjunkturkomponente des strukturellen Finanzierungssaldos für Deutschland, die sich auf Basis aller Varianten der umfragegestützten Methode für die Jahre von 2010 bis 2020 ergeben, und vergleicht sie mit den Revisionen nach der EU-Methode. Dabei lassen sich folgende Ergebnisse festhalten.

Zum einen weisen die Schätzungen der Konjunkturkomponente bei der umfragegestützten Methode eine hohe ex-post-Verlässlichkeit auf. So sind nach Abschluss eines Jahres  $t$  nachträgliche Veränderungen der Konjunkturkomponente bei allen Varianten der umfragegestützten Methode vernachlässigbar und ergeben sich nur durch die Anpassungen des Mittelwertes  $\overline{CU}_t$ . Bei den Konjunkturkomponenten, die im Rahmen der EU-Methode geschätzt werden, fallen diese Revisionen durch regelmäßige Neuschätzungen und Anpassungen des Produktionspotenzials deutlich größer aus.

Zum anderen weisen die Schätzungen der Konjunkturkomponente bei der umfragegestützten Methode eine geringfügig höhere ex-ante-Verlässlichkeit auf als die Schätzungen nach der EU-Methode. So fallen vor Abschluss eines Jahres  $t$  die Revisionen der Konjunkturkomponente bei der umfragegestützten Methode etwas kleiner aus. Die ex-ante-Revisionen ergeben sich vor allem durch Anpassungen der Konjunkturprognose, was sich bei beiden Methoden in ähnlicher Weise auf die Schätzung der Konjunkturkomponente niederschlagen dürfte. Je stärker jedoch die umfragegestützte Methode bei der Prognose der Kapazitätsauslastungslücke auf bereits realisierte Kapazitätsauslastungen zurückgreifen kann und je mehr unterjährige Information damit die Prognosegüte verbessern, desto stabiler werden die Ergebnisse. Entsprechend fallen sowohl bei der zweiten Variante mit Jahresdaten (Abbildung 4 B.) als auch bei der Vierteljahresvariante (Abbildung 4 C.) die Revisionen für das laufende Jahr  $t$  geringer aus. Im Mittel sind sie bei diesen Varianten fast Null.

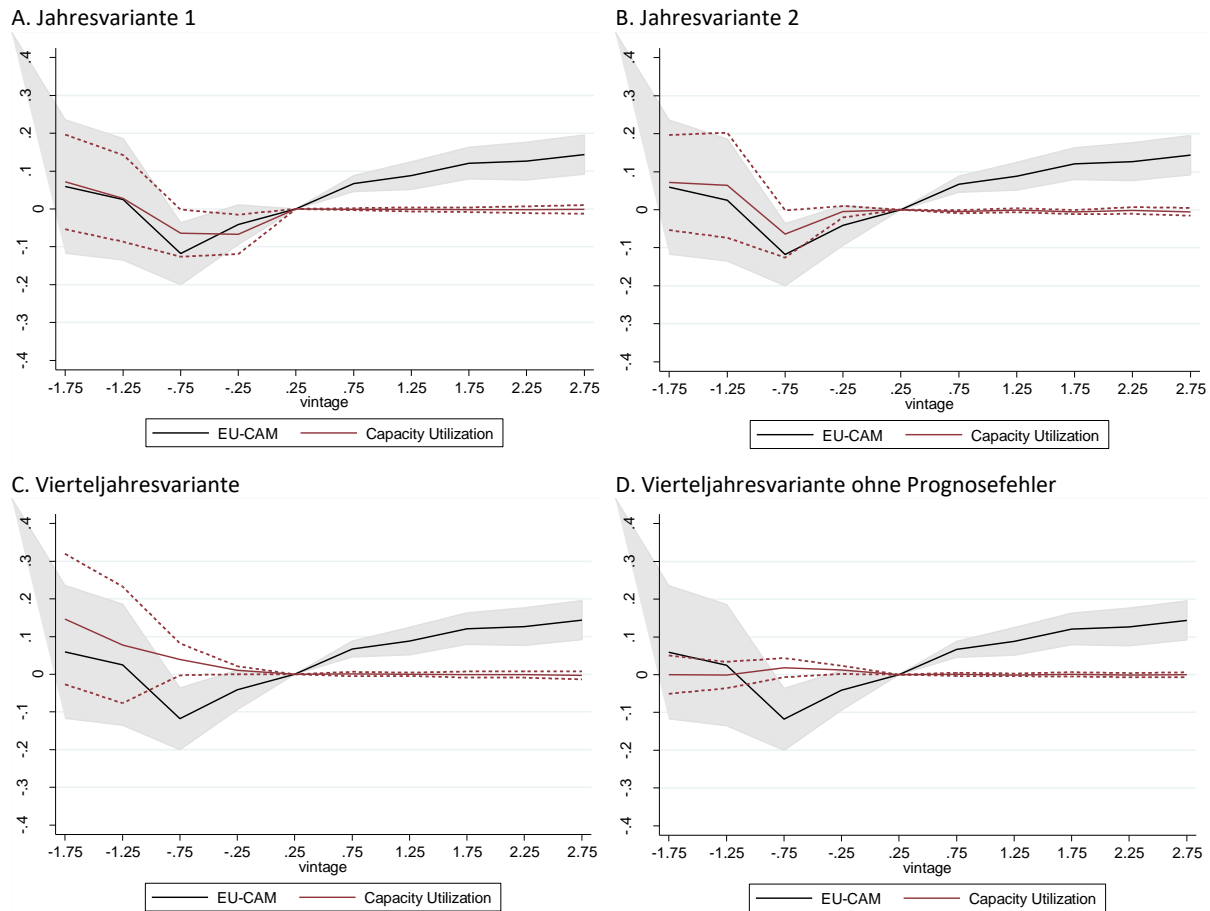
Zudem veranschaulicht Abbildung 4 D., dass ein Großteil der ex-ante-Revisionen bei der umfragegestützten Methode auf Fehler bei Prognose des Verlaufs des preisbereinigten Bruttoinlandsprodukts zurückzuführen sind. Im Unterschied zur Variante mit Vierteljahresdaten, in denen die Prognosen des preisbereinigten Bruttoinlandsprodukts der Projektgruppe Gemeinschaftsdiagnose einfließen (Abbildung 4 C.), wird nun ein perfekter Prognostiker unterstellt, der immer die korrekten Verlaufsrate des preisbereinigten Bruttoinlandsprodukts voraussagt. Im Ergebnis fallen die durchschnittlichen ex-ante-Revision der Konjunkturkomponente deutlich geringer

---

<sup>7</sup> Die halbjährlichen Prognosen der Projektgruppe Gemeinschaftsdiagnose dienen der Bundesregierung als Referenz für ihre eigenen Projektionen. Dies dürfte einer der Gründe dafür sein, warum es in den vergangenen Jahren zumindest für die jahresdurchschnittlichen Veränderungsraten des preisbereinigten Bruttoinlandsprodukts nur zu geringen Unterschieden gekommen ist, die zudem nicht systematisch, sondern eher zufälliger Natur waren (vgl. Wollmershäuser 2017). Deshalb kann davon ausgegangen werden, dass es auch bei den prognostizierten vierteljährlichen Verläufen des preisbereinigten Bruttoinlandsprodukts keine systematischen Abweichungen gibt.

aus. Die verbleibenden Revisionen sind auf den Fehler zurückzuführen, der aus der Schätzung des Zusammenhangs zwischen der Kapazitätsauslastung und der Veränderung des preisbereinigten Bruttoinlandsprodukts auf Basis von Gleichung (13) resultiert. Mit diesem hypothetischen Prognostiker soll lediglich gezeigt werden, dass im Rahmen der umfragegestützten Methode Verbesserungspotenzial vor allem im Hinblick auf das vierteljährliche Prognosemodell besteht.

Abbildung 4: Durchschnittliche Revisionen der Konjunkturkomponenten

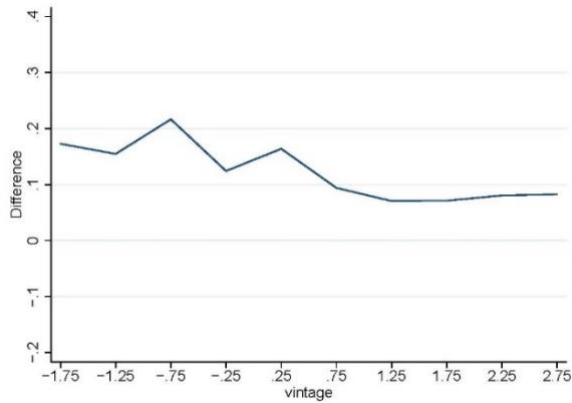


Anmerkung: Die durchschnittliche Revision der Konjunkturkomponente wird in Prozentpunkten des Produktionspotenzials gemessen. Die rote Linie zeigt die Revision der Konjunkturkomponente bei der umfragegestützten Methode, die schwarze Linie die der EU-Methode. Die rot gestrichelten Linien und die grau schraffierte Fläche zeigen das 95%-Konfidenzintervall der jeweiligen Schätzungen.

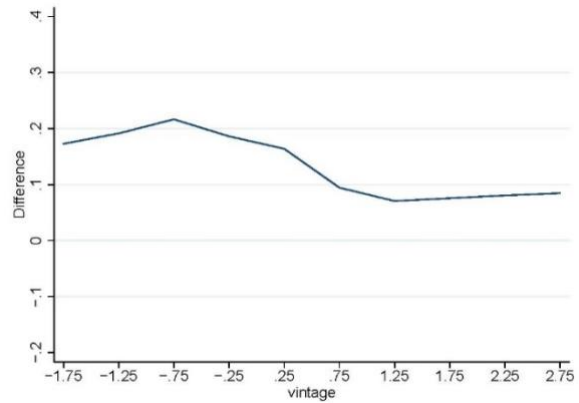
Neben der geringeren Anfälligkeit der Schätzungen der Konjunkturkomponente für Revisionen weist die umfragegestützte Methode auch keine systematischen Revisionen auf. Insbesondere zeigt Abbildung 5, dass sich die nach der EU-Methode geschätzte Konjunkturkomponente im Zeitablauf jener der umfragegestützten Methode annähert. Dies deutet darauf hin, dass die Kapazitätsauslastungslücke bereits frühzeitig ein verlässliches und insbesondere unverzerrtes Signal über die aktuelle konjunkturelle Lage liefert und dass sich die nach der EU-Methode geschätzten Konjunkturkomponenten im Zuge ihrer systematischen Aufwärtsrevision nur allmählich dieser Diagnose anschließen. Im Schnitt reduziert sich im betrachteten Vierjahres-Fenster (von  $\tau = -1,75$  bis  $\tau = +2,75$ ) die Überschätzung des strukturellen Finanzierungssaldos und damit des fiskalischen Handlungsspielraums um etwa 0,1 Prozentpunkt des Produktionspotenzials.

Abbildung 5: Differenz der Konjunkturkomponente zwischen der EU-Methode und der umfragegestützten Methode

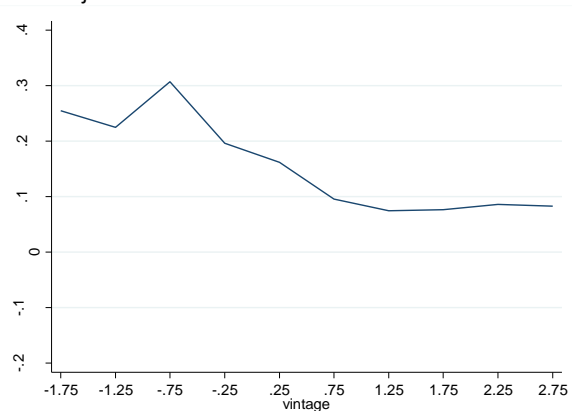
A. Jahresvariante 1



B. Jahresvariante 2



C. Vierteljahresvariante



Anmerkung: Die Differenz der Konjunkturkomponente zwischen der EU-Methode und der umfragegestützten Methode wird in Prozentpunkten des Produktionspotenzials gemessen.

## Fazit

Im vorliegenden Beitrag vergleichen wir eine neue umfragegestützte Methode, die die konjunkturelle Lage anhand der Kapazitätsauslastung misst, mit der bisher im Rahmen der EU-Haushaltsüberwachung verwendeten Produktionsfunktionsmethode. Wir zeigen, dass die Revisionen der Konjunkturkomponente bei der umfragegestützten Methode deutlich geringer sind.

Insbesondere tendieren die nachträglichen Anpassungen der Schätzungen bereits abgelaufener Jahre (sog. ex-post-Revisionen) nahezu gegen Null. Aber auch die Anpassungen der Prognosen zukünftiger Konjunkturkomponenten (sog. ex-ante-Revisionen) sind geringfügig kleiner als bei der Produktionsfunktionsmethode. Sie resultieren in erster Linie aus den Fehlern bei der Prognose des preisbereinigten Bruttoinlandsprodukts, auf Basis derer die Kapazitätsauslastung fortgeschrieben wird, und nicht aus der umfragegestützten Methode an sich. Zudem weist die umfragegestützte Methode im Gegensatz zur Produktionsfunktionsmethode keine systematischen Revisionen auf. Somit liefert die Kapazitätsauslastung bereits frühzeitig ein verlässliches und insbesondere unverzerrtes Signal über die aktuelle konjunkturelle Lage.

Es spricht vieles dafür, bei der Bestimmung der strukturellen Finanzierungssalden vom traditionellen Produktionsfunktionsansatz auf einen umfragegestützten Ansatz zu wechseln. Neben der geringeren Revisionsanfälligkeit der Konjunkturkomponente des staatlichen Finanzierungssaldos hat das Konzept der umfragegestützten strukturellen Finanzierungssalden den Vorteil, dass es leicht zu kommunizieren ist und keine Möglichkeiten der politischen Einflussnahme bietet. Die Berechnung der Kapazitätsauslastung basiert seit vielen Jahrzehnten auf einer einfachen und innerhalb der EU



harmonisierten Aggregation der Ergebnisse einer repräsentativen Unternehmensbefragung. Da die Unternehmen über ihre eigene Lage berichten und eine ausreichend große Anzahl an Umfrageteilnehmern (in Deutschland circa 9 000 Meldungen) gewährleistet ist, ist eine Beeinflussung der Ergebnisse schwer vorstellbar. Im Gegensatz dazu beruhen Schätzungen der strukturellen Finanzierungssalden mit der Produktionsfunktionsmethode auf einer Reihe von Annahmen bezüglich der modelltheoretischen und empirischen Spezifikation. Zwar werden die Schätzergebnisse des Produktionsfunktionsansatzes durch die EU-Kommission ausführlich erläutert (vgl. Blondeau et al. 2021). Allerdings führt die Vielzahl der zu treffenden Annahmen zwangsläufig dazu, dass die Richtigkeit der gewählten Spezifikation in Frage gestellt wird und durch eine Änderung einzelner Annahmen das Ergebnis, möglicherweise auch zu politischen Zwecken, verändert werden kann (vgl. Fioramanti und Waldmann 2017, sowie Heimberger et al. 2020). Es ist daher wenig überraschend, dass das Herzstück der Haushaltsüberwachung in der EU über die Jahre hinweg an Glaubwürdigkeit verloren hat.

## Literatur

Ademmer, M., Boysen-Hogrefe, J., Carstensen, K., Hauber, Ph., Janssen, N., Kooths, St. Rossian, Th. und Stolzenburg, U. (2019). Schätzung von Produktionspotenzial und -lücke: Eine Analyse des EU-Verfahrens und mögliche Verbesserungen. Kieler Beiträge zur Wirtschaftspolitik 19, Kiel Institute for the World Economy (IfW Kiel).

Blondeau, F., Planas, Ch. und Rossi, A. (2021). Output Gap Estimation Using the European Union's Commonly Agreed Methodology: Vade Mecum and Manual for the EUCAM Software. European Economy - Discussion Papers Nr. 148, Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Deutsche Bundesbank (2014). Zur Verlässlichkeit der Schätzungen internationaler Organisationen zur Produktionslücke. Monatsbericht April, 13 – 38.

ECOFIN (2017). Specifications on the Implementation of the Stability and Growth Pact and Guidelines on the Format and Content of Stability and Convergence Programmes. <http://data.consilium.europa.eu/doc/document/ST-9344-2017-INIT/en/pdf>, [letzter Zugriff: 18.01.2022].

Europäische Kommission (2018). Vade Mecum on the Stability and Growth Pact. European Economy Institutional Paper - March 2018 075, European Commission.

Europäische Kommission (2021). The Joint Harmonised EU Programme of Business and Consumer Surveys – User Guide. Februar 2021, European Commission.

Europäische Kommission (2022). Communication on orientations for a reform of the EU economic governance framework. COM (2022) 583 final.

Europäische Zentralbank (2007). Kapazitätsauslastung im Euroraum: Messgrößen und Auswirkungen. EZB Monatsbericht Oktober, 50 – 54.

Feld, L. P., Nöh, L., Reuter, W. H. und Yeter, M. (2021). Von der Corona-bedingten Schuldenaufnahme zur Wiedereinhaltung der Schuldenbremse. Perspektiven der Wirtschaftspolitik, 22 (4) 330 – 349.

Fioramanti, M. und Waldmann, R. J. (2017). The Econometrics of the EU Fiscal Governance: Is the European Commission Methodology still Adequate? *Rivista Internazionale Di Scienze Sociali*, 125 (4), 389 – 404.

Götttert, M. und Wollmershäuser, T. (2021). Survey-Based Structural Budget Balances. CESifo Working Paper Nr. 8911.

Heimberger, Ph., Huber, J. und Kapeller, J. (2020). The Power of Economic Models: the Case of the EU's Fiscal Regulation Framework, *Socio-Economic Review*, 18 (2), 337 – 366.

Marcellino, M. and Musso, A. (2011). The Reliability of Real-Time Estimates of the Euro Area Output Gap. *Economic Modelling*, 28 (4), 1842 – 1856.

Orphanides, A., Porter, R. D., Reifschneider, D., Tetlow, R. and Finan, F. (2000). Errors in the measurement of the output gap and the design of monetary policy. *Journal of Economics and Business*, 52 (1-2), 117 – 141.

Wollmershäuser, T. (2017). Die Konjunkturprognosen der Gemeinschaftsdiagnose und des Bundeswirtschaftsministeriums im Vergleich. *ifo Schnelldienst*, 70 (3), 40 – 45.

## **Vorschlag zur Bestimmung des Produktionspotenzials im Rahmen der grundgesetzlich verankerten Schuldenregel**

### **Inhaltsverzeichnis**

Einleitung	2
Zielsetzung	2
Die Berechnung des Produktionspotenzials bei „normaler“ Auslastung der Kapazitäten	2
Konjunkturbereinigung	3
Abweichung der Ergebnisse der eigenen Berechnungen von denen des BMWK nach der EU-Methode	4
Relevanz der Berechnungsmethode des Produktionspotenzials für die Wirtschaftspolitik, die Konjunktur, den Arbeitsmarkt, die Entwicklung des öffentlichen Finanzierungssaldos und die Preisstabilität	5
Konjunkturkomponente	7
Anhang: Grafiken basierend auf aktuellstem Stand der tatsächlichen Werte, abgerufen: 5. Mai 2023	8
Endnoten	11

## Einleitung

Die hier vorgeschlagene Methode zur Bestimmung des gesamtwirtschaftlichen Produktionspotenzials baut nicht auf dem gängigen Produktionsfunktionsansatz der gemeinsamen EU-Methode (EU-CAM) auf. Der Vorschlag ist jedoch – wie der Ansatz der gemeinsamen EU-Methode – auf Grundlage der Daten der Volkswirtschaftlichen Gesamtrechnung (VGR) operationalisierbar und quantifizierbar. Die angewandte Methode basiert auf Überlegungen von Claus Köhler, ehemaliges Mitglied des Sachverständigenrats zur Begutachtung der gesamtwirtschaftlichen Entwicklung und des Direktoriums der Deutschen Bundesbank.

## Zielsetzung

Ziel des Vorschlags ist ein Berechnungsverfahren, das eine realistische, verlässliche und nachvollziehbare Bestimmung des Produktionspotenzials auf Basis von Originalwerten gewährleistet und die Anwendbarkeit und Ausrichtung auf wirtschaftspolitische Ziele (Einhaltung der Schuldenregel, Vollbeschäftigung, Preisstabilität) sicherstellt.

## Die Berechnung des Produktionspotenzials bei „normaler“ Auslastung der Kapazitäten

Das Produktionspotenzial nach Köhler ist das Produkt aus Erwerbspersonen\*Arbeitsproduktivität je Erwerbstätigen.<sup>1</sup> Die Erwerbspersonen setzen sich aus den Erwerbstätigen und den Erwerbslosen zusammen. Diese Größen sind Teil der VGR.

Für Deutschland ließen sich alternativ zur Zahl der Erwerbslosen die Zahl der registrierten Arbeitslosen aus der Statistik der Bundesagentur für Arbeit zugrunde legen<sup>2</sup>, die Zahl der Unterbeschäftigten<sup>3</sup> oder auch das „unausgeschöpfte Arbeitskräftepotenzial“ („labour market slack“<sup>4</sup>). Auch die Einbeziehung der so genannten Stillen Reserve, die das Statistische Bundesamt und das Institut für Arbeitsmarkt und Berufsforschung (IAB) schätzen<sup>5</sup>, unfreiwillige Teilzeit<sup>6</sup> und eine höhere Erwerbsbeteiligung/Erwerbsquote von Frauen<sup>7</sup> könnten berücksichtigt werden. Das Arbeitspotenzial und mit diesem auch das Produktionspotenzial würden sich entsprechend verändern.

Aus Gründen der Vergleichbarkeit mit den Ergebnissen der gemeinsamen EU-Methode wird für diesen Vorschlag die Zahl der Erwerbspersonen zugrunde gelegt und mit Werten gerechnet, auf denen auch die Berechnungen des Bundesministeriums für Wirtschaft und Klimaschutz (BMWK)<sup>8</sup> basieren (siehe Grafiken unten und die Excel-Datei „Thorsten Hild – Angaben – Berechnungen – Produktionspotenzial – 12-05-2023“).

Die Erwerbspersonen bilden das Arbeitspotenzial. Multipliziert mit der Arbeitsproduktivität (je Erwerbstätigen oder je Erwerbstätigenstunde; beide Größen sind Bestandteil der VGR) ergibt sich das Produktionspotenzial. Die Arbeitsproduktivität beinhaltet das technische Niveau der Volkswirtschaft.

Köhler selbst grenzt seine Methode von der gemeinsamen EU-Methode zur Schätzung des Produktionspotenzials ab:

„Üblich ist nicht eine so einfache Definition des Produktionspotenzials. Häufig werden komplizierte mathematische Modelle entwickelt, um diese Größe zu schätzen...Das Bundesministerium für Wirtschaft und Technologie schätzt das Produktionspotential ´aus den Produktionsfaktoren Arbeit, Kapital und der so genannten Totalen Faktorproduktivität (TFP).´ Das Ministerium fügt hinzu: ´Das Produktionspotenzial ist eine in der Realität nicht beobachtbare Größe. Sie wird vielmehr mithilfe statistischer Verfahren geschätzt.“<sup>9</sup>

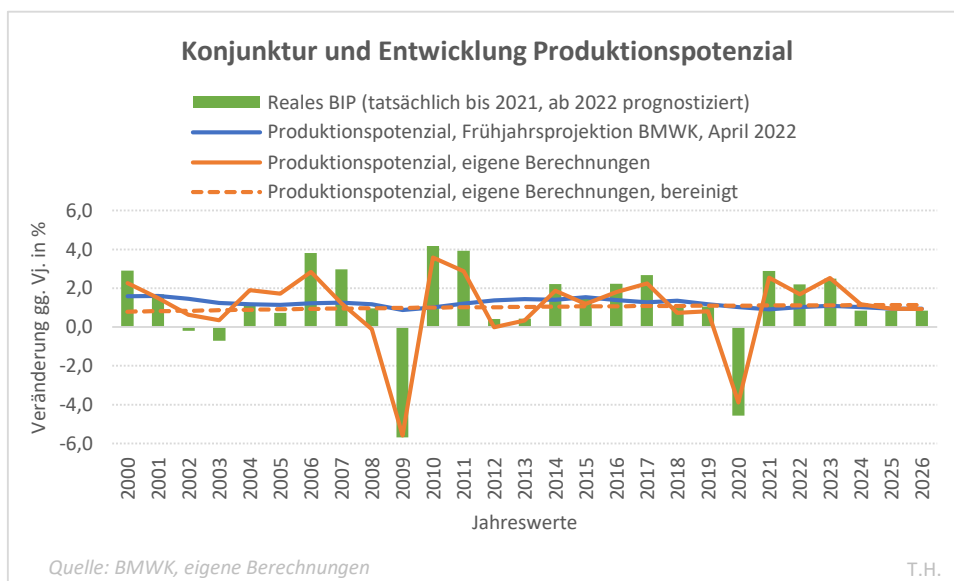
Um eine „normale“ Auslastung der Kapazitäten zu berechnen, berücksichtige ich – anders als Köhler – bei der Berechnung des Produktionspotenzials eine Arbeitslosen- bzw. Erwerbslosenquote, die Vollbeschäftigung signalisiert.

Köhler begründet – losgelöst von seiner Berechnung zum Produktionspotenzial – eine unvermeidbare Arbeitslosigkeit mit saisonaler (2%) und friktioneller Arbeitslosigkeit (1%).<sup>10</sup> Pascal Michailat, Emmanuel Saez haben wiederum erst 2022 Vollbeschäftigung („efficient unemployment“) als „geometrisches Mittel von Arbeitslosenquote und der Quote offener Stellen“ definiert.<sup>11</sup> Nach eigenen Berechnungen auf der Grundlage der Methode von Michailat und Saez, bewegte sich die „efficient unemployment rate“ – im Folgenden: Vollbeschäftigungsquote – für Deutschland in den Jahren 2004 bis 2022 zwischen 2,3 und 3,1 Prozent. Bei einem Mittelwert von 2,8 Prozent bestätigt sich annähernd die Annahme Köhlers, der aufgrund von saisonaler und friktioneller Arbeitslosigkeit Vollbeschäftigung bei einer Arbeitslosenquote von 3 Prozent annimmt. Da sich die in der VGR enthaltene Erwerbslosigkeit von der registrierten Arbeitslosigkeit unterscheidet, ist entsprechend der Berechnung auf der Grundlage der Methode von Michailat und Saez die Arbeitslosenquote durch die Erwerbslosenquote zu ersetzen. Das Ergebnis ist eine Vollbeschäftigungsquote von 2,5 Prozent (die Berechnung der Vollbeschäftigungsquote ist in der Excel-Datei „Thorsten Hild – Angaben – Berechnungen – Produktionspotenzial – 12-05-2023“ enthalten).

### Konjunkturbereinigung

Die Entwicklung der Arbeitsproduktivität, der Erwerbstätigen und der Erwerbslosen steigt und fällt mit der Konjunktur, der Zuwachsrates des realen Bruttoinlandsprodukts (BIP). Diese Schwankungen bilden sich auch in der Entwicklung des Produktionspotenzials ab (siehe Linie „Produktionspotenzial, eigene Berechnungen“ in Grafik 1).

Grafik 1:



Um die Entwicklung des Produktionspotenzials von diesen Schwankungen zu bereinigen wird das folgende Verfahren gewählt. Es sei hier aber bereits darauf verwiesen, dass in dem Abschnitt „Konjunkturkomponente“ (unten Seite 7) mit den Originalwerten gearbeitet werden kann, da die zentrale Grundlage die Differenz aus der Zuwachsrates des realen BIP und der Zuwachsrates des Produktionspotenzials ist.

1. Gegeben sind für den jeweiligen Beobachtungszeitraum das reale BIP (tatsächlich und prognostiziert), die Zahl der Erwerbstätigen und der Erwerbslosen (tatsächlich und prognostiziert). Damit ist auch die Arbeitsproduktivität je Erwerbstätigen gegeben bzw. lässt sich diese Größe berechnen (reales BIP/Zahl der Erwerbstätigen).

2. Für die Konjunkturbereinigung wird der Beobachtungszeitraum gewählt (hier 1999-2026) und für diesen Zeitraum die durchschnittliche jährliche Wachstumsrate für die Größen Erwerbstätige, Erwerbslose und Arbeitsproduktivität je Erwerbstätigen berechnet. Ausgehend von den Ausgangswerten 1999 werden jene Größen nun mit der durchschnittlichen jährlichen Wachstumsrate für den Beobachtungszeitraum (hier bis 2026) neu berechnet und ergeben die um die Konjunkturschwankungen bereinigten Werte. Auf Basis dieser bereinigten Werte werden schließlich das bereinigte reale BIP und das bereinigte Produktionspotenzial berechnet, aus deren Differenz sich wiederum die bereinigte Produktionslücke ergibt.

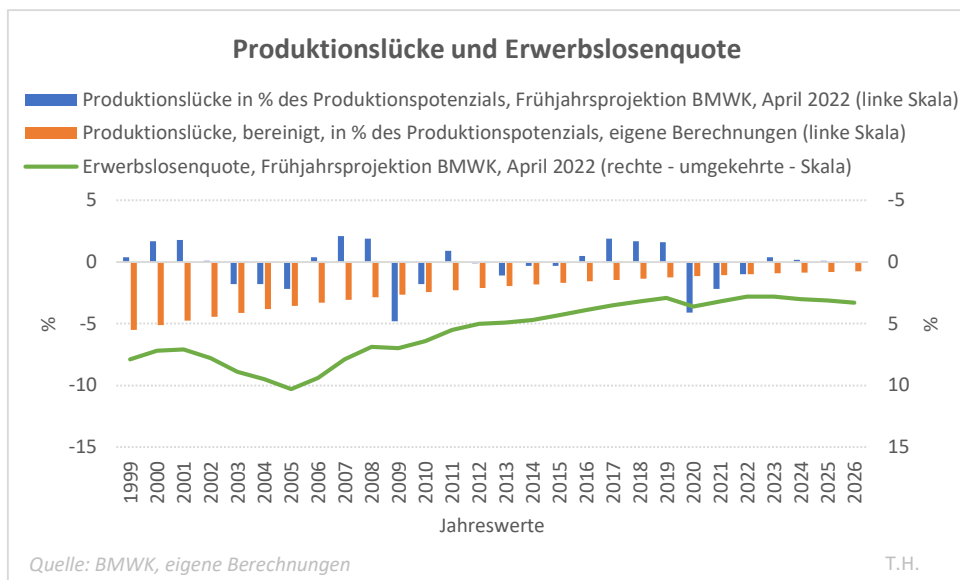
Grafik 1 zeigt, dass die Zuwachsrate „Produktionspotenzial, eigene Berechnungen, bereinigt“ nicht wesentlich von der Zuwachsrate „Produktionspotenzial, Frühjahrsprojektion BMWK, April 2022“ abweicht.

### Abweichung der Ergebnisse der eigenen Berechnungen von denen des BMWK nach der EU-Methode

Erhebliche Abweichungen zwischen den Berechnungen/Schätzungen des BMWK und den eigenen Berechnungen zeigen sich allerdings in der Produktionslücke (siehe Grafiken 2 und 3).

Die bereinigte Produktionslücke nach eigenen Berechnungen (Grafik 2) zeigt, erstens, einen stabileren Verlauf, als die nach Berechnungen/Schätzungen des BMWK. Zweitens, von der Entwicklung im Zuge schwerer Wirtschaftskrisen abgesehen (2009-2010, 2020-2022), fällt die Produktionslücke nach eigenen Berechnungen größer aus, als die nach Berechnungen/Schätzungen des BMWK. Drittens, ist die Produktionslücke nach eigenen Berechnungen immer negativ, weil im Beobachtungszeitraum Vollbeschäftigung nicht erreicht wird; die Produktionslücke nach Berechnungen/Schätzungen des BMWK fällt demgegenüber in einigen Jahren selbst bei hoher Erwerbslosenquote positiv aus (1999-2002; 2006-2008) oder aber im negativen Bereich geringer (2003-2005, 2013-2015).

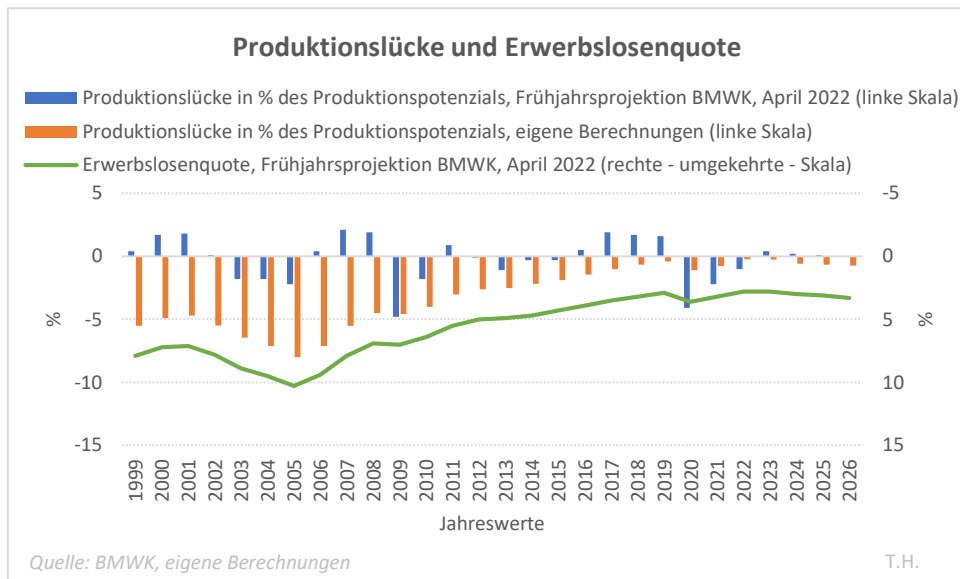
Grafik 2



Grafik 3 zeigt die unbereinigte Produktionslücke in Prozent des Produktionspotenzials nach eigenen Berechnungen. Sie entspricht annähernd der Erwerbslosenquote (umgekehrte Skala) minus der Vollbeschäftigungsquote. Das auf Basis der EU-Methode geschätzte Produktionspotenzial des BMWK weist demgegenüber nur einen sehr schwachen Bezug zu den Erwerbslosen/zur Erwerbslosenquote auf. Auffallend ist auch in dieser Betrachtung, dass sich die Produktionslücke des BMWK in den schweren Wirtschaftskrisen 2009-2010 und 2020-2022 der Produktionslücke nach eigenen Berechnungen annähert und damit auch der Erwerbslosenquote (umgekehrte Skala), sonst aber selbst bei hoher bzw.

sehr hoher Erwerbslosenquote positiv ist oder nur sehr niedrige negative Werte ausweist. Das ist seitens der NutzerInnen der EU-Methode erklärungsbedürftig.

Grafik 3:

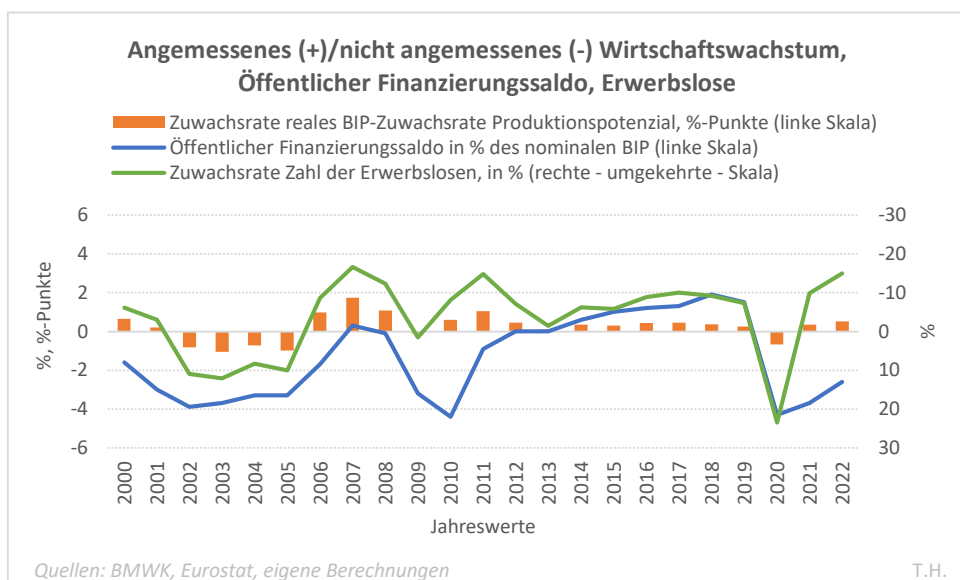


### Relevanz der Berechnungsmethode des Produktionspotenzials für die Wirtschaftspolitik, die Konjunktur, den Arbeitsmarkt, die Entwicklung des öffentlichen Finanzierungssaldos und die Preisstabilität

Eine Unterschätzung (Überschätzung) des Produktionspotenzials führt dazu, dass das Produktionspotenzial nicht voll beschäftigt (überlastet) wird, sofern sich die Wirtschaftspolitik daran orientiert. Eine unterbeschäftigte Volkswirtschaft bedeutet weniger Einnahmen aus Steuern und Abgaben und höhere Sozialausgaben. Das hat Auswirkungen auf das öffentliche Defizit und die öffentlichen Schulden.

Die Schuldenregel lässt sich unter den Bedingungen eines „normal“ ausgelasteten Produktionspotenzials und eines darauf zielenden, angemessenen, potenzialorientierten Wirtschaftswachstums einhalten (siehe Grafik 4).

Grafik 4:

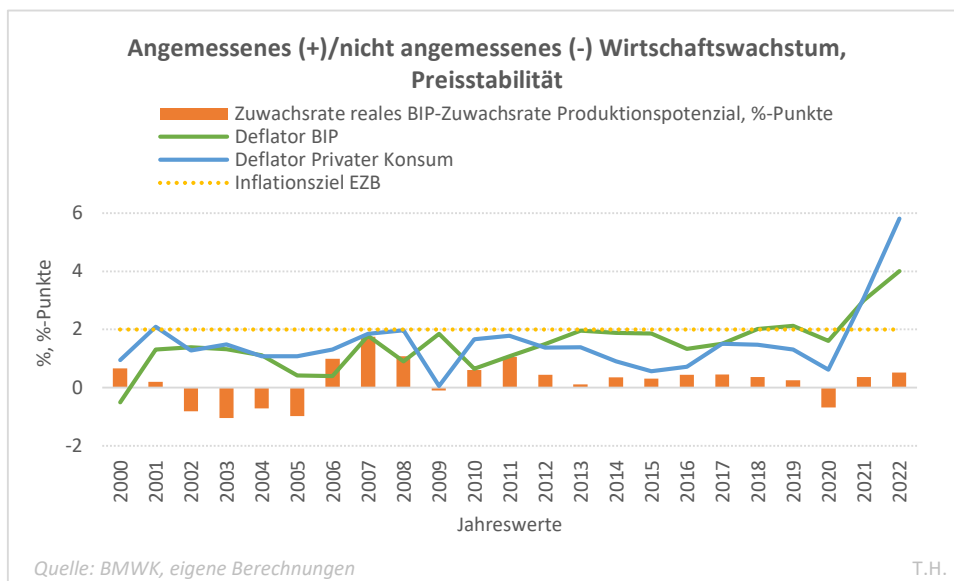


Grafik 4 zeigt, dass bei einem angemessenen Wirtschaftswachstum nicht nur die Zahl der Erwerbslosen sinkt (umgekehrte Skala), sondern auch der negative öffentliche Finanzierungssaldo, der sich unter dieser Voraussetzung sogar von einem Defizit in einen Überschuss verändern kann (siehe die Jahre 2011-2019 oder auch die Jahre 2006-2008).

Liegt die Arbeitslosenquote/Erwerbslosenquote über der Vollbeschäftigungsquote ist das Wirtschaftswachstum angemessen, wenn die Zuwachsrate des realen BIP über der des Produktionspotenzials liegt. Das ist die Voraussetzung, um die Arbeitslosigkeit zu senken. Ist der öffentliche Finanzierungssaldo negativ, ist das Wirtschaftswachstum ebenfalls dann angemessen, wenn die Zuwachsrate des realen BIP über der des Produktionspotenzials liegt. Unter dieser Voraussetzung kann das öffentliche Defizit gesenkt bzw. ein Haushaltsüberschuss erzielt werden.

Gewährleistet ist auch die Preisstabilität. Selbst, als die Zuwachsrate des realen BIP die des Produktionspotenzials deutlich übertraf (siehe Grafik 5, die Jahre 2006-2008 und die Jahre 2010-2011) oder über einen längeren Zeitraum (siehe Grafik 5, die Jahre 2010-2019), bewegten sich die Indikatoren für die Preisentwicklung, die auch das BMWK heranzieht<sup>12</sup>, nicht über das Inflationsziel der EZB hinaus, von der einmaligen, marginalen Abweichung im Jahr 2019 (Deflator BIP: 2,1%) abgesehen. Besonders schwach hat sich der Deflator für die Privaten Konsumausgaben entwickelt. Für die Jahre 2021 und 2022 wissen wir, dass die Entwicklung Sonderfaktoren geschuldet ist.<sup>13</sup>

Grafik 5:

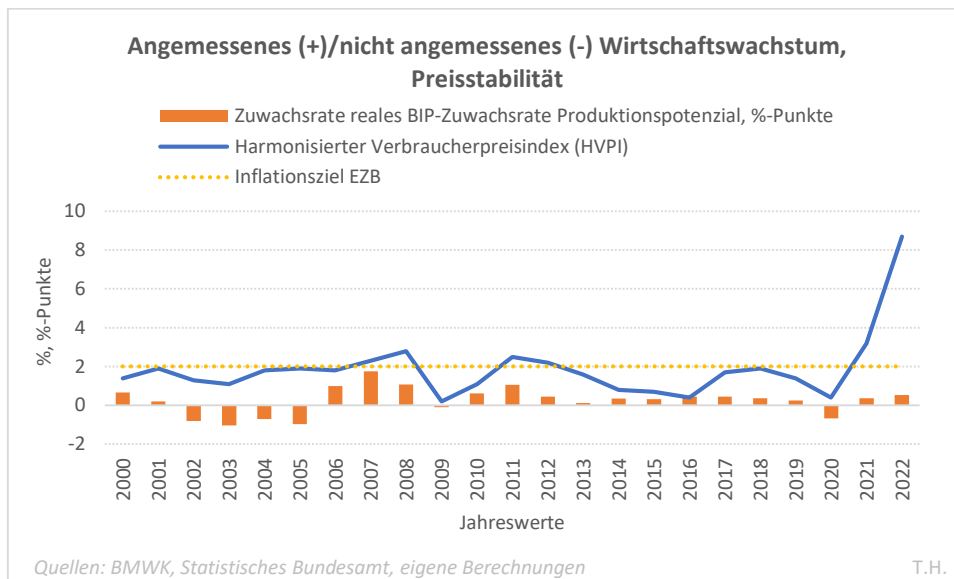


Ein Blick auf die Entwicklung des Harmonisierten Verbraucherpreisindex zeigt ebenfalls, dass die Inflationsrate im Beobachtungszeitraum lediglich in drei Jahren leicht vom Inflationsziel nach oben abgewichen ist (Grafik 6). Für die Jahre 2021 und 2022 gilt wiederum der Einfluss von Sonderfaktoren.<sup>14</sup>

Auch in den Jahren, in denen die Zuwachsrate des realen BIP über einen längeren Zeitraum durchgehend die des Produktionspotenzials überstieg (2010-2019), blieb die Inflationsrate mit Ausnahme des Jahres 2011 unter dem Inflationsziel der EZB, zum Teil sehr deutlich.



Grafik 6:



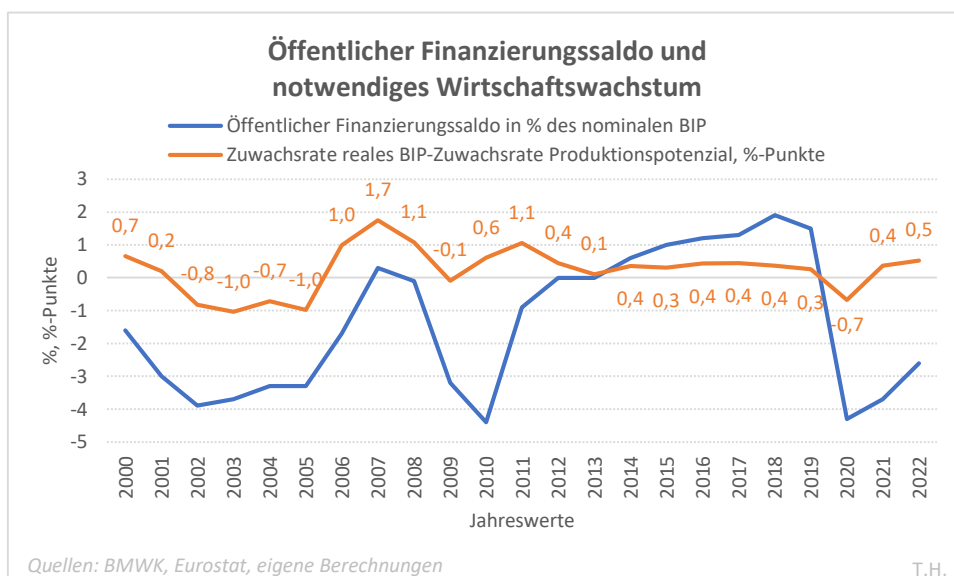
### Konjunkturkomponente

Eine Konjunkturkomponente sollte sich vor dem aufgezeigten Hintergrund an der Differenz aus der Zuwachsrate des realen BIP und der Zuwachsrate des Produktionspotenzials orientieren.

Für einen ausgeglichenen öffentlichen Haushalt oder den Abbau eines negativen öffentlichen Finanzierungssaldos ist ein bestimmtes Wirtschaftswachstum *notwendig*. Der oben vorgenommene empirische Rückblick zeigt, dass die Zuwachsrate des realen BIP über der des Produktionspotenzials liegen muss, um das Ziel eines ausgeglichenen öffentlichen Haushalts oder des Abbaus eines negativen öffentlichen Finanzierungssaldos zu erreichen.

Im Beobachtungszeitraum ist das Ziel in den Jahren erreicht worden, als die Differenz der Zuwachsraten 0,3 Prozentpunkte betrug oder darüber lag. Zuletzt ist das öffentliche Defizit im Jahr 2022 gesunken, als die Zuwachsrate des realen BIP um 0,5 Prozentpunkte über der Zuwachsrate des Produktionspotenzials lag (siehe Grafik 7).

Grafik 7:



Eine Konjunkturkomponente sollte also berücksichtigen, dass die Zuwachsrate des realen BIP um mindestens 0,3 Prozentpunkte über der Zuwachsrate des Produktionspotenzials liegt, solange sich der Haushalt im Defizit befindet. Ausgehend von einem ausgeglichenen Haushalt, könnte sich für die Konjunkturkomponente auch schon eine Differenz von 0,1 Prozentpunkten als ausreichend erweisen (siehe Grafik 7, das Jahr 2013).

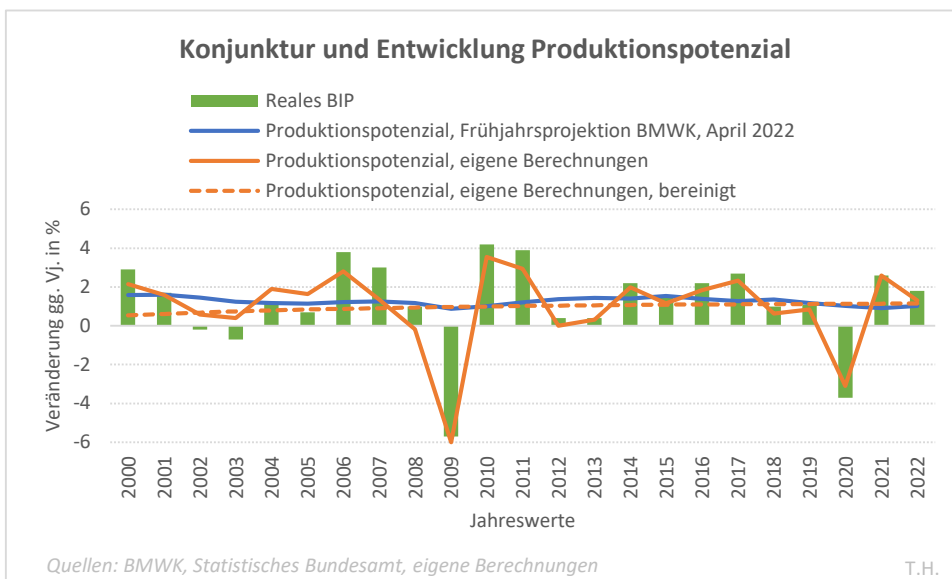
Ist das reale Wirtschaftswachstum für das Jahr 2023 bspw. mit 2,5% prognostiziert (Frühjahrsprojektion des BMWK vom 27. April 2022) und die Zuwachsrate des Produktionspotenzials ebenfalls mit 2,5% (eigene Berechnungen), müsste die Zuwachsrate des realen BIP um 0,3 Prozentpunkte höher ausfallen, die Zuwachsrate des realen BIP also 2,8% betragen. Bei einem Wirtschaftswachstum von 2,5% im Jahr 2023 würde das reale BIP um 81 Mrd. Euro zulegen. Notwendig aber wäre ein Wirtschaftswachstum von 2,8%. Dies entspricht einem Zuwachs von 91,6 Mrd. Euro. Die Differenz zwischen tatsächlichem/prognostiziertem Wirtschaftswachstum und dem notwendigen Wirtschaftswachstum beträgt -10,6 Mrd. Euro. Multipliziert man dieses Ergebnis mit der Budgetsemielastizität des BMWK von 0,203 (die das BMWK in seinen Berechnungen mit der Produktionslücke multipliziert), ergibt sich eine negative Konjunkturkomponente von -2,1 Mrd. Euro. Das heißt übersetzt, dass die Differenz von -10,6 Mrd. Euro den öffentlichen Finanzierungssaldo konjunkturbedingt um 2,1 Mrd. geringer ausfallen lässt. Da der Finanzierungssaldo im kommenden Jahr voraussichtlich ein Defizit ausweisen wird, kann dies nicht gewünscht sein.

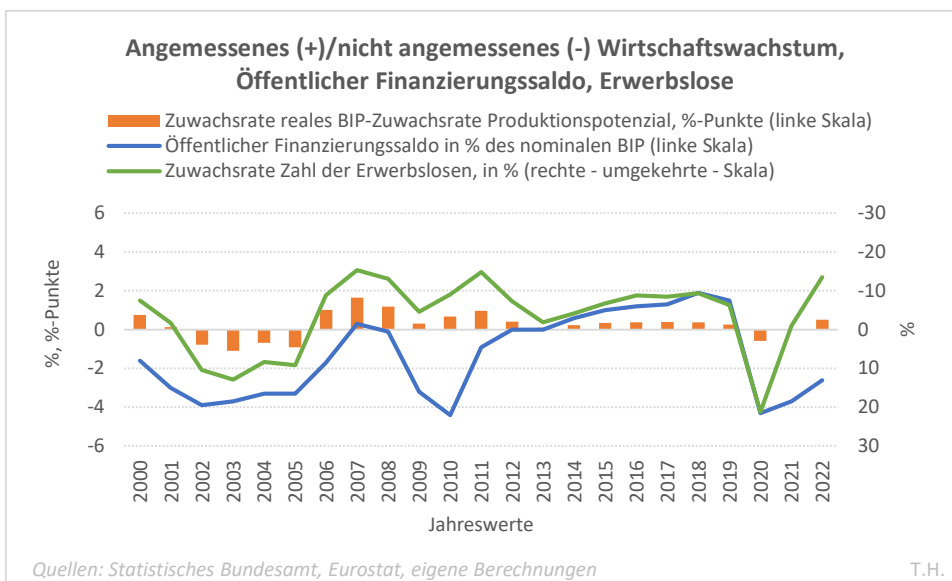
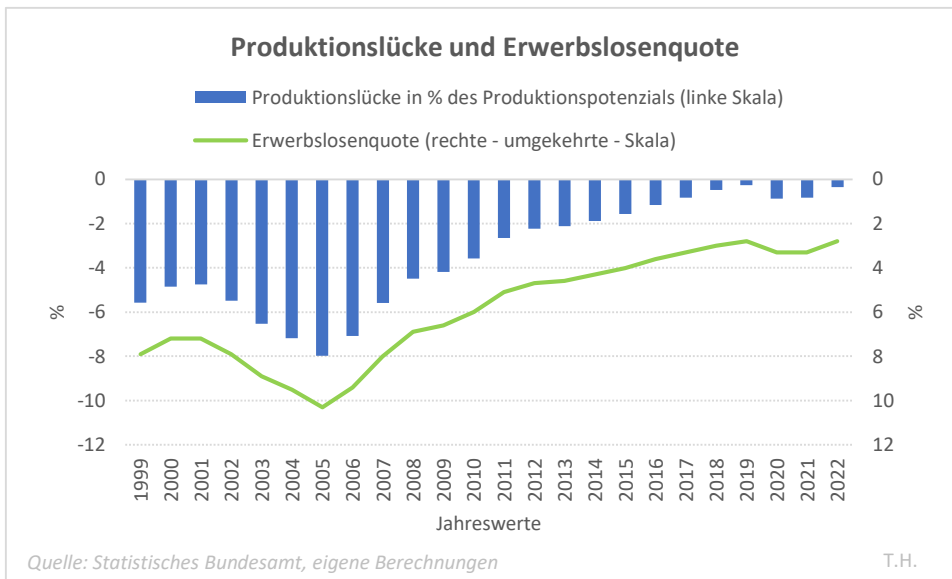
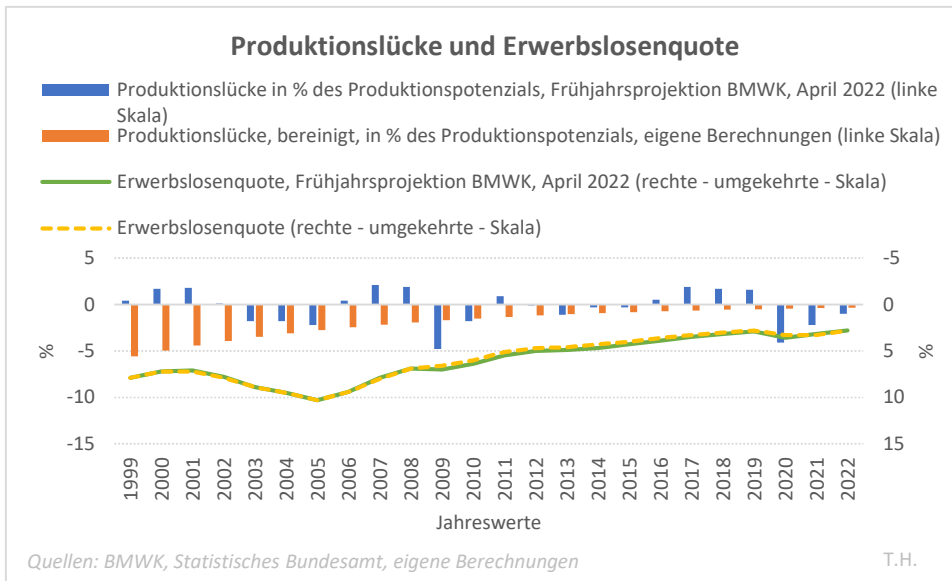
Um die Mehr-/Minderausgaben zu schätzen, die notwendig sind, um jene konjunkturell bedingte Verschlechterung des Finanzierungssaldos zu verhindern, wird als Gewicht eine Staatsquote von 45 Prozent herangezogen. Multipliziert man die Konjunkturkomponente von -2,1 Mrd. Euro mit 0,45, ist das Ergebnis 1 Mrd. Euro. Um diese 1 Mrd. Euro sollten die bisher geplanten Ausgaben erhöht werden, um ein konjunkturell bedingtes, steigendes Haushaltsdefizit zu vermeiden, respektive das Ziel, ein sinkendes Haushaltsdefizit zu erzielen, zu erreichen (die Berechnungen sind in der Excel-Datei „Thorsten Hild – Angaben – Berechnungen – Produktionspotenzial – 12-05-2023“ enthalten).

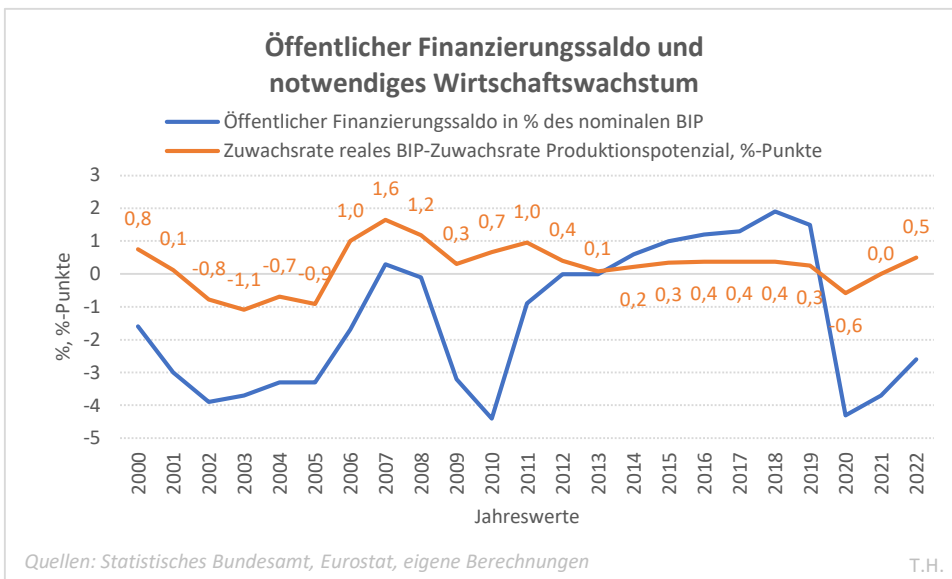
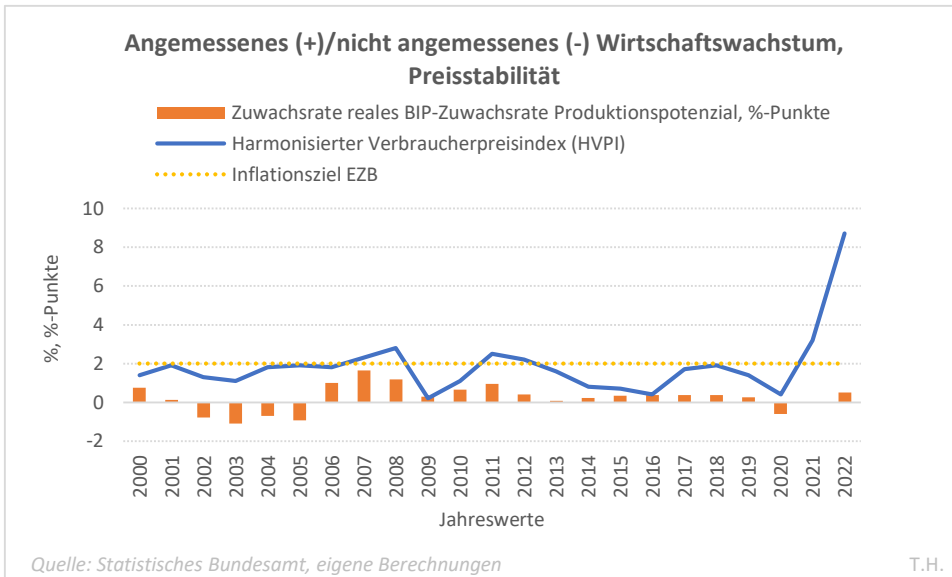
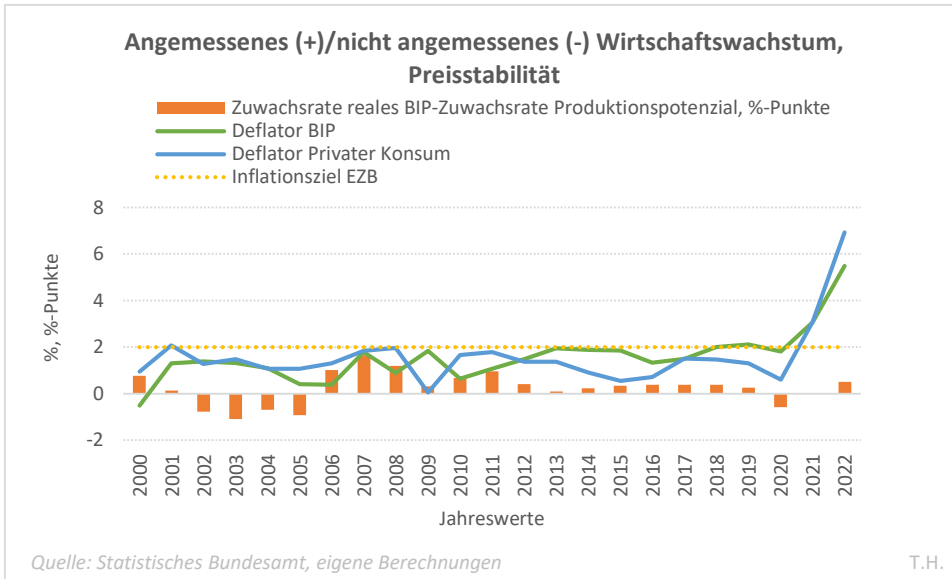
---

#### Anhang:

Grafiken basierend auf aktuellstem Stand der tatsächlichen Werte, abgerufen: 5. Mai 2023<sup>15</sup>







<sup>1</sup> Claus Köhler, Wirtschaftspolitisches Umdenken in der globalen Welt, Volkswirtschaftliche Schriften, Band 570, Berlin, 2019, S. 25; ders., Wirtschaftspolitische Ziele in der globalen Welt, Volkswirtschaftliche Schriften, Band 564, Berlin, 2013, S. 28ff.; ders., Orientierungshilfen für die Wirtschaftspolitik, Veröffentlichungen des Instituts für Empirische Wirtschaftsforschung, Band 41, Berlin, 2004, S. 19f.

<sup>2</sup> Vgl. zum Unterschied Arbeitslosigkeit und Erwerbslosigkeit: Bundesagentur für Arbeit, Arbeitslosigkeit und Erwerbslosigkeit, <https://statistik.arbeitsagentur.de/DE/Navigation/Grundlagen/Definitionen/Arbeitslosigkeit-Unterbeschaeftigung/Arbeitslosigkeit-Erwerbslosigkeit-Nav.html>

<sup>3</sup> Vgl. Bundesagentur für Arbeit, Unterbeschäftigung, <https://statistik.arbeitsagentur.de/DE/Navigation/Grundlagen/Definitionen/Arbeitslosigkeit-Unterbeschaeftigung/Unterbeschaeftigung-Nav.html>

<sup>4</sup> Vgl. Eurostat, Labour market slack: what's it all about?, [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Labour\\_market\\_slack\\_-\\_employment\\_supply\\_and\\_demand\\_mismatch#Labour\\_market\\_slack:\\_what.27s\\_it\\_all\\_about.3F](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Labour_market_slack_-_employment_supply_and_demand_mismatch#Labour_market_slack:_what.27s_it_all_about.3F)

<sup>5</sup> Vgl. Statistisches Bundesamt, Stille Reserve am Arbeitsmarkt im Jahr 2021 bei gut 3,1 Millionen Menschen, Pressemitteilung Nr. 035 vom 27. Januar 2023, [https://www.destatis.de/DE/Presse/Pressemitteilungen/2023/01/PD23\\_035\\_13.html](https://www.destatis.de/DE/Presse/Pressemitteilungen/2023/01/PD23_035_13.html); IAB, IAB-Kurzbericht, 5/2023, IAB-Prognose 2023, Rekordarbeitskräftebedarf in schwierigen Zeiten, Tabelle 1, Entwicklung von Wirtschaft und Arbeitsmarkt 2016 bis 2023, <https://doku.iab.de/kurzber/2023/kb2023-05.pdf>; zu den unterschiedlichen Methoden: Martina Rengers, Johann Fuchs, Stille Reserve in Deutschland: Gemeinsamkeiten und Unterschiede zweier Konzepte, 28. Juli 2022, <https://link.springer.com/article/10.1007/s11943-022-00309-3>

<sup>6</sup> Vgl. Statistisches Bundesamt, Unfreiwillige Teilzeitbeschäftigte, <https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Qualitaet-Arbeit/Dimension-3/unfreiwillig-teilzeitbeschaeftige.html#:~:text=Zahl%20der%20Betroffenen%20geht%20zur%C3%BCck,erfolgte%20zwischen%202002%20und%202006.>

<sup>7</sup> Vgl. Statistisches Bundesamt, Teilhabe von Frauen am Erwerbsleben, <https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Qualitaet-Arbeit/Dimension-1/teilhabe-frauen-erwerbsleben.html#:~:text=Der%20Abstand%20bei%20der%20Erwerbsbeteiligung,sich%20der%20Anstieg%20allerdings%20verlangsamt.>

<sup>8</sup> Bundesministerium für Wirtschaft und Klimaschutz, Bundesministerium der Finanzen, Gesamtwirtschaftliches Produktionspotenzial und Konjunkturkomponenten, Datengrundlagen und Ergebnisse der Bundesregierung, Stand: Frühjahrsprojektion der Bundesregierung vom 27. April 2022, [https://www.bmwk.de/Redaktion/DE/Downloads/G/gesamtwirtschaftliches-produktionspotenzial-fruehjahrsprojektion-2022.pdf?\\_\\_blob=publicationFile&v=1](https://www.bmwk.de/Redaktion/DE/Downloads/G/gesamtwirtschaftliches-produktionspotenzial-fruehjahrsprojektion-2022.pdf?__blob=publicationFile&v=1)

<sup>9</sup> Claus Köhler, 2013, a.a.O., S. 28

<sup>10</sup> Claus Köhler, 2019, a.a.O., S. 22

<sup>11</sup> Pascal Michailat, Emmanuel Saez,  $u^* = \sqrt{uv}$ , June 2022, <https://arxiv.org/pdf/2206.13012.pdf>; siehe auch: <https://arxiv.org/abs/2206.13012>

<sup>12</sup> Vgl. Bundesministerium für Wirtschaft und Klimaschutz, Bundesministerium der Finanzen, 27. April 2022, a.a.O., Tabelle 8

<sup>13</sup> Vgl. zur Inflation im Jahr 2021: Statistisches Bundesamt, Pressemitteilung Nr. 025 vom 19. Januar 2022, [https://www.destatis.de/DE/Presse/Pressemitteilungen/2023/01/PD23\\_022\\_611.html#:~:text=Pressemitteilung%20Nr.,Januar%202023&text=WIESBADEN%20%E2%80%93%20Die%20Verbraucherpreise%20in%20Deutschland,%20C9%20%25%20gegen%C3%BCber%202021%20erh%C3%B6ht.](https://www.destatis.de/DE/Presse/Pressemitteilungen/2022/01/PD22_025_611.html#:~:text=Inflationsrate%202021%3A%20%2B3%2C1,gegen%C3%BCber%20dem%20Vorjahr%20%2D%20Statistisches%20Bundesamt; vgl. zur Inflation im Jahr 2022: Statistisches Bundesamt, Pressemitteilung Nr. 022 vom 17. Januar 2023, <a href=)

<sup>14</sup> Vgl. Endnote 13.

<sup>15</sup> Die Berechnungen sind in der Excel-Datei „Thorsten Hild – Angaben – Berechnungen – Produktionspotenzial – tatsächliche Werte bis 2022 – 12-05-2023“ enthalten.

# Modellierung der NAWRU im D\*-Modell von Kiel Economics

Carsten-Patrick Meier  
Kiel Economics  
Juni 2023

## I. Einleitung

Die im Grundgesetz verankerte Schuldenbremse fordert für die Berechnung der zulässigen Neuverschuldung die Bereinigung des Bundeshaushalts um Positionen, die sich im Konjunkturverlauf automatisch erhöhen bzw. verringern. Der konjunkturbedingte Anteil des Haushaltssaldos ergibt sich nach den derzeit geltenden Berechnungsvorschriften als Produkt aus einer festen Budgetsemielastizität und der laufenden gesamtwirtschaftlichen „Produktionslücke“, d. h. der relativen Abweichung des Bruttoinlandsprodukts vom Produktionspotential. Letzteres wird mittels eines Produktionsfunktionsansatzes geschätzt, der sich eng am EU-Verfahren für die Haushaltsüberwachung im Rahmen des Europäischen Stabilitäts- und Wachstumspaktes (Havik et al. 2014) anlehnt. Die Bundesregierung hat sich in ihrem Koalitionsvertrag auf eine Kontrolle dieser Vorschriften verständigt. Der vorliegende Beitrag beschreibt einen Vorschlag, der allein durch eine methodisch verbesserte Umsetzung des EU-Ansatzes zu einer ökonomisch plausibleren und transparenteren Schätzung der Produktionslücke kommt, die außerdem einen geringeren Revisionsbedarf aufweist.

Der Vorschlag orientiert sich auf konzeptioneller Ebene sehr eng am Produktionsfunktionsansatz der EU-Kommission, inklusive dessen maßgeblichen Bausteins, des Ansatzes zur Schätzung der nicht-konjunkturellen Erwerbslosenquote (NAWRU). Anders als es für die Daten der Eurozone oder der EU der Fall zu sein scheint, erlauben die ab den 1950er Jahren verfügbaren langen Zeitreihen für Deutschland bzw. das Frühere Bundesgebiet zusammen mit der langjährigen Stabilitätsorientierung der bundesdeutschen Geld- und Finanzpolitik, eine empirische Umsetzung des Ansatzes, die deutlich näher am theoretischen Grundmodell bleibt als der Standardschätzansatz. Dadurch ist es – anders als mit letzterem – möglich, jene strukturellen Faktoren zu quantifizieren, die die nicht-konjunkturelle Erwerbslosenquote treiben, insbesondere den Anspruchslohn der Beschäftigten, und auf diese Weise zu einer nicht nur statistisch, sondern ökonomisch begründeten Konjunkturbereinigung der Erwerbslosenzahlen zu kommen.

Es lässt sich zeigen, dass die Teilkomponenten der geschätzten NAWRU eine ökonomisch plausible Entwicklung im Zeitablauf aufzeigen. Insbesondere korreliert der geschätzte Lohnaufschlagsfaktor eng mit Daten zur Lohnersatzleistungsquote und lässt sich insofern als Maß für den Anspruchslohn der Beschäftigten interpretieren. Weiterhin lässt sich zeigen, dass Abweichungen der Erwerbslosenquote von der geschätzten NAWRU, wie von der makroökonomischen Theorie gefordert, stationär sind und einen statistisch signifikanten und stabilen Vorlauf von einem Jahr vor dem gesamt- bzw. binnenwirtschaftlichen Preisauftrieb aufweisen und dass umfragebasierte Daten zum Kapazitätsaus-

lastungsgrad unter diesen Umständen keinen signifikanten weiteren Beitrag zur Erklärung des Preisanstiegs zu leisten vermögen. Für Abweichungen der Erwerbslosenquote von der mit dem Verfahren der EU-Kommission geschätzten NAWRU lässt sich für Deutschland weder deren Stationarität noch überhaupt ein signifikanter Beitrag zur Erklärung oder zur Prognose des gesamtwirtschaftlichen Preisauftriebs nachweisen; streng genommen erfüllen die Schätzwerte zur NAWRU die Anforderungen, die aus theoretischer Sicht an eine NAWRU-Schätzung zu stellen sind, nicht. Das hier vorgestellte Schätzverfahren für die NAWRU kann insofern als klar überlegen gelten.

## II. Schätzung der NAWRU

Im Folgenden wird der modifizierte Schätzansatz im Einzelnen beschrieben. Er betrifft alle Komponenten des EU-Produktionsfunktionsansatzes, d. h. die Schätzung der nicht-konjunkturellen Erwerbslosenquote NAWRU, die Schätzung der übrigen Aspekte des konjunkturbereinigten Arbeitsangebots und die Schätzung der nicht-konjunkturellen TFP. Zentral ist allerdings die Modifikation des Schätzansatzes für die NAWRU. Deren Ergebnisse sind Input der Bereinigungsverfahren für die übrigen Komponenten des Arbeitsvolumens und der TFP.

### *Theoretische Konzeption*

Ausgangspunkt der Schätzung der NAWRU ist, Havik et al. (2014) folgend, ein Modell unvollständiger Konkurrenz auf dem Gütermarkt und dem Arbeitsmarkt, angelehnt an Blanchard und Katz (1999), Nickell und Layard (1999), Bårdsen et al. (2005) sowie Bårdsen und Nymoen (2003, 2009). Die Unternehmen verfügen über eine begrenzte Preissetzungsmacht, die von der Intensität des Wettbewerbs auf dem Gütermarkt abhängt; je höher ihr Preisaufschlag auf die Grenzkosten ausfällt, umso geringer ist ihre Güterproduktion und umso weniger Arbeitskräfte fragen sie nach. Die Arbeitnehmer/Gewerkschaften verfügen über einen Lohnsetzungsspielraum, der von Arbeitsschutzgesetzgebung und der Ausgestaltung der finanziellen Absicherung bei Arbeitslosigkeit im engeren Sinn (Arbeitslosenunterstützung) und im weiteren (Sozialstaat, Rentenversicherung, Zugriff auf öffentliche Güter etc.) abhängt; je besser die Arbeitnehmer gegen Entlassungen und im Fall von Arbeitslosigkeit abgesichert sind, umso höher fällt der Lohnaufschlag auf den markträumenden Lohn aus und umso weniger Arbeit fragen die Unternehmen nach. Das Modell besteht aus den folgenden Bestandteilen:

$$Y = A * (L E)^\alpha * K^{1-\alpha} \quad \text{Produktionsfunktion} \quad (1)$$

$$G = P * Y - W * L - r * K \quad \text{Gewinn} \quad (2)$$

$$\frac{\partial Y}{\partial L} = \alpha \frac{Y}{L} + Z^p = \frac{W}{P} + Z^w \quad \text{Optimum bei unvollständigen Produkt- und Arbeitsmärkten} \quad (3)$$

$$p = w - (y - l) - \ln \alpha + z^p \quad \text{Preissetzung} \quad (4)$$

$$\begin{aligned} w &= \ln \alpha + (y - l) + p + z^w \\ &= \ln \alpha + (y - l) + p^c - \beta u + z^{\tilde{w}} \end{aligned} \quad \text{Lohnsetzung} \quad (5)$$

$$\bar{u} = \frac{1}{\beta} [z^{\tilde{w}} + z^p + (p^c - p)] \quad \text{NAWRU} \quad (4)$$

$$\bar{U} = e^{[z^{\tilde{w}} + z^p + (p^c - p)]/\beta} \quad (7)$$

Die Arbeitslosenquote, die die Lohnsetzung der Arbeitnehmervertreter und die Preissetzung der Unternehmen miteinander ins Gleichgewicht bringt, die NAWRU  $\bar{u}$ , ist gemäß (6) bzw. (7) umso höher, je höher der Lohnaufschlag der Arbeitnehmer bzw. Gewerkschaften auf den markträumenden Reallohn, je höher der Preisaufschlag der Unternehmen auf die Grenzkosten und je weiter das Niveau der Konsumentenpreise das der Produzentenpreisen übersteigt.

Bis auf den letzten Term entspricht die theoretische Spezifikation der NAWRU exakt der von Havik et al. (2014, Annex 1, Gleichung 4). Der letzte Term ergibt sich insbesondere dadurch, dass in einer offenen Volkswirtschaft wie der deutschen für Unternehmen und Arbeitnehmer unterschiedliche Preisgrößen entscheidungsrelevant sind, für die Unternehmen die Preise ihrer Bruttowertschöpfung, für die Arbeitnehmer die Konsumentenpreise. Außenwirtschaftliche Einflüsse durch Wechselkursschwankungen und Veränderungen der Weltmarktpreise importierter Erzeugnisse können einen Keil zwischen die relevanten Preisgrößen treiben und dadurch Bedeutung für die NAWRU bekommen: Ziehen beispielsweise infolge steigender Weltmarktpreise für importierte Energieträger oder einer Abwertung des Euro die Konsumentenpreise stärker an als die Erzeugerpreise, so werden die Arbeitnehmer bei gegebenen Knappheitsverhältnissen auf dem Arbeitsmarkt höhere Löhne fordern, die Unternehmen werden aber nicht bereit sein, diese zu zahlen und infolgedessen steigt die – gleichgewichtige – Erwerbslosigkeit.

Dieser Kanal dürfte allerdings nicht der einzige sein, über den die Außenwirtschaft Einfluss auf die NAWRU nimmt. In einer offenen Volkswirtschaft stehen die inländischen Unternehmen sowohl im Inland als auch im Ausland mit ausländischen Anbietern im Wettbewerb, die häufig ähnliche Varianten derselben Produkte anbieten. Es ist plausibel, dass die Unternehmen ihren Preisaufschlag den Wettbewerbsbedingungen auf dem Weltmarkt anpassen.<sup>1</sup> Sie erhöhen ihn, wenn ihre (Grenz-) Kosten nach Umrechnung in ausländische Währung unter den (Grenz-) Kosten der Wettbewerber liegen, sie verringern ihn, wenn sie in Auslandswährung kostenmäßig ungünstiger dastehen. Der Preisaufschlag  $z^p$  ist also umso höher, je niedriger der reale Außenwert der inländischen Währung  $q$  ist, d. h. es gilt  $z^p = z^{\tilde{p}} - \lambda q$ , wobei  $z^{\tilde{p}}$  einen konstanten Basisaufschlag und  $\lambda (> 0)$  einen Proportionalitätsfaktor angeben. Die Preissetzungsgleichung (4) lässt sich jetzt modifizieren zu

$$p = w - (y - l) - \ln \alpha + z^{\tilde{p}} - \lambda q \quad (8)$$

und für die NAWRU  $\bar{u}$  ergibt sich damit

$$\bar{u} = \frac{1}{\beta} [z^{\tilde{w}} + z^{\tilde{p}} - \lambda q + (p^c - p)] \quad (9)$$

$$\bar{u} = e^{[z^{\tilde{w}} + z^{\tilde{p}} - \lambda q + (p^c - p)]/\beta} \quad (10)$$

Die beiden außenwirtschaftlichen Kanäle wirken dabei tendenziell in dieselbe Weise auf die NAWRU. Eine reale Abwertung der inländischen Währung führt beispielsweise sowohl zu einem höheren Preisaufschlag der Unternehmen als auch zu niedrigeren Importpreisen und damit zu einem Rückgang der Verbraucherpreise in Relation zu den Erzeugerpreisen; beides lässt die gleichgewichtige Erwerbslosigkeit steigen. Die beiden Kanäle sind gleichwohl nicht identisch, da die Konsumentenpreise

---

<sup>1</sup> Dornbusch (1987) und Krugman (1987) zeigen, dass die genaue Art der Preisanpassung von der angenommenen Form des Wettbewerbs und weiterer Annahmen abhängig ist. Hintergrund ihrer Analyse ist das Preissetzungsverhalten deutscher (und japanischer) Autohersteller in den Vereinigten Staaten während der Phase der Hochbewertung des US-Dollar in der ersten Hälfte der 1980er Jahre.



auch durch andere Faktoren beeinflusst werden können, z. B. die Preise von Energierohstoffen und durch die indirekten Steuern.

### *Empirische Spezifikation*

Gemäß (9) bzw. (10) gilt es für eine Schätzung der NAWRU zum einen die als konstant unterstellten, nicht direkt beobachtbaren („latenten“) Aufschlagsfaktoren  $z^{\tilde{p}}$  und  $z^{\tilde{w}}$  empirisch zu bestimmen, zum anderen den Parameter  $\lambda$ . Hinsichtlich der theoretisch ebenfalls unbestimmten Elastizität des Lohnsatzes hinsichtlich der Erwerbslosenquote,  $\beta$ , gibt es umfangreiche empirische Vorkenntnisse, die berücksichtigt werden können. Blanchflower und Oswald (1994, 2005) schätzen diesen Parameter für eine große Zahl von Lohnsetzungsgleichungen in der Art von (5) mit regionalen Daten für die Vereinigten Staaten und stellen fest, dass er wie theoretisch zu erwarten negativ in die Beziehung eingeht und typischerweise bei 0,1 liegt. Bellmann und Blien (1996) bestätigen diesen Schätzwert für Deutschland mit Daten des IAB-Betriebspanels. Vorläufige Schätzungen mit dem hier verwendeten gesamtwirtschaftlichen Datensatz legen ebenfalls einen (Absolut-) Wert für  $\beta$  in der Nähe von 0,1 nahe. Daher wird dieser Parameter im Folgenden auf diesen Wert kalibriert.

Zur Schätzung der übrigen Parameter werden langen Zeitreihen für Deutschland bzw. das Frühere Bundesgebiet verwendet. Da diese integriert vom Grade 1 sind, ihre Niveaus sind also nichtstationär, ihre Differenzen hingegen stationär, können die Langfristparameter verzerrungsfrei mit der Methode der Kleinsten Quadrate geschätzt werden, sofern die theoretisch motivierten Linearkombinationen zwischen den Größen stationär, die Variablen also kointegriert sind. In kleinen Stichproben – die Untersuchung hier basiert auf Jahresdaten ab 1950, umfasst also nicht mehr als reichlich 70 Beobachtungen – ist es zweckmäßig, die Spezifikationen durch die Berücksichtigung von nachlaufenden und vorlaufenden (erwarteten) Beobachtungen angemessen zu dynamisieren. Für die beispielhafte Darstellung hier wird unterstellt, dass dazu jeweils ein Nachlauf von zwei Jahren und die Erwartung für das Folgejahr ausreichen; bei der empirischen Umsetzung wird die Anzahl der Verzögerungen so bestimmt, dass die Residuen der Gleichungen seriell unkorreliert sind.

In Gleichgewichtskorrekturform lassen sich die dynamischen Versionen von (8) und (5) schreiben als

$$\Delta p_t = \mu [p_{t-1} + (w - y + l)_{t-1} + \ln \alpha - z^{\tilde{p}} + \lambda q_t] + \sum_{i=0}^1 v_i \Delta q_{t-i} \quad (11)$$

$$+ \theta_p E_t(\Delta p_{t+1}) + \sum_{i=0}^1 \delta_i \Delta(w - y + l)_{t-i} + (1 - \theta_p - \sum_{i=0}^1 \delta_i) \Delta p_{t-1} + \varepsilon_{1,t}^p$$

$$\Delta(w - p^c)_t = \gamma [(w - p^c)_{t-1} - (y - l)_{t-1} - \ln \alpha + \beta u_{t-1} + z^{\tilde{w}}] + \sum_{i=0}^1 \mu_i \Delta u_{t-i} \quad (12)$$

$$+ \theta_w E_t(\Delta p_{t+1}^c) + \sum_{i=0}^1 \varphi_i \Delta(y - l)_{t-i} + (1 - \theta_w - \sum_{i=0}^1 \varphi_i) \Delta(w - p^c)_{t-1} + \varepsilon_{1,t}^w$$

dabei stellen die Terme in eckigen Klammern die theoretisch motivierten Gleichgewichtsbeziehungen dar.  $\Delta$  steht für den Differenzenoperator ( $\Delta x_t = x_t - x_{t-1}$ ) und die Parameter  $\mu$ ,  $\gamma$ ,  $v$ ,  $\theta_p$ ,  $\theta_w$ ,  $\delta$  und  $\varphi$  beschreiben die Dynamik der Zusammenhänge.

Zu beachten ist, dass die erste Gleichung unter Vernachlässigung des realen Wechselkurses ( $\lambda = 0$ ) unter bestimmten Annahmen über die Dynamik der hybriden Neu-Keynesianischen Phillips-Kurve

von Gali und Gertler (1999) entspricht.<sup>2</sup> Für  $\lambda > 0$  stimmt sie – unter denselben Annahmen über die Dynamik – konzeptionell mit der Formulierung einer Neu-Keynesianischen Phillips-Kurve für offene Volkswirtschaften von Batini et al. (2005) überein. Beide Spezifikationen lassen sich mithilfe von Gleichung (11) durch die Restriktion  $\delta_0 = \delta_1 = 0$  darstellen. Empirisch wird diese Beschränkung der Dynamik für deutsche Daten allerdings abgelehnt.<sup>3</sup> Die zweite Gleichung ist das Pendant dazu für die Lohnsetzung. In beiden Gleichungen stellen die Restriktionen auf den jeweils letzten Koeffizienten sicher, dass im Rahmen der Schätzung die theoretischen Homogenitätsannahmen nicht nur im langfristigen Gleichgewicht, sondern permanent erfüllt sind, wie es für unverzerrte Schätzung der Aufschlagterme  $z^p$  und  $z^{\tilde{w}}$  und des Parameters  $\beta$  erforderlich ist (Currie 1981).

Eine vorläufige Schätzung der beiden Gleichungen über den Zeitraum von 1954 bis 2022 auf der Basis von Jahresdaten der Volkswirtschaftlichen Gesamtrechnung für Deutschland bzw. das frühere Bundesgebiet<sup>4</sup> sowie Prognosen der Gemeinschaftsdiagnose vom Herbst für den Preisanstieg im Folgejahr zu Modellierung der Inflationserwartungen  $E_t(\Delta p_{t+1})$  bzw.  $E_t(\Delta p_{t+1}^c)$  ergibt, dass die Hypothese der Nicht-Kointegration für die Preissetzungsgleichung (11) abgelehnt werden kann. Für die Lohnsetzungsgleichung (12) gilt dies für den Teilzeitraum von 1970 bis 2000 ebenfalls.<sup>5</sup> Werden allerdings Daten davor und/oder danach in die Analyse einbezogen, so gilt es nicht. Offenbar blieb der Lohnaufschlagfaktor  $z^{\tilde{w}}$  zwischen 1970 und 2000 ungefähr gleich, während er im Zeitraum davor und danach deutlichen Veränderungen unterworfen war. Tatsächlich wurden 1969 das Arbeitsförderungsgesetz (AFG), das der „Bundesanstalt für Arbeit“ umfangreiche Aufgaben im Bereich der aktiven Beschäftigungsförderung und -sicherung zuwies (Kühl 1982, S. 258) sowie eine deutliche Ausweitung der Leistungen des Arbeitslosengelds und der Arbeitslosenhilfe beschlossen (Boss 2009), die deutliche Auswirkungen auf den Anspruchslohn bzw. den Lohnaufschlagfaktor gehabt haben dürften. Für den Zeitraum nach dem Jahr 2000 hat eine Reihe von Untersuchungen gezeigt, dass die institutionellen Reformen des Arbeitsmarkts und des Systems der sozialen Sicherung im Rahmen der „Agenda 2010“ („Hartz-Reformen“) dämpfend auf den Anspruchslohn bzw. den Lohnaufschlagfaktor gewirkt haben dürften (Hochmuth et al. 2021, Burda und Seele 2017, Dustmann et al. 2014).<sup>6</sup>

Grundsätzlich ließe sich ein zeitvariabler Anspruchslohn in einem Ansatz mit konstanten Parametern wie (12) durch geeignete Daten zu den Reformen beschreiben und schätzen. Allerdings berühren institutionelle Reformen des Arbeitsmarkts und der sozialen Sicherung wie das Arbeitsförderungsgesetz von 1969 oder die „Hartz-Reformen“ in der Regel mehrere oder sogar eine Vielzahl von Regelungen, wobei noch dazu häufig für die jeweilige Einzelmaßnahme abgesehen vom Zeitpunkt der Reform keine oder nur sehr wenige Veränderungen im Zeitablauf vorliegen, was eine solche Modellie-

<sup>2</sup> Der Term  $p_{t-1} + (w - y + l)_{t-1}$  in (4a) entspricht mit den realen Lohnstückkosten jener Größe, die Gali und Gertler (1999) in ihrer empirischen Analyse zur Approximation der Entwicklung der Grenzkosten der gesamtwirtschaftlichen Produktion einsetzen.

<sup>3</sup> Bårdsen et al. (2005, S. 137) zeigen, dass es sich im Euroraum ähnlich verhält.

<sup>4</sup> Eine genaue Beschreibung des Datensatzes findet sich im Anhang.

<sup>5</sup> Dieses Resultat deckt sich mit der Untersuchung von Franz (2005).

<sup>6</sup> So wurden im Rahmen der „Agenda 2010“ u. a. die Voraussetzungen, unter denen ein Anspruch auf Arbeitslosengeld besteht restriktiver gefasst, die Kriterien für die Zumutbarkeit einer neuen Beschäftigung für Arbeitslose wurden erweitert und, die Ablehnung eines Arbeitsangebots, unzureichende Eigenbemühungen des Arbeitslosen und Meldesäumnisse konnten zu Sanktionen bis zum Erlöschen des Leistungsanspruchs führen. Außerdem wurden die Höhe und die maximale Bezugsdauer des Arbeitslosengeldes verringert. Ab Jahresbeginn 2005 ersetzte außerdem das Arbeitslosengeld II die Arbeitslosenhilfe, die Bedürftige nach Ablauf ihres Anspruchs auf Arbeitslosengeld beanspruchen konnten und die Sozialhilfe für Erwerbsfähige – mit erheblichen Leistungseinbußen für höherqualifizierte Arbeitnehmer (Deutsche Bundesbank 2006, Boss 2009). Hinzu kam die Lockerung des Kündigungsschutzes in Betrieben mit bis zu zehn Beschäftigten und die Ausweitung der Möglichkeit von befristeten Arbeitsverträgen für ältere Arbeitnehmer (Boss 2009, Steffen 2022).

zung vor Herausforderungen stellt. Alternativ wird der latente Anspruchslohn als zeitvariable Größe in einem Zustandsraummodells dargestellt (Havik et al. 2014, Staiger et al. 1997, Franz 2005). Diesem Ansatz wird hier gefolgt; dazu wird die Lohnsetzungsgleichung modifiziert zu:

$$\Delta(w - p^c)_t = \gamma [(w - p^c)_{t-1} - (y - l)_{t-1} - \ln \alpha + \beta u_{t-1} + z_t^{\tilde{w}}] + \sum_{i=0}^1 \mu_i \Delta u_{t-i} \quad (13)$$

$$+ \theta_w E_t(\Delta p_{t+1}^c) + \sum_{i=0}^1 \varphi_i \Delta(y - l)_{t-i} + (1 - \theta_w - \sum_{i=0}^1 \varphi_i) \Delta(w - p^c)_{t-1} + \varepsilon_{1,t}^w$$

$$z_t^{\tilde{w}} = \phi_{\tilde{w}} z_{t-1}^{\tilde{w}} + \varepsilon_{2,t}^{\tilde{w}} \quad (14)$$

wobei  $z_t^{\tilde{w}}$  für den nunmehr zeitvariablen Aufschlagsfaktor steht, dessen Veränderungen im Zeitablauf durch den Prozess (14) beschrieben werden sollen. Die beiden Gleichungen können gemeinsam mithilfe des Kalman Filters mit der Maximum-Likelihood-Methode geschätzt werden. Dabei wird allein  $z_t^{\tilde{w}}$  als zeitvariabel, alle anderen Parameter werden als im Zeitablauf konstant unterstellt; für sie gelten weiter die statischen und dynamischen Homogenitätsrestriktionen, die sich aus dem theoretischen Modell ergeben.

### Schätzergebnisse

Die Schätzergebnisse für die beiden Gleichungen sind in Tab. 1 angegeben. Eine genaue Beschreibung der Konstruktion des Datensatzes findet sich im Anhang. Durch die gängigen Tests wurde sichergestellt, dass beide Gleichungen statistisch und ökonomisch wohlspezifiziert sind in dem Sinn,<sup>7</sup>

- dass die Modellresiduen keine dynamische Struktur aufweisen,
- dass sie frei von Strukturbrüchen sind,
- dass für die Langfristbeziehungen die Hypothese der Nicht-Kointegration zurückgewiesen werden kann,
- dass die geschätzten Langfristparameter, sofern sie nicht kalibriert wurden, die theoretisch zu erwartenden Vorzeichen aufweisen,
- dass sie einen wesentlichen Teil der Varianz der jeweiligen abhängigen Variablen erklären.

Beide Gleichungen sind darüber hinaus frei von Strukturbrüchen, sieht man von der beabsichtigten zeitvariablen Schätzung des Anspruchslohns ab. Vereinzelt mussten dafür vom Modell besonders schlecht getroffene Beobachtungspunkte, die eine signifikante Wirkung auf die Parameterschätzungen aufweisen – Leverage Points im Sinne von Belsley et al. 1980 – durch 0/1-Dummyvariablen  $d...$  modelliert werden.

---

<sup>7</sup> Gegenüber der ursprünglichen Spezifikation (11) bzw. (13) wurden folgende Änderungen oder Ergänzungen vorgenommen: Die maximale Anzahl von Verzögerungen (in Niveaus) wurde in beiden Gleichungen von zwei auf fünf Jahre erhöht. In der Preissetzungsgleichung wurde zur Verbesserung der Anpassung anstelle der ersten Differenzen des Indikators der preislichen Wettbewerbsfähigkeit  $\Delta q_t$  die ersten Differenzen des Importdeflators  $p^m$  in Relation zum Deflator des Gesamtabsatzes  $\tilde{p}$  verwendet. In der Lohnsetzungsgleichung wurde der geschätzte zeitvariable Lohnaufschlagsfaktor  $z_t^{\tilde{w}}$  auf das Niveau der Erwerbslosenquote reskaliert und als  $\bar{u}_t^w$  bezeichnet ( $\bar{u}_t^w = z_t^{\tilde{w}}/\beta$ ), was für den (anspruchs-) lohnbedingten Teil der NAWRU  $\bar{u}_t$  stehen soll.

---

Tab. 1: Schätzergebnisse für die Preissetzungs- und die Lohnsetzungsfunktion

---

Preissetzung gemäß (11)

$$\begin{aligned} \Delta p_t = & -0,09 \left[ p_{t-1} - (w - y + l)_{t-1} + \ln \alpha + 0,41 q_{t-4} \right] + 0,57 E_t(\Delta p_{t+1}) \\ & \quad (4,03) \qquad \qquad \qquad (4,50) \qquad \qquad \qquad (12,95) \\ & + 0,25 \Delta p_{t-1} + (1 - 0,57 - 0,25) \Delta(w - y + l)_t + - 0,06 \Delta \Delta(p^m - \tilde{p})_t \\ & \quad (4,55) \qquad \qquad \qquad (8,25) \qquad \qquad \qquad (5,47) \\ & + 0,01 dum_{1970} + 0,01 dum_{1971} - 0,01 dum_{1972} + 0,01 dum_{1921} \\ & \quad (3,03) \qquad \qquad \qquad (3,05) \qquad \qquad \qquad (3,65) \qquad \qquad \qquad (2,66) \end{aligned}$$

Stützzeitraum: 1954 - 2022 ( $T=69$ );  $\overline{R^2}$ (ohne Jahre mit  $dum$ ) = 0,87  
 $DW = 2,08$ ;  $AR(1) = 0,49$ ;  $AR(4) = 0,48$ ;  $AP(1981) = 0,67$

Lohnsetzung gemäß (13)

$$\begin{aligned} \Delta(w - p^c)_t = & -0,30 \left[ (w - p^c)_{t-1} - (y - l)_{t-1} - \ln \alpha + 0,1 u_{t-1} + z_t^{\tilde{w}} \right] + E_t(\Delta p_{t+1}^c) \\ & \quad (10,41) \\ & + 0,29 \Delta(w - p^c)_{t-1} + 0,43 \Delta(y - l)_{t-1} + (1 - 0,29 - 0,43) \Delta(y - l)_{t-4} \\ & \quad (6,25) \qquad \qquad \qquad (7,22) \qquad \qquad \qquad (4,42) \\ & - 0,04 \Delta(mwst_t - mwst_{t-2}) - 0,05 \Delta \Delta u_{t-2, < 1975} \\ & \quad (2,52) \qquad \qquad \qquad (4,79) \\ & - 0,03 dum_{1958} - 0,02 dum_{1967} - 0,02 dum_{2022} \\ & \quad (4,18) \qquad \qquad \qquad (3,35) \qquad \qquad \qquad (3,25) \end{aligned}$$

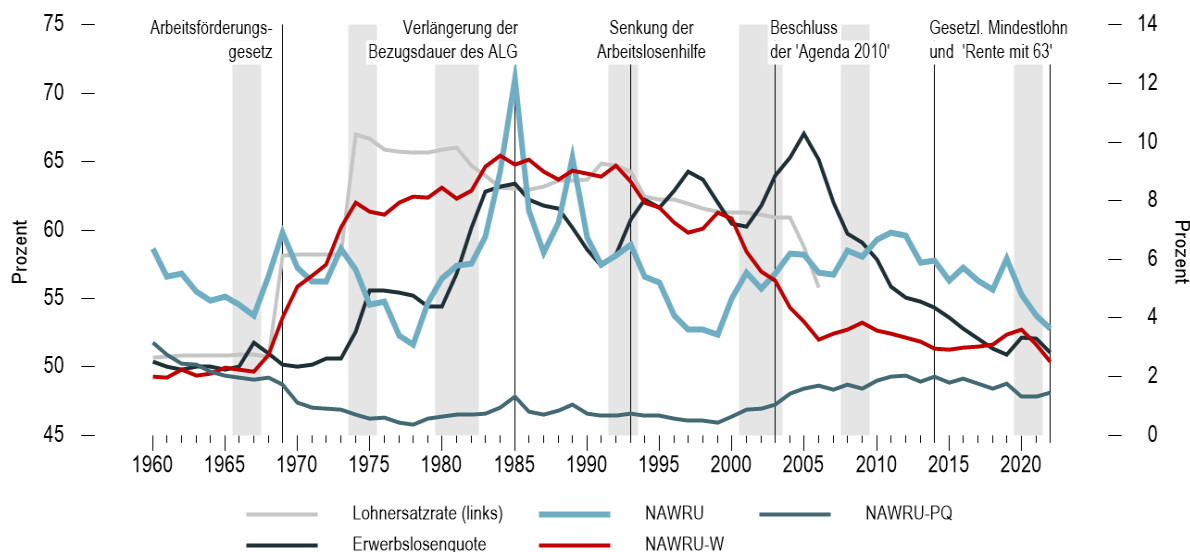
Stützzeitraum: 1956 - 2022 ( $T=67$ );  $\overline{R^2}$ (ohne Jahre mit  $dum$ ) = 0,94  
 $DW = 2,4$ ;  $AR(1) = 0,14$ ;  $AR(4) = 0,16$ ;  $AP(1971) = 0,23$

Anmerkung:  $T$  steht für die Anzahl der Beobachtungen,  $\overline{R^2}$  steht für das korrigierte Bestimmtheitsmaß; bei der Berechnung wurden Jahre, für die eine 0/1-Dummy-Variable gesetzt wurde, nicht berücksichtigt.  $DW$  ist die Durbin-Watson-Teststatistik,  $AR(1)$  und  $AR(4)$  sind die marginalen Signifikanzniveaus zweier Lagrange-Multiplikatortests auf Residuenautokorrelation erster bzw. vierter Ordnung,  $AP$ (Jahreszahl) das marginale Signifikanzniveau des Andrews-Ploberger-Tests auf einen Strukturbruch im Jahr mit der höchsten Teststatistik.  $dum$  steht für eine Dummy-Variable, die im angegebenen Jahr den Wert 1 annimmt und 0 sonst. Die Variable  $\tilde{u}_t^w$  ist der durch Division durch  $\beta$  auf das Niveau der Erwerbslosenquote reskalierte zeitvariable Lohnaufschlagsfaktor  $z_t^{\tilde{w}}$ .

Quelle: Eigene Berechnungen.

---

Abb.1: Erwerbslosenquote, geschätzte NAWRU und Lohnersatzrate 1960-2021



Anmerkung: Daten bis 1994 für das frühere Bundesgebiet. NAWRU gemäß Gleichung (7). Rezessionsphasen grau hinterlegt.  
Quelle: Bundesagentur für Arbeit; eigene Berechnungen.

Mit den Schätzergebnissen für  $\lambda$  und  $z_t^w$  liegen alle empirischen Informationen vor, um die NAWRU gemäß der Rechenanweisung (10) zu bestimmen. Das Ergebnis zeigt Abb. 1 zusammen mit der Erwerbslosenquote und den Teilkomponenten, d. h. dem (reskalierten) geschätzten Anspruchslohn und dem „außenwirtschaftlichen Faktor“, zu dem für die Abbildung der „Wettbewerbsfähigkeitseffekt“  $Q^{\lambda}$  und das Verhältnis aus Konsumenten- und Erzeugerpreisindex  $P^C/P$  multiplikativ zusammengefasst wurden. Außerdem enthält die Abbildung eine Zeitreihe zur Höhe der (durchschnittlichen) Lohnersatzleistungsrate zwischen 1960 und 2005. Sie geht auf Berechnungen von Boss (2009) zurück. Boss ermittelt das Verhältnis aus den Unterstützungsleistungen, auf die ein Facharbeiter in unterschiedlichen Familienzusammenhängen – ledig, verheiratet, verheiratet und zwei Kinder – unter bestimmten Bedingungen in Form von Arbeitslosengeld, Arbeitslosenhilfe oder Sozialhilfe Anspruch gehabt hätte, zu dessen Nettolohn; dargestellt ist der einfache Mittelwert der Quoten für alle drei Haushaltstypen und alle drei Arten von Unterstützungsleistungen.<sup>8</sup>

Der Schätzung zufolge ist der Anspruchslohn zwischen 1968 und 1975 stark gestiegen (Abb. 1). In diesem Zeitraum wurden die Lohnersatzrate stark, von etwas mehr als 50% auf 67% des Nettolohns, erhöht (Boss 2009, S. 226ff.). Der Anspruchslohn erreicht seinen Höhepunkt Mitte der 1980er Jahre – damals wurde die Bezugsdauer des Arbeitslosengelds von 12 auf bis zu 24 Monate für ältere Erwerbstätige ausgeweitet – und ist dann stetig gefallen. Dabei spielte auch eine Rolle, dass die Bezugsbedingungen für die Arbeitslosenhilfe im Zuge dauernder Reformen ständig restriktiver gefasst

<sup>8</sup> Ab Jahresbeginn 2005 ersetzte das Arbeitslosengeld II die Arbeitslosenhilfe, die Bedürftige nach Ablauf ihres Anspruchs auf Arbeitslosengeld beanspruchen konnten und die Sozialhilfe für Erwerbsfähige. Dabei kam es zu erheblichen Leistungseinbußen für höherqualifizierte Arbeitnehmer, während sich für weniger qualifizierte Arbeitnehmer wenig änderte. Die Wirkung auf die durchschnittliche Lohnersatzrate lässt sich aufgrund des Systembruchs nur grob schätzen. Die dargestellten Werte für die Jahre 2005 und 2006 fußen auf Berechnungen von Boss (2009) und von der Deutschen Bundesbank (2006).

wurden, was durch die Lohnersatzrate allein nur ungenügend abgebildet wird (Boss 2009, S. 231ff.).<sup>9</sup> Dieser Abwärtstrend erreicht seinen Tiefpunkt mit den Reformen der „Agenda 2010“, die unter anderem zu einer vollständigen Entkopplung der Unterstützungsleistung für erwerbsfähige Personen durch den Ersatz der Arbeitslosenhilfe durch das Arbeitslosengeld II führte, aber auch die Bezugsbedingungen abermals restriktiver fasste. Seither ist der Anspruchslohn – trotz der Einführung des gesetzlichen Mindestlohns und der „Rente mit 63“ für langjährig Erwerbstätige – weitgehend unverändert auf dem damals erreichten niedrigen Niveau geblieben. Der Vergleich des Verlaufs des geschätzten Anspruchslohns mit der Lohnersatzrate gemäß Boss (2009) lässt die Schätzung alles in allem – zumindest bis 2006, dem letzten Jahr für das Daten für letztere vorliegen – plausibel erscheinen.

Die gesamte NAWRU schwankt deutlich weniger als der Anspruchslohn. Hier wirkt sich aus, dass der „außenwirtschaftliche Faktor“, also das Produkt aus dem von der internationalen preislichen Wettbewerbsfähigkeit abhängigen Preisaufschlag der Unternehmen und dem Verhältnis der Konsumenten zu den Erzeugerpreisen, sich häufig gegenläufig zum Anspruchslohn entwickelt haben. So wurde Anfang der 2000er Jahre der Effekt des Rückgangs des Anspruchslohns auf die NAWRU überkompensiert durch die Erhöhung der Preisaufschläge der Unternehmen infolge der nach Einführung des Euro eingetretenen verbesserten internationalen preislichen Wettbewerbsfähigkeit und die trotz stagnierender Erzeugerpreise steigenden Verbraucherpreise. Seither ist die NAWRU lange Zeit weitgehend konstant bei knapp 6% geblieben, bevor sie sich ab dem Jahr 2020 mit dem gleichzeitigen Rückgang von Anspruchslohn und preislicher Wettbewerbsfähigkeit auf etwa 4% verminderte. Auffällig ist, dass die geschätzte NAWRU zweifach sehr weit und lange anhaltend von der tatsächlichen Erwerbslosenquote entfernt gelegen hat, während sie in anderen Perioden, etwa in den zurückliegenden zehn Jahren in deren Nähe lag. Ihr Verlauf unterscheidet sich damit sehr deutlich von der mit dem EU-Verfahren geschätzten NAWRU, die sich, einem gleitenden Durchschnitt durch die Erwerbslosenquote nicht unähnlich, kaum weit von dieser entfernt.

### III. Plausibilitätsprüfung

Das Produktionspotential und seine Komponenten wie die NAWRU sind grundsätzlich keine beobachtbaren Größen. Bei ihrer Schätzung handelt es sich insofern – in der Diktion des maschinellen Lernens – um ein „unüberwachtes Lernen“, d. h. um eine Schätzung ohne vorgegebenen Ziel- oder „wahren“ Wert, den es durch die Modellierung zu treffen gilt. Der oben vorgenommenen Prüfung der ökonomischen Plausibilität des theoretischen Ansatzes, des Schätzverfahrens und ggf. auch der Verläufe der geschätzten Teilkomponenten kommt daher für die Bewertung des Ergebnisses eine hohe Bedeutung zu.

Ein explizites Kriterium zur Beurteilung der Schätzung liefert darüber hinaus die makroökonomische Theorie. In ihrer Folge gehen Konjunkturschwankungen und damit von null verschiedene Produktions- und Arbeitsmarktlücken – Abweichungen zwischen der Erwerbslosenquote und der NAWRU – maßgeblich darauf zurück, dass sich die Preise für die meisten Güter nach Schocks auf die gesamtwirtschaftliche Nachfrage nur mit erheblichen Verzögerungen anpassen. Wenn die Lücke aber maßgeblich auf verzögerte Preisanpassungen zurückgeht, sollte sie in der Lage sein, zukünftige Preisänderungen dem Vorzeichen und dem Ausmaß nach zu prognostizieren. Beispielsweise sollte eine große

---

<sup>9</sup> Dustmann et al. (2014) betonen hingegen den Druck, den die deutsche Wiedervereinigung und der für 2004 beschlossene EU-Beitritt der osteuropäischen Länder auf den deutschen Arbeitsmarkt ausübten als maßgebliche Faktoren hinter dem Rückgang des Anspruchslohns in den 1990er und frühen 2000er Jahren.

positive Lücke einer starken Beschleunigung, eine negative kleine Lücke hingegen einer leichten Abschwächung des Preisauftriebs vorauslaufen.

Dies lässt sich mithilfe der Schätzung einer einfachen Phillips-Kurve untersuchen. Auf der linken Seite der Regression steht dabei als zu erklärende Größe der gesamtwirtschaftliche oder – in offenen Volkswirtschaften – der binnenwirtschaftliche Preisauftrieb  $\Delta p$ , als erklärende Größe kommen die geschätzte Arbeitsmarktlücke, idealerweise mit Verzögerung, sowie ggf. Verzögerungen der zu erklärenden Variablen zum Einsatz:

$$\Delta p_t = \gamma_0 + \gamma_1(u - \bar{u})_{t-1} + \gamma_3\Delta p_{t-1} + \epsilon_t \quad (15)$$

Soll die Untersuchung auch Daten der 1970er und frühen 1980er Jahre umfassen, so ist der Preisauftrieb um den geldpolitisch „beabsichtigten“ oder zugelassenen Preisanstieg zu korrigieren, da der Preisauftrieb ohne diese Korrektur nichtstationär ist:

$$(\Delta p - \overline{\Delta p})_t = \gamma_0 + \gamma_1(u - \bar{u})_{t-1} + \gamma_3(\Delta p - \overline{\Delta p})_{t-1} + \epsilon_t \quad (16)$$

Ein alternatives gesamtwirtschaftliches Auslastungsmaß steht mit den Umfragedaten des ifo Instituts zum Auslastungsgrad der Produktionskapazitäten im verarbeitenden Gewerbe und im Bauhauptgewerbe zur Verfügung. Um zu prüfen, welchen zusätzlichen Beitrag dieses Maß zur Erklärung des zukünftigen Preisauftriebs leistet, wird mit der Variable *CUBS* ein solches, ebenfalls mit einer Verzögerung von einem Jahr, in die Phillips-Kurven-Analyse aufgenommen:

$$(\Delta p - \overline{\Delta p})_t = \gamma_0 + \gamma_1(u - \bar{u})_{t-1} + \gamma_2CUBS_{t-1} + \gamma_3(\Delta p - \overline{\Delta p})_{t-1} + \epsilon_t \quad (17)$$

Als Maß für den binnenwirtschaftlichen Preisauftrieb wird der Anstieg des Deflators der aggregierten Bruttowertschöpfung verwendet. Dieser ist über den Zeitraum von 1985 bis zum aktuellen Rand stationär, so dass die Spezifikation (15) verwendet werden kann. Für längere Untersuchungszeiträume – die Analyse soll 1971 beginnen – bedarf es des Korrekturfaktors, sodass Spezifikation (16) zum Einsatz kommt. Die Variante mit dem umfragebasierten Auslastungsgrad wird für beide Zeiträume untersucht.

Es zeigt sich, dass die auf der Basis des Benchmark-Verfahrens (Havik et al. 2014) berechnete Arbeitsmarktlücke in keinem der beiden Zeiträume – und damit auch der Spezifikation der zu erklärenden Variablen – einen statistisch signifikanten Beitrag zu Erklärung des zukünftigen binnenwirtschaftlichen Preisauftriebs zu leisten vermag (Tab. 2). Dies gilt auch, wenn, wie in Gleichung (17), der Kapazitätsauslastungsgrad zusätzlich in die Phillips-Kurve aufgenommen wird. Der Auslastungsgrad selbst leistet dann einen signifikanten Erklärungsbeitrag, wenn auch nur bei einem Konfidenzniveau von 5%.

Wird statt der NAWRU des Benchmark-Verfahrens, die des hier vorgestellten Ansatzes verwendet, ändert sich das Ergebnis deutlich. In beiden Zeiträumen lässt sich die Hypothese, dass die Lücke nicht zur Erklärung des nachfolgenden Preisauftriebs beiträgt, auf dem 1%-Konfidenzniveau verwerfen. Auch insgesamt haben die Phillips-Kurven einen höheren Erklärungsgehalt, die Bestimmtheitsmaße sind unabhängig vom Stützzeitraum um rund 15 Prozentpunkte höher als bei der Benchmark-Lücke. Wird zusätzlich der umfragebasierte Kapazitätsauslastungsgrad zur Erklärung herangezogen, so zeigt sich für beide Untersuchungszeiträume, dass dieser bei konventionellen Signifikanzniveaus keinen weiteren Beitrag zu leisten vermag.

Tab. 2: Empirische Phillips-Kurven für alternative NAWRU-Schätzungen 1971-2021

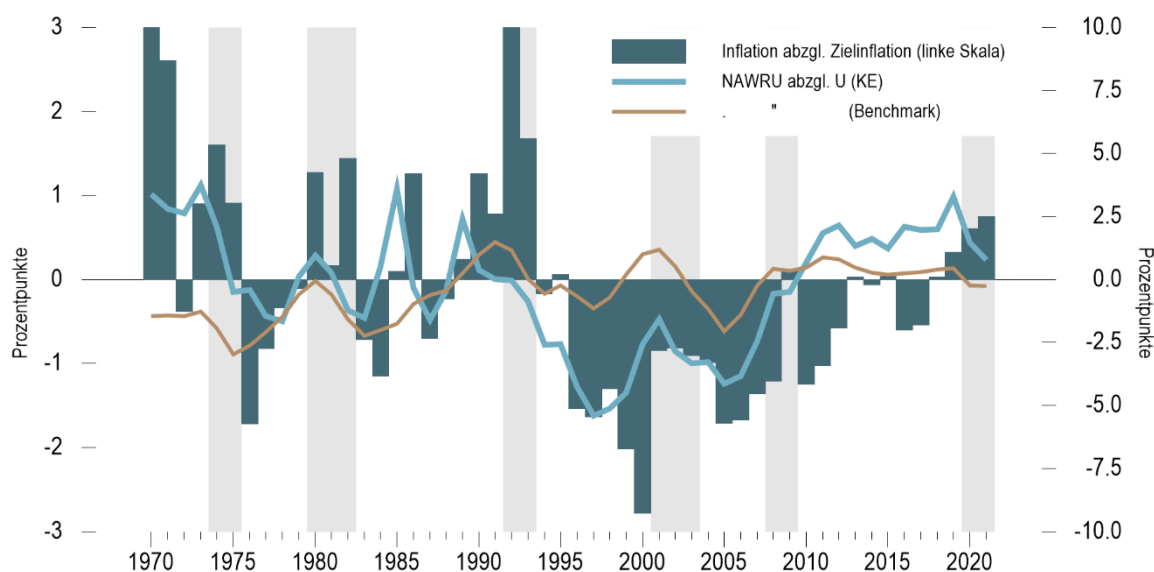
$$\text{Spezifikation: } (\pi - \bar{\pi})_t = \gamma_0 + \gamma_1(u - \bar{u})_{t-1} + \gamma_2 CUBS_{t-1} + \gamma_3(\pi - \bar{\pi})_{t-1} + \epsilon_t$$

Stützzeitraum	1971-2021 ( $\bar{\pi}$ zeitvariabel)			1985-2021 ( $\bar{\pi} = 0$ )		
	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$R^2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$R^2$
Benchmark- Verfahren	0,11 (0,93)	-	0,44	0,15 (1,00)	-	0,59
	-0,01 (0,02)	0,08 (2,36*)	0,49	0,05 (0,36)	0,01 (2,39*)	0,65
Kiel Economics	0,27 (5,05**)	-	0,63	0,22 (4,62**)	-	0,74
	0,24 (4,30**)	0,04 (1,66)	0,64	0,19 (3,72**)	0,05 (1,43)	0,75

Anmerkung: Betragsmäßige  $t$ -Werte in Klammern unter den Koeffizientenschätzern. :  $\bar{R}^2$  steht für das korrigierte Bestimmtheitsmaß. Abhängige Variable ist für den Zeitraum von 1971-2021 die Veränderungsrate des Deflators der gesamtwirtschaftlichen Bruttowertschöpfung gegenüber dem Vorjahr korrigiert um den „unvermeidlichen Preisanstieg“ gemäß Ausweis der Deutschen Bundesbank bei der Begründung ihrer jährlichen Geldmengenziele bzw. (ab 1999) um die Obergrenze für den Preisniveauanstieg im Euroraum der EZB. Für den Zeitraum von 1985-2021 ist es nur die Veränderungsrate des Deflators. \* bzw. \*\* indizieren Signifikanz auf dem 5%- bzw. 1%-Konfidenzniveau.

Quelle: Eigene Berechnungen.

Abb.2: Abweichung zwischen Erwerbslosenquote und geschätzter NAWRU und Abweichungen des gesamtwirtschaftlichen Preisniveauanstiegs vom angestrebten Wert 1960-2018



Anmerkung: Daten bis 1994 für das frühere Bundesgebiet. Inflation und „Zielinflation“ gemessen am Deflator der Bruttowertschöpfung. „Zielinflation“ gemäß den jährlichen Ableitungen des Geldmengenziel der deutschen Bundesbank, davor Anstiegsrate des US-BIP-Deflators; ab 1999 gemäß der Obergrenze für den Preisniveauanstieg der EZB von 2%.

Quelle: Statistisches Bundesamt, Bundesagentur für Arbeit, Deutsche Bundesbank, Europäische Zentralbank; eigene Berechnungen.



Der insbesondere auch im Vergleich zum Benchmark-Ansatz hohe Erklärungsgehalt des hier vorgestellten Ansatzes zur Schätzung der NAWRU ergibt sich nicht zuletzt in den beiden Dekaden zwischen 1990 und 2010. Im Nachgang der deutschen Vereinigung und der Gründung der Europäischen Währungsunion (EWU) stagnierte das gesamtwirtschaftliche Preisniveau ab Mitte der 1990er Jahre in der Tendenz und blieb damit um bis zu 2½ Prozentpunkte hinter dem von der Deutschen Bundesbank bzw. ab 1999 der Europäischen Zentralbank angestrebten Preisauftrieb zurück (Abb. 2). Die Arbeitsmarktlücke auf der Basis des Benchmark-Verfahrens vermag diese Entwicklung nicht nachzuzeichnen; sie zeigt in diesem Zeitraum keine starke Unterauslastung des Arbeitsmarkts an. Anders die Lücke auf der Basis des hier vorgestellten Verfahrens, die bereits in der ersten Hälfte der 1990er Jahre in den negativen Bereich rutscht und im Jahr 1997 mit rund 8 Prozent ihren Tiefpunkt durchschreitet. Danach verharrt sie weiter lange im negativen Bereich. Erst nach der Mitte der 2000er-Jahre bewegt sie sich wieder in Richtung Nulllinie – und „prognostiziert“ damit die Normalisierung des Preisauftriebs im Jahrzehnt nach 2010. Das hier vorgestellte Verfahren diagnostiziert für die Zeit nach der Wiedervereinigung und den Beginn der EWU somit gemessen an der Lage auf dem Arbeitsmarkt eine starke Unterauslastung der gesamtwirtschaftlichen Kapazitäten, die erst ab dem Jahr 2010 wieder abgebaut war.

#### IV. Schlussfolgerungen

Die Bedeutung einer belastbaren Schätzung der NAWRU für die Konjunkturbereinigung geht über die zyklische Bereinigung der Erwerbslosenquote hinaus. Da das Ausmaß der Erwerbslosigkeit Rückwirkungen auf das Arbeitsangebotsverhalten hat, kann die geschätzte Aufteilung der Erwerbslosenquote in eine konjunkturelle und eine nicht-konjunkturelle Komponente auch zur Konjunkturbereinigung der Arbeitszeit und der Erwerbsneigung eingesetzt werden. Die am aktuellen Rand immanent revisionsanfälligen Filterverfahren, die im Standardansatz bislang zur Bereinigung des Arbeitsangebots eingesetzt werden, werden dadurch möglicherweise entbehrlich. Darüber hinaus zeigt sich empirisch, dass der geschätzte Anspruchslohn, die wichtigste Komponente der NAWRU, eine statistisch signifikante Wirkung auf die Totale Faktorproduktivität hat. Dies dürfte sich dadurch erklären lassen, dass der aggregierte Anspruchslohn darüber entscheidet, in welchem Umfang Personen mit geringer Arbeitsproduktivität ihre Fähigkeiten am Arbeitsmarkt einbringen können, statt von Sozialleistungen leben zu müssen. Er bestimmt somit maßgeblich über die qualifikatorische Komposition des Arbeitsinputs, die notwendigerweise in der TFP mitgemessen wird, da der Produktionsfunktionsansatz den Arbeitsinput als qualifikatorisch homogen unterstellt (Sachverständigenrat 2015, Tz. 599).

Das hier vorgestellte Modell zur Schätzung der NAWRU entspricht in der theoretischen Konzeption nahezu exakt dem von Havik et al. (2014) und damit dem von der EU-Kommission verwendeten Verfahren. Havik et al. (2014) setzen bei der empirischen Umsetzung allerdings einige Restriktionen auf die Dynamik, die sich empirisch für Deutschland nicht bestätigen lassen. Verzichtet man auf diese Restriktionen, so erhält man – in der Konzeption von Blanchard und Katz (1999), Bårdsen et al. (2005) etc. – eine NAWRU, die nicht auf dem Phillips-Kurven-Ansatz basiert, sondern auf dem Lohnkurven-Ansatz, also auf der Basis von Niveaugrößen an Stelle von Veränderungsraten modelliert ist. Aus der Zeitreihenökonomie ist bekannt, dass eine Modellierung in Veränderungsraten Informationen verschonkt und geschätzte Dynamiken verzerrt, wenn die Daten eine Modellierung in Niveaugrößen nahelegen.

Die mit der geschätzten NAWRU abgeleitete Arbeitsmarktlücke ist deutlich persistenter, als der auf der Basis von Havik et al. (2014) ermittelte. Da allerdings der Preisauftrieb (bzw. dessen Abweichun-

gen von der geldpolitisch angestrebten Rate) ebenfalls eine hohe Persistenz aufweist, ist dies ökonomisch plausibel. Die hier geschätzte NAWRU könnte daher ein wichtiger Input in eine valide Schätzung des gesamtwirtschaftlichen Auslastungsgrades als Basis für die Ausrichtung der Finanzpolitik und der Geldpolitik der EZB darstellen. Da die Schätzung auf einem strukturellen Modell basiert, ist sie zudem transparent und inhaltlich nachvollziehbar. Da das Schätzverfahren konsistent in das D\*-Modell von Kiel Economics eingebunden ist, steht es ferner direkt für die Erstellung von Konjunkturprognosen und Mittelfristprojektionen zur Verfügung.

## Anhang – Datensatz

Die gesamtwirtschaftliche Produktion  $y$  wird durch das reale Bruttoinlandsprodukt gemessen, der Arbeitseinsatz  $l$  durch die geleisteten Arbeitsstunden der Erwerbstätigen, der Kapitaleinsatz  $k$  durch den gesamtwirtschaftlichen Kapitalstock. Als Lohnsatz  $w$  wird das Arbeitnehmerentgelt je Beschäftigtenstunde, als Erzeugerpreis  $p$  der Deflator der gesamtwirtschaftlichen Bruttowertschöpfung, als Konsumentenpreis  $p^c$  der Konsumdeflator verwendet und mit  $p^m$  wird der Deflator der Einfuhr und mit  $\tilde{p}$  der Deflator des Gesamtabsatzes bezeichnet. Diese Daten werden der Volkswirtschaftlichen Gesamtrechnung (VGR) und beginnen standardmäßig im Jahr 1950. Ältere Daten beziehen sich auf das frühere Bundesgebiet, sie wurden mit Ausnahme der Erwerbslosenquote  $u$  im Jahr 1991 mit gesamtdeutschen Daten verkettet. Für die Erwerbslosenquote werden Daten für das frühere Bundesgebiet bis zum Jahr 1994 verwendet; Hintergrund des späteren Übergangs bei der Erwerbslosenquote ist die Einschätzung, dass der Anstieg der Erwerbslosigkeit in Ostdeutschland zwischen 1991 und 1994 maßgeblich auf den Systemumbruch dort zurückgehen und insofern nicht-konjunkturelle Natur sein dürfte. Da dieser – historisch mutmaßlich einmalige – Effekt im Rahmen des theoretischen Modells nicht berücksichtigt wird, würden die ostdeutschen Daten jener Jahre die Schätzung der übrigen nicht-konjunkturellen Treiber der Erwerbslosenquote, d. h. der Preis- und Lohnaufschlagsfaktoren, verzerren.

Die internationale preisliche Wettbewerbsfähigkeit der deutschen Wirtschaft (bzw. der reale Außenwert der heimischen Währung),  $q$ , wird ab dem Jahr 1994 durch den Indikator der Deutschen Bundesbank gegenüber einem Kreis von 37 Ländern auf der Basis von Deflatoren des Gesamtabsatzes abgebildet. Von 1961 bis 1993 wird der entsprechende Indikator gegenüber ausgewählten Industrieländern verwendet. Für den Zeitraum davor wird der Indikator als konstant unterstellt. Alle genannten Daten werden in natürlich-logarithmischer Transformation verwendet.

Zur Modellierung der erwarteten Veränderungsraten der beiden Preisgrößen im Folgejahr,  $E_t(\Delta p_{t+1})$  bzw.  $E_t(\Delta p_{t+1}^c)$ , werden Prognosen verwendet (vgl. Fuhrer 2017, Zhang et al. 2009). Zum Einsatz kommen ab dem Jahr 1968 die Prognosen dieser Größen für das Folgejahr, die von der Gemeinschaftsdiagnose jeweils in ihren Herbstgutachten veröffentlicht werden. Da es sich bei der Gemeinschaftsdiagnose regelmäßig um ein Übereinkommen von – je nach Modus des beauftragenden Bundeswirtschaftsministeriums – vier bis sechs unabhängigen Wirtschaftsforschungsinstituten handelt, dürften deren Prognosen dem Ideal einer Konsensprognose durchaus nahekommen und Repräsentativität für die Erwartungsbildung unter Experten zum jeweiligen Zeitpunkt besitzen. Für die Zeit vor 1968 stehen keine veröffentlichten Prognosen der Gemeinschaftsdiagnose zur Verfügung, allerdings weisen Vogel und Duelli (1968) für die Jahre 1962 bis 1967 nicht veröffentlichte Prognosen der Gemeinschaftsdiagnose aus und ab 1964 veröffentlichte der Sachverständigenrat zur Beurteilung der gesamtwirtschaftlichen Entwicklung Prognosen im Rahmen seiner Jahresgutachten. Schließlich wurden – allerdings nicht im Herbst, sondern im Frühjahr eines jeden Jahres – zwischen 1958 und 1966

und Federführung des Bundeswirtschaftsministeriums von einem interministeriellen Arbeitskreis „Gesamtwirtschaftliche Vorausschätzungen“ unter Beteiligung des Statistischen Bundesamts und der Deutschen Bundesbank regierungsinterne Prognosen erstellt, die von Schlesinger (1967) ausgewertet und veröffentlicht wurden. Die Erwartungsdaten hier basieren jeweils auf dem Mittelwert der jeweils verfügbaren Prognosen, beinhalten für den Zeitraum von 1958 bis 1961 also nur auf Prognosen des interministeriellen Arbeitskreises. Für die Jahre 1951 bis 1957 werden schlicht die Anstiegsraten im Vorjahr als Erwartungsgrößen angesetzt. Statt durch prognostizierten Anstieg des Deflators der gesamtwirtschaftlichen Bruttowertschöpfung wird der erwartete Erzeugerpreisanstieg für die Jahr vor 1968 durch den (implizit, über die ausgewiesenen Anstiegsraten des nominalen und realen Bruttoinlandsprodukts) prognostizierten Anstieg des Deflators des Bruttoinlandsprodukts gemessen.

Für die Produktionselastizität der Arbeit  $\alpha$  wird ein Wert von 0,65 unterstellt.

## Literatur

- Andrews, D., und W. Ploberger (1994). Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative. *Econometrica* 62 (6): 1383–1414.
- Bårdsen, G und R. Nymoen (2003). Testing Steady-State Implications for the NAIRU, *The Review of Economics and Statistics* 85 (4), 1070-1075.
- Bårdsen, G und R. Nymoen (2009). Macroeconometric Modelling for Policy. In *Palgrave Handbook of Econometrics*. Bd. 2, hg. v. Terence C. Mills und Kerry D. Patterson, 851–916.
- Bårdsen, G., O. Eitrheim, E. Jansen und R. Nymoen (2005). *The Econometrics of Macroeconomic Modelling*. Oxford.
- Batini, N., B. Jackson und S. Nickell (2005). An Open-Economy New Keynesian Phillips Curve for the U.K., *Journal of Monetary Economics* 52, 1061-1071
- Bayoumi, T., D. T. Coe und E. Helpman (1999). R&D Spillovers and Global Growth. *Journal of International Economics* 47(2), 399-428.
- Belsley, D., E. Kuhn und R. Welsh (1980): *Regression Diagnostics - Identifying Influential Data and Sources of Collinearity*. New York.
- Bellmann, L. und U. Blien (1996). Die Lohnkurve in den neunziger Jahren - Der Zusammenhang zwischen regionalen Durchschnittslöhnen und regionaler Arbeitslosigkeit in einer Mehrebenenanalyse mit dem IAB-Betriebspanel. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* 29 (3), 465-470.
- Blanchard, O. und L. Katz (1999). Wage Dynamics: Reconciling Theory and Evidence, *American Economic Review* 89, 69-74.
- Blanchflower, D. G. und A. J. Oswald (1994). *The Wage Curve*. Cambridge, MA.
- Blanchflower, D. G. und A. J. Oswald (2005). *The Wage Curve Reloaded*. NBER Working Paper No. 11338. Cambridge MA.
- Boss, A. (2009). Zur Entwicklung des Anspruchslohns in Deutschland. *Zeitschrift für Wirtschaftspolitik* 58 (2): 222-254.

- Burda, M. C. und S. Seele (2017). Das deutsche Arbeitsmarktwunder: Eine Bilanz. *Perspektiven der Wirtschaftspolitik* 18(3): 179–204.
- Currie, D. (1981). Some Long-Run Features of Dynamic Time Series Models. *Economic Journal* 91, 704-715.
- Deutsche Bundesbank (2006). Monatsbericht September. Frankfurt am Main.
- Dustmann, C., B. Fitzenberger, U. Schönberg und A. Spitz-Oener (2014). From Sick Man of Europe to Economic Superstar: Germany's Resurgent Economy. *Journal of Economic Perspectives* 28(1): 167-188.
- Franz, W. (2005). Will the (German) NAIRU please Stand up? *German Economic Review* 6(2), 131-153.
- Havik, K., K. Mc Morrow, F. Orlandi, C. Planas, R. Raciborski, W. Röger, A. Rossi, A. Thum-Thysen, V. Vandermeulen (2014), The Production Function Methodology for Calculating Potential Growth Rates & Output Gaps. EU-Kommission Working Paper, Brüssel.
- Hochmuth, B., B. Kohlbrecher, C. Merkl und H. Gartner (2021). Hartz IV and the Decline in German Unemployment – A Macroeconomic Evaluation. *Journal of Economic Dynamics and Control* 127(1), 1-33.
- Layard, R., S. Nickell und R. Jackman (1991). Unemployment. *Macroeconomic Performance and the Labour Market*. Oxford.
- Layard, R., und S. Nickell (1999). Labor Market Institutions and Economic Performance. In O. Ashenfelter und D. Card (eds.), *Handbook of Labor Economics*. Vol. 3. Amsterdam.
- Fuhrer, J. (2017). Expectations as a Source of Macroeconomic Persistence: Evidence from Survey Expectations in a Dynamic Macro Model, *Journal of Monetary Economics* 86, 22-35
- Grossman, G. und E. Helpman (1991). *Innovation and Growth in the Global Economy*. Cambridge.
- Manning, A. (1993). Wage Bargaining and the Phillips Curve: The Identification and Specification of Aggregate Wage Equations. *The Economic Journal* 103 (January): 98–118.
- Pissarides, C.A. (1998). The Impact of Employment Tax Cuts on Unemployment and Wages. The Role of Unemployment Benefits and Tax Structure. *European Economic Review* 42(1): 155–183.
- Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (2015). *Zukunftsfähigkeit in den Mittelpunkt. Jahresgutachten 2015/2016*. Wiesbaden.
- Schlesinger, H. (1967). Wert und Grenzen gesamtwirtschaftlicher Vorausschätzungen. *Allgemeines statistisches Archiv* 51 (2/3): 401-416.
- Staiger, D., J. H. Stock und M. W. Watson (1997). The NAIRU, unemployment and monetary policy. *Journal of Economic Perspectives* 11(1): 33-49.
- Steffen, J. (2022). *Sozialpolitische Chronik – Die wesentlichen Änderungen in der Arbeitslosenversicherung*. Bremen.
- Vogel, O. und L. Duelli (1968). Prognosewerte und Ist-Werte. *Beiträge des deutschen Industrieinstituts* 6 (2/3).
- Wickens, M.R., und T.S. Breusch (1989). Dynamic Specification, the Long Run and the Estimation of Transformed Regression Models. *Economic Journal* 98 (Supplement): 189–205.
- Zhang, C., D. R. Osborne und D. H. Kim (2009). Observed Inflation Forecasts and the New Keynesian Phillips Curve, *Oxford Bulletin of Economics and Statistics* 71(3), 375-398.

Christian R. Proaño

## Für eine ökonomischere Fundierung der Trend-TFP in der Ermittlung des Produktionspotenzials

Obwohl die Schuldenbremse seit 2009 im deutschen Grundgesetz verankert ist, bleibt ihre Sinnhaftigkeit und ihr spezifisches Design umstritten (siehe z.B. Bardt et al., 2019). Neben konzeptioneller Vorbehalte, die schon im Vorfeld ihrer Einführung diskutiert wurden (siehe z.B. Horn et al., 2008) und die weiterhin thematisiert werden (siehe z.B. Sachverständigenrat, 2019), wird insbesondere der Ansatz des verwendeten Konjunkturbereinigungsverfahrens, die sogenannte Konjunkturkomponente der Schuldenbremse, weiterhin kritisiert (siehe auch Heimberger und Truger, 2020).

Wie wohl bekannt, soll die Konjunkturkomponente die geschätzte Lage der deutschen Wirtschaft im Konjunkturzyklus widerspiegeln und somit eine „konjunkturge-rechte“ Fiskalpolitik insbesondere im Hinblick auf die Begrenzung der Staatsverschuldung ermöglichen. Dabei wird die konjunkturelle Lage der Wirtschaft durch die sogenannte Outputlücke, also die prozentuale Abweichung des gegenwärtigen Produktionsniveaus zum sogenannten Potenzialniveau, gemessen. Unter der Annahme, dass die Outputlücke und somit die wahre konjunkturelle Lage richtig eingeschätzt oder sogar gemessen wird, würde die Konjunkturkomponente maßgeblich zur Einhaltung der Schuldenbremse und somit zur Beschränkung bzw. zur Reduzierung der Staatsverschuldung (siehe z.B. Sachverständigenrat, 2019, 298 ff.) in der mittleren bis langen Frist beitragen, da sich rezessionsbedingte Defizite durch fiskalische Konsolidierungen in Aufschwungsphasen ausgleichen würden. Somit ist eine richtige oder zumindest nicht stark verzerrte Einschätzung oder Messung der gegenwärtigen konjunkturellen Lage, sprich Outputlücke, eine quasi notwendige Bedingung für die Einhaltung der Schuldenbremse.

© Der/die Autor:in 2022. Open Access: Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht ([creativecommons.org/licenses/by/4.0/deed.de](https://creativecommons.org/licenses/by/4.0/deed.de)).

Open Access wird durch die ZBW – Leibniz-Informationszentrum Wirtschaft gefördert.

**Prof. Dr. Christian R. Proaño** lehrt monetäre Makro-  
ökonomie und empirische Wirtschaftsforschung an der  
Otto-Friedrich-Universität Bamberg.

Was auf den ersten Blick richtig und schlüssig klingt, steht jedoch theoretisch und empirisch aus verschiedenen Gründen auf wackeligen Füßen. Zum einen gibt es genügend überzeugende theoretische Argumente, um vom Gegenteil auszugehen (siehe z.B. Aiyar und Voigts, 2019). Zum anderen stellt sich die grundsätzliche Frage, ob eine Fiskalregel, die stark von einer unbeobachtbaren Variable – dem Potenzialniveau der Produktion – abhängt, eine so zentrale Rolle in der Gestaltung der Fiskalpolitik hätte einnehmen sollen.

Nicht nur seit der Debatte vor der Einführung der Schuldenbremse 2009 ist jedoch wohlbekannt, dass die Nutzung der Outputlücke nicht unproblematisch ist. Während auf der theoretischen Ebene der „Modellbauer“ diese – in der Realität unbeobachtbare – Größe kennt,<sup>1</sup> ist der Potenzialoutput in der Realität nicht direkt beobachtbar. Dies macht eine Schätzung desgleichen, egal wie ausgefallen die angewandte ökonometrische Methode sein mag, angreifbar und noch wichtiger, nur indirekt und nicht hinreichend eindeutig verifizierbar.

Trotz dieser und anderer wohlbekannter Einwände beinhaltet die deutsche Schuldenbremse eine Konjunkturkomponente, die auf der Schätzung des Potenzialniveaus beruht. Genauer gesagt hat der von der Europäischen Kommission verwendete Produktionsfunktionsansatz (siehe Havik et al., 2014) als Ausgangspunkt die aggregierte Cobb-Douglas-Produktionsfunktion<sup>2</sup>

$$Y_t = TFP_t \times L_t^\alpha \times K_t^{1-\alpha}$$

wo  $Y_t$  der aggregierte Output,  $L_t$  und  $K_t$  die Produktionsfaktoren Arbeit und Kapital und  $TFP_t$  die sogenannte totale Faktorproduktivität, welche als Residualgröße (das Solow-Residuum) ermittelt wird, darstellen.

- 1 Es ist anzumerken, dass die Outputlücke in dem zurzeit prädominanten DSGE-Ansatz (dynamic stochastic general equilibrium) als prozentuale Abweichung zum Outputniveau unter völlig flexiblen Preisen gilt, während in der praxisnahen Debatte das Potenzialoutput als das langfristige Trendniveau der Produktion verstanden wird. Ausnahmen, in denen die „praxisnahe“ Definition der Outputlücke in DSGE-Modellen benutzt und anhand eines von Hodrick und Prescott (1997) entwickelten Filters approximiert wird, sind Born und Pfeiffer (2014, 2020), siehe auch Proaño (2022) für ein makroökonomisches Modell mit einem ähnlichen Ansatz, aber unter der Annahme von beschränkt rationalen Agenten.
- 2 Die Benutzung der Cobb-Douglas-Produktionsfunktion selbst ist aus theoretischen und empirischen Gründen umstritten, siehe z.B. Gechert et al. (2022).

Das Produktionspotenzial  $YPOT_t$  wiederum ist definiert als

$$YPOT_t = TFPT_t \times LP_t^\alpha \times K_t^{1-\alpha}$$

wo  $LP_t$  das potenzielle Arbeitsvolumen und  $TFPT_t$  die Trend-TFP darstellen.

Wie ausführlich in Ademmer et al. (2019) diskutiert, wird die Trend-TFP durch einen diffusen Kalman-Filter-Ansatz, welcher die zyklische Komponente der TFP mit dem Auslastungsgrad der Volkswirtschaft verknüpft, modelliert und anhand eines bayesianischen Verfahrens geschätzt. Somit wird die Trend-TFP indirekt und nur auf Basis zyklischer Dynamiken ermittelt, was im Kontrast zum eher längerfristigen Charakter dieser Variablen steht. Somit erscheint es sinnvoll, nach einer besseren ökonomischen Fundierung für die Modellierung der Trend-TFP zu suchen.<sup>3</sup>

Die endogene Wachstumstheorie (siehe Lucas, 1988; Romer, 1990; und Aghion und Howitt, 1992) ist eine natürliche Kandidatin für eine solche ökonomische Fundierung: Demnach ist der technologische Fortschritt einer Volkswirtschaft (empirisch durch die residual ermittelte TFP approximiert) nicht durch exogene Zufallsschocks – wie z.B. in Real-Business-Cycles-Modellen (RBC) angenommen (siehe z.B. Kydland und Prescott, 1982) – getrieben, sondern ist das Ergebnis von Investitionen in Forschung und Entwicklung von neuen Technologien, welche das Wissen einer Gesellschaft und somit die totale Faktorproduktivität erhöhen. Während in den Arbeiten von Lucas und Romer sowie Aghion und Howitt angenommen wurde, dass der Marktwettbewerb der treibende Faktor für diesen Innovationprozess war, ist die zentrale Rolle des Staates in diesem Prozess in den vergangenen Jahren immer mehr in den Vordergrund der öffentlichen Debatte gerückt. In ihrem Buch „The Entrepreneurial State“ dokumentiert Mazzucato (2013), wie der US-amerikanische Staat die Entwicklung von bahnbrechenden Technologien wie dem Smartphone oder dem Suchalgorithmus von Google massiv förderte. Antolin-Díaz und Surico (2022), zeigen anhand eines vektorautoregressiven Modells und US-Quartalsdaten in einer Zeitspanne von 125 Jahren, dass staatliche Ausgaben in Forschung und Entwicklung zu einer Erhöhung in Innovation und Produktivität führen, die den aggregierten Output in der mittleren Frist steigern. Auch Krebs et al.

3 Wie in den anderen Beiträgen dieses Zeitgesprächs (siehe z.B. Schuster und Sigl-Glöckner) diskutiert wird, ist die Ermittlung des potenziellen Arbeitsvolumens umstritten, da es auf der Schätzung der nicht minder umstrittenen NAWRU basiert. In diesem Beitrag wird jedoch nur auf die Modellierung der Trend-TFP eingegangen, um einen punktuellen und „minimalinvasiven“ Vorschlag zur Veränderung des gegenwärtigen Verfahrens zur Ermittlung des Produktionspotenzials zu liefern.

Tabelle 1  
Regression mit Wachstumsrate der TFP (2005-2021)

Variable	Koeffizient	SEM	t-Statistik	Prob.
C	0,141	0,032	4,390	0,001
PAT_PC(-2)	0,234	0,106	2,212	0,049
PAT_PC(-3)	0,352	0,097	3,632	0,004
PAT_PC(-4)	0,509	0,104	4,881	0,001
FuEAUSG_HSH(-1)	0,369	0,103	3,577	0,004
FuEAUSG_WIRT(-1)	-0,163	0,037	-4,401	0,001
R <sup>2</sup>	0,806	Mittel abh. Variable		0,005
R <sup>2</sup> adj.	0,718	SEM abh. Variable		0,017
Standardfehler (SEM)	0,009	AIC		-6,328
Quadrat-Σ der Resid.	0,001	SIC		-6,034
Log likelihood	59,79	HQC		-6,299
F-Statistik	9,145	Durbin-Watson-Test		2,765
Prob (F-Statistik)	0,001			

Quelle: Statistisches Bundesamt; eigene Berechnungen.

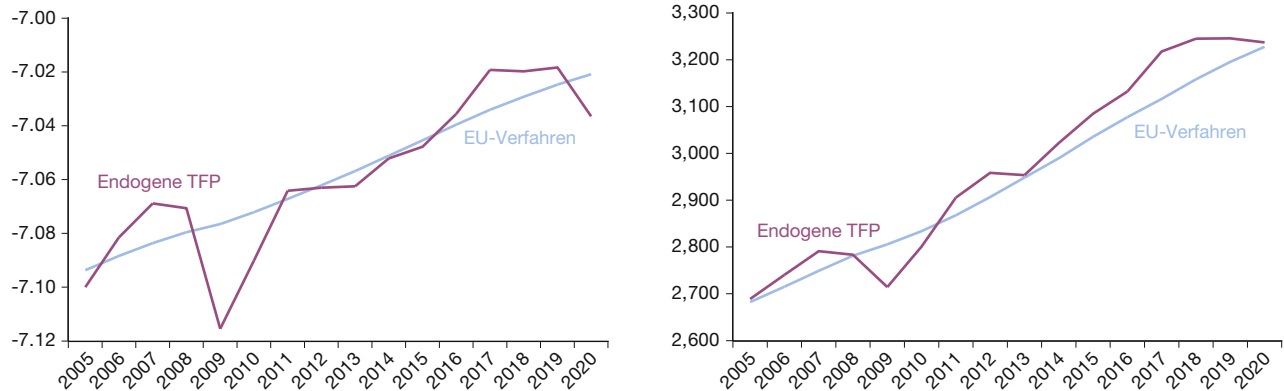
(2021) argumentieren, dass eine dauerhafte Steigerung der öffentlichen Investitionsausgaben im Infrastruktur- und Bildungsbereich das Produktionspotenzial signifikant erhöhen kann. Da der Staat in der Humankapitalakkumulation in Volkswirtschaften wie Deutschland eine zentrale Rolle spielt, ist es naheliegend zu versuchen, diesen Transmissionsmechanismus staatlicher Aktivität explizit in die Ermittlung des Produktionspotenzials miteinzubeziehen.

Benutzt man z.B. frei verfügbare Daten des Statistischen Bundesamts zu Patentanmeldungen und Ausgaben in Forschung und Entwicklung (im Staats- und Privatsektor) für die Erklärung der Wachstumsrate der TFP für den Zeitraum von 2005 bis 2021, bekommt man anhand einer einfachen linearen Regression mit der Methode der kleinsten Quadrate die folgenden Schätzergebnisse für die Wachstumsrate der TFP (vgl. Tabelle 1). Dabei stellen PAT\_PC die jährliche Wachstumsrate der Patentanmeldungen, FuEAUSG\_HSH die FuE-Ausgaben als Anteil der Gesamtausgaben der deutschen Hochschulen und FuEAUS\_WIRT die FuE-Ausgaben als Anteil der Gesamtausgaben des privaten Firmensektors dar.

Aufgrund der begrenzten Zahl von verfügbaren Beobachtungen ist eine ausführlichere Modellspezifikation nicht ohne Nachteile durchführbar. Trotz der relativ niedrigen Zahl von erklärenden Variablen hat jedoch die obige Schätzgleichung einen akzeptablen *fit*, wobei anzumerken ist, dass die geschätzten Residuen eine gewisse Au-

Abbildung 1

## Logarithmierte TFP nach EU-Verfahren und nach vorgestellten Verfahren und resultierende Potenzialoutputs



Quelle: eigene Berechnungen.

tokorrelation erster Ordnung aufweisen. Wichtiger jedoch ist die Tatsache, dass die geschätzten Koeffizienten der Patentmeldungen und der staatlichen FuE-Ausgaben statisch signifikant, ökonomisch sinnvoll und von einer ähnlichen Größenordnung wie die der European Commission (2016) sind. Der negative Koeffizient der privaten FuE-Ausgaben ist jedoch kontraintuitiv und weitere Spezifikationen sollten daher untersucht werden.

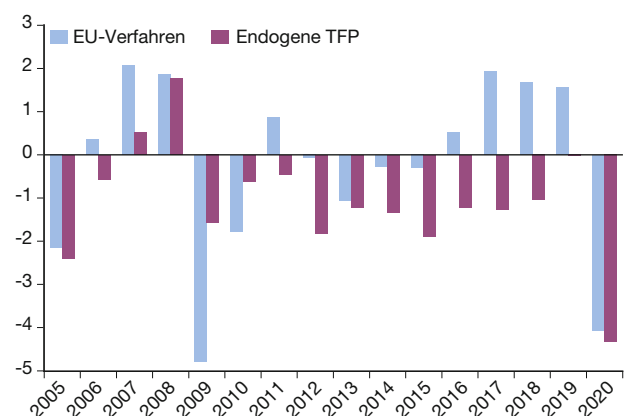
Die linke Grafik in Abbildung 1 stellt die Trend-TFP nach dem EU-Verfahren der geschätzten TFP gegenüber, welche hier als Trend-TFP interpretiert wird, im Zeitraum 2005 bis 2021 (dem Schätzzeitraum der Letzteren), während die rechte Grafik in der gleichen Abbildung die daraus resultierenden Potenzialniveaus gemäß beiden Ansätzen illustriert. Zwei Punkte sind hierbei hervorzuheben: *Erstens*, die „EU“-Log-Trend-TFP hat einen nahezu linearen Verlauf im dargestellten Zeitraum, während die „endogene“ Trend-TFP mehr Volatilität aufweist, welche jedoch auf die Entwicklung der Patentanmeldungen und der FuE-Ausgaben im staatlichen und privaten Sektor zurückzuführen ist. *Zweitens* ist das sich daraus ergebende Potenzialniveau bei der vorgeschlagenen Modellierung der Trend-TFP seit 2010 höher als das Potenzialniveau nach dem EU-Verfahren, wie die rechte Grafik deutlich darstellt.

Wie Abbildung 2 zeigt, hat dies einen signifikanten Effekt auf die geschätzte Outputlücke und somit auf die Einschätzung der konjunkturellen Lage: Während die Outputlücke in Deutschland nach dem EU-Verfahren z.B. seit 2015 positiv eingeschätzt war und somit eine eher restriktive Fiskalpolitik gemäß der Konjunkturkomponente implizierte, hätte die vorgestellte Methode auf eine negative Outputlücke (aufgrund eines höheren Potenzialoutputs) hingedeutet, was wiederum eine eher expansive Fiskalpolitik ermöglicht hätte.

Diese einfache empirisch-ökonomische Übung zeigt zum einen deutlich, wie leichte Änderungen in der Ermittlung des Potenzialoutputs zu grundlegend anderen Implikationen der Schuldenbremse für die Gestaltung der Fiskalpolitik führen können und somit auf welch losen Fundamenten eine Fiskalregel basiert, die auf unbeobachtbaren Größen beruht. Das Ziel dieses Beitrags ist jedoch dezidiert konstruktiv, denn eine „ökonomischere“ Fundierung der Trend-TFP erscheint als ein vielversprechender Weg, um hoffentlich eine bessere (weil fundiertere) Einschätzung der wahren konjunkturellen Lage zu liefern, möchte man trotz wohlbekannter konzeptioneller Vorbehalte immer noch an der Schuldenbremse festhalten.

Abbildung 2

## Outputlücken nach EU-Verfahren und nach vorgestelltem Verfahren



Quelle: eigene Berechnungen.

## Literatur

- Ademmer, M., S. Boysen-Hogrefe, K. Carstensen, P. Hauber, N. Jannsen, S. Kooths, T. Rossian und U. Stolzenburg (2019), Schätzung von Produktionspotenzial und -lücke: Eine Analyse des EU-Verfahrens und mögliche Verbesserungen, *Kieler Beiträge zur Wirtschaftspolitik*.
- Aghion, P. und P. Howitt (1992), A Model of Growth through Creative Destruction, *Econometrica*, 60, 323-351.
- Antolin-Díaz, J. und P. Surico (2022), The Long-Run Effects of Government Spending, *CEPR Press Discussion Paper*, 17433.
- Aiyar, S. und S. Voigts (2019), The negative mean output gap, *Working Paper*, 19/183, International Monetary Fund.
- Bardt, H., S. Dullien, M. Hüther und K. Rietzler (2019), Für eine solide Finanzpolitik. Investitionen ermöglichen!, *IMK Report*, 152.
- Born, B. und J. Pfeifer (2014), Policy risk and the business cycle, *Journal of Monetary Economics*, 68, 68-85.
- Born, B. und J. Pfeifer (2020), Uncertainty-Driven Business Cycles: Assessing the Mark-Up Channel, *Quantitative Economics*, 12(2), 587-623.
- Gechert, S., T. Havranek, Z. Irsova und D. Kolcunova (2022), Measuring Capital-Labor Substitution: The Importance of Method Choices and Publication Bias, *Review of Economic Dynamics*, 45 (Juli), 55-82.
- European Commission (2016), Drivers of total factor productivity growth in the EU: The role of firm entry and exit, *Quarterly Report on the Euro Area*, 15(1), 33, Februar.
- Havik, K., K. Mc Morrow, F. Orlandi, C. Planas, R. Raciborski, W. Röger, A. Rossi, A. Thum-Thysen und V. Vandermeulen (2014), The Production Function Methodology for Calculating Potential Growth Rates & Outputs Gaps, *European Economy Economic Papers*, 535, November.
- Heimberger, P. und A. Truger (2020), Der Outputlücken-Nonsense gefährdet Deutschlands Erholung von der Corona-Krise, <https://makronom.de/der-outputluecken-nonsense-gefaehrdet-deutschlands-erholung-von-der-corona-krise-36125> (2. November 2022).
- Hodrick, R. J. und E. C. Prescott (1997), Postwar U.S. business cycles: An empirical investigation, *Journal of Money, Credit, and Banking*, 29, 1-16.
- Horn, G. A., T. Niechoj, C. R. Proaño, A. Truger, D. Vesper und R. Zwiener (2008), Die Schuldenbremse – eine Wachstumsbremse?, *IMK Report*, 29.
- Krebs, T., J. Steitz und P. Graichen (2021), Öffentliche Finanzierung von Klima- und anderen Zukunftsinvestitionen, *Agora Energiewende und New Economy Forum*, 244/16-1-2021/DE, November.
- Kydland, F. E. und E. C. Prescott (1982), Time to Build and Aggregate Fluctuations, *Econometrica*, 50(6), 1345-1370.
- Lucas, R. (1988), On the Mechanics of Economic Development, *Journal of Monetary Economics*, 22, 3-42.
- Mazzucato, M. (2013), *The Entrepreneurial State: Debunking Public vs. Private Sector Myths*, Anthem Press.
- Proaño, C. R. (2022), Output Gap Uncertainty, Fiscal Policy and Risk Premia under Behavioral Expectations, Universität Bamberg, mimeo.
- Romer, P. (1990), Endogenous Technical Change, *Journal of Political Economy*, 98, 71-102.
- Sachverständigenrat (2019), Den Strukturwandel meistern, Jahresgutachten des Sachverständigenrats zur Begutachtung der gesamtwirtschaftlichen Entwicklung 2019/20.
- Schuster, F. und P. Sigl-Glöckler (2022), Behebung eines Webfehlers: Wie die Konjunkturkomponente demokratisch legitimiert und nachhaltiger werden könnte, *Wirtschaftsdienst*, 102(11), 826-829.

### Title: *For a More Economics-Based Foundation of Trend-TFP in the Estimation of Potential Output*

**Abstract:** *In the production function-based approach proposed by the European Commission for the estimation of the potential output, trend total factor productivity (TFP) is indirectly estimated on the basis of cyclical factors. Instead, this article proposes a more economics-based foundation of trend-TFP that is based on the arguments of the endogenous growth theory. A simple econometric exercise shows how the application of this approach may lead to a significantly different assessment of the potential output and, by extension, of the economy's output gap. This may have far-reaching consequences for the conduct of monetary and fiscal policy, e. g. in the context of the German debt brake.*



# Robust Real-Time Estimates of the German Output Gap based on a Multivariate Trend-Cycle Decomposition\*

Tino Berger<sup>†1</sup> and Christian Ochsner<sup>‡2</sup>

<sup>1</sup>University of Göttingen, Chair of Empirical International Economics

<sup>2</sup>German Council of Economic Experts

August 2, 2022

## Abstract

The German economy is an important economic driver in the Euro-area in terms of gross domestic product, labour force and international integration. We provide a state of the art estimate of the German output gap between 1995 and 2022 and present a nowcasting scheme that accurately predicts the German output gap up to three months prior to a gross domestic product data release. To this end, we elicit a mixed-frequency vector-autoregressive model in the spirit of [Berger, Morley, and Wong \(forthcoming\)](#) who propose to use monthly information to form an expectation about the current-quarter output gap. The mean absolute error of our nowcast compared to the final estimate is very small (0.28 percentage points) after only one month of observed data. Moreover, we show that business and consumer expectations, international trade and labour market aggregates consistently explain large shares of variation in the German output gap. Finally, our procedure is very reliable, as it implies an output gap that is hardly revised ex post. This is particularly important for policymakers.

**Keywords:** output gap, Germany, nowcast, mixed frequency, vector-autoregression

**JEL classification:** E32, E37, C53

---

\*We thank David Boll, Lasse Trienens and Lieve Vanhooren for research assistance. We are grateful to Elmar Mertens, Julia Richter and Maik Wolters for great discussions. Moreover, we thank workshop participants at University of Göttingen, Deutsche Bundesbank, German Council of Economic Experts, the 15<sup>th</sup> Ruhr Graduate School Doctoral Conference in Economics, the 30<sup>th</sup> Annual Symposium of the Society for Non-Linear Dynamics and Econometrics and at the 8<sup>th</sup> Annual Meeting of the International Association for Applied Econometrics for helpful comments. We acknowledge financial support from Deutsche Bundesbank. The usual disclaimers apply. In particular, the views expressed in the article at hand represent the authors' personal opinions and do not necessarily reflect the views of the German Council of Economic Experts, Deutsche Bundesbank or the Eurosystem. Note that the draft was written before Christian Ochsner joined the the scientific staff of the German Council of Economic Experts.

<sup>†</sup>Email: [tino.berger@wiwi.uni-goettingen.de](mailto:tino.berger@wiwi.uni-goettingen.de);

<sup>‡</sup>Email: [christian.ochsner@uni-goettingen.de](mailto:christian.ochsner@uni-goettingen.de); corresponding author.

# 1 Introduction

The German economy is the fifth largest economy in the world and with a GDP weight of about 28% in 2021 the single largest economy in the Euro area. Therefore, the German economic situation is of key importance for economic flourishing and stability in the European Union. For the conduct of monetary policy in the Euro area, the European Central Bank relies on measurements of a Euro area business cycle. The latter is driven to a large extent by the German business cycle, the focus of the paper at hand.

Central banks and other policy institutions frequently use the output gap, i.e. the deviation of actual from potential real gross domestic product (GDP) as a measure of the business cycle. Being intrinsically unobserved, the output gap has to be estimated. Numerous models and filtering approaches have been proposed to this end (most prominently, [Hodrick and Prescott \(1997\)](#) and more recently [Hamilton \(2018\)](#)). Moreover, estimating the output gap with the help of production-functions is wide-spread among policymakers across the globe. This approach is appealing, as it provides information on the structural determinants of the output gap. However, the vast majority of available procedures only yield retrospective insights into the output gap due to a significant delay in GDP data availability. As a recent exception, [Berger, Morley, and Wong \(forthcoming\)](#) ([BMW \(forthcoming\)](#), henceforth) propose to nowcast the output gap using a multivariate Beveridge-Nelson decomposition based on a mixed-frequency Bayesian vector-autoregressive model (VAR). They show that a model comprising economic aggregates available at monthly frequency implies a reasonable output gap for the U.S. economy well in advance current quarter GDP data is released.

We estimate the German output gap by means of the output gap using the approach of [BMW \(forthcoming\)](#). Going beyond their work, we present a reliable procedure to select the most relevant variables for estimating the German output gap in a multivariate model. Using an informational decomposition allows us to quantify the relative importance of each variable. Moreover, we analyze the contribution of each variable in each month within a quarter. The accuracy of the models nowcasts are evaluated by comparing the nowcast after each month of a given quarter to the final estimate obtained using the full information set. In addition, we extensively discuss the role of data, parameter and specification revisions. In particular, we present a detailed analysis in the spirit of [Orphanides and van Norden \(2002\)](#) to demonstrate the reliability of our approach.

The contribution is threefold. First, we provide a state-of-the-art estimate of the German output gap. German output gap fluctuations are monitored by the German council of economic experts ('Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung') at the behest of the federal government. As part of its mandate, the council presents a comprehensive summary of the main economic developments. The biannual expert reports make an important contribution to understanding the German economy. To approximate the output gap, the council estimates a production function. However, academic accounts of the German output gap are sparse. We reconcile traditional production-function identification of the German output gap with the Beveridge-Nelson decomposition as a tool to disentangle trends and cycles. The Beveridge-Nelson decomposition is particularly appealing, as it provides a straightforward characterization of the cyclical component as well as the possibility to interpret the output gap in terms of the aggregates that inform cycle estimation. Second, by tracking the German output

gap up to three months prior to GDP data release, we provide timely information about the stance of the largest economy in the Euro area that are crucial for the conduct of monetary and fiscal policy. The impact of large economic shocks, such as the COVID-19 pandemic or recent oil-market disturbances can be straightforwardly tracked within a month within a given quarter, rather than six weeks after the previous quarter. Finally, we contribute to the literature on output gap estimation in data-rich environments. It is well known that univariate approaches suffer from unreliability in real-time (see [Orphanides and van Norden \(2002\)](#)). On the contrary, recently proposed multivariate models of the output gap (e.g. [Jarocinski and Lenza \(2018\)](#), [Morley and Wong \(2020\)](#) and [Barigozzi and Luciani \(2021\)](#)) take advantage of a large set of information which improves not only economic interpretability, but also real-time forecasting performance. In the vein of this literature, we show that our model is reliable: The nowcasting scheme is very accurate and the output gap estimate is hardly revised ex post.

Our estimate of the output gap lines up reasonably well with established filter and production-function based measures. Furthermore, we find that fluctuations in the German output gap are mainly transmitted by expectations. Moreover, in line with the related literature we show that external relations play a key role for the German output gap (see e.g. [Eickmeier \(2007\)](#)), whereas labour market aggregates are informative in times of large deviations from potential output. Regarding the nowcast accuracy, we find that after the first month of a quarter, the nowcast has a mean absolute error of 0.28 percentage points from the final estimate of the output gap. Even in times of substantial volatility, our model yields robust results under real-world conditions without observing current quarter GDP data. For instance, the output gap in the COVID-19 induced recession in 2020Q2 was nowcasted to be  $-8.0\%$  after the release of the May 2020 monthly data. It turned out to be  $-8.8\%$  after the release of the entire quarter data (including revised GDP) at the end of August 2020.

Sections 2 and 3 present our empirical approach and the data, respectively. In Sections 4, 5 and 6, we discuss our the estimate and the nowcasting performance in real-time. Section 7 examines robustness of our results to an even-larger information set and Section 8 concludes.

## 2 Methodology

We estimate a model in the spirit of [BMW \(forthcoming\)](#), who propose a mixed-frequency Bayesian vector-autoregression (MF-BVAR) to obtain the output gap  $c_t$  as the cyclical component of output from a multivariate Beveridge-Nelson decomposition. Using the Beveridge-Nelson decomposition in a multivariate setting enables us to interpret the resulting estimate in terms of the information that is most relevant for disentangling trend (i.e. potential output) and cycle (i.e. output gap). More precisely, if  $y_t$  is a  $K \times 1$  vector of macroeconomic observables with  $K \times 1$  drift component  $\mu$ , its Beveridge-Nelson trend  $\tau_t$  is given as  $\lim_{h \rightarrow \infty} \mathbb{E}_t(y_{t+h} - h\mu)$  and the cycle can be obtained as  $c_t = y_t - \tau_t$ . Assuming stationarity and zero mean, [Morley and Wong \(2020\)](#) show that the cycle  $c_t$  is given by

$$c_t = -\mathbf{F}(\mathbf{I}_K - \mathbf{F})^{-1} \mathbf{X}_t, \quad (1)$$

where  $\mathbf{I}_K$  is the identity matrix of rank  $K$  and the remaining quantities derive from the vector-autoregressive model  $\mathbf{X}_t = \mathbf{F}\mathbf{X}_{t-1} + \mathbf{H}\epsilon$  with  $\epsilon \sim \mathcal{N}(0, \Sigma)$  (BMW forthcoming). In particular,  $\mathbf{X}_t$  and  $\epsilon$  are  $K \times 1$  vectors of macroeconomic observables and innovations in  $t$ , respectively.  $\mathbf{F}$  is a companion-form coefficient matrix and  $\mathbf{H}$  maps the innovations into companion form. We stack all high-frequency (HF) variables above the lower frequency (LW) variables. For instance, assume one low-frequency period can be subdivided into  $d = 1, \dots, D$  equidistant high-frequency-periods, then we obtain  $\mathbf{X}_t$  as

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{x}_{t-d-1/D}^{HF} \\ \vdots \\ \mathbf{x}_{t-1/D}^{HF} \\ \mathbf{x}_t^{HF} \\ \mathbf{x}_t^{LF} \end{bmatrix},$$

where  $\mathbf{x}_t$  indicate partitions of  $\mathbf{X}_t$  for the high-frequency series in  $d = 1, \dots, D$  in  $t$ . To contain parameter proliferation, we follow Morley and Wong (2020) in adopting Bayesian shrinkage. We use a standard Minnesota prior for location and scale parameters, where the Minnesota-prior shrinkage parameter  $\lambda$  is chosen such that the one-step-ahead root mean squared error of output is minimized.<sup>1</sup>

The higher (in our case, monthly) frequency information is exploited in the spirit of Waggoner and Zha (1999) to update the vector-autoregression for the subsequent period. To this end, we note that (by positive-definiteness) the innovation covariance  $\Sigma$  obeys a representation  $\Sigma = \mathbf{B}\mathbf{B}'$ , where  $\mathbf{B}$  is the lower-triangular Choleski factor (alternative decompositions might be employed).  $\mathbf{B}$ , by virtue of its triangular structure, is used as the contemporaneous impact multiplier for the new high-frequency information. Put differently, by pre-multiplying the relevant parameters in  $\mathbf{B}$  to the observed high-frequency shocks, we can track their propagation through the system in time  $t$  during  $D$  high-frequency periods. More precisely, we observe the upper partitions of  $\epsilon_{T+1}$  and use  $\mathbf{B}$  to form an expectation about future innovations during the *entire* low-frequency period ahead. Hence, the expected innovations are non-zero *conditionally* on information observed in a given high-frequency interval. By means of the subsequent evaluation of the vector-autoregression, we obtain a forecast of  $T + 1$  for the entire system. Taking advantage of this technique, BMW (forthcoming) elicit a nowcast of the output gap by iterating on Eq. (1):

$$c_{t+1} = -\mathbf{s}_k \mathbf{F} (\mathbf{I} - \mathbf{F})^{-1} [\mathbf{F}\mathbf{X}_t + \mathbf{H}\epsilon_{t+1}],$$

where  $\mathbf{s}$  is a selection column vector. Finally, we can trace out variation in the cycle  $c_{ij,t}$  of the higher-frequency series  $i$  to surprises in variable  $j$  in time  $t$  by means of

$$c_{ij,t} = - \sum_{d=1}^D \sum_{l=0}^{t-1} \mathbf{s}_k \mathbf{F}^{l+1} (\mathbf{I} - \mathbf{F})^{-1} \mathbf{H} \mathbf{s}'_j \mathbf{s}_j \epsilon_{t-1}. \quad (2)$$

We emphasize that this ‘informational decomposition’ (Morley and Wong 2020) is not

---

<sup>1</sup>In our final specification, we obtain  $\lambda = 0.18$ . In order to obtain a stable shrinkage parameter, we estimate  $\lambda$  only with data until 2019Q4.

structural as the innovations  $\varepsilon$  are not necessarily orthogonal and economically interpretable. However, even though Eq. 2 does not permit a causal interpretation, it is a convenient instrument to shed light on the *transmitters* of variation in the output gap.

### 3 Data and model selection

Currently, the output gap literature is evolving from univariate and filter-based approaches to estimation in data-rich environments. For instance, [Morley and Wong \(2020\)](#) give an account of the US output gap using a similar approach as the paper at hand. Moreover, [Jarocinski and Lenza \(2018\)](#) and [Barigozzi and Luciani \(2021\)](#) estimate the output gap in a big data environment. While multivariate models have the advantage that more information can be analysed, they may be subject to over-fitting. Thus, we face a trade-off between eliciting a model that exploits all relevant information and a parsimonious specification.

Our variable selection procedure seeks to reconcile traditional economic approaches to estimating potential output with statistical information on variable relevance. To this end, we estimate a medium scale model including economic aggregates in the spirit of the production function approach to potential output estimation. That is, we assume output growth  $y_t$  (and thus, implicitly, the output gap) obeys a linearized Cobb-Douglas regime of the form

$$y_t = \alpha k_t + (1 - \alpha)l_t + a_t,$$

where  $a_t$  is the Solow residual,  $k_t$  is capital formation,  $l_t$  denotes labour inputs and  $\alpha$  is the substitution elasticity of capital. As the Solow residual cannot be subjected to direct analysis, we treat it as stochastic innovation. However, we can approximate capital  $k_t$  and labour  $l_t$  inputs by means of observable economic aggregates. Our choice of candidate variables is inspired by the literature on production function estimation. In particular, the included labour market aggregates largely derive from the EU commission’s procedure on potential output estimation ([Havik, McMorrow, Orlandi, Planas, Raciborski, Roeger, Rossi, Thum-Thysena, and Vandermeulen 2014](#)). For approximating innovations to the capital stock, we propose to use a larger set of economic indicators broadly related to capital formation (e.g. investment and industrial production) and its costs (e.g. exchange and interest rates). Thus, we include various sectors of the real economy, the German labour market, external relations and financial markets. Subsequently, we estimate a candidate model and reduce the number of variables in accordance with a statistical criterion.

Table 1 depicts candidate variables, transformations and sampling frequencies. We sample data for the period of 1995Q1 until 2022Q1 in monthly frequency. If an indicator is not available on a monthly basis, we obtain it at quarterly frequency. The time period is constrained by data availability and by the German reunification (1990) which possibly caused a business cycle regime change that we omit from the model for our purposes. If not stated otherwise in Table 1, we obtain data from the Deutsche Bundesbank database. We partition our data-set into five variable blocks, which are ordered as shown above. All monthly series are ordered before all quarterly series. We emphasize that model invertibility is indispensable for our purposes (see Eq. 1). Thus, we apply convenient transformations to secure stationarity of each time series.

Variable	Transformation	Frequency
CAPITAL		
<i>External Relations</i>		
CPB World Trade Monitor: World Trade Volume	rolling demean	M
Current Account: Exports	growth rates	M
Capital account Balance: Portfolio Investment	growth rates	M
Capital Account Balance: Direct Investment	growth rates	M
Real Effective Exchange Rate of the Euro against EERK-42	rolling demean	M
<i>Finance</i>		
Interbank Rate for Germany (obtained from FRED database)	growth rates	M
Total Share Prices for All Shares for Germany (obtained from FRED)	rolling demean	M
Term spread (1 year over 10 year government bonds)	growth rates	M
Non-financial private sector credit (obtained from BIS)	growth rates	Q
<i>Fiscal Activity</i>		
Government Consumption	growth rates	Q
Government Investment	growth rates	Q
<i>Sentiment and Expectations</i>		
OECD Consumer Opinion Surveys (obtained from FRED)	growth rates	M
ifo Business Climate Index (obtained from ifo Institute)	growth rates	M
ifo Business Expectations Index (obtained from ifo Institute)	growth rates	M
<i>Real Economy</i>		
Consumer Price Index	growth rates	M
Construction Permits	growth rates	Q
Industrial Production	growth rates	M
New Orders of Consumption Goods	growth rates	Q
New Orders of Investment Goods	growth rates	Q
New Orders of Input Goods	growth rates	Q
New Orders of Industrial Goods	growth rates	Q
Real Gross Domestic Product	growth rates	Q
Resource Price Index (excl. Energy)	growth rates	Q
Resource Price Index (only Energy)	growth rates	Q
LABOUR		
Hours in Construction	growth rates	M
Labour Compensation Index	growth rates	Q
Labour Unit Costs	growth rates	Q
Labour Market Stabilization Policy ('Kurzarbeit' policy)	growth rates	M
Unemployment Rate	growth rates	M
Working Population	growth rates	M

Table 1: Variable blocks and data transformations. ‘Growth rates’ denotes the transformation  $100 \times$  first differences of natural logarithms. ‘rolling demean’ denotes a rolling demean (backward moving average) filter with a 40-quarters window. ‘M’ and ‘Q’ denote monthly and quarterly frequency, respectively.

Parsimony is the second most important priority after economic plausibility. To reduce the size of the model in the interest of parsimony, we proceed as follows. Subsequent to estimation of the model implied by Table 1, we compute the standard deviations of the

informational decomposition contributions (see Eq. 2) to approximate the explanatory relevance of the variables for the output gap (following Morley and Wong (2020)). Higher standard deviations imply higher relevance of a given economic aggregate. Moreover, model selection is a reoccurring problem, as the model requires re-estimation after each quarterly period. Thus, it seems reasonable to stick to specifications that consistently implied decent estimates in the past. Therefore, we estimate the full model as shown in Table 1 for every quarter since 2008Q4. We then average over the time dimension of the normalized standard deviations of the contributions to explained variation for each variable.<sup>2</sup> Thereafter, we estimate a model that includes all variables with an average, normalized standard deviation of explained contribution to variance equal to or larger than that of GDP growth (i.e. unity). Thus, we drop all variables from the model with relatively low explanatory power.

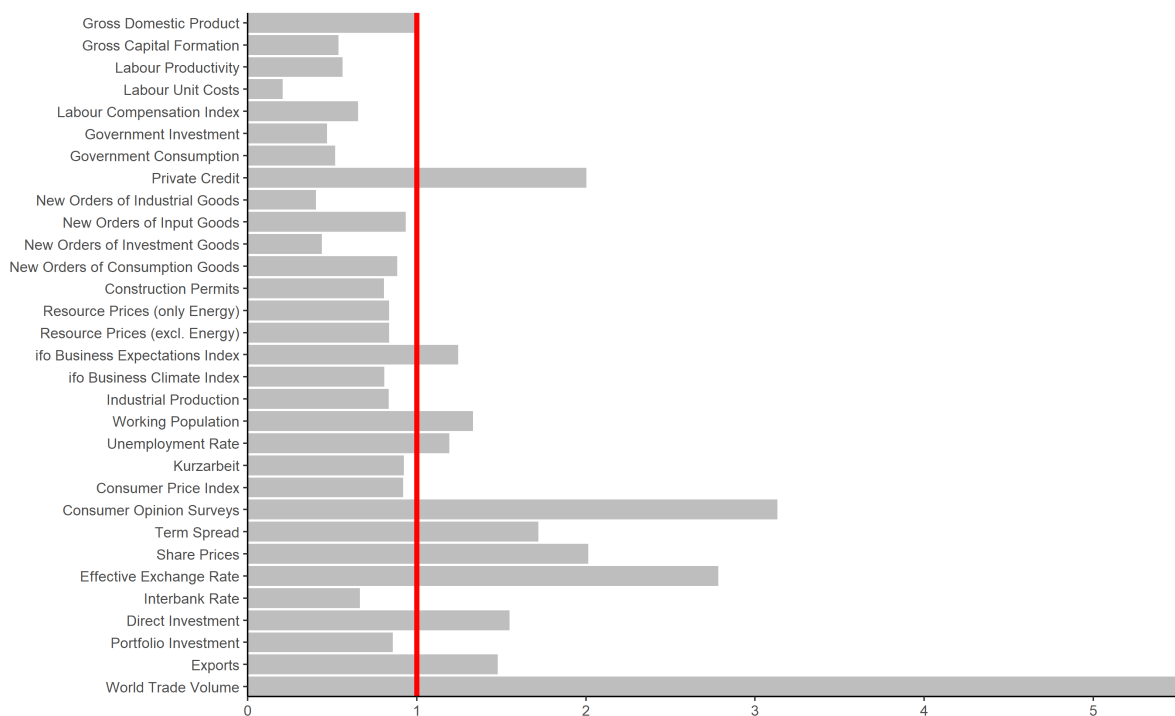


Figure 1: Normalized time-averages of standard deviations of informational decomposition. The red line indicates the standard deviation of the GDP growth contribution to output gap variation.

## 4 The German output gap

We now turn discussing the economic properties of the proposed output gap. In a first step, we discuss our estimate in more detail and assess its plausibility. Subsequently, we examine the contributions of the individual variables to informing our estimate in a reduced-form framework.

<sup>2</sup>We normalize by means of the standard deviation of GDP growth.

## 4.1 Estimate of the output gap

Figure 2 depicts the German output gap for 1995Q1 until 2022Q1.

As can be seen, the German economy suffered from substantial slack at the prior to the beginning of the millennium. This was likely due to high adjustment costs resulting from German reunification and the global millennium recession. Subsequent to the millennium recession and after the Dotcom-bubble burst in 2002, we observe strong overheating (+4%) prior to the financial crisis (2005 – 2008). Unsurprisingly, the European sovereign debt and banking crisis (2010 – 2015) coincides with sluggish mean-reversion tendencies of the output gap which fluctuated slightly below zero at the time.

As a result of the COVID-19 shock the output gap dropped from the pre-pandemic level of 1.17 to  $-8.78\%$  at the end of 2020Q2. Thus the  $-9.95\%$  decline of the output gap is similar to the decline of German GDP growth ( $-10\%$ ) in 2020Q2. This implies that the COVID-19 shock only marginally affected potential output but is accounted for by a massive decline in the output gap. Interestingly, even the second and third ‘lock-down’ episodes (2020M11 – 2021M3 and 2021M4 – 2021M5) exerted – relatively – small contractionary pressures of about  $-1.5\%$  on the German output gap. In the second half of 2021, the German output gap is still well below the zero mean while we see massive overheating at the beginning of 2022. The latter is most likely due to the Russian attack on Ukraine and the encompassing pressures on global energy, food and material supplies as well as due to inflationary pressures.

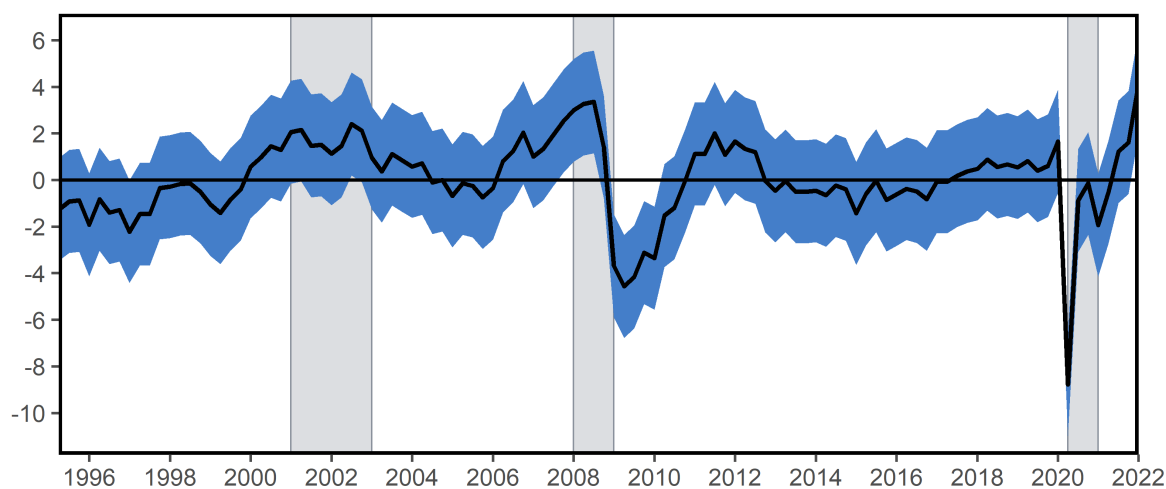


Figure 2: German output gap in percentage deviation from potential output, 1995Q1 - 2022Q1 (black). Blue and orange shaded areas indicate 90% credible sets following [Kamber et al. \(2018\)](#) for the mean estimate and the nowcast, respectively. Grey areas indicate recessions according to the German council of economic experts and the COVID-19 pandemic.



## 4.2 Is our output gap estimate plausible?

We turn to discussing the plausibility of our results. To this end, Figure 3 depicts the comparison of our output gap estimate (black) with the GDP-based output gaps implied by the Hodrick and Prescott (1997) filter (green) with the smoothing parameter set to 1600 (as is common for quarterly data) and the Hamilton (2018) filter (orange) with  $p = 4$  lags. Moreover, Figure 3 depicts a comparison of our estimate (black) to the official output gaps estimated by the German council of economic experts (dashed red) and AMECO (dashed blue). The latter two estimates are obtained from models that use a production function approach to approximate potential output. Both are only available at yearly frequency.

First of all, we note that all estimates are reasonable similar during most of the sample period. Our estimate differs in magnitude compared to the Hamilton regression filter during and prior to the two large recessions (2008 and 2020), as do all alternative estimates. In particular, we note that the magnitudes for the Hamilton-filtered estimate appear relatively large. For instance during 2016 – 2018, the Hamilton filter indicates an overheating almost as substantial as prior to the Great Recession in 2008. Given that no economic narrative is available to support this conjecture, this seems surprising. Thus, the Hamilton filter seems to yield an ‘outer bound’ estimate of the German business cycle.<sup>3</sup>

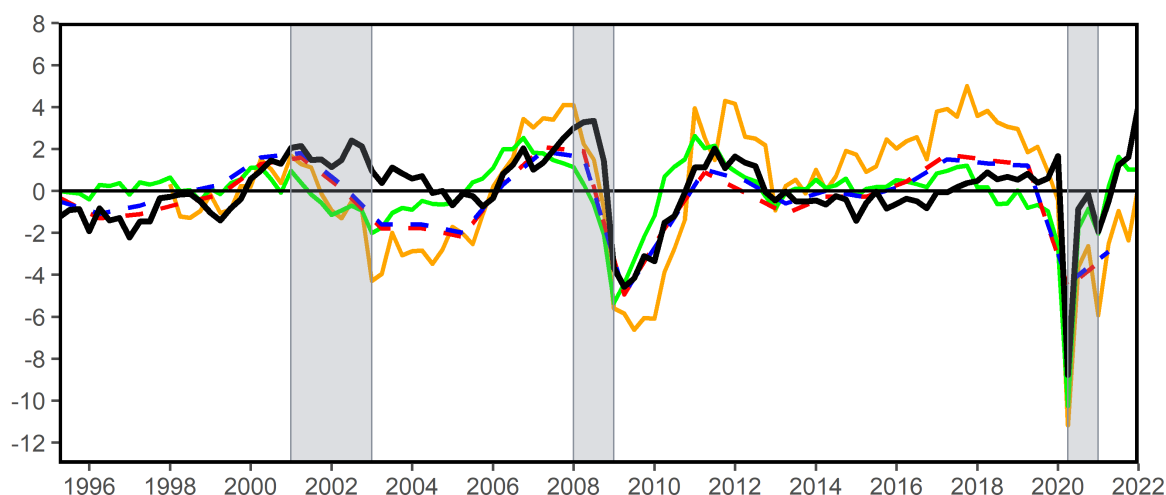


Figure 3: Comparison of our output gap measure (black) with established alternatives: One-sided Hodrick and Prescott (1997) filter (green) and Hamilton (2018) filter (orange). All estimates are reported in growth rates. Red and blue dashed lines are the (yearly-frequency) estimates by the German council of economic expert and AMECO, respectively. For further notes see Figure 2.

Moreover, we observe that the HP-filtered estimates have a significant tendency to indicate recessions and more subtle economic downturns earlier (about one year, generally) than alternative estimates. This is exemplified prior to the financial crisis in 2008

<sup>3</sup>Quast and Wolters (2022) offer a convincing explanation for the possibly spurious dynamics implied by the baseline Hamilton filter.

and even more in advance of the COVID-19 health crisis. In case of the latter, the HP-filter produces results that indicate mean-reversal in mid 2017, which seems debatable (at least with respect to the shown magnitudes) in absence of a plausible economic narrative. For instance, German GDP grew between 2017 and 2018 about 0.5% to 1% almost each quarter compared to the previous quarter (except 2018Q1 and 2018Q3 with about  $-0.4\%$  each). Given the HP-filtered negative growth rates of the output gap, this would imply substantial and implausibly high growth rates for potential GDP in 2017 and 2018. [Morley and Wong \(2020\)](#) suggest that a decent output gap estimate should be correlated positively with future inflation and negatively with future output growth. Thus, we compute correlations between the output gap estimate and the future quarter-on-quarter growth rates of output and consumer price index. We find that for the Hamilton filter, results are inconclusive (Pearson correlation coefficients of  $-0.11$  for inflation and  $-0.26$  for output growth). The coefficients of the HP-filtered estimate line up a little better with the economic expectation (coefficients of  $0.09$  and  $-0.16$  with inflation and output growth, respectively). Our model compares well to the HP-filter (correlations of  $0.01$  and  $-0.50$  with inflation and output growth, respectively). Summing up, we conclude that our estimate is economically at least as plausible as the output gaps obtained from filtering GDP by means of the procedures proposed in [Hamilton \(2018\)](#) and [Hodrick and Prescott \(1997\)](#).

Moreover, we compare our proposed output gap to production-function based estimates. Our estimate implies about the same overheating tendencies prior to the financial crisis (2007) and prior to the COVID-19 shock (2019) as do both output gaps by AMECO (blue dashed) and the German council of economic experts (red dashed). Furthermore, all three models indicate sluggish regression to the mean in 2020 and 2021 at about the same pace and to about the same levels. We take this as evidence that our model yields reasonably similar approximations to production-function based approaches. However, note that our model indicates slightly more overheating during and less slack before the financial crisis of 2008. We conjecture that both differences are partly be explained with reference to the underlying conceptions of the output gap. Whereas our estimate includes information on financial markets, this is not incorporated in the production function approaches. This aspect likely explains the positive output gap during the 2001-2003 recession. The aforementioned alternative output gaps focus on real economic activity (without considering financial transactions and imbalances), whereas our estimate is best understood as a real indicator that takes into account *all* economic activity in Germany, including finance. Nevertheless, incorporating financial information in the course of estimating the output gap is important (as pointed out by [Borio, Disyatat, and Juselius \(2013\)](#) and [Berger, Richter, and Wong \(2022\)](#)) when it comes to judging the sustainability of output growth, e.g. due to financial imbalances. Overall, we are confident that our model yields a plausible estimate for the German output gap.

### 4.3 Informational decomposition

The results described above raise a number of questions regarding the key determinants of German business cycle fluctuations. The German output gap is much less researched than, say, the United States output gap. Therefore, even non-structural information is valuable to understand German business cycle fluctuations. Subsequently, we aim to contribute

towards closing this research gap. Figure 4 shows the informational decomposition of the German output gap.

Although we refrain from drawing causal conclusions, Figure 4 yields interesting insights into the reduced-form contributions of our five variable blocks to business cycle variation. The capital inputs block accounts for roughly 86% of the variation (expectations 24%; external relations 35%; the real economy, i.e. GDP, 5%; financial 22%), whereas the labour block explains about 14% of the variation in the German output gap. The high relevance of the international aggregates does not come as a surprise, as the (very open) German economy is shaped by its external relations (Eickmeier 2007). In fact, Figure 4 unambiguously shows that international economic aggregates shape the output gap dominantly throughout the entire sample period. That is, their importance does not seem to be regime-dependent.

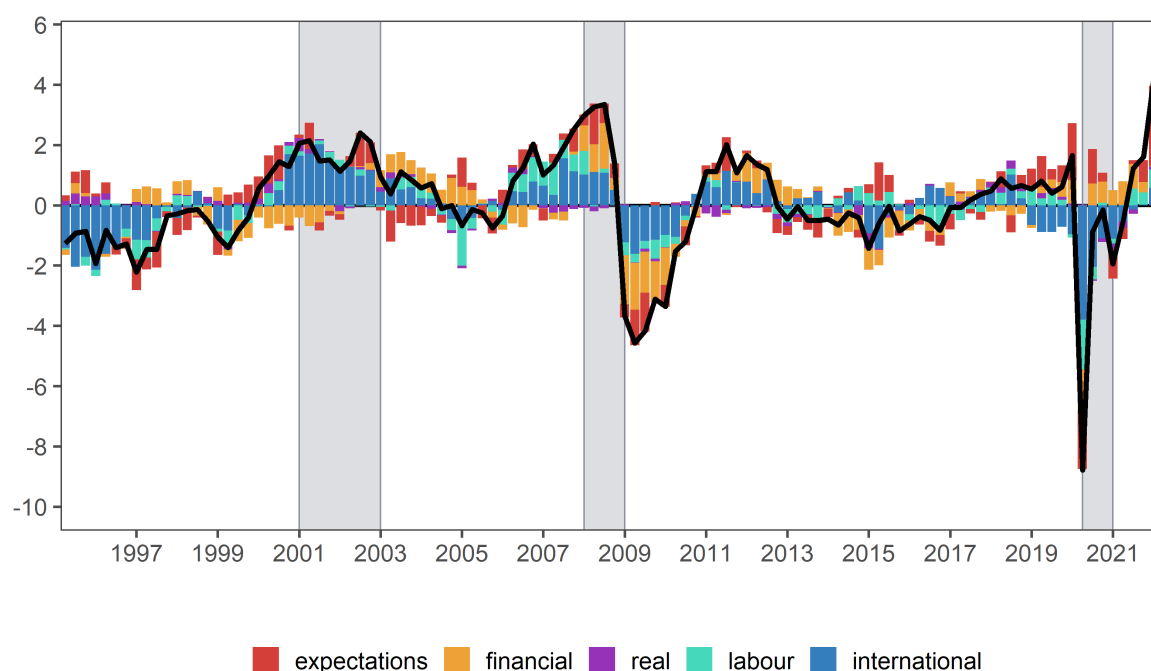


Figure 4: Informational decomposition of the German output gap. The financial block comprises private credit, share prices, the term spread, the real economy block summarizes contributions from gross domestic product, the labour block is made up of unemployment and the working population, the expectations block comprises consumer sentiment and business expectations and the international block contains the direct investment, the exchange rate, exports and world trade. For further notes see Figure 2.

Financial market information is mostly relevant in times of large overheating (2006-07) or substantial economic slack (e.g. during the late European banking and sovereign debt crisis in 2012-15), less so when the output gap reverts to its mean. Similar results for the US have been reported by Berger, Richter, and Wong (2022). However, this does not imply that the underlying *structural* real or financial shocks are irrelevant for the German business cycle, as the informational decomposition remains silent on this issue. Moreover, expectations explain large amounts of variation in the output gap. In

particular, the strong increase in 2022Q1 can largely be explained by variation that is transmitted via expectations. Furthermore, we observe that labour market innovations contribute large amounts of variation in times of large and spontaneous contraction (e.g. 2020Q2). However, we emphasize that this interpretation does point to labour market shocks in a structural sense.<sup>4</sup>

To assess the cyclicity of the variable blocks' contributions to the German output gap more broadly, we compute correlations of the contributions to explained variation and the estimated output gap. The results are summarized in Table 2. Any sensible output gap estimate should be correlated positively on average with its explained variation by both capital and labour inputs, as displayed in Figure 4. As we see in Table 2, this is the case. In fact, the correlation coefficients are relatively large. In particular, the shares of variation in the output gap explained by the labour market and international as well as financial aggregates behave (strongly) pro-cyclically. In the short-run this also holds for expectations. Moreover, if expansionary shocks are transmitted via these variable blocks, the next-quarter output gap will increase on average. Vice versa, as we cannot infer to causal chains at this point: As the output gap widens (closes), larger (smaller) shocks are transmitted by expectations as well as international, financial and labour aggregates. The variation explained by GDP itself is hardly informative for the current or near-future output gap.

	CAPITAL				LABOUR
	<i>Expectations</i>	<i>External Relations</i>	<i>Finance</i>	<i>Real economy</i>	
$c_t$	0.61*	0.78*	0.47*	-0.01	0.64*
$c_{t+1}$	0.19*	0.58*	0.22*	-0.04	0.44*
$c_{t+2}$	0.12	0.48*	0.07	-0.09	0.35*
$c_{t+3}$	0.07	0.40*	-0.03	-0.09	0.27*
$c_{t+4}$	-0.08	0.27*	-0.20*	-0.05	0.14

Table 2: Pearson correlations between average contributions to explained variation with the current and next-four-quarters German output gaps  $c_t$  as well as  $c_{t+1}$ ,  $c_{t+2}$ ,  $c_{t+3}$  and  $c_{t+4}$ . \* marks correlation coefficients in excess of  $2/\sqrt{T}$ , which roughly corresponds to a significance level of 95%.

Interestingly, shock transmission from the international block exhibits quite substantial correlations with the four-quarters-ahead output gap. This finding is unique to international aggregates. For example, shock transmission from the labour market (which is associated with the largest contemporary and one-quarter-ahead correlations) are not too informative about the output gap beyond three horizons. We interpret this finding to point to the special relevance – and potential vulnerability – of the German economy to shocks transmitted by international aggregates. With regard to expectations, finance and the real economy, correlations at farther horizons confirm our previous conclusions.

<sup>4</sup>We present a structural historical decomposition on Figure 10 in Appendix A.

## 5 Nowcasting performance in the baseline model

In times of economic disruption, a fast economic policy response is asked for. Waiting after GDP data has been released one month after the end of a quarter can be economically costly. Therefore, a timely estimate of the output gap is needed. We assess the nowcasting abilities of our approach. We proceed by analysing the nowcasting qualities of our model rigorously.

Table 3 shows the mean absolute forecast errors (MAE) for our model and given monthly indicators (first row for each indicator) compared to the end-of-quarter output gap estimate for the baseline model. Moreover, in order to assess the relevance of the individual variables, we compare two forecasts by means of the Diebold-Mariano procedure (Diebold and Mariano 1995): one forecast obtained from a unrestricted model with all variables and another forecast from a restricted model with the monthly indicator of interest omitted. More precisely, we test whether the forecast of the latter is superior in terms of mean absolute error than the forecast of the former model.

	1 <sup>st</sup> month	2 <sup>nd</sup> month	3 <sup>rd</sup> month
Share prices	0.39	0.28	0.12**
Term spread	0.38	0.26*	0.11**
Consumer opinion surveys	0.34**	0.16**	0.06**
Business expectations	0.28	0.15	0.04
Unemployment rate	0.31*	0.15**	0.05**
Working population	0.29*	0.15	0.05
World trade volume	0.43**	0.28	0.15
Exports	0.41	0.28	0.15
Direct investment	0.40	0.28	0.14
Effective exchange rate	0.40	0.27	0.14

Table 3: Within-quarter mean absolute forecast errors associated with monthly variables rounded to two decimal places for all four models, given the full-sample parameters. \*\* and \* indicate Diebold-Mariano p-values (based on mean absolute error) equal to or smaller than 0.05 and 0.10, respectively. Variables are ordered by expected release.

From Table 3 we see that the mean absolute forecast error is moderate throughout compared to an unconditional forecast with within-quarter information. The unconditional forecast implies a mean absolute forecast error of 0.51. Effectively, the additional information incorporated throughout the course of the quarter cuts the nowcast error almost in half (from 0.51 to 0.28 percentage points). At the end of the third month, the mean absolute forecast error is negligible (0.04). We emphasize that this still is several weeks prior to a GDP data release in the subsequent quarter. Observing the world trade volume, the working population, unemployment and consumer opinion help to improve forecast precision in the first month. Subsequently, consumer opinion, the term spread and the unemployment rate help to improve the forecast in the second month, whereas

consumer opinion, share prices, the term spread and the unemployment rate do so in the third month of a given quarter.

Following [BMW \(forthcoming\)](#), Table 4 depicts correlations of the within-quarter output gap nowcasts (unconditionally and after a given month) and the final estimate (left block) as well as model-implied and realized output growth (right block).

	Output gap	Output growth
No information	0.92	0.91
First month	0.97	0.94
Second month	0.99	0.96
Third month	1.00	0.96

Table 4: Correlations of the within-quarter nowcasts with the final estimate, model-implied and real-time output growth

As can be seen, the model benefits from the high degree of persistence in the output gap (see first row of Table 4), but our estimate substantially improves upon an unconditional forecast after only a single month. Observing GDP hardly increases the correlations after seeing data for three complete months. This picture hardly changes when we consider correlations between model-implied and realized output growth: Our specification implies plausible output growth rates and observing GDP data after three month adds only little information.

## 6 Real-time reliability

In this Section, we analyze the real-time performance of our model. As [Orphanides and van Norden \(2002\)](#) point out, real-time estimates of the output gap are chronically unreliable. We extensively discuss whether our baseline model is subject to this charge. We begin with assessing our model’s nowcasting abilities in a real-world setting (i.e. including model selection). Subsequently, we investigate estimation revisions and decompose the ex post output gap revision into effects resulting from quarter-on quarter changes in model selection (i.e. different variables), parameter revisions (i.e. different sample periods for the baseline specification) and data revisions<sup>5</sup> (for a given specification and given parameters).

### 6.1 Nowcasting in real-time

The COVID-19 health crisis has brought about the most devastating economic shock since World War II to many advanced economies across the world. Germany is no exception. According to our estimates, the German output gap was at historical low (-8.8%). In this situation it is crucial for policy makers to obtain real-time insight into the state of the economy to adjust or maintain stabilization measures. In times of such large (exogenous)

<sup>5</sup>Unfortunately, we cannot obtain real-time data for the majority of the monthly indicators from Deutsche Bundesbank. Thus, we limit ourselves to investigating the effects of GDP data revisions, which are available since 2005.

shocks neither the unconditional forecast nor the ex post estimate are particularly useful for the conduct of stabilization policy. The former can – by construction – not forecast large shocks, and the latter is obtained far too late for a timely policy response. The mixed frequency sampling of our approach allows us to nowcast the output gap in a timely matter.

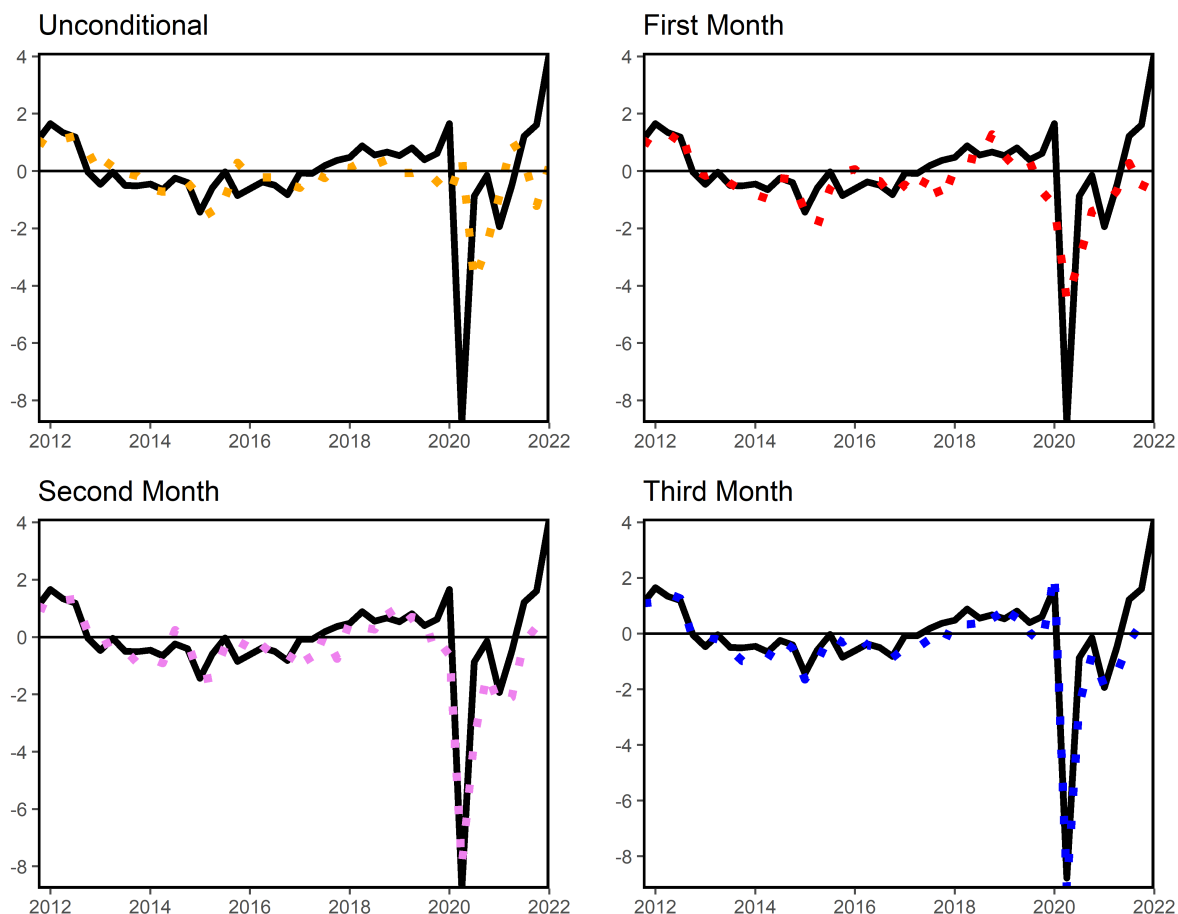


Figure 5: Nowcasts since 2012. Ex post estimate (black), unconditional forecast (orange, upper left panel) and nowcasts from the perspective of month 1 (red, upper right panel), month 2 (pink, lower left panel), month 3 (blue, lower right panel). Vintage GDP data is used, parameters are estimated and the specification is chosen in real-time (i.e. every quarter).

Figure 5 compares the unconditional forecast, the nowcasts after one month, two months, three months, and the final estimate of the output gap since 2012 (coinciding with the European banking and sovereign debt crisis) in real-time. That is, we estimate the model using vintage GDP data, automatically update the specification (i.e. the set of included aggregates) and re-estimate the parameters in every quarter. The orange dotted line (upper left panel) is the unconditional forecast after observing all data until the previous quarter. As can be seen, it is mean-reverting – and for large shocks (e.g. COVID-19) distant from the final estimate. However, considering the case of 2020Q2, once data for April 2020 is brought in (red dotted line, upper right panel), we note that the nowcast improves substantially. Once the nowcast incorporates data for May 2020, the difference

is even smaller and with all monthly, but no quarterly information (blue dotted line), the real-time estimate is  $-9.2\%$  and very close to the final (ex post, i.e. full sample) estimate ( $-8.8\%$ ). Thus the model provided a reasonable quantification of the COVID-19 shock after observing all 2020Q2 monthly data, without seeing private credit nor GDP. These conclusions generalize to the entire sample period under investigation with the interesting exception of 2022Q1. We conjecture that this behaviour occurs due to the relatively short sample period. Put differently, the model has been specified and estimated on data that does not include such sudden and large upward output gap expansions. Overall, our model yields stable nowcasts, even under real-world conditions including revised data, revised parameters and revised model specification.

## 6.2 Examining output gap revisions

A reliable model of the output gap not only implies decent nowcasts, but also small estimate revisions. To investigate how much the estimated output gap is revised ex post, we pseudo-decompose the revisions into contributions from specification revisions (i.e. different cross-sectional information that is chosen automatically over time), parameter revisions (i.e. the effect of more information in the time dimension on the baseline specification) and GDP data revisions. We evaluate the associated revisions in the spirit of [Orphanides and van Norden \(2002\)](#). That is, we obtain the difference between the final full-sample estimate (i.e. the benchmark) and (pseudo-)real-time estimates. Subsequently, we compute various loss statistics. For instance, if revisions are small, the mean and the standard deviation of the difference should be small as well, it should not be autocorrelated (to make sure the revision is not systematic) and no extreme deviations should occur. The results are summarized in [Table 5](#).

	Mean	Sd	Min	Max	RMSD	AR(1)
$y_t - y_t^1$	-0.01	0.05	0.04	-0.07	0.07	-0.01
$y_t - y_t^2$	0.01	0.04	0.04	-0.13	0.07	0.01
$y_t - y_t^3$	0.58	0.64	0.61	-0.33	1.53	0.52

[Table 5](#): Summary and loss statistics in the spirit of [Orphanides and van Norden \(2002\)](#) of the difference of the baseline output gap  $y_t$  and the result taking into account parameter revisions ( $y_t^1$ ), GDP data-revisions ( $y_t^2$ ) as well as real-time model selection ( $y_t^3$ ). From left to right: mean, standard deviation, minimum, maximum, root mean squared difference and the AR(1) regression coefficient of  $y_t - y_t^{\{1,2,3\}}$ , none of which is significantly different from zero at conventional levels.

[Table 5](#) yields two key results. First of all, revisions in GDP data and the baseline model parameters are, on average, irrelevant for output gap revisions. That is, given our baseline specification, the estimated gap is hardly, if at all, revised ex post, indicated by the small mean revisions ( $-0.01$  and  $0.01$ , respectively). Secondly, if a different model specification is chosen, the output gap may revise. However, the revision due to specification changes is still quite small and on average hardly different from zero. In the next step, we examine the revisions due to GDP data updates, parameter estimations and model selection in more detail across the time dimension.



### 6.2.1 Model selection in real-time

Our model selection procedure is designed to yield a stable specification. It ensures that only aggregates that consistently explain at least as much reduced-form variation over several periods in the output gap as GDP growth enter the model. However, as we run model selection since 2008Q4, the implied cross-section might in fact change from time to time. Figure 6 shows the implications for output gap revisions. As is apparent, they are larger than the previously discussed sources for revisions. Especially at the beginning of 2022, the revision is obvious. However, even in times of large economic stress (e.g. 2020), the specification is very stable. Overall, the revisions due to changes in the specification are small.

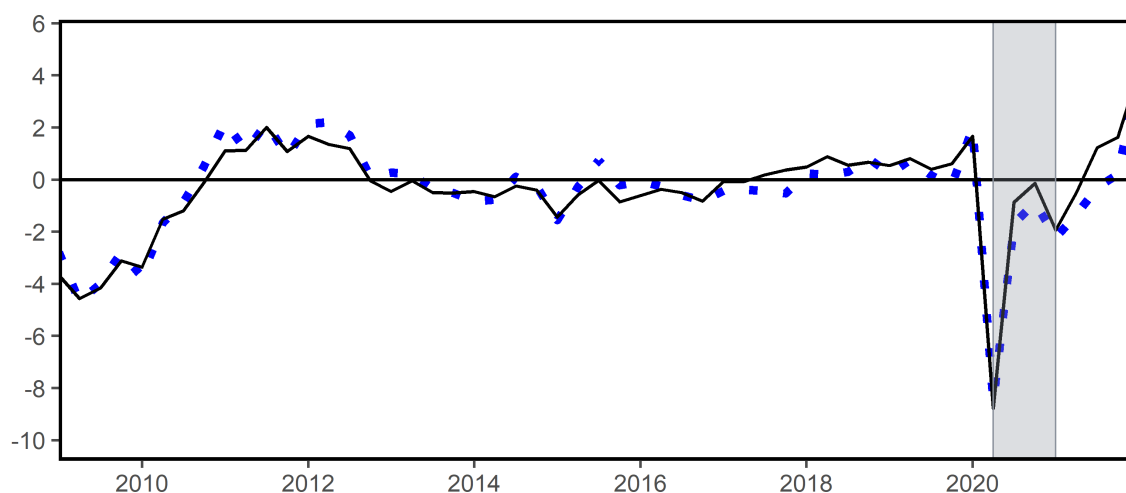


Figure 6: Black solid line: Ex post estimate. Blue dotted line: Estimate with real-time model selection. For further notes, see Figure 2.

### 6.2.2 Parameter revisions in pseudo real-time

After the previous quarter has been observed, our model needs updating in order to incorporate all available information. That is to say, even if the specification does not change, the parameters of the autoregressive vector as well as the shrinkage factor  $\lambda$  need to be estimated again. In the following, we examine the effect of parameter revisions on our estimate and nowcasting performance. To this end, we estimate the model until 2008Q3 as an initialization. Then we elicit a pseudo real-time update (ignoring data-revisions and specification changes for the moment) by adding full-quarter new information of the subsequent quarters in sequentially. Thus, we re-estimate the model every quarter conditionally on ex post data, the specification and the entire-quarter information. Figure 7 depicts the pseudo real-time output gap obtained from this procedure. Clearly, the difference between the real-time estimate and the ex post output gap is negligible in the entire period under investigation.

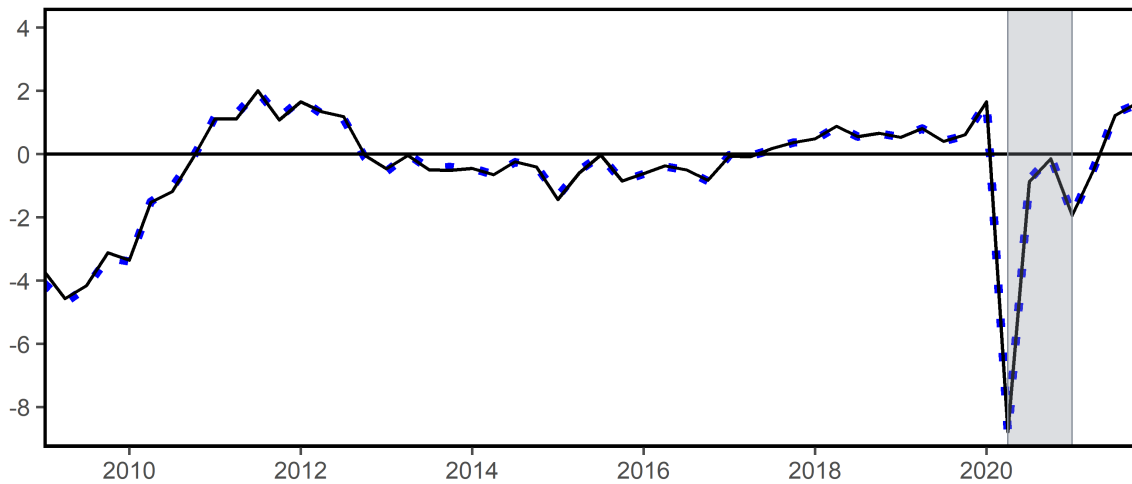


Figure 7: Black sold line: Ex post estimate. Blue dotted line: Estimate with real-time parameters in the baseline specification. For further notes, see Figure 2.

### 6.2.3 GDP data revisions in pseudo-real-time

In the final real-time analysis, we investigate the relevance of data revisions. We re-estimate the model for each quarter, given the baseline specification and the full-sample monthly information. Note that there are two sources of revisions in GDP. First of all, as we use GDP data chained in previous year prices, the level of GDP in the past years is adjust in each first quarter of a given year. We expect this effect to be small, as we employ GDP in growth rates. Moreover, since 2005, GDP data is revised in the next quarter after the initial release. From Figure 8, we clearly see that these revisions do not play a role.

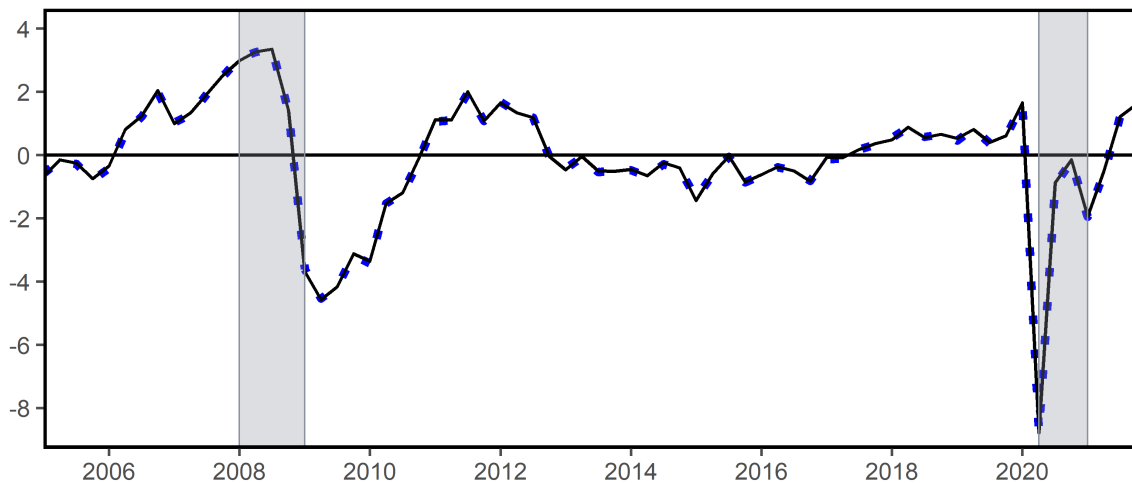


Figure 8: Black sold line: Ex post estimate. Blue dotted line: Estimate with GDP revisions. For further notes, see Figure 2.

## 7 Robustness: Larger information set

In this section, we investigate the effects of including a larger set of variables in our model. We estimate a model that comprises all variables of the baseline specification and all variables with a standard deviation of explained shares of variation larger than the median standard deviation of explained shares of variation of the aggregates that are not included in the baseline model. That is to say, in addition to the baseline variables, we include the consumer price index, Kurzarbeit and indices for new order for consumption goods and input goods. Figure 9 depicts the German output gap estimated from the alternative model. We note that for most of the period under examination, the two estimates are very similar. The difference for 2022Q1 is, most likely, due to the inclusion of the consumer price index.

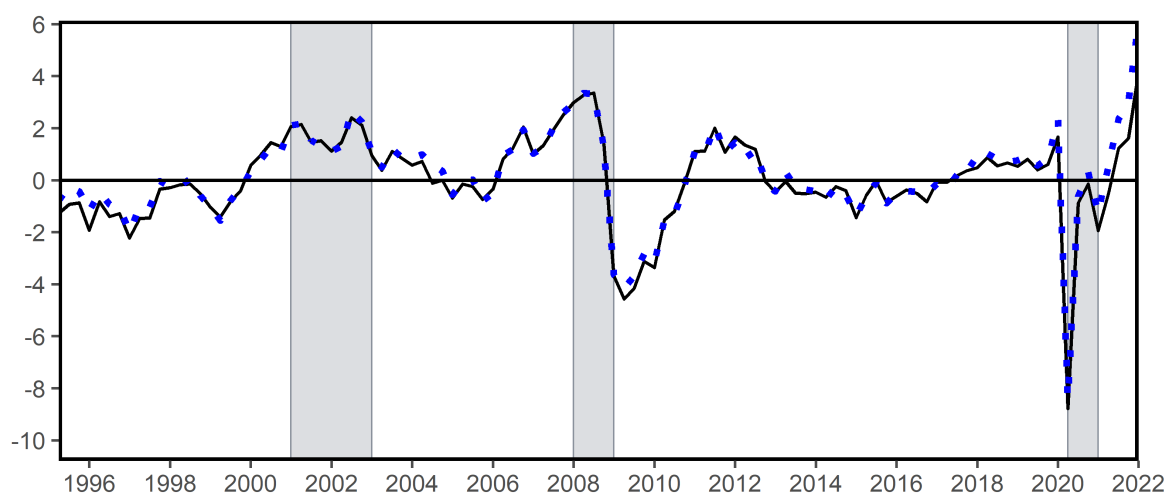


Figure 9: Black solid line: Ex post estimate. Blue dotted line: Estimate based on larger information set. For further notes, see Figure 2.

## 8 Conclusion

We have provided an in-depth analysis of the international and domestic determinants of the German output gap between 1995 and 2022 by means of a medium-size mixed-frequency vector-autoregressive model that exploits monthly information to evaluate the expectation associated with a multivariate Beveridge-Nelson decomposition. We showed that substantial shares of variation in the German business cycle are explained by expectations, the labour market and international economic aggregates. Moreover, we demonstrated that our model fairly accurately predicts the German output gap up to three months prior to a German gross domestic product data release. In particular, observing consumer sentiment and the labour market allows to produce a decent nowcast of the German output gap.

## References

- BARIGOZZI, M. AND M. LUCIANI (2021): “Measuring the Output Gap using Large Datasets,” *The Review of Economics and Statistics*, 1–45.
- BERGER, T., J. MORLEY, AND B. WONG (forthcoming): “Nowcasting the output gap,” *Journal of Econometrics*.
- BERGER, T., J. RICHTER, AND B. WONG (2022): “A unified approach for jointly estimating the business and financial cycle, and the role of financial factors,” *Journal of Economic Dynamics and Control*, 136.
- BORIO, C., F. P. DISYATAT, AND M. JUSELIUS (2013): “Rethinking potential output: Embedding information about the financial cycle,” BIS Working Papers 404, Bank for International Settlements.
- DIEBOLD, F. AND R. MARIANO (1995): “Comparing Predictive Accuracy,” *Journal of Business and Economic Statistics*, 13, 253–263.
- EICKMEIER, S. (2007): “Business cycle transmission from the US to Germany—A structural factor approach,” *European Economic Review*, 51, 521–551.
- HAMILTON, J. D. (2018): “Why You Should Never Use the Hodrick-Prescott Filter,” *The Review of Economics and Statistics*, 100, 831–843.
- HAVIK, K., K. MCMORROW, F. ORLANDI, C. PLANAS, R. RACIBORSKI, W. ROEGER, A. ROSSI, A. THUM-THYSENA, AND V. VANDERMEULEN (2014): “The Production Function Methodology for Calculating Potential Growth Rates Output Gaps,” *Economic Papers*.
- HODRICK, R. J. AND E. C. PRESCOTT (1997): “Postwar U.S. Business Cycles: An Empirical Investigation,” *Journal of Money, Credit and Banking*, 1–16.
- JAROCINSKI, M. AND M. LENZA (2018): “An Inflation-Predicting Measure of the Output Gap in the Euro Area,” *Journal of Money, Credit and Banking*, 50, 1189–1224.
- KAMBER, G., J. MORLEY, AND B. WONG (2018): “Intuitive and Reliable Estimates of the Output Gap from a Beveridge-Nelson Filter,” *The Review of Economics and Statistics*, 100, 550–566.
- MORLEY, J. AND B. WONG (2020): “Estimating and accounting for the output gap with large Bayesian vector autoregressions,” *Journal of Applied Econometrics*, 35, 1–18.
- ORPHANIDES, A. AND S. VAN NORDEN (2002): “The Unreliability of Output-Gap Estimates in Real Time,” *The Review of Economics and Statistics*, 84, 569–583.
- QUAST, J. AND M. H. WOLTERS (2022): “Reliable real-time output gap estimates based on a modified Hamilton filter,” Tech. rep.
- WAGGONER, D. AND T. ZHA (1999): “Conditional Forecasts In Dynamic Multivariate Models,” *The Review of Economics and Statistics*, 81, 639–651.

# Appendix

## A Structural historical decomposition

In this Appendix, we briefly present a structural historical decomposition. Structural identification of our model is a challenging task, as the mere size due to the mixed-frequency setup, complicates the analysis tremendously. Traditional identification schemes based on sign restrictions and more recent data-driven alternatives can hardly cope with a system of this size. For instance, the rotation space for identification based on sign restrictions is vast even for a relatively small number of restrictions. Thus, we base our brief analysis on a Choleski factor of the reduced form error covariance and leave more sophisticated structural identification to future research. Figure 10 depicts a structural decomposition.

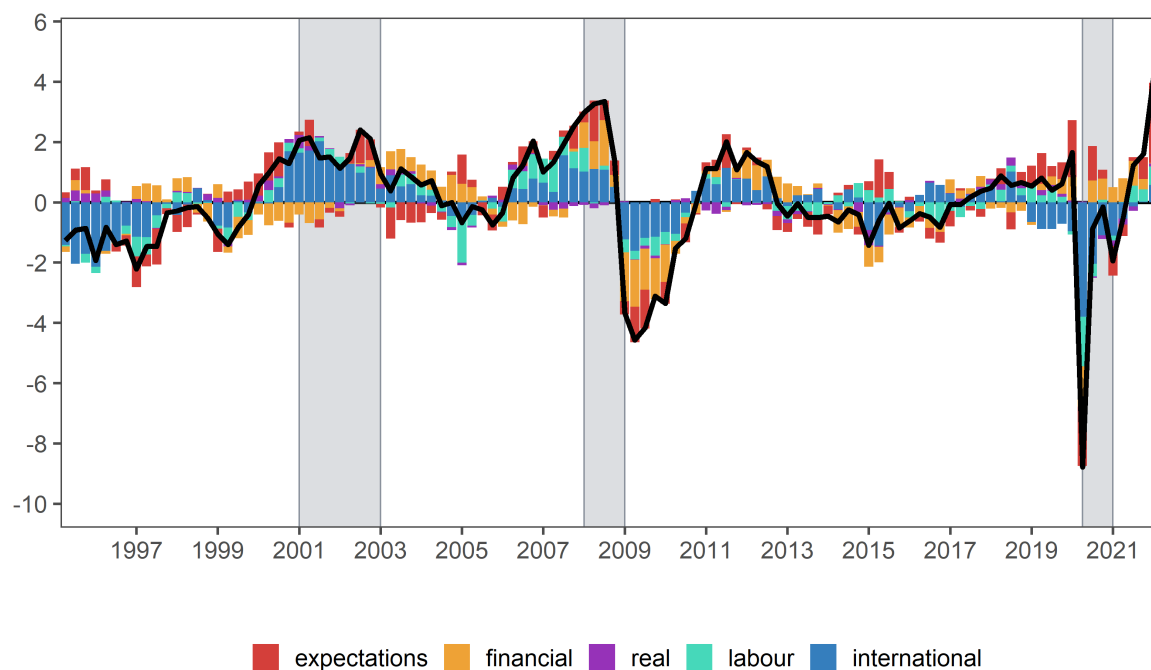


Figure 10: Structural historical decomposition based on Choleski decomposition of the German output gap. For further notes see Figure 4.

## **Zur Potenzialschätzung im Rahmen der Schuldenregel des Bundes – Ein Beitrag zum Beteiligungsprozess**

Zentralbereich Volkswirtschaft der Deutschen Bundesbank<sup>1</sup>

Anlässlich des „Calls for Proposals“ zum Potenzialverfahren möchten wir im Folgenden einige grundlegende Überlegungen einbringen, die wir hinsichtlich der Evaluierung eines Potenzialverfahrens sowie des allgemeinen Beteiligungsprozesses als wichtig erachten.

Es gibt zahlreiche Potenzialverfahren, die jeweils Vor- und Nachteile haben. Bei ökonomischen Schätzungen ist es üblich, je nach Fragestellung unterschiedliche Modelle heranzuziehen. Jeder einzelne Ansatz basiert auf Annahmen und bildet nur Teilaspekte ökonomischer Zusammenhänge ab. Dies bedeutet auch: Ein Potenzialverfahren, das für eine Fiskalregel vorteilhaft ist, muss nicht auch das beste Verfahren sein, um etwa eine spezielle Strukturform zu evaluieren.

Bei der Wahl eines Potenzialverfahrens ist deshalb zu berücksichtigen, für welchen Zweck es genutzt werden soll. Hier ist der Zweck die Konjunkturbereinigung in der Schuldenregel. Dafür ergeben sich besondere Anforderungen:

- Es ist elementar, die Symmetrievorgabe des Grundgesetzes zu berücksichtigen. Das gewählte Potenzialverfahren sollte die Symmetrie der Konjunkturkomponenten möglichst gut absichern. In jedem Fall wäre es wichtig, über die konjunkturbedingten Überschüsse und Defizite auf einem Kontrollkonto Buch zu führen. Im Zeitverlauf entstandene größere Abweichungen in die eine oder andere Richtung könnten dann gegebenenfalls korrigiert werden.
- Fiskalregeln sollten transparent und nachvollziehbar sein. Idealerweise sollte dafür das Potenzialverfahren stabile Ergebnisse liefern, ohne dass die Modellierung oder die Schätzparameter regelmäßig angepasst werden müssen. Eine gewisse Komplexität erscheint nicht per se problematisch. Allerdings sollten Verfahren und Ergebnisse öffentlich gut dokumentiert und nachprüfbar sein. Dies erleichtert eine Kontrolle durch Parlament und Öffentlichkeit, was für die Schuldenregel besonders wichtig ist.
- Eine spezielle Herausforderung ergibt sich für die gesamtstaatliche Haushaltsüberwachung, wenn die genutzten nationalen Schätzverfahren vom EU-Verfahren abweichen. Denn dann kann es eher zu Verfehlungen von EU-Vorgaben kommen. Bei sich stärker unterscheidenden Verfahren wäre es noch wichtiger, dass der nationale Stabilitätsrat die Haushalte gut fundiert überwacht und bei absehbaren Verfehlungen der EU-Vorgaben aktiv wird.

---

<sup>1</sup> Kontakt-Adresse: [oeffentliche.finanzen@bundesbank.de](mailto:oeffentliche.finanzen@bundesbank.de)

Unabhängig vom letztlich gewählten Potenzialverfahren empfehlen wir, die Konjunkturbereinigung in der Schuldenbremse um eine Fehlerkomponente zu ergänzen.<sup>2</sup> Diese haben wir bereits in der digitalen Auftaktkonferenz skizziert. Durch die Fehlerkomponente erhält die Haushaltspolitik mehr Anpassungszeit bei Revisionen der erwarteten Wirtschaftsentwicklung (und damit des Potenzialverfahrens) und Schätzfehlern bei den Steuern. Die ergänzende Fehlerkomponente kann einen wesentlichen Beitrag leisten, um innerhalb der Schuldenbremse eine stetige Haushaltspolitik und eine antizyklische Wirkung der Finanzpolitik zu fördern. Es ist sinnvoll, die Fehlerkomponente auf die Konjunkturkomponente insgesamt zu beziehen (nicht nur auf das Potenzial). Denn die Schätzfehler bei den konjunkturbereinigten Budgetsalden resultieren zum großen Teil aus instabilen Zusammenhängen zwischen gesamtwirtschaftlichen und fiskalischen Entwicklungen.

Abschließend möchten wir auf das in der Bundesbank verwendete Konjunkturbereinigungsverfahren hinweisen.<sup>3</sup> Dieses fokussiert auf den Zusammenhang zwischen Makro- und Fiskalentwicklung. Es enthält allerdings kein Potenzialverfahren und entspricht daher nicht der Stoßrichtung des Workshops. Gesamtwirtschaftliche Schwankungen werden vielmehr mit einem einfachen Filterverfahren geschätzt. Insgesamt ist der Ansatz aus unserer Sicht besonders gut für die Fiskalanalyse geeignet.

- Unser Verfahren geht bei der Trend-Zyklus-Zerlegung gesamtwirtschaftlicher Entwicklungen pragmatisch vor: Wir verwenden den HP-Filter. Das Endpunktproblem des HP-Filters adressieren wir mit einer Fortschreibungsregel unter Einbezug der Bundesbank-Projektion und der Bundesbank-Potenzialschätzung.
- Der HP-Filter liefert per Konstruktion symmetrische Trendabweichungen. Verzerrungen in Echtzeit ergeben sich im Wesentlichen nur durch Revisionen von Daten, Projektionen und Potenzial im Zeitraum der Fortschreibung. Zudem ist der HP-Filter einfach und ohne Ad-Hoc-Eingriffe in Modellierung und Parameter anzuwenden. Auch die Ergebnisse sind plausibel: Die BIP-Trend-Abweichung liegt in der Regel in der Bandbreite gängiger Schätzungen der Produktionslücke.
- Wir bestimmen dann separate Konjunkturkomponenten für einzelne Budgetkategorien (Lohnsteuer, Sozialbeiträge, Umsatzsteuer etc.) anhand der Trend-Abweichungen der jeweiligen gesamtwirtschaftlichen Bemessungsgrundlagen. Die Bemessungsgrundlagen und Fiskalelastizitäten entsprechen dabei weitestgehend unseren Prognosemodellen.
- Solch ein disaggregiertes Verfahren kann konjunkturelle Schwankungen im Staatshaushalt detaillierter abbilden. Denn die Entwicklung zum Beispiel der Löhne muss nicht mit der des BIP übereinstimmen. Ein aggregiertes Verfahren unterstellt dagegen immer einen (aus der Vergangenheit abgeleiteten) konstanten Zusammenhang zwischen BIP und Makrobezugsgröße.

---

<sup>2</sup> Vgl. ausführlich: Bundesbank (2022), Die Schuldenbremse des Bundes: Möglichkeiten einer stabilitätsorientierten Weiterentwicklung, Monatsbericht April 2022, S.55-65.

<sup>3</sup> Vgl. ausführlich: Deutsche Bundesbank (2006), Ein disaggregierter Ansatz zur Analyse der Staatsfinanzen: die Entwicklung der öffentlichen Finanzen in Deutschland in den Jahren 2000 bis 2005, Monatsbericht März 2006, S. 63 ff.

# Konjunkturbereinigungsverfahren

## Verbesserung der Schätzung, aber keine Abschwächung der Fiskalregeln durch die Hintertür

Veronika Grimm

FAU Erlangen-Nürnberg & Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung

1. **Fiskalregeln** legen Grenzen für fiskalische Größen fest, wie etwa für das Defizit öffentlicher Haushalte und die Staatsverschuldung, und **helfen** so, die langfristige **Tragfähigkeit der öffentlichen Finanzen sicherzustellen** (SVR 2020 Ziffern 297 ff.). Eine umfangreiche empirische Literatur belegt dabei die Effektivität von glaubwürdigen und verbindlichen Fiskalregeln insbesondere die politökonomischen Anreize für höhere Staatsverschuldungen einzudämmen (siehe SVR 2019 Ziffern 433 f. und dort zitierte Quellen). Die gute finanzpolitische Ausgangslage Deutschlands hat es im Zusammenspiel mit der Ausnahmeregelung der Schuldenbremse ermöglicht, dass die Bundesregierung während der Corona-Krise mit umfangreichen fiskalpolitischen Stabilisierungsmaßnahmen reagieren konnte, ohne dass die Tragfähigkeit der deutschen Staatsfinanzen in Zweifel gezogen wurde.
2. **Damit** der Staat **in zukünftigen Krisensituationen** in ausreichendem Maße **handlungsfähig** ist **und resilient** auf adverse makroökonomische Schocks reagieren kann, sollte die öffentliche Verschuldungsquote in konjunkturell guten Zeiten auf ein Niveau zurückgefahren werden, das die Tragfähigkeit sicherstellt. Insbesondere wird damit ein Puffer für den Fall erneuter krisenbedingter Schocks beziehungsweise für konjunkturell schlechtere Zeiten bereitgestellt (SVR 2021, Ziffern 125 ff.). Um solch eine **stabilisierende, antizyklische Fiskalpolitik** im Rahmen der Fiskalregeln zu ermöglichen und die Wirkung automatischer Stabilisatoren nicht einzuschränken, werden die Größen zur Bestimmung der fiskalischen **Grenzwerte um konjunkturelle Einflüsse bereinigt**. Im Vergleich zu einer nicht-Berücksichtigung konjunktureller Einflüsse wird der Spielraum in konjunkturell schlechten Zeiten erhöht und in konjunkturell guten Zeiten verringert. Die gesetzliche Grundlage für die deutsche Schuldenbremse bildet Artikel 109 GG, der eine „im Auf- und Abschwung symmetrische [...] Berücksichtigung der Auswirkungen einer von der Normallage abweichenden konjunkturellen Entwicklung“ vorsieht.
3. Die dafür verwendeten Konjunkturbereinigungsverfahren orientieren sich auf Bundesebene an den Verfahren im Rahmen des europäischen Stabilitäts- und Wachstumspakts. Hierbei wird als Referenz für die Evaluierung der konjunkturellen Entwicklung das **gesamtwirtschaftliche Produktionspotenzial** genutzt. Konzeptionell soll das Produktionspotenzial das zu erwartende Produktionsniveau einer Volkswirtschaft unter den aktuell gegebenen wirtschaftspolitischen Rahmenbedingungen bei Normalauslastung der



Produktionsfaktoren, und ohne Inflationsdruck zu erzeugen, messen. Das Produktionspotenzial ist nicht direkt beobachtbar und muss dementsprechend mit Hilfe statistischer Verfahren geschätzt werden. Aufgrund verschiedener Probleme und notwendiger Annahmen im Zusammenhang mit der **verwendeten Schätzmethode**, steht das aktuelle Verfahren **vermehrt in der Kritik** und verschiedene Reformvorschläge werden diskutiert (siehe beispielsweise Ademmer et al., 2019 sowie SVR 2019 Ziffer 496 ff. und dort zitierte Quellen). Die Vorschläge konzentrieren sich dabei vor allem auf methodische Änderungen an der Schätzung des Potenzialniveaus, und zielen damit auf eine verbesserte Berücksichtigung der konjunkturellen Lage im Rahmen bestehender Fiskalregeln ab.

4. **Vorschläge**, die insbesondere die **Echtzeitschätzung** des Auslastungsgrads des gesamtwirtschaftlichen Produktionspotenzials **verbessern**, sind **zu begrüßen**. Allen voran stellen betragsmäßig große und prozyklische Revisionen sowie eine unsaubere Trennung zwischen strukturellen und konjunkturellen Veränderungen der Wirtschaftsleistung kritische Herausforderungen für die bestehenden statistischen Methoden zur Konjunkturbereinigung dar (siehe beispielsweise Ademmer et al., 2019 sowie SVR 2019 Ziffer 499 ff. und dort zitierte Quellen). **Zentrale Stellschrauben** für methodische Verbesserungen und eine Erhöhung der Verlässlichkeit der Schätzverfahren dürften die Verwendung **zusätzlicher Indikatoren** für bestimmte Komponenten des Produktionspotenzials (u.a. Ademmer et al., 2019; Göttert & Wollmershäuser, 2021; Weiske, 2018) **und alternativer Filterverfahren**, die insbesondere das Randwertproblem adressieren (u.a. Quast und Wolters, 2022), sein.
5. Demgegenüber sind Vorschläge, die auf eine **Umgehung der bestehenden Fiskalregeln** durch eine systematische Ausweitung der Neuverschuldungsmöglichkeiten im Zuge der Überarbeitung der Konjunkturbereinigungsverfahren abzielen, **kritisch zu bewerten**. Allen voran sollte das Konjunkturbereinigungsverfahren nicht durch politische Zielsetzungen verwässert werden. Dies wäre nicht nur dem Ziel einer antizyklischen Fiskalpolitik im Konflikt, sondern dürfte zudem den Anforderungen von Artikel 109 GG nicht gerecht werden.
6. Darüber hinaus könnte eine solche **Aufweichung der Fiskalregeln** die **Glaubwürdigkeit und die Verbindlichkeit** der bestehenden institutionellen Rahmenbedingungen **beeinträchtigen**. Vor dem Hintergrund deutlich höherer Schuldenstände in vielen EU-Mitgliedstaaten wäre es sinnvoll, die Komplexität des Regelsystems zu reduzieren, die Transparenz hinsichtlich der Einhaltung und Durchsetzung zu verbessern, sowie Prozyklizität der Regelwirkung zu vermeiden. Eine mögliche Ausweitung der Neuverschuldungsmöglichkeiten im Zuge einer Überarbeitung des Konjunkturbereinigungsverfahrens wäre in dieser Hinsicht kontraproduktiv (SVR 2021 Ziffern 215 ff.). Gerade mit Blick auf den bevorstehenden Strukturwandel müssen Fiskalregeln effektiv die notwendigen Anreize für eine politische Abwägung zwischen **zukunftsorientierten Ausgabenarten** gegenüber rein konsumtiven Ausgaben setzen (Feld et al., 2021; SVR 2021 Ziffern 216 ff.).

## Literatur:

- Ademmer, M. et al. (2019), Schätzung von Produktionspotenzial und -lücke: Eine Analyse des EU-Verfahrens und mögliche Verbesserungen, Kieler Beiträge zur Wirtschaftspolitik Nr. 19, Institut für Weltwirtschaft, Kiel.
- Feld, L.P., V. Grimm und W.H. Reuter (2021), Zukunftsperspektiven sichern durch Reformen, nicht durch Schulden, Wirtschaftsdienst 101 (6), 418–424.
- Götttert, M. und T. Wollmershäuser (2021), Survey-Based Structural Budget Balances, CESifo Working Paper, 8911.
- Quast, J. und M. H. Wolters (2022), Reliable Real-Time Output Gap Estimates Based on a Modified Hamilton Filter, Journal of Business & Economic Statistics, 40(1): 152-168.
- SVR (Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung) (2021), Transformation gestalten: Bildung, Digitalisierung und Nachhaltigkeit, Jahresgutachten 2021/22.
- SVR (Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung) (2020), Corona-Krise gemeinsam bewältigen, Resilienz und Wachstum stärken, Jahresgutachten 2020/21.
- SVR (Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung) (2019), Den Strukturwandel meistern, Jahresgutachten 2019/20.
- Weiske, S. (2018), Indicator-based estimates of the output gap in the euro area, Arbeitspapier 12/2018, Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung, Wiesbaden.

# Der SVR-Ansatz zur Bestimmung des Produktionspotenzials

Von Christian Ochsner, Lars Other und Christopher Zuber

1. Der **Sachverständigenrat** zur Begutachtung der gesamtwirtschaftlichen Entwicklung (SVR) erstellt jährlich im Rahmen des Jahresgutachtens eine Mittelfristprojektion für die deutsche Volkswirtschaft. Hierfür **schätzt** der SVR das **Produktionspotenzial mithilfe** einer **Produktionsfunktion**, die sich an der EU-Methode (Havik et al., 2014) orientiert. Die Methodik des SVR wurde in den vergangenen Jahren kontinuierlich weiterentwickelt (Breuer und Elstner, 2020; SVR 2014 Ziffern 202 ff.; SVR 2017 Ziffer 320; SVR 2020 Ziffern 89 f.). Wie die EU-Methode basiert das vom SVR verwendete Verfahren auf der Cobb-Douglas-Produktionsfunktion, bei der die Entwicklung des Produktionspotenzials auf die Faktoren Arbeit und Kapital sowie auf die Totale Faktorproduktivität (TFP) zurückgeführt wird. Hinsichtlich der verwendeten Daten und der verwendeten Schätzmethoden weicht das SVR-Verfahren aber in vielerlei Hinsicht von der EU-Methode ab. Im Folgenden werden die wesentlichen Unterschiede beleuchtet. Eine ausführliche Diskussion des bisherigen SVR-Verfahrens findet sich in Breuer und Elstner (2020) und SVR (2020 Ziffern 89 f.). Im Anschluss wird ein Ausblick über Verbesserungsmöglichkeiten gegeben, an denen der wissenschaftliche Stab des SVR aktuell arbeitet.
2. Der **Faktor Arbeit** wird bei der Methode des SVR und der EU aus vier Zeitreihen aggregiert: der Anzahl der Personen im erwerbstätigen Alter, der Partizipationsquote, der durchschnittlichen Arbeitszeit pro erwerbstätiger Person und einem Maß für die langfristig inflationsneutrale Arbeitslosenrate. Der SVR misst diese gleichgewichtige Arbeitslosenrate durch die Non Accelerating Inflation Rate of Unemployment (NAIRU), während die EU die Non Accelerating Wage Rate of Unemployment (NAWRU) schätzt. Die NAIRU-Schätzung des SVR basiert dabei auf einem eklektischen Ansatz, bei dem die Ergebnisse von verschiedenen Spezifikationen mit unterschiedlichen Inflationsindikatoren gemittelt werden.

Die Aggregation des SVR unterscheidet sich weiterhin in zwei wesentlichen Aspekten von der EU-Methode. Erstens verwendet der SVR zur Berechnung der Bevölkerungszahl ein **Kohortenmodell**, das auf detaillierten Bevölkerungsdaten und Vorausberechnungen durch das Statistische Bundesamt basiert. Dies erlaubt eine realistische Modellierung der Partizipationsquote, die relative Verschiebungen zwischen Gruppen von Erwerbstätigen mit strukturell unterschiedlichen gruppenspezifischen Partizipationsquoten berücksichtigen kann. So ist davon auszugehen, dass aufgrund eines steigenden Anteil älterer Erwerbstätiger in den nächsten Jahren die durchschnittliche Partizipationsquote sinken dürfte. Zudem berücksichtigt das verwendete Kohortenmodell weiterhin bestehende Unterschiede zwischen Männern und Frauen sowie die niedrigere Partizipationsquote von Geflüchteten in den ersten Jahren nach dem Zuzug.

Zweitens wird die **Arbeitszeit je erwerbstätiger Person** zudem **getrennt nach Teil- und Vollzeitbeschäftigten sowie Selbstständigen** mithilfe von AR-Modellen fortgeschrieben.

3. Für die Bestimmung des **Faktors Kapital** verwendet der SVR das Konzept des **Kapitaleinsatzes** (capital services). **Gegenüber** der in der EU-Methode vorgesehenen Bestimmung über den preisbereinigten **Nettokapitalstock** (Anlagevermögen) berücksichtigt der aggregierte Kapitaleinsatz die Kapitalnutzungskosten, die die Nutzungsintensität und Qualitätseigenschaften der einzelnen Vermögensklassen widerspiegeln sollten. Aufgrund der größeren konjunkturellen Schwankungen beim aggregierten Kapitaleinsatz findet für die Bestimmung der Trendkomponente eine Konjunkturbereinigung der einzelnen Anlagearten statt.
4. Bei der **TFP** folgt der SVR, ähnlich zur NAIRU-Schätzung, einem eklektischen Ansatz (SVR 2020 Kasten 6). Neben einem **Zustandsraummodell**, dass verschiedene Konjunkturindikatoren zur Identifikation der Konjunkturzykluskomponente der TFP in verschiedenen Spezifikationen nutzt und dem in der EU-Methode verwendeten Modell ähnelt, wird ein statisches **Faktormodell** (Weiske, 2018) **und der modifizierte Hamilton-Filter** (Hamilton, 2018; Quast und Wolters, 2020) zur Bestimmung der Trend-TFP genutzt. Die Schätzergebnisse der verschiedenen Modelle werden dann über einen ungewichteten Durchschnitt aggregiert.
5. Im Detail weicht die bisherige Methode des SVR auch in der Verwendung der **Konjunkturbereinigung** der einzelnen Komponenten **mittels des HP-Filters** von der EU-Methode ab. Neben der Konjunkturbereinigung des Trend-Kapitals fixiert der SVR etwa den Glättungsparameter in dem Zustandsraummodell für die TFP auf  $\lambda=100$  statt ihn wie bei der EU-Methode zu schätzen.
6. Die Anwendung einfacher Filterverfahren, wie etwa des HP-Filters, zur Trend-Zyklus-Zerlegung wird sowohl aus theoretischer Perspektive (Hamilton, 2018) als auch unter dem Gesichtspunkt der **Revisionsanfälligkeit** (Orphanides und van Norden, 2002) kritisiert. Aus diesen Gründen arbeitet der wissenschaftliche Stab des SVR aktuell an einer **alternativen Möglichkeit** zur Konjunkturbereinigung von makroökonomischen Zeitreihen. Insbesondere soll durch die Verwendung eines flexiblen, bayesianisch geschätztem Zustandsraummodells, zunächst bei den Trendkomponenten Arbeit und Kapital, die Sensitivität der Schätzung bezogen auf neue Beobachtungen (Randwertproblem) reduziert werden.

## Literatur

- Ademmer, M. et al. (2019), Schätzung von Produktionspotenzial und -lücke: Eine Analyse des EU-Verfahrens und mögliche Verbesserungen, Kieler Beiträge zur Wirtschaftspolitik Nr. 19, Institut für Weltwirtschaft, Kiel.
- Breuer, S. und S. Elstner (2020), Germany's growth prospects against the backdrop of demographic change, *Jahrbücher für Nationalökonomie und Statistik* 240 (5), 565–605.
- Hamilton, J.D. (2018), Why you should never use the Hodrick-Prescott filter, *Review of Economics and Statistics* 100 (5), 831–843.
- Havik, K. et al. (2014), The production function methodology for calculating potential growth rates & output gaps, *European Economy – Economic Paper* 535, Europäische Kommission, Generaldirektion Wirtschaft und Finanzen (DG ECFIN), Brüssel.
- Orphanides, A. und S. van Norden (2002), The Unreliability of Output-Gap Estimates in Real Time, *The Review of Economics and Statistics* 84 (4), 569–583.
- Quast, J. und M. H. Wolters (2022), Reliable Real-Time Output Gap Estimates Based on a Modified Hamilton Filter, *Journal of Business & Economic Statistics*, 40(1): 152–168.
- SVR (Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung) (2020), Corona-Krise gemeinsam bewältigen, Resilienz und Wachstum stärken, *Jahresgutachten 2020/21*.
- SVR (Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung) (2017), Für eine zukunftsorientierte Wirtschaftspolitik, *Jahresgutachten 2017/18*.
- SVR (Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung) (2014), Mehr Vertrauen in Marktprozesse, *Jahresgutachten 2014/15*.
- Weiske, S. (2018), Indicator-based estimates of the output gap in the euro area, *Arbeitspapier* 12/2018, Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung, Wiesbaden.

## Vorschlag zur Bestimmung des Produktionspotenzials im Rahmen der grundsätzlich verankerten Schuldenregel

Ansprechpartner

Prof. Dr. Torsten Schmidt

Leiter Kompetenzbereich Wachstum, Konjunktur, Öffentliche Finanzen

torsten.schmidt@rwi-essen.de

Fon +49 201-81 49-287

Essen, 15.08.2022

Das aktuelle Verfahren zur Berechnung des Produktionspotenzials basiert auf einer Cobb-Douglas Produktionsfunktion, in die Arbeit und Kapital sowie die totale Faktorproduktivität als Faktoren einfließen (Havik et al. 2014). Die Faktoren Arbeit und Kapital werden dabei entsprechend der ökonomischen Theorie mit Hilfe weiterer Variablen erklärt. In die Prognose dieser Variablen fließen für die ersten zwei Prognosejahre die Ergebnisse aus detailliert erstellten Konjunkturprognosen mit ein. Für die folgenden Jahre der mittleren Frist kommen Zeitreihenverfahren zum Einsatz. Eine Ausnahme ist die Projektion der Bevölkerung im erwerbsfähigen Alter, die von den Bevölkerungsvorausberechnungen von EUROSTAT übernommen wird.

Im Unterschied dazu ist die totale Faktorproduktivität entsprechend der klassischen Wachstumstheorie exogen. Die Fortschreibung dieser Größe erfolgt mit Hilfe eines Kalman Filters, in den eine Variable für die Kapazitätsauslastung eingefügt wird, um eine Bereinigung um die Konjunkturschwankungen zu erhalten. Diese auf historischen Daten basierende Fortschreibung der TFP könnte in den kommenden Jahren zu einer Unter- oder Überschätzung des Produktionspotenzials führen. Wie in dem Schwerpunktthema in der Gemeinschaftsdiagnose vom Frühjahr (Projektgruppe Gemeinschaftsdiagnose 2022) gezeigt wurde, wird eine starke Produktivitätsentwicklung eine entscheidende Größe sein, um die klimapolitischen Ziele der Bundesregierung zu erreichen. Dadurch könnte die Entwicklung der totalen Faktorproduktivität in den kommenden Jahren von den historischen Mustern abweichen.

Modelle der endogenen Wachstumstheorie bieten Ansatzpunkte für die Endogenisierung der TFP, die auch im Rahmen der Potenzialschätzung genutzt werden könnten. Diese Ansätze betonen die Bedeutung der Forschungs- und Entwicklungsaktivitäten sowie der Bildung von Humankapital für die Entwicklung der TFP. Comin und Gertler (2006) haben in einem Konjunkturmodell die TFP endogenisiert indem sie einen FuE-Sektor eingeführt haben. Darin wird ein Teil der Produktion eingesetzt, um Innovationen zu entwickeln und daraus marktfähige Produkte zu generieren. Durch diesen Mechanismus erzeugen konjunkturelle Schocks kurz- und mittelfristige Schwankungen in der Produktion. Dieser Zusammenhang ist für die Schätzung des Produktionspotenzials in der mittleren Frist von besonderem Interesse.

Empirische Studien zu den Determinanten der TFP bestätigen weitgehend den Zusammenhang zwischen Forschungs- und Entwicklungsaktivitäten und der TFP (European Commission 2017). Das Gleiche gilt für den Zusammenhang zwischen TFP und Humankapital. Darüber hinaus zeigen sich Zusammen-

hänge mit weiteren Variablen, wie der Offenheit der Volkswirtschaft und der demografischen Entwicklung. Zudem zeigen empirische Studien, dass die FuE-Ausgaben einen deutlichen Vorlauf haben. So zeigen Arce-Alfaro und Schmidt (2022), dass die FuE-Ausgaben die BIP-Prognosen über einen Horizont von fünf Jahren noch deutlich verbessern. Da insbesondere im Bereich der FuE-Aktivitäten für die kommenden Jahre eine Intensivierung zu erwarten ist, sollte dieser Zusammenhang in Zukunft bei der Berechnung und der Fortschreibung des Produktionspotenzials berücksichtigt werden.

Die für diese Erweiterung notwendigen Daten stehen im Rahmen der jährlichen VGR zur Verfügung. Da in der VGR keine Daten zu FuE-Investitionen quartalsweise ausgewiesen werden, müssten die FuE-Investitionen aus den sonstigen Investitionen aus den prognostizierten Jahreswerten herausgerechnet werden, um eine Konsistenz der Kurzfristprognose zu gewährleisten. Für die mittlere Frist wird eine Projektion der FuE-Investitionen vorgenommen.

Ein weitgehenderer Schritt wäre eine Berücksichtigung des Produktionsfaktors Energie zusammen mit der Berücksichtigung des energiesparenden technischen Fortschritts. Die Bedeutung der Energie für die Entwicklung des Produktionspotenzials ist in der Gemeinschaftsdiagnose vom Frühjahr (Projektgruppe Gemeinschaftsdiagnose 2022) herausgearbeitet worden. Dazu müsste der Produktionsfaktor Energie in den Ansatz der EU-Kommission integriert werden. Auf diese Weise wäre es möglich, den energiesparenden technischen Fortschritt vom allgemeinen technischen Fortschritt zu trennen.

Wird das EU-Verfahren zur Schätzung und Fortschreibung des Produktionspotenzials durch die Einführung zusätzlicher Variablen modifiziert, die selbst wieder prognostiziert werden müssen, werden die diskretionären Gestaltungsspielräume zwangsläufig erhöht. Diese Spielräume könnten reduziert werden, indem die neuen Variablen wiederum durch formalisierte Verfahren fortgeschrieben werden. Eine weitere Möglichkeit, die Unverzerrtheit der mittelfristigen Fortschreibung des Produktionspotenzials zu gewährleisten, ist die Einbindung dieses Prozesses in das Begutachtungsverfahren durch die Gemeinschaftsdiagnose. Dazu müssten zusätzliche Variablen in das Prognoseprogramm aufgenommen werden.

## Literatur

Arce-Alfaro, G. und T. Schmidt (2022), Forecasting Medium-Term GDP, mimeo.

Comin, D. und M. Gertler (2006), Medium-Term Business Cycles, *American Economic Review* 96(3): 523-551.

European Commission (2017), Determinants of Trend TFP Growth and Key Policies that influence it, *Quarterly Report on the Euro Area* 16(2): 31-41.

Havik, K., K. Mc Morrow, F. Orlandi, C. Planas, R. Raciborski, W. Roeger, A. Rossi, A. Thum-Thysen und V. Vandermeulen (2014), The Production Function Methodology for Calculating Potential Growth Rates & Output Gaps, *European Economy. Economic Papers* 535, Brüssel.

Projektgruppe Gemeinschaftsdiagnose (2022), *Gemeinschaftsdiagnose Frühjahr 2022: Von der Pandemie zur Energiekrise – Wirtschaft und Politik im Dauerstress*, Kiel.