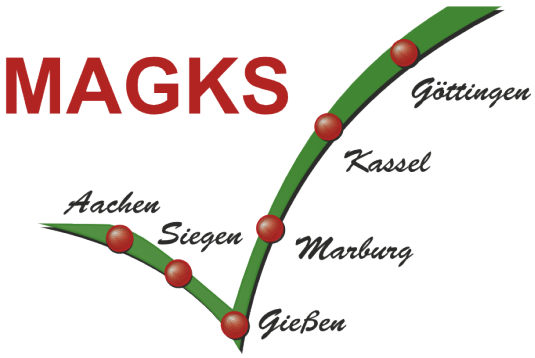


Joint Discussion Paper Series in Economics

by the Universities of
Aachen • Gießen • Göttingen
Kassel • Marburg • Siegen

ISSN 1867-3678

No. 21-2025

Bernd Hayo and Johannes Zahner

Fiscal Talks: Parliamentary Debates and Government Expenditure

Coordination: Bernd Hayo
Philipps-University Marburg
School of Business and Economics
Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Fiscal Talks: Parliamentary Debates and Government Expenditure

Bernd Hayo[a,b,*] and Johannes Zahner[b,c]

Abstract

We investigate the relationship between parliamentary debates and public expenditure by mapping legislative speeches to fiscally relevant topics and examining their connection in both long-term trends and short-term adjustments. Our analysis draws on transcripts of federal legislative discussions and federal government spending data in Germany (1950–2020), classified into nine policy functions (e.g. Social Security, National Defence and Education). We apply a state-of-the-art natural language processing technique – a structural topic model – to match identified debate topics to corresponding spending functions. Using cointegration analysis and error-correction models, we find (i) significant long-term equilibria between parliamentary debates and corresponding fiscal expenditure and (ii) that in cases of short-term disequilibrium, adjustments occur through government expenditure; that is, parliamentary debates are weakly exogenous.

JEL: *E62, C32, D78*

Keywords: *Fiscal expenditure, parliamentary debate, Bundestag, text analysis, structural topic model, error-correction model*

Version: 06 October 2025

a School of Business and Economics, Macroeconomic Research Group, Philipps-Universität Marburg, Germany.
b Marburg Center for Institutional Economics and Sustainability (MACIES), Philipps-Universität Marburg, Germany.
c Chair of Macroeconomics and Finance, Goethe University Frankfurt.

* Corresponding author: Bernd Hayo, Philipps-Universität Marburg, Macroeconomic Research Group, Universitätsstr. 24, 35037 Marburg, Germany. E-mail: hayo@wiwi.uni-marburg.de.

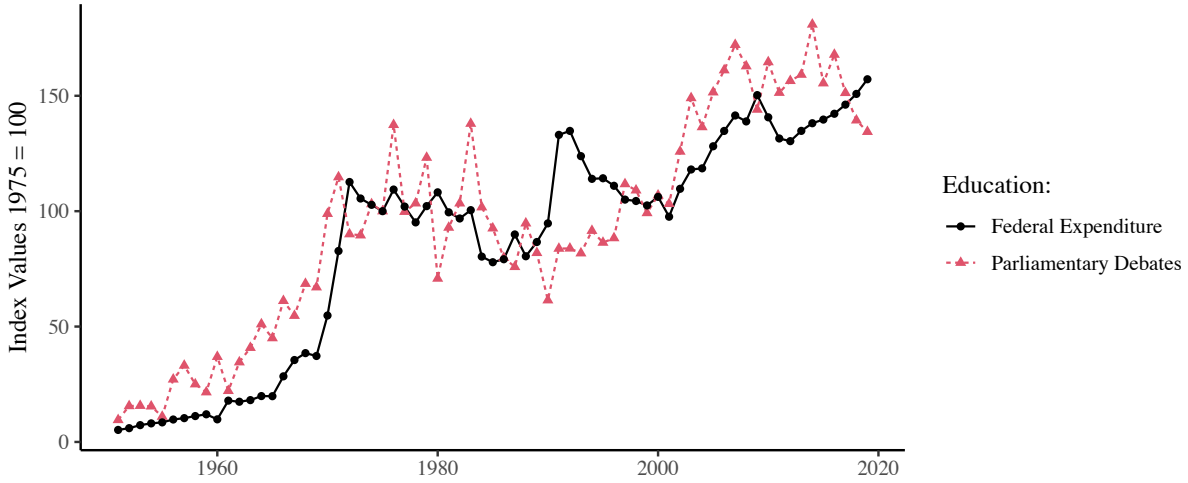
Acknowledgements: We thank the participants of the 2025 ZEW Public Finance Conference and the 2025 MACIES Workshop for their helpful comments.

1 Introduction

In Germany, any federal budget proposal must undergo a formal legislative process, ultimately requiring approval by the Bundestag, the federal parliament. Since 1949, the Bundestag has held the *'power of the purse'*, mandating that all federal expenditure is debated and ratified by elected legislators. This central role gives parliamentarians direct influence over fiscal policy, enabling them to propose expansions, scrutinise spending and signal shifts in priorities.

Consequently, changes in expenditure, such as expanding social security or reducing defence spending, are typically preceded by extensive and publicly accessible political debates. Parliamentary deliberations may therefore serve as early indicators of budgetary shifts. Historically, major policy shifts – such as the expansion of Education spending in the 1970s (see Figure 1) – have coincided with heightened debate in the Bundestag on that topic, illustrating how parliamentary discussion can signal fiscal priorities.

Figure 1: Co-movement of Education spending and parliamentary debates (1950–2020).



Note: Federal education expenditure is inflation-adjusted to 2010 euros (Section 2.1); parliamentary debates measure the share of debates related to *Education* relative to total debates (Section 3). Both series are indexed to 1975 for comparability.

A growing interdisciplinary literature highlights the links between political discourse and policy outcomes. Svaleryd (2009) demonstrates that both the identity of parliamentary speakers and the content of their arguments influence budgetary decisions. Macroeconomists have employed qualitative approaches to identify fiscal policy signals. Romer and Romer (2010) analyse official documents, including parliamentary transcripts and news reports, to detect exogenous policy changes. Hayo and Uhl (2014) apply this method to post-war German tax legislation.

Recently, computational methods for analysing legislative texts have gained prominence (Abercrombie and Batista-Navarro, 2020). Applications include Lieb et al. (2025), who show that monthly variations in US presidential 'tax talk' predict upcoming legislation and Latifi et al. (2024), who analyse German Bundestag transcripts and report that shifts in 'fiscal sentiment' have tangible macroeconomic consequences. While these studies demonstrate the influence of political debates, several gaps remain.

First, most research focuses on short- to medium-term horizons, leaving the long-run dynamics of parliamentary discourse largely unexplored. Second, dictionary-based methods or low-

dimensional topic models are prone to noise (Hayo and Zahner, 2022) and underperform relative to more advanced techniques (Baumgärtner and Zahner, 2025). Third, most studies concentrate on aggregate government spending (Latifi et al., 2024; Tillmann et al., 2024), and thus do not account for variation across distinct expenditure functions or differentiate spending decisions that take place at different levels of government.

We address these gaps by combining two novel datasets to construct a 70-year panel of *federal* parliamentary debate transcripts and *federal* government spending data, each classified by policy function. The dataset is unique in its historical scope and level of granularity, allowing us to examine how shifts in *function-specific* federal parliamentary debates relate to the corresponding federal expenditure over multiple decades. We analyse both long- and short-term relationships using cointegration techniques and vector error-correction models (VECM), testing hypotheses for the federal budget as a whole (Federal Budget) and across eight distinct federal spending categories closely related to the classification of the functions of government (COFOG) by the OECD: Education, Healthcare, Housing, National Defence, Public Safety, Regional & Structural Policy, Science & Research, and Social Security.

Our first hypothesis posits a long-term relationship between debates on a topic and the corresponding fiscal expenditure, which we find strongly supported across all functions. Second, we examine the direction of this relationship by testing whether increases in debate intensity lead to subsequent changes in the corresponding expenditure function or whether spending adjustments prompt changes in parliamentary discussion. Deviations from the long-run equilibrium are mainly corrected through short-term adjustments in expenditure that respond to preceding Bundestag debates. Finally, we formulate function-specific directional hypotheses: greater emphasis on Education, National Defence, Science & Research, Housing, and Public Safety is associated with higher expenditure, whereas the Federal Budget and Social Security show the expected negative association. No clear a priori hypothesis can be derived for Healthcare or Regional & Structural Policy; the evidence indicates a negative long-run relationship.

The remainder of the paper is structured as follows. Section 2 introduces the datasets: the Bundestag debate corpus and federal expenditure data. Section 3 outlines the Structural Topic Model (STM) used to quantify debates. Section 4 develops hypotheses on both the existence and direction of relationships between debates and fiscal expenditure. Section 5 presents the empirical analysis using cointegration and error-correction models. Section 6 concludes.

2 Data

This section describes the construction of our two novel datasets: (i) a long-term, annual time series of German federal expenditure by policy function, and (ii) a corpus of Bundestag debates.

2.1 Federal Government Spending

To the best of our knowledge, no single source currently provides a comprehensive, long-term and disaggregated view of German government expenditure. Consequently, we integrate and harmonise data from three distinct sources: (i) Fachreihe 14 3.1 (published by the Statistisches Bundesamt), (ii) annual government expenditure data compiled by Spoerer (2022), and (iii) the German National Accounts. Regarding the functional classification, we follow Spoerer's (2022) modified version of the OECD's Classification of the Functions of Government (COFOG). We align these three sources through semantic and structural mapping, impute missing values, and separate federal from state and municipal spending within each function.

This separation is crucial in a federal country such as Germany. For instance, in 2020 only about 5% (€7 billion) of total education expenditure came from the federal government, whereas nearly all national defence spending is federal. Appendix A1 details the harmonisation process.

The resulting harmonised dataset covers the period from 1950 to 2020, providing long-term real German federal government expenditure across eight distinct spending functions: Education, Healthcare, Housing, National Defence, Public Safety, Science & Research, Social Security, and Regional & Structural Policy. These unified series form the basis of our analysis of the relationship between fiscal policy decisions and parliamentary debates. The descriptive statistics for the dataset are presented in Table A1.

2.2 Parliamentary Debate Corpus

The second dataset is a cleaned corpus of Bundestag debate transcripts covering all legislative periods between 1950 and 2020 based on the Bundestag's data service and enriched with metadata such as dates, speaker identity and political affiliations.[1] We segment the transcripts into individual speeches to capture speaker-level contributions.

In order to improve the expressiveness of the textual information, we preprocess the transcripts using standard natural language processing (NLP) procedures such as text normalisation (lowercasing and handling of Umlaut), artefact removal, elimination of punctuation and numbers, and German stop-word filtering. As we are interested in economic content, we exclude speeches that do not contain economic terms as defined by the *Lexikon der Wirtschaft* of the Bundeszentrale für Politische Bildung.[2] We further restrict the vocabulary to terms occurring in at least 0.01% but no more than 10% of the speeches to balance coverage and noise.

The final corpus comprises approximately 480,000 speeches, 63,000 distinct terms and 61 million words. Figure 3 illustrates an upward trend in the number of speeches over the past 70 years, with considerable year-to-year variation.

Figure 3: Time series of number of speeches

[1] The raw data are available here: https://www.bundestag.de/services/opendata (access: 1 September 2023).
[2] The raw data are available here: https://www.bpb.de/kurz-knapp/lexika/lexikon-der-wirtschaft/ (access: 1 September 2023).

3 Identifying Topics in Parliamentary Debates

We analyse the evolution of the content of Bundestag debates over time using a structural topic model (STM) developed by Roberts et al. (2016). This unsupervised machine-learning model identifies latent *topics*, which we subsequently map to the respective expenditure functions introduced earlier.

3.1 Introduction to Structural Topic Modelling

Structural topic modelling extends commonly used topic models, such as Latent Dirichlet Allocation (LDA; Blei et al., 2003), by incorporating document-level metadata into the estimation process. STM allows a more nuanced analysis of corpora and is increasingly applied in text-based macroeconomic research (e.g., Ferrara et al., 2022; Campiglio et al., 2025; Kanelis et al., 2025). For a comprehensive introduction to advanced topic models, see Ash and Hansen (2023).

For a corpus with d documents (speeches), v words, and k topics, STM estimates two distributions:

1. **Document-topic distribution ($\boldsymbol{\theta_{d,k}}$):** the probability that topic k is discussed in document d is modelled as a logistic normal distribution. STM allows for the inclusion of document-level covariates, such as speaker-specific dummies, in the estimation process. We include year and quarter fixed effects to capture long-term and cyclical variation (e.g., budget debates typically occur in autumn sessions), and party affiliation fixed effects to address ideological influences. These controls mitigate potential biases arising from procedural or institutional regularities in parliamentary activity.[3]

2. **Word-topic distribution ($\boldsymbol{\beta_k}$):** the probability of word v occurring in topic k follows a logistic-normal distribution. We include year, quarter, and party fixed effects in the estimation process to address biases stemming from political ideology or temporal variation. For instance, the term '*defence'* primarily referred to Germany's post-war reconstruction in the 1950s, NATO in the 1970s, and counterterrorism in the 2000s. Time fixed effects allow the STM to capture such shifts, improving topic identification.

The number of topics (k) is a key hyperparameter in topic modelling. Previous studies on topic modelling have used 10–80 topics (Hansen and McMahon, 2016; Hansen et al., 2019; Larsen et al., 2021; Ferrara et al., 2022; Bohl et al., 2023). However, most studies apply topic modelling to corpora covering shorter time horizons and narrower topic scopes. In contrast, our 70-year dataset spanning multiple spending functions requires broader coverage. As a result, we opt for a model with a comparatively large number of topics ($k = 150$). While large, the supervised machine learning literature often employs models with hundreds of dimensions (e.g., Baumgärtner and Zahner, 2025). Robustness checks with lower topic counts yield consistent results.

[3] Hayo and Zahner (2023) highlight the sensitivity of text-based approaches to speaker-specific variation, which is computationally infeasible in our study. As a feasible alternative, we include time and party fixed effects.

3.2 Identifying Topics

Following standard procedure (e.g., Hansen et al., 2018; Ferrara et al., 2022; Bohl et al., 2023), each topic is labelled according to its word distribution (β_k). Specifically, topics are labelled according to their most prevalent terms. For example, a topic with the following high-probability terms is labelled *Regional Development* (we indicate debate topics by italics and aggregated topics, constructed to match spending functions, by capital letters) and mapped to the spending function Regional & Structural Policy:

funding, programme, development, investment, region, regional.[4]

We manually label and map 34 of the 150 STM topics to the eight government spending functions identified earlier and to a general budget category (*Federal Budget*). This conservative strategy prioritises precision by focusing exclusively on topics that can be confidently linked to distinct fiscal policy functions. Table 1 provides the mapping.

Table 1: Mapping of STM topics to spending functions

Spending Functions	STM Topics
Federal Budget	*Budget planning; Budgeting and debt*
Education	*Vocational training; Public school education; Financing higher education*
Healthcare	*Health insurance; Healthcare system; Medical coverage; Covid-19; Sports*
Housing	*Housing*
National Defence	*German armed forces; NATO; Afghanistan mission; Care for war victims*
Public Safety	*Data protection; National security; Law enforcement; Right-wing extremism*
Regional & Structural Policy	*Heavy industry; Small and medium-sized enterprises; Regional development*
Science & Research	*Science; Research*
Social Security	*Unemployment; Minimum wage; Family policy; Pensions; Poverty and distribution; Inclusion policy*

Note: Column 1 lists the spending functions defined in Section 2.1, while column 2 presents the STM topics mapped to each function, as described in Section 3.2. Topic labels and their mappings were manually assigned to ensure conceptual alignment with the expenditure categories. For detailed descriptions of the topic modelling and mapping procedure, refer to the Online Appendix.

Next, we turn to the document-topic scores (θ_d) to analyse the content of individual speeches. For example, a 1976 speech by Georg Gölter (CDU/CSU) begins as follows (we underline the important parts of the section):

'Instead of escalating this dispute, it would be much more sensible for us to agree today on the model for financing vocational education and training […] to develop the system so that vocational education and training pathways open up access to responsible professions in the state and society everywhere'.[5]

[4] In German: förderung, programm, verbesserung, investitionen, gefördert, region, regional.

[5] In German: 'Statt diesen Streit auf die Spitze zu treiben, wäre es doch viel vernünftiger, wir würden uns heute zur Finanzierung der beruflichen Bildung auf das Modell einigen, […] das berufliche Bildungssystem so auszubauen, dass auch berufliche Bildungswege überall den Zugang zu verantwortlichen Berufen in Staat und Gesellschaft eröffnen' .

Our STM assigns a high probability (52%) to the topic *financing higher education,* with additional weight on *vocational training* (8%) and *public school education* (7%). It also detects small overlaps with related themes, such as *poverty and distribution* (3%) of the topic prevalence, which becomes clearer as the speech continues:

'Ultimately, we will only be able to develop vocational education and training into an equivalent and equally valuable alternative to the Abitur and university studies if we complement the increasingly refined permeability of the education system with greater permeability in vocational and career pathways.[6]

We then combine these topics to the broader debate topic *Education*. This nuanced assignment highlights the model's ability to identify the multidimensional nature of political discourse, where a *single* speech often addresses *multiple* topics.

To construct topic-specific time series, we average annual θ values across speeches. Table 2 presents the descriptive statistics. On average, the economic topics account for between 0.4% (*Education*) and 0.9% (*Federal Budget*) of Bundestag discourse, consistent with the expected value for a 150-topic model ($1/150 \approx 0.6\%$). Despite these balanced averages, Table 2 indicates considerable temporal variation. For instance, *Education* ranges from 0.04% (in 1951) to 0.8% (in 2014), representing a twenty-fold increase.

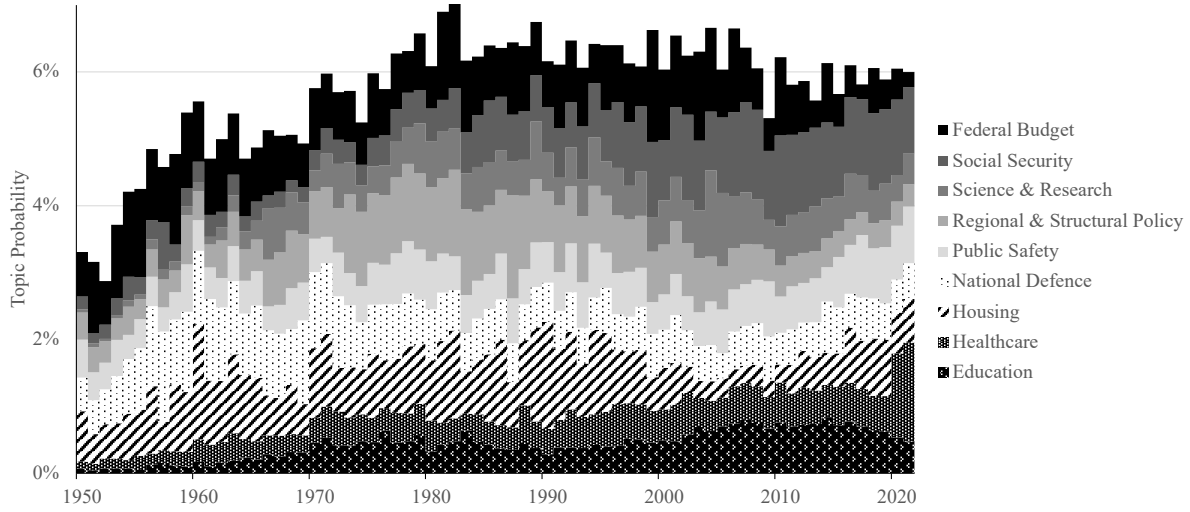
Table 2: Descriptive statistics of annualised STM topics in parliamentary debates

STM topic	Obs.	Mean	St. Dev.	Min	Max
Federal Budget	73	0.9%	0.3%	0.2%	1.7%
Education	73	0.4%	0.2%	0.04%	0.8%
Healthcare	73	0.4%	0.2%	0.1%	1.5%
Housing	73	0.8%	0.3%	0.2%	1.7%
National Defence	73	0.7%	0.2%	0.4%	1.4%
Public Safety	73	0.6%	0.1%	0.4%	0.9%
Regional & Structural Policy	73	0.7%	0.3%	0.2%	1.5%
Science & Research	73	0.5%	0.3%	0.03%	1.3%
Social Security	73	0.7%	0.4%	0.1%	1.4%

Figure 4 shows the temporal variation over time. Two patterns stand out. First, the relative share of economic topics consistently makes up approximately 6% of total Bundestag debates, indicating that variation is due to shifts across economics topics, rather than between economic and non-economic topics. Second, individual topics exhibit substantial shifts over time. For instance, *National Defence* has declined markedly in relevance, falling from about 1.5% of parliamentary speeches in the 1950s to less than 0.05% in the last years of our sample, whereas *Social Security* has risen substantially from 0.1% of speech topics in the 1950s to around 1% today. These time series form the basis for our dynamic analysis of fiscal policy debates in the Bundestag and actual federal expenditure.

[6] In German: 'Wir werden die berufliche Bildung letztlich nur zu einer gleichrangigen und gleichwertigen Alternative zu Abitur und Studium entwickeln können, wenn wir endlich die immer mehr perfektionierte Durchlässigkeit im Bildungswesen durch mehr Durchlässigkeit in den Berufs- und Laufbahnstrukturen ergänzen'. 'Abitur' is the German university-entrance qualification, roughly equivalent to A-levels in the UK or the completion of high school with advanced coursework in the US.

Figure 4: Relative importance of topics in parliamentary debates over time



4 Hypothesis Development

In this section, we develop hypotheses on the long-term relationship between parliamentary debates and federal spending in Germany. We first present overarching hypotheses before turning to topic-specific ones.

4.1 General Hypotheses

Our first hypothesis addresses the long-run relationship between Bundestag debates and the respective federal spending function. Since fiscal expenditure is typically the outcome of a political decision-making process, we propose:

> ***H1****: A long-term relationship exists between political debates on a topic and the corresponding fiscal expenditure.*

In Section 5, we conduct cointegration tests to examine the possible existence of long-term equilibria between these two variables.

The second hypothesis addresses the dynamic interaction between Bundestag debates and federal spending. Recent literature, such as Latifi et al. (2024) and Lieb et al. (2025), has adopted an input-output perspective of political decision-making by identifying shocks to debates or speeches to capture sudden changes in the government budget.[7] However, the reverse relationship is also plausible, as parliamentary discussion may be triggered by the current budget. For instance, a notable increase in an expenditure function might raise questions about

[7] Whether the government actually changes its budget plans following the debate, or implements the pre-debate plans, is an interesting question in itself. However, this is of secondary importance here and is therefore left for future research. Although our focus is on parliamentary debates and their influence on public spending, it should be noted that expenditure may be determined for various reasons, including political economy considerations, such as signalling competence and preferences through political budget cycles (e.g., Rogoff, 1990; Drazen and Eslava, 2010).

fiscal sustainability. Thus, it is unclear, *ex ante,* whether speeches drive expenditure or *vice versa*. To address this, we specify two competing hypotheses:

> ***H2a****: Speeches on a topic lead to changes in the corresponding expenditure function.*
>
> ***H2b****: Developments in a specific fiscal expenditure function lead to changes in parliamentary debates on this topic.*

In Section 5, we use weak exogeneity tests to investigate the direction of adjustment.

4.2 Specific Hypotheses

Conditional on the existence of a long-term relationship between expenditure and speeches, the sign of this relationship may vary by function. Thus, while hypotheses *H1* and *H2a/b* apply to all identified topics, we refine our analysis by formulating specific hypotheses for each of the nine spending topics. To derive hypotheses regarding the direction of the long-term relationship, we identify representative speeches for each spending function based on the topic distributions, following the fiscal policy topic modelling literature (e.g., Lieb et al., 2025). For example, Gülistan Yüksel's (SPD) speech in 2015 has a document-topic score of $\theta_d = 0.86$ for the topic *National Defence*, multiple standard deviations above its mean (see Table 2). We use the content of this representative speech to derive a topic-specific hypothesis regarding the direction of the relationship between debates on *National Defence* and federal spending on National Defence (discussed in Section 4.5). After systematically reviewing the most representative speeches for each of the nine topics, we present a synopsis of their content below. Extended quotations are available in the Online Appendix.

Federal Budget

Bundestag debates about the federal budget typically centre on issues of fiscal discipline, debt sustainability and budgetary constraints, often in response to economic downturns, revenue shortfalls or rising government expenditure. For example, Norbert Barthle (CDU/CSU) in 2005 explicitly addresses the tension between revenue generation and expenditure control:

'The federal budget has a structural funding gap of 50 billion euros. This represents 20 per cent of the approved total expenditure of 261 billion euros. […] Neither increasing revenues nor cutting expenditure alone is sufficient to consolidate the budget'.[8]

Such speeches highlight the recurring concern of uncontrolled spending growth on budgetary stability, necessitating corrective action in the form of tax increases or, more relevant to our focus, expenditure cuts. Given the emphasis on fiscal restraint, we expect Bundestag debates on the federal budget to be *negatively* related to government spending in the long run.

Education

Education spending in Germany is primarily determined and executed at the state level due to the *Kooperationsverbot*, a legal restriction preventing federal–state co-operation in education

[8] In German: 'Beim Bundeshaushalt haben wir eine strukturelle Deckungslücke von 50 Milliarden Euro. Das sind 20 Prozent der beschlossenen Gesamtausgaben in Höhe von 261 Milliarden Euro. [...] Nur durch eine Verbesserung der Einnahmebasis oder allein durch Ausgabenkürzungen kann man diesen Haushalt nicht konsolidieren'.

policy. Despite this, debates occasionally push for greater federal involvement, as in Rosemarie Hein's (Die Linke) speech in 2013:

'[…] about lifting the ban on cooperation between the federal and state governments in matters of education […] We need schools where social work is an integral part, where learners can seek counselling just as easily as parents and teachers. We need schools where every child is individually supported and receives the help they need'.[9]

This recurring theme of a stronger federal role in education in combination with calls for increased federal investment suggests a *positive* long-term relationship between Bundestag debates on education and federal education spending.

Healthcare

Only a small share of health spending in Germany is executed at the federal level (4% in 2019). Yet federal debates often address higher-order concerns such as cost control, service expansions, solidarity and coverage, rather than concrete policy proposals. For instance Ulrike Flach (FDP) stresses in 2010:

'We do not see individual responsibility and solidarity as opposites. […] Solidarity is necessary for large and expensive operations, such as heart surgery. On the other hand, however, we also need more individual responsibility'.[10]

Moreover, many of the debates on health care concern expenditures at the state or municipal level, even though they are discussed in the Bundestag. Given the prevalence of systematic concerns and the mismatch between debates and spending at the federal level, we remain agnostic about the expected long-term relationship between debates on healthcare and federal healthcare spending.

Housing

Bundestag discussions on housing policy consistently highlight the importance of financial assistance for low-income households. For example, Hildebrecht Braun (FDP) notes in 1994:

'For this reason, housing subsidies will also be targeted, in fiscal terms, at so-called threshold households – those who, under the previous subsidy framework, were barely or not at all able to afford homeownership'.[11]

As debates frequently call for expanding social housing programmes, we expect a *positive* long-run relationship between housing debates and federal spending on housing.

---

[9] In German: '[...] über die Aufhebung des Kooperationsverbotes zwischen Bund und Ländern in Bildungsfragen [...] Wir brauchen eine Schule, in der Schulsozialarbeit zur Selbstverständlichkeit gehört, wo sich Lernende ebenso beraten können wie Eltern und Lehrende, eine Schule, in der jedes Kind individuell gefördert wird und jedes Kind die Hilfen erhält, die es benötigt'.

[10] In German: 'Wir sehen in Eigenverantwortung und Solidarität keinen Gegensatz. Ich will das an einem Beispiel deutlich machen. Solidarität ist notwendig bei großen und teuren Operationen, wie zum Beispiel bei Herzoperationen. Auf der anderen Seite brauchen wir aber auch mehr Eigenverantwortung'.

[11] In German: 'Deswegen wird die Wohnbauförderung auch in steuerlicher Hinsicht auf die sogenannten Schwellenhaushalte konzentriert werden, also diejenigen, die nach der bisherigen Förderkonzeption nur mit größter Mühe oder gar nicht in der Lage waren, Eigentum zu schaffen'.

National Defence

Debates on national defence are driven by three major themes. The first is victim compensation (*Kriegsopferversorgung)* in the post-World War II period, as addressed by Eugen Glombig (SPD) in 1962:

'[...] the SPD parliamentary group presented a draft bill in December last year providing for a 10% increase in the basic pensions for disabled persons and widows, as well as in old-age pensions'.[12]

The remaining two themes centre on geopolitical shifts during the Cold War and international terrorism, both of which emphasise the costly nature of military engagement, as illustrated by Gülistan Yüksel (SPD) in 2015:

'We strongly support the federal government's efforts to intensify its actions against international terrorism, especially ISIS [...] In the 2016 federal budget, we have increased the allocation for [...] civilian crisis prevention'.[13]

All three themes emphasise an expansion in government spending. Therefore, we expect a *positive* long-run relationship between Bundestag debates on national defence and government expenditure.

Public Safety

Debates on public security have shifted over time. In the 1970s and mid-1980s, they focus on internal security and counterterrorism, as highlighted in the speech by Horst Eylmann (CDU/CSU) in 1986:

'The wave of terrorism in our country remains unbroken. The recent murders of Beckurts, Groppler and von Braunmühl reveal an almost unparalleled disregard for human life'.[14]

In the 2010s, debates have turned to data security and the protection of democratic institutions, as stressed by Niema Movassat (Die Linke) in 2019:

'For several years now, the rule of law and fundamental rights have been under attack by right-wing populists and extremists in Germany and Europe. We must engage young people in upholding the values of our constitution – equality, freedom and democracy'.[15]

Arguably, this engagement of young people has clear budgetary implications. We therefore expect a *positive* long-term relationship between debates and spending on public security.

---

[12] In German: '[…] hat die SPD-Fraktion schon im Dezember vorigen Jahres den Entwurf eines Gesetzes […] vorgelegt, der eine 10%ige Erhöhung der Grundrenten der Beschädigten und Witwen sowie der Altersrenten vorsieht'.

[13] In German: 'Wir unterstützen die Bundesregierung ausdrücklich darin, ihre Aktivitäten gegen den internationalen Terrorismus im Allgemeinen und gegen ISIS im Besonderen zu verstärken. [...] Im Haushalt 2016 haben wir den Ansatz für […] die zivile Krisenprävention erhöht'.

[14] In German: 'Die Terrorwelle in unserem Land ist ungebrochen. Die letzten Mordfälle Beckurts, Groppler und von Braunmühl zeigen eine kaum noch zu überbietende Geringschätzung des menschlichen Lebens'.

[15] In German: 'Seit einigen Jahren wird der Rechtsstaat und werden die Grundrechte von Rechtspopulisten und Rechtsextremisten in Deutschland und Europa wieder angegriffen. Wir müssen vor allem junge Menschen für die Werte des Grundgesetzes – Gleichheit, Freiheit und Demokratie – gewinnen'.

Regional & Structural Policy

Discussions on *Regional & Structural Policy* historically focus on industrial policy and regional economic growth. In the 1950s and 1960s, during the *Wirtschaftswunder* (economic miracle), debates highlighted support for structurally weak regions, as in the speech by Parliamentary Secretary Grüner in 1979:

'Under the Regional Action Programme "Ostbayerisches Fördergebiet", the government has allocated DM 238.8 million in investment subsidies and DM 232.4 million in GA funds to support structurally weak regions'.[16]

The 1980s marked a shift in debates towards industrial restructuring, particularly in the steel and coal industries. In 1987, Norbert Lammert (CDU/CSU) remarked:

'It is not possible to restore the profitability of steel companies and thereby secure jobs in the long term. Further reductions in capacity, including in German steel firms, are therefore painful but unavoidable'.[17]

Following reunification in the early 1990s, debates again focused on regional economic integration. Klaus Beckmann, Parliamentary Secretary, emphasised federal support in 1992:

'The entire territory of Mecklenburg-Western Pomerania is designated as an assisted area under the Joint Federal–State Programme for the Improvement of Regional Economic Structures. In addition, for 40% of this assisted area, a special programme within the "Aufschwung-Ost" initiative will provide a total of DM 300 million in additional funds in 1991 and 1992 – with equal contributions from the federal and state governments'.[18]

In light of these conflicting perspectives – debates emphasising expansionary policies in the 1960s and 1990s, but lacking in the 1970s and 1980s – we do not derive a hypothesis for the long-run relationship between parliamentary discussions on *Regional & Structural Policy* and corresponding federal spending.

Science & Research

From the 1950s to the 1980s, debates on research were largely framed as an economic necessity for advancing industrial modernisation. Gerold Benz (CDU/CSU) states this in 1973:

'The Federal Government intends to develop the Fraunhofer Society into a high-performing umbrella organisation for institutes of applied research. It is pursuing the following goals: it aims to provide industry and the state with a versatile research institution comprising qualified

[16] In German: 'Nach dem Regionalen Aktionsprogramm „Ostbayerisches Fördergebiet" […] sind] insgesamt 238,8 Millionen DM an Investitionszulagen und 232,4 Millionen DM an GA-Mitteln eingeplant'. 'GA' stands for 'Gemeinschaftsaufgabe Verbesserung der regionalen Wirtschaftsstruktur', the 'Joint Federal–State Programme for the Improvement of Regional Economic Structures'.

[17] In German: 'Die Wiederherstellung der Rentabilität der Stahlunternehmen und damit die dauerhafte Sicherung der Arbeitsplätze ist nicht möglich. Der weitere Abbau von Kapazitäten auch in deutschen Stahlunternehmen ist daher schmerzhaft, aber unvermeidlich'.

[18] In German: 'Das gesamte Gebiet des Landes Mecklenburg-Vorpommern ist Fördergebiet der Gemeinschaftsaufgabe „Verbesserung der regionalen Wirtschaftsstruktur". Darüber hinaus werden für 40% dieses Fördergebiets mit einem Sonderprogramm des Gemeinschaftswerkes „Aufschwung-Ost" in den Jahren 1991 und 1992 insgesamt zusätzliche Mittel in Höhe von 300 Millionen DM (Bund und Land je zu 50%) zur Verfügung gestellt'. 'Aufschwung-Ost' was a joint federal–state initiative launched after German reunification to support the economic development and modernisation of eastern Germany

institutes for contract research, in order to support technological development and fulfil public tasks'.[19]

In the 2000s, the focus shifts towards innovation-led growth in high-tech industries, with research linked to competition. Annette Schavan, Minister of Education and Research, states in 2005:

'Our research policy is based on three principles: First, excellence. We measure ourselves nationally and internationally against the best. [...] Therefore, we rely on competition in research funding. Second, priority for innovation. This applies to the entire innovation chain, from idea to product. Third, we will pool efforts across politics, science and industry, in both university and non-university research, spanning the humanities and natural sciences'.[20]

Given the continuous emphasis on expanding research funding in political debates, we expect a *positive* long-run relationship between Bundestag discussions on research and government investment in the sector.

Social Security

Debates on social security reflect disputes over the welfare state. On one side, left-wing parties advocate an expansion of social security such as in 2021 by Sören Pellmann (Die Linke):

'Participation is strengthened by improving participation benefits and guaranteeing the right to self-determination. […] Participation benefits for people with disabilities and chronic illnesses must […] fully meet needs and be independent of income and assets'. [21]

These positions are regularly countered by fiscally conservative arguments from centre-right parties or the governing coalitions, as in Tankred Schipanski's (CDU/CSU) speech in 2011:

'A social market economy means that economic rationality and social cohesion go hand in hand. [...] But the prosperity to be distributed must also be generated. You want to distribute it without generating it. That this does not work [...] should have become clear in light of the numerous cases of government debt in recent months'.[22]

Since the latter group holds decisive influence over budgetary decisions, we expect a *negative* long-run relationship between debates on social security and social security expenditure.

---

[19] In German: 'Die Bundesregierung beabsichtigt, die Fraunhofer-Gesellschaft zu einer leistungsfähigen Trägerorganisation für Institute der angewandten Forschung auszubauen. Sie verfolgt dabei folgende Ziele: Sie will der Wirtschaft und dem Staat zur Sicherung der technologischen Entwicklung und zur Erfüllung öffentlicher Aufgaben eine vielseitige Forschungseinrichtung mit qualifizierten Instituten für die Vertragsforschung zur Verfügung stellen'.

[20] In German: 'Unsere Forschungspolitik setzt auf drei Prinzipien: Erstens auf Exzellenz. Wir messen uns national und international an den Besten. [...] Deshalb setzen wir auf den Wettbewerb in der Forschungsförderung [...] Zweitens: Vorrang für Innovation. Das gilt für die gesamte Innovationskette von der Idee bis zum Produkt [...] Drittens: Wir werden Kräfte bündeln, und zwar in Politik, Wissenschaft und Wirtschaft, in universitärer und ausseruniversitärer Forschung, in den Geistes- und Naturwissenschaften'.

[21] In German: 'Teilhabe stärkt man, indem man Teilhabeleistungen verbessert und das Selbstbestimmungsrecht garantiert. […] Teilhabeleistungen für Menschen mit Behinderungen und chronischen Erkrankungen müssen […] bedarfsdeckend und Einkommens- und Vermögens unabhängig [sein]'.

[22] In German: 'Soziale Marktwirtschaft bedeutet, dass wirtschaftliche Vernunft und sozialer Zusammenhalt der Gesellschaft zusammengehören. [...] Der Wohlstand, der verteilt werden soll, muss aber auch erwirtschaftet werden. Sie wollen ihn verteilen, ohne ihn zu erwirtschaften. Dass das nicht funktioniert, [... muss] doch spätestens angesichts der zahlreichen Fälle von Staatsverschuldung in den vergangenen Monaten deutlich geworden sein'.

Summary of Specific Hypotheses

Our directional hypotheses for the long-run relationship between parliamentary debates and government expenditure are summarised in Table 3. Based on our reading of the debate content in the Bundestag, we expect positive relationships for *Education, National Defence, Science & Research, Housing* and *Public Security*, where debates call for increased spending. Conversely, we expect a negative relationship for the *Federal Budget* and *Social Security,* where debates centre on fiscal discipline. Due to inconsistencies in the discourse over time, we do not formulate a clear hypothesis for *Healthcare* and *Regional & Structural Policy*.

Table 3: Directional hypotheses for the long-run relationship between parliamentary debates and federal expenditure

STM topic	Hypotheses	STM topic	Hypotheses
Federal Budget	–	*Public Safety*	+
Education	+	*Regional & Structural Policy*	?
Healthcare	?	*Science & Research*	+
National Defence	+	*Social Security*	–
Housing	+		

5 Cointegration and error-correction models

Next we test, for each spending function, whether parliamentary debates (as measured in Section 4) and government spending (as measured in Section 3) form a long-term equilibrium – that is, by employing cointegration and error-correction models.

5.1 Econometric Approach

Even though we analyse a 70-year period, the use of annualised data constrains the total number of observations per spending function. This limitation reduces the precision of parameter estimates, particularly for capturing dynamic adjustments. Consequently, we focus on the bivariate relationship between each identified topic and its corresponding expenditure function. Immediate adjustments between spending and debates are unlikely due to decision-making lags, institutional delays and administrative processes. We therefore account for temporal dependencies, recognising that the effect of debates on spending unfolds over multiple periods. Consequently, we conduct our analysis within the framework of dynamic time-series models (Hendry et al. 1984).

First, we examine whether there is a long-term relationship between the topics of Bundestag debates and the respective federal government spending by conducting cointegration testing and estimation. The premise is that both time series are non-stationary, specifically, integrated of order one (I(1)), but share a common stochastic trend, which is captured by a linear cointegration vector. Unit root tests (Table A2 of the Appendix) show that our variables are I(1). The cointegration vector makes the combination of these variables stationary and describes their long-term equilibrium.

We determine the cointegration rank and the associated cointegrating vector using Johansen's (1988) method for estimating long-run relationships with vector autoregressions (VARs). Given

our extended observation period, this method should provide considerable statistical power.[23] In addition to capturing short-term dynamics in the estimation of cointegration vectors, vector error-correction models (VECMs) enable us to test for weak exogeneity of the variables of interest, as defined by Engle et al. (1983). This, in turn, indicates the direction of causality between the two variables following a deviation from the long-term equilibrium.

Second, we switch to the short-run dynamics and adjustment processes, that is, the rate of change of the variables in the VECM. Conditioning on weakly exogenous variables tends to improve the stochastic properties of the model, especially in smaller samples. If either the debates or the spending can be determined weakly exogenous, the two-equation VECM can be reduced to a one-equation error-correction model (ECM), a partial dynamic model conditional on the long-term equilibrium (Harbo et al. 1998). This simplification improves efficiency while preserving the relevant system dynamics. To optimise the models, we systematically reduce the number of lags and variables in a data-admissible way, following Hendry's (1993) general-to-specific methodology. These congruent reduced models form the basis of our analysis.

Finally, we conduct a series of tests to ensure the validity of our inferences. We use Breusch–Pagan tests for autocorrelation and LM tests for heteroskedasticity in their respective F-test forms, which are particularly suited for small samples (Kiviet 1987). We assess temporal stability by evaluating the VECMs in-sample and out-of-sample stability using 1-step-ahead Chow tests (Chow 1960) and recursive estimation of the adjustment parameter on the error-correction term (initialisation phase: 20 periods). For the Chow out-of-sample test, we exclude the last five years and use 2016–2020 to examine parameter constancy. Finally, we investigate the impact of government revenue on our VECMs to assess whether omitting the government's revenue side biases our models. To do so, we include the current value, along with its first and second lags, in our reduced VECMs/ECMs, testing whether they are jointly significant.

5.2 Long-Run Relationships: Cointegration Testing

We begin the cointegration analysis with a VAR model for the two variables of interest: federal expenditure and parliamentary debate on the corresponding function. Including four lags allows us to capture delayed dependencies up to the full duration of a legislative period in the German Bundestag. We include an impulse dummy for 1991 and a step dummy for 1950–1969. The former captures structural changes associated with reunification, while the latter accounts for the post-war recovery period.[24] A constant is included in the cointegrating vector (i.e. the long-run part of the model), which allows the cointegrating relationships to be trend stationary and to have non-zero intercepts. The cointegration tests address our Hypothesis 1, while the signs of the relationship assess the validity of the specific hypotheses listed in Table 3.

Omitting the deterministic components to avoid notational clutter, our VAR model is specified in equation (1):

$$(1)\ \begin{bmatrix} LnExp_t \\ LnDebate_t \end{bmatrix} = A_1 \begin{bmatrix} LnExp_{t-1} \\ LnDebate_{t-1} \end{bmatrix} + A_2 \begin{bmatrix} LnExp_{t-2} \\ LnDebate_{t-2} \end{bmatrix} + A_3 \begin{bmatrix} LnExp_{t-3} \\ LnDebate_{t-3} \end{bmatrix} + A_4 \begin{bmatrix} LnExp_{t-4} \\ LnDebate_{t-4} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

[23] In cointegration analysis, it is the time span covered by the data that is important, not the number of observations per se. Put differently, sampling data at a higher frequency does not facilitate the identification of cointegration relationships (Engle and Granger 1987).

[24] We do not include the dummy for 1950 to 1969 in the case of Federal Budget, Regional & Structural Policy, and Social Security, as we did not find significant cointegrating vectors when including the dummy.

This can be transformed into:

$$(2)\ \begin{bmatrix} \Delta LnExp_t \\ \Delta LnDebate_t \end{bmatrix} = \Pi \begin{bmatrix} LnExp_{t-1} \\ LnDebate_{t-1} \end{bmatrix} + \Gamma_1 \begin{bmatrix} \Delta LnExp_{t-1} \\ \Delta LnDebate_{t-1} \end{bmatrix} + \Gamma_2 \begin{bmatrix} \Delta LnExp_{t-2} \\ \Delta LnDebate_{t-2} \end{bmatrix} + \Gamma_3 \begin{bmatrix} \Delta LnExp_{t-3} \\ \Delta LnDebate_{t-3} \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix},$$

where $\Gamma_1 = -(A_2 + A_3 + A_4)$, $\Gamma_2 = -(A_3 + A_4)$, $\Gamma_3 = -(A_4)$, and $\Pi = -(I - A_1 - A_2 - A_3 - A_4)$.

Assuming a rank of matrix $\Pi = 1$, that is, a single cointegrating relationship, we can state the following:

$$(3)\ \Pi \begin{bmatrix} LnExp_{t-1} \\ LnDebate_{t-1} \end{bmatrix} = \alpha\beta' \begin{bmatrix} LnExp_{t-1} \\ LnDebate_{t-1} \end{bmatrix},$$

here $\beta' = [\beta_{LnExp}, \beta_{LnDebate}]$ is the 2×1 cointegrating vector capturing the long-term relationship between federal expenditure and parliamentary debates. To facilitate interpretation, we standardise $\beta_{LnExp} = 1$ so that $\beta_{LnDebate}$ describes the long-term elasticity between debates and expenditure. The corresponding adjustment coefficients $\alpha = \begin{bmatrix} \alpha_{LnExp} \\ \alpha_{LnDebate} \end{bmatrix}$ indicate the speed and direction in which a variable corrects its deviation from the long-term equilibrium. Negative values indicate movement towards equilibrium, and larger values in absolute terms indicate faster correction. The relative size of α helps identify the adjustment process. If α_{LnExp} is large and significant, while $\alpha_{LnDebate}$ is close to zero, expenditure adjusts to debates rather than the reverse. The outcome of the weak exogeneity tests on the estimated adjustment coefficients therefore directly informs Hypotheses 2a and 2b.

Take, for instance, the recent surge in *National Defence*-related debates following the invasion of Ukraine. Spending and parliamentary debates on defence exhibit a positive long-term relationship (see below). The sharp rise in defence discussions represents a short-term deviation from the long-term equilibrium, signalling a shift in policy priorities and resulting in a gradual increase in defence spending. This adjustment is captured by the model, with α_{LnExp} measuring the speed at which expenditure realigns with the long-term equilibrium defined by the debates.

Since we restrict the constant to the long-term part of the VAR, the right-hand side of equation (3) becomes:

$$(4)\ \Pi \begin{bmatrix} LnExp_{t-1} \\ LnDebate_{t-1} \end{bmatrix} = \begin{bmatrix} \alpha_{LnExp} \\ \alpha_{LnDebate} \end{bmatrix} [\beta_1 LnExp_{t-1} + \beta_2\ LnDebate_{t-1} + \beta_3\ Constant]$$

To facilitate interpretation, we standardise the cointegrating vectors with respect to government expenditure, as indicated above:

$$(5)\ \Pi \begin{bmatrix} LnExp_{t-1} \\ LnDebate_{t-1} \end{bmatrix} = \begin{bmatrix} \tilde{\alpha}_{LnExp} \\ \tilde{\alpha}_{LnDebate} \end{bmatrix} [\tilde{\beta}_{LnExp} LnExp_{t-1} + \tilde{\beta}_{LnSpeech}\ LnDebate_{t-1} + \tilde{\beta}_{Constant}\ Constant]$$

where $\tilde{\alpha}_{LnExp} = \alpha_{LnExp}\, \beta_{LnExp}$, $\tilde{\alpha}_{LnDebate} = \alpha_{LnDebate}\, \beta_{LnExp}$, $\tilde{\beta}_{LnExp} = 1$, $\tilde{\beta}_{LnDebate} = \beta_{LnDebate}/\beta_{LnExp}$, $and\ \tilde{\beta}_{Constant} = \beta_{Constant}/\beta_{LnExp}$

The results for estimating equation (2) with the standardisation proposed in equation (5) are reported in Table 4, where each row corresponds to a specific spending function. The second column reports the trace test statistics of the cointegration test, which suggest significant long-term relationships between Bundestag debates and the corresponding expenditure across all spending functions, supporting Hypothesis 1.[25]

Table 4: Cointegration vectors tests and estimates (estimation period: 1954–2015)

Spending Function	Trace test	$\tilde{\beta}_{LnExp}$	$\tilde{\beta}_{LnSpeech}$	$\tilde{\beta}_{Constant}$
Federal Budget	27.9**	1	−1.20	0.3
Education	30.5**	1	0.86	6.1
Healthcare	36.7**	1	−0.24	1.1
Housing	32.5**	1	0.38	2.5
National Defence	20.2*	1	2.75	17.5
Public Safety	24.2^{+}	1	1.03	0.03
Regional & Structural Policy	22.8*	1	−0.49	0.2
Science & Research	27.8**	1	0.26	3.5
Social Security	28.4**	1	−0.54	3.2

Note: * and ** indicate significance at a 5% and 1% level, respectively. The column 'trace test' gives the trace test statistics when testing two cointegration vectors. Here, the hypothesis of one cointegration vector is never rejected and we omit the information to economise on space. $^{+}$: A trend is included in the cointegration vector, which is only significant at the 10% level.

The cointegration vector is shown in the last three columns of Table 4. Due to the standardisation, the estimated parameters of the cointegrating vector are in the following form (with $\tilde{\beta}_{LnExp} = 1$):

$$(6)\ \tilde{\beta}_{LnExp} LnExp_{t-1} = \tilde{\beta}_{LnDebate}\, LnDebate_{t-1} - \tilde{\beta}_{Constant}\, Constant$$

We find the following. First, there is considerable variability in the absolute magnitude of the estimated elasticities of federal expenditure with respect to parliamentary debates, ranging from being inelastic in the case of Science & Research spending (almost 0.3) to highly elastic in the case of National Defence spending (almost 3). In other words, an increase in parliamentary debates about National Defence (Science & Research) by 1%, raises National Defence spending (Science & Research spending) by almost 3% (0.3%) in the long run. This contrast reflects the differing nature of these expenditure functions, where most of research spending is tied to established long-term funding mechanisms (such as universities), whereas national defence spending responds to immediate external threats, leading to a more pronounced reaction to changes in political debates. Importantly, the cointegration vectors are stable and well-behaved over time, as illustrated in the upper-left panels of Figures 5–13.

Second, we find both positive and negative long-term relationships. In five of the nine spending functions, debates positively relate to government spending in the long-term, consistent with the notion that increased parliamentary attention supports higher expenditures. In contrast, four categories – Federal Budget, Healthcare, Regional & Structural Policy, and Social Security – exhibit a negative long-term relationship with their respective policy debates, suggesting that increased Bundestag debates drive contractionary measures in those functions. However, except for the federal budget, these negative relationships are inelastic. Overall, these results

[25] The trace test is generally preferable to the maximum likelihood statistic (Johansen 1996), as a sequence of trace tests yields a consistent test procedure, whereas there is no comparable result for the maximum eigenvalue test.

support the function-specific hypotheses listed in Table 3, with the additional finding that the unclear relationships in the case of Healthcare and Regional & Structural Policy are estimated to be negative.

Next, we systematically simplify our two-equation VECM, following Hendry's (1993) general-to-specific modelling approach. Testing for weak exogeneity of either government expenditure or parliamentary debate, the hypothesis of weak exogeneity of debate topics (i.e. $\Delta LnDebate_t$) cannot be rejected, providing support for Hypothesis 2a. Weak exogeneity for debate topics is relevant for two reasons. First, parliamentary debates do not adjust to correct for long-term disequilibria between debates and spending. Second, as a result, the debate equation (second row of the system in equation 2) does not contain information about the long-run parameters in the cointegrating vector (Banerjee et al. 1993). Consequently, the system can be simplified to a single-equation ECM for $\Delta LnExp_t$.

5.3 Short-Run Relationships: Error Correction

We start again with the full VAR equation for $\Delta LnExp_t$ and then reduce the model consistently from general to specific. By imposing the weak exogeneity condition on $\Delta LnDebate_t$, we can estimate the short-term adjustment dynamics more efficiently. Table 5 summarises the results for the final ECMs. Several general results emerge.

First, since the testing-down restriction (see column 4 of Table 5) is never rejected, we can simplify the models to the reduced ECMs (columns 5 and 6). Second, most models pass the diagnostic tests for autocorrelation (column 8) and heteroscedasticity (column 9), indicating well-behaved residuals. In cases of rejection of either test, we employ Newey–West standard errors to conduct our significance tests in the ECMs.[26] Third, the reduced ECMs are stable and consistently pass out-of-sample Chow tests (column 10) for the one-step forecast covering the years 2016–2020. Fourth, most models contain relatively few dynamic terms and the error-correction term appears to play an important role in shaping the short-term dynamics. Finally, the inclusion of federal government revenue (column 11) has no significant impact on the relative rates of change across the different federal government expenditure functions, indicating that estimates are not affected by omitted variable bias arising from the revenue side of the government budget.

[26] Although the Newey-West estimator is biased, it is consistent even in the presence of non-spherical errors. If a rejection of a diagnostic test triggers the use of Newey-West standard errors, we base our modelling choices on these results.

Table 5: Error-correction models for government expenditure $\Delta LnExp$ (estimation period: 1954−2015)

Spending functions	LnDebate weakly exogenous?	$\tilde{\alpha}_{LnExp}$	Testing-down restriction	Lags i: $\Delta LnExp_{t-i}$	Lags i: $\Delta LnDebate_{t-i}$	Deterministic terms	AR(3) test	Heteroske-dasticity test	1-step forecast test 2016-2020	F-test γ_i: $\sum_{i=0}^{2}\gamma_i\,\Delta LnGovI$
Federal Budget	Yes	-0.05*	F(3,52)=1.51	2, 3	2	Const., D91	F(3,52)=3.26*	F(8,51)=0.87	F(5,55)=0.61	F(3,51)=0.92
Education	Yes	-0.23*	F(6,51)=0.88		1	D50-69, D91	F(3,54)=0.66	F(5,54)=1.1	F(5,57)=0.05	F(3,54)=1.02
Healthcare	Yes[a]	-0.28**	F(8,52)=1.07				F(3,57)=0.11	F(2,58)=0.27	F(5,60)=0.88	F(3,57)=1.14
Housing	Yes	-0.60**	F(7,51)=1.98	1	1		F(3,55)=2.03	F(6,54)=0.33	F(5,58)=0.58	F(3,55)=1.40
National Defence	Yes[i]	-0.05*	F(7,51)=1.07		1	D50-69	F(3,55)=0.53	F(5,55)=3.37*	F(5,58)=0.33	F(3,54)=1.85
Regional & Structural Policy	Yes	-0.13**	F(6,52)=0.44		1	D91	F(3,55)=0.24	F(4,55)=1.2**	F(5,58)=0.66	F(3,55)=0.06
Public Safety	Yes	-0.17*	F(5,51)=0.78	3	2	Const.	F(3,54)=1.17	F(6,54)=3.68**	F(5,56)=0.23	F(3,53)=1.50
Science & Research	Yes	-0.14**	F(7,51)=1.12	1	2		F(3,54)=0.87	F(6,54)=1.2	F(5,59)=0.56	F(3,55)=0.22
Social Security	Yes	-0.04**	F(7,51)=1.30			D50-69, D91	F(3,55)=3.39*	F(3,56)=3.50*	F(5,58)=0.25	F(3,55)=1.01

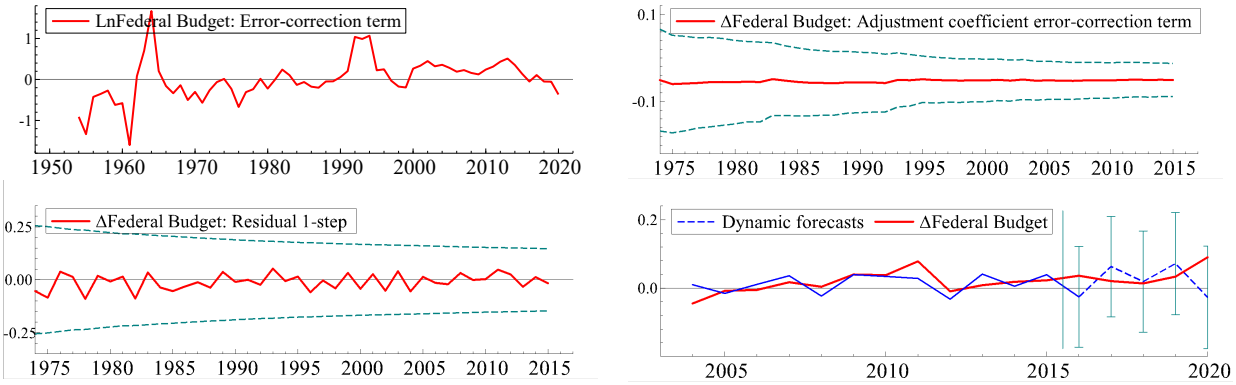
Note: * and ** indicate significance at a 5% and 1% level, respectively. [i]: In the $\Delta LnDebate$ equation, the error-correction term takes on a positive value. [a]: Weakly exogenous when considering $\Delta LnGovRev$ in the $\Delta LnDebate$ equation. Newey-West standard errors are used when the tests for autocorrelation and/or heteroscedasticity are significant.

Building on these general findings, we next present specific results for the federal budget and our eight spending functions.

Federal Budget

We find a negative long-run relationship between parliamentary debates and Federal Budget. Overall federal spending declines in a proportionately elastic manner (elasticity: −1.2), suggesting that *Federal Budget* becomes a focal topic in Parliament during periods of anticipated fiscal constraint. However, short-term adjustment is sluggish: only 5% of the deviation from the long-term equilibrium is corrected through spending cuts each year. Figure 5 confirms that the cointegration term is stationary, the adjustment coefficient stable, the one-step residuals are random, and the out-of-sample predictions lie within the 95% confidence bands.

Figure 5: Federal Budget: Error-correction term, recursive estimation of adjustment coefficient on error-correction term, 1-step residual, and 1-step out-of-sample dynamic forecasts.

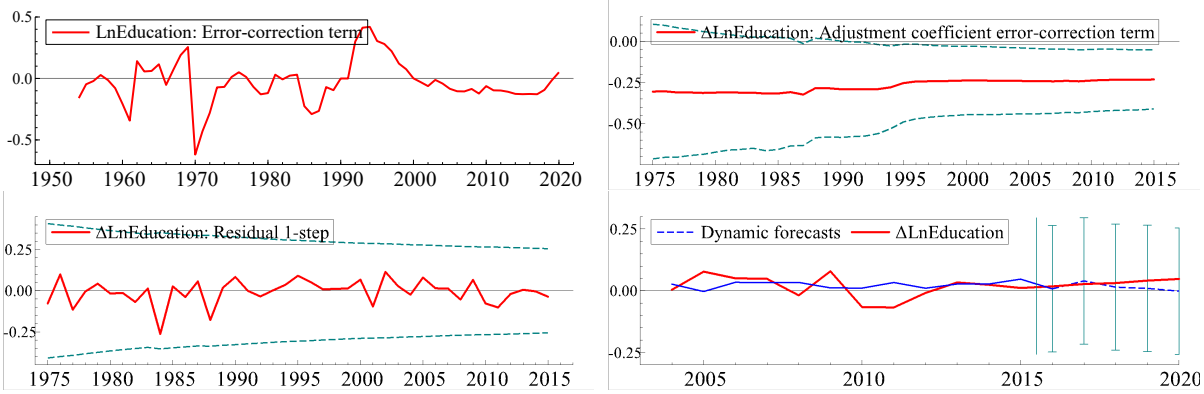


Note: Confidence bands are based on +/- 2 standard errors.

Education

Education is positively but inelastically related to spending on education (elasticity: 0.9). The speed of adjustment is relatively fast: almost 25% of the deviation from the long-term equilibrium is corrected within a year. Figure 6 shows stability in the adjustment coefficient, a stationary cointegration term, random residuals, and robust out-of-sample forecasts.

Figure 6: Education: Error-correction term, recursive estimation of adjustment coefficient on error-correction term, 1-step residual, and out-of-sample dynamic forecasts.

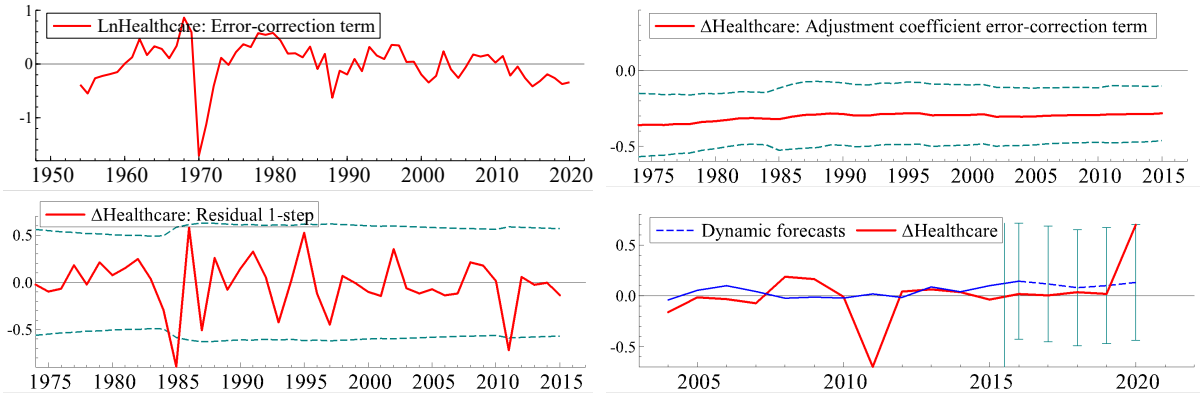


Note: Confidence bands are based on +/- 2 standard errors.

Healthcare

The long-run relationship between Healthcare and parliamentary debates is negative and inelastic (elasticity: −0.2). The resulting spending cuts are implemented relatively quickly, as almost 30% of the discrepancy between current spending and the long-term equilibrium is adjusted within one year. The general-to-specific approach yields an ECM that contains only the error-correction term and no other short-term dynamics, though significant residuals arise twice in-sample (Figure 7, third panel). Nevertheless, the cointegration vector is stationary, the adjustment coefficient is stable, and the predictions lie within the confidence bands. The jump in the growth rate of Healthcare in the last out-of-sample observation reflects the beginning of the COVID-19 pandemic.

Figure 7: Healthcare: Error-correction term, recursive estimation of adjustment coefficient on error-correction term, 1-step residual, and 1-step out-of-sample dynamic forecasts.

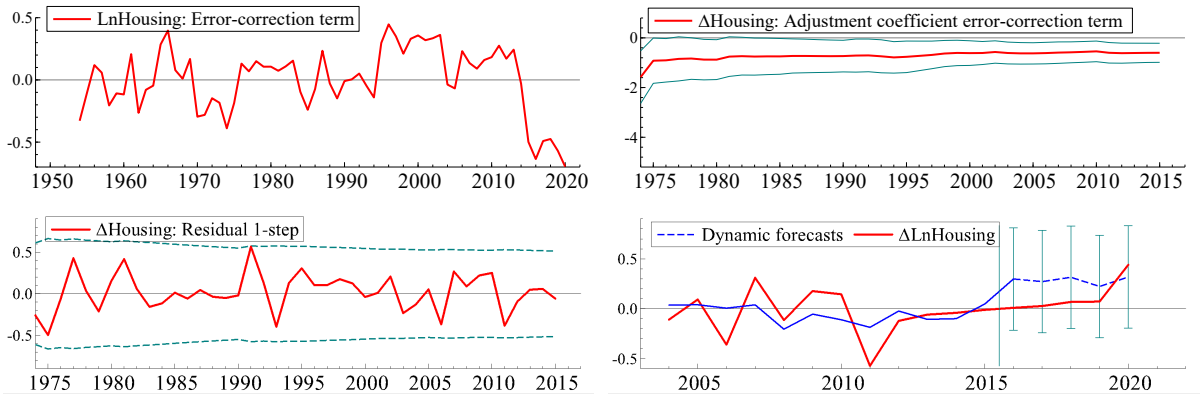


Note: Confidence bands are based on +/- 2 standard errors.

Housing

Housing expenditure and parliamentary debates on housing are positively related, albeit inelastically (elasticity: 0.4). In the short term, a deviation from this long-term equilibrium leads to a rapid adjustment of Housing, with more than 60% of the deviation being corrected within one year. Housing exhibits the fastest adjustment speed of all our expenditure functions. Figure 8 illustrates the diagnostic properties of the model. The cointegration vector appears stationary, except during the out-of-sample period. The adjustment coefficient is very stable over time, the residuals are well-behaved, and the dynamic forecasts fall within their confidence bands.

Figure 8: Housing: Error-correction term, recursive estimation of adjustment coefficient on error-correction term, 1-step residual, and 1-step out-of-sample dynamic forecasts.

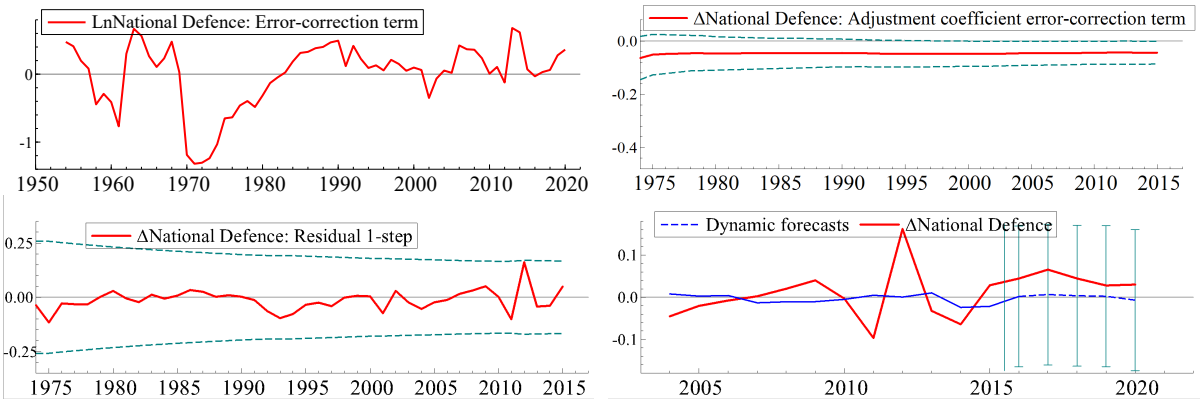


Note: Confidence bands are based on +/- 2 standard errors.

National Defence

Defence spending is positively related to debates on defence and highly elastic (elasticity: 2.8), though the adjustment speed is slow, with a 5% annual correction. The parsimonious ECM includes only the error-correction term. Figure 9 highlights stationarity, stability, and strong forecast performance of this specification.

Figure 9: National Defence: Error-correction term, recursive estimation of adjustment coefficient on error-correction term, 1-step residual, and 1-step out-of-sample dynamic forecasts.

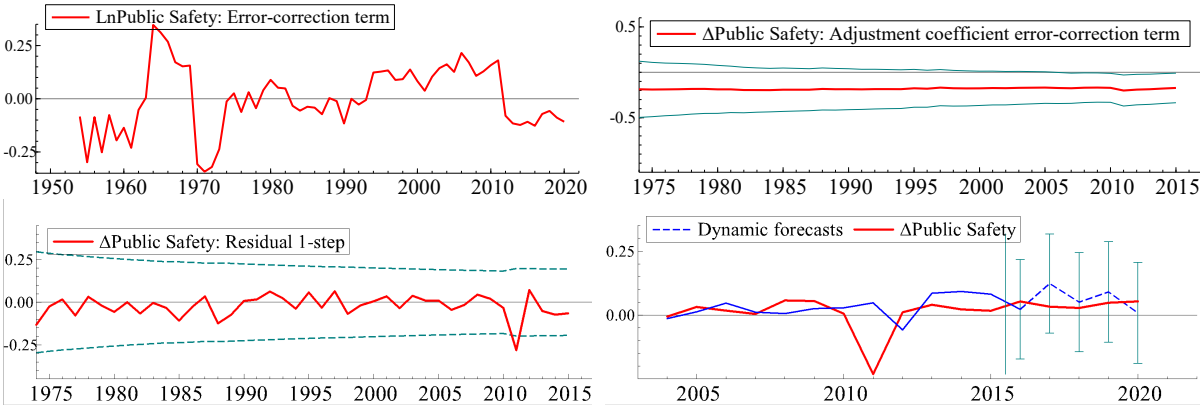


Note: Confidence bands are based on +/- 2 standard errors.

Public Safety

Public Safety responds positively and close to elastic (elasticity: 1) to parliamentary debates on the subject. With a correction rate of less than 20% per year, the short-term adjustment is somewhat slow. The diagnostic analysis set out in Figure 10 indicates no major issues. The error-correction term is stationary and its influence on short-term expenditure dynamics is constant over time, residuals are random except in 2011, and the dynamic forecasts remain well within the 95% confidence bands.

Figure 10: Public Safety: Error-correction term, recursive estimation of adjustment coefficient on error-correction term, 1-step residual, and 1-step out-of-sample dynamic forecasts.



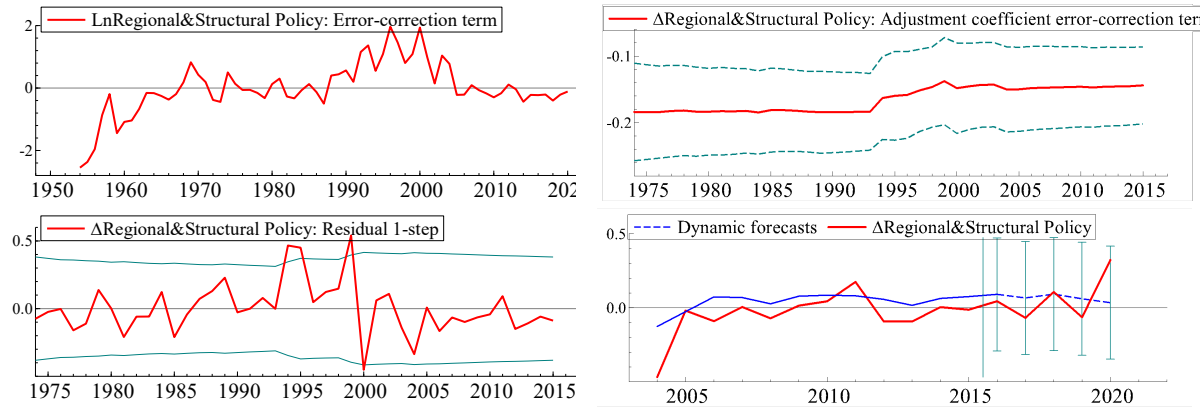
Note: Confidence bands are based on +/- 2 standard errors.

Regional & Structural Policy

In the long run, Regional & Structural Policy is negatively and inelastically associated with parliamentary debates on this topic (elasticity: −0.5). Adjustment is slow at 13% per year. Deviations from the long-term equilibrium are evident during the 1990s, reflecting the

turbulence caused by German reunification, which was a particularly important event for Regional & Structural Policy (see Chapter 4.2). However, following this period, the model again behaves well, including in its out-of-sample performance.

Figure 11: Regional & Structural Policy: Error-correction term, recursive estimation of adjustment coefficient on error-correction term, 1-step residual, and 1-step out-of-sample dynamic forecasts.

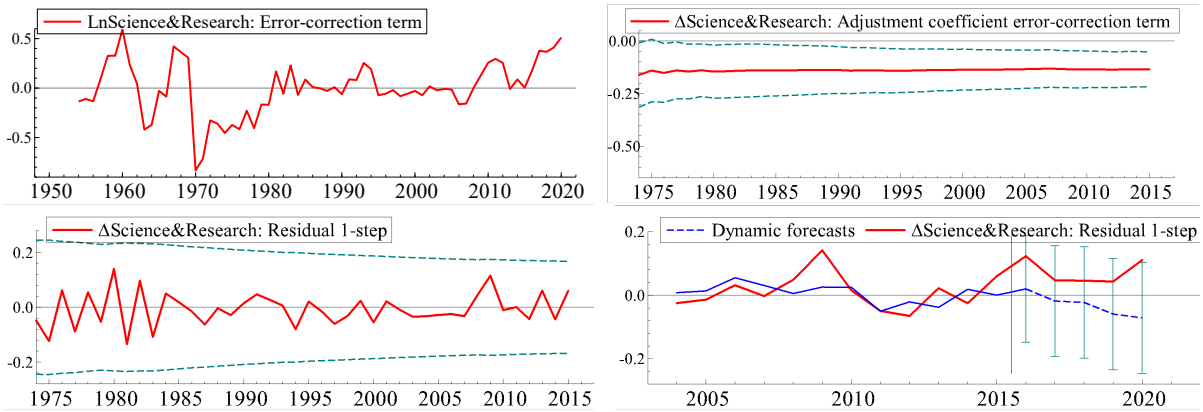


Note: Confidence bands are based on +/- 2 standard errors.

Science & Research

The long-term relationship between Science & Research and parliamentary debates is positive, but inelastic (elasticity: 0.3). The rate of adjustment to the long-run equilibrium is relatively slow with 15% per year. Figure 12 shows that the error-correction term is stationary, although there is considerable variation at the beginning of the series and, to a lesser extent, at the end. From 1980 to 2005 debate and expenditure are closely aligned with their long-term equilibrium. The adjustment coefficient exhibits very little variation, residuals are random and stable, and out-of-sample performance is satisfactory, apart from the 2020 observation.

Figure 12: Science & Research: Error-correction term, recursive estimation of adjustment coefficient on error-correction term, 1-step residual, and 1-step out-of-sample dynamic forecasts.



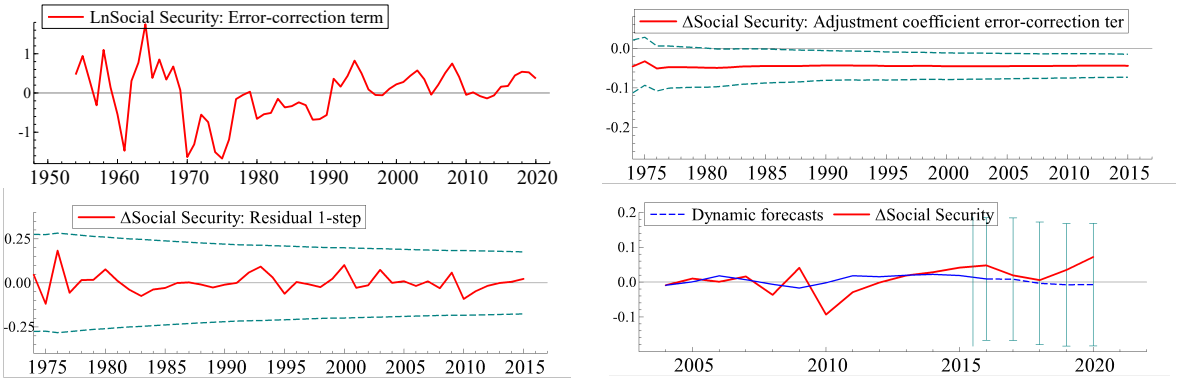
Note: Confidence bands are based on +/- 2 standard errors.

Social Security

We find a negative and inelastic long-run relationship for Social Security (elasticity: −0.8). The rate of adjustment is slow at around 4% per year. The long-term equilibrium appears to be

stationary (except 1970–1990, when there was a downward deviation), the adjustment coefficient is constant, and forecasts are reliable (Figure 13).

Figure 13: Social Security: Error-correction term, recursive estimation of adjustment coefficient on error-correction term, 1-step residual, and 1-step out-of-sample dynamic forecasts.



Note: Confidence bands are based on +/- 2 standard errors.

5.4 Comparative Findings

To facilitate a comparison across government functions, we summarise the main findings of our analysis in Table 6. Overall, the results support our general and function-specific hypotheses: Education, Housing, National Defence, Public Safety and Science & Research all exhibit positive long-term relationships with parliamentary debates, whereas the Federal Budget, Healthcare and Social Security show negative long-term relationships. For Regional & Structural Policy, for which no *a priori* hypothesis was formulated, we also observe a negative relationship.

Table 6: Summary cointegration results: Sign and speed of adjustment across spending functions

		Long-term relationship	
		Positive	Negative
Adjustment	Fast	– Education – Housing	– Healthcare
	Slow	– National Defence* – Public Safety* – Science & Research	– Federal Budget* – Regional & Structural Policy – Social Security

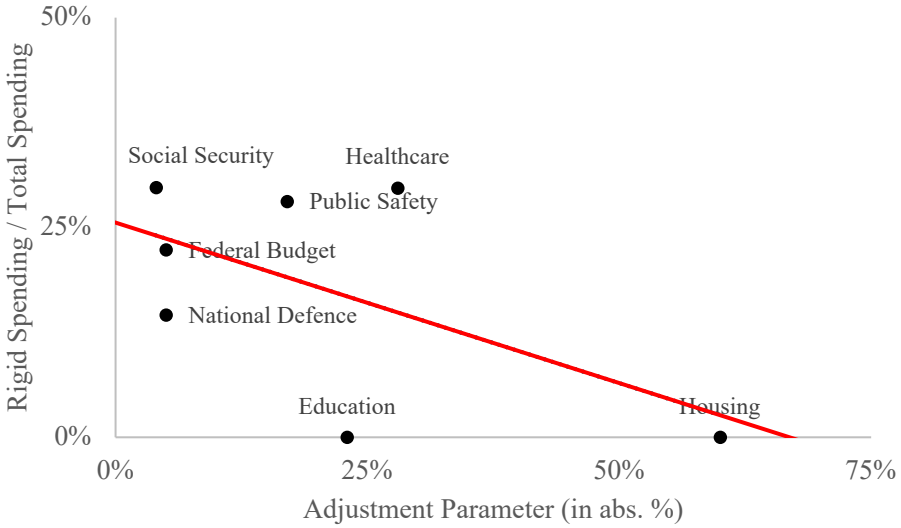
Note: Adjustment speed is categorised as fast (>25% correction per year) and slow (<25%). 'Long-term relationship' refers to the sign of the estimated elasticity between speeches and expenditure for each government function; (*) denotes an elastic response of expenditure to debates.

As Table 6 shows, elasticities are strongest in traditional core government functions – Federal Budget, National Defence and Public Safety – suggesting high responsiveness to long-term shifts in political attention, although expenditure adjustment is slow. In contrast, for publicly salient areas – Housing and Healthcare – we find faster adjustment dynamics, suggesting that political discourse in these domains quickly translates into expenditure changes.

Evidence from the Eurostat COFOG data reinforces this interpretation. For a subset of functions, we match function-specific spending to subcategories and classify them into *rigid* components (hard to adjust within a year, e.g. compensations, pensions and statutory transfers) and *flexible* components (easier to adjust, e.g. investment, subsidies and intermediate consumption).

Figure 14 shows that the functions' share of rigid spending negatively correlates with the corresponding speed of fiscal adjustment (slope = –0.38, R^2 = 0.33). Thus, policy functions dominated by rigid items, such as Social Security and Public Safety, display systematically lower adjustment parameters, whereas more flexible functions such as housing and healthcare adjust more quickly. These findings support our claim that the short- and long-run effects of parliamentary debates vary systematically across policy domains.

Figure 14: Rigidity in spending vs VECM adjustment parameter



Note: Rigid and flexible components are defined using Eurostat COFOG expenditure subcategories. Spending functions are matched to subcategories where possible. Expenditure values are computed as averages over 2015–2020. Rigid spending includes compensation of employees, social transfers (in kind and other than in kind), property income, other taxes on production, current taxes on income and wealth, and the adjustment for the change in pension entitlements. The adjustment parameters are taken from Table 5.

6 Conclusion

This paper contributes to the growing body of literature on the role of parliamentary communication in fiscal policymaking. Prior work (e.g., Latifi et al., 2024; Lieb et al., 2025) has focused on how shocks to political debates influence future expenditure decisions in the short to medium term. However, the long-term relationship between parliamentary debate and public spending, as well as the adjustment dynamics linking them, has yet to be examined empirically.

Addressing this gap, we explicitly estimate the long-run relationship between parliamentary debates and federal spending in Germany over a 70-year period. Specifically, we analyse the co-movement between discourse in the national parliament and spending across multiple government functions, assessing whether debates drive or respond to changes in fiscal outcomes.

To accomplish this, we have constructed two novel datasets: a harmonised annual series of German federal government expenditure spanning 70 years, disaggregated by policy function; and a parliamentary speech corpus spanning the same period, categorised using state-of-the-art natural language processing to assign each speech to the relevant fiscal function. Using (vector) error-correction models, we obtain three main findings.

First, we find evidence of a long-term equilibrium between parliamentary debates and expenditure across all analysed policy functions. The strength and direction of this relationship is function-specific: topics such as Education, Housing and National Defence exhibit positive relationships, reflecting debates focusing on an expansion of government activity, while topics such as Healthcare or the Federal Budget itself show a negative relationship, representing a parliamentary discourse that is centred on reining in what is considered excessive spending.

Second, adjustments to deviations from this long-term equilibrium occur through government expenditure rather than shifts in parliamentary debates. In other words, parliamentary debates are weakly exogenous with respect to spending. This suggests that political discourse in the short term shapes rather than reacts to fiscal expenditure.

Finally, we find that the speed of adjustment varies across government functions. Our results suggest that spending in politically salient areas, such as Housing and Healthcare, tends to adjust more rapidly, whereas spending on National Defence and Public Safety reacts more slowly, potentially due to long-term budgetary planning structures.

Overall, our findings provide new evidence that parliamentary discourse is not merely rhetorical but influential for fiscal outcomes. These results highlight the importance of studying political communication as a substantive component of the policy process rather than as merely symbolic action.

References

Abercrombie, G. & Batista-Navarro, R. (2020). Sentiment and position-taking analysis of parliamentary debates: a systematic literature review. *Journal of Computational Social Science*, *3*(1), 245–270.

Ash, E. & Hansen, S. (2023). Text algorithms in economics. *Annual Review of Economics*, *15*, 659–688.

Banerjee, A., Dolado, J., Galbraith, J.W. & Hendry, D.F. (1993). *Co-integration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.

Baumgärtner, M. & Zahner, J. (2025). Whatever it takes to understand a central banker: Embedding their words using neural networks. *Journal of International Economics*, forthcoming.

Blei, D. M., Ng, A.Y. & Jordan, M.I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Bohl, M.T., Kanelis, D. & Siklos, P.L. (2023). Central bank mandates: How differences can influence the content and tone of central bank communication. *Journal of International Money and Finance*, *130*, 102752.

Campiglio, E., Deyris, J., Romelli, D. & Scalisi, G. (2025). Warning words in a warming world: Central bank communication and climate change. *European Economic Review*, 178, 105101.

Chow, G.C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28, 591–605.

Engle, R.F. & Granger, C.W.J. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica*, 55, 251–276.

Engle, R.F., Hendry, D.F. & Richard, J.-F. (1983). Exogeneity. *Econometrica*, 51, 277–304.

Eslava, M. (2011). The political economy of fiscal deficits: A survey. *Journal of Economic Surveys*, 25, 645–673.

Ferrara, F.M., Masciandaro, D., Moschella, M. & Romelli, D. (2022). Political voice on monetary policy: Evidence from the parliamentary hearings of the European Central Bank. *European Journal of Political Economy*, *74*, 102143.

Hansen, S. & McMahon, M. (2016). Shocking language: Understanding the macroeconomic effects of central bank communication. *Journal of International Economics*, *99*, S114–S133.

Hansen, S., McMahon, M. & Tong, M. (2019). The long-run information effect of central bank communication. *Journal of Monetary Economics*, *108*, 185–202.

Harbo, I., Johansen, S., Nielsen, B. & Rahbek, A. (1998). Asymptotic inference on the cointegrating rank in partial systems. *Journal of Business and Economic* Statistics, 16, 388–399.

Hayo, B. & Zahner, J. (2023). What is that noise? Analysing sentiment-based variation in central bank communication. *Economics Letters*, *222*, 110962.

Hayo, B. & Uhl, M. (2014). The effects of legislated tax changes in Germany, *Oxford Economic Papers,* 66, 397–418.

Hendry, D.F. (1993). *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers.

Hendry, D.F., Pagan, A. R., & Sargan, J.D. (1984). Dynamic specification. In Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, Vol. 2, 1023–1100. Amsterdam: North-Holland.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and* Control, 12, 231–254.

Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, 2nd ed. Oxford: Oxford University Press.

Kanelis, D., Kranzmann, L., & Siklos, P.L. (2025). The financial instability–monetary policy nexus: Evidence from the FOMC minutes. *Bundesbank Discussion Paper* 13/2025.

Kiviet, J.F. (1987). *Testing Linear Econometric Models*. Amsterdam: University of Amsterdam.

Larsen, V.H., Thorsrud, L.A. & Zhulanova, J. (2021). News-driven inflation expectations and information rigidities. *Journal of Monetary Economics*, *117*, 507–520.

Latifi, A., Naboka-Krell, V., Tillmann, P. & Winker, P. (2024). Fiscal policy in the Bundestag: Textual analysis and macroeconomic effects. *European Economic Review*, *168*, 104827.

Latifi, A., Naboka-Krell, V., Tillmann, P. & Winker, P. (2025). Disagreement about fiscal policy. *Mimeo* (SSRN 5084219).

Lieb, L., Jassem, A., Almeida, R.J., Baştürk, N. & Smeekes, S. (2025). Min (d) ing the President: A text analytic approach to measuring tax news. *American Economic Journal: Macroeconomics,* forthcoming.

Roberts, M.E., Stewart, B.M. & Airoldi, E.M. (2016). A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*, *111*(515), 988–1003.

Rogoff, K. (1990). Equilibrium political budget cycles. *American Economic Review*, 80, 21–36.

Svaleryd, H. (2009). Women's representation and public spending. *European Journal of Political Economy*, *25*(2), 186–198.

Appendix

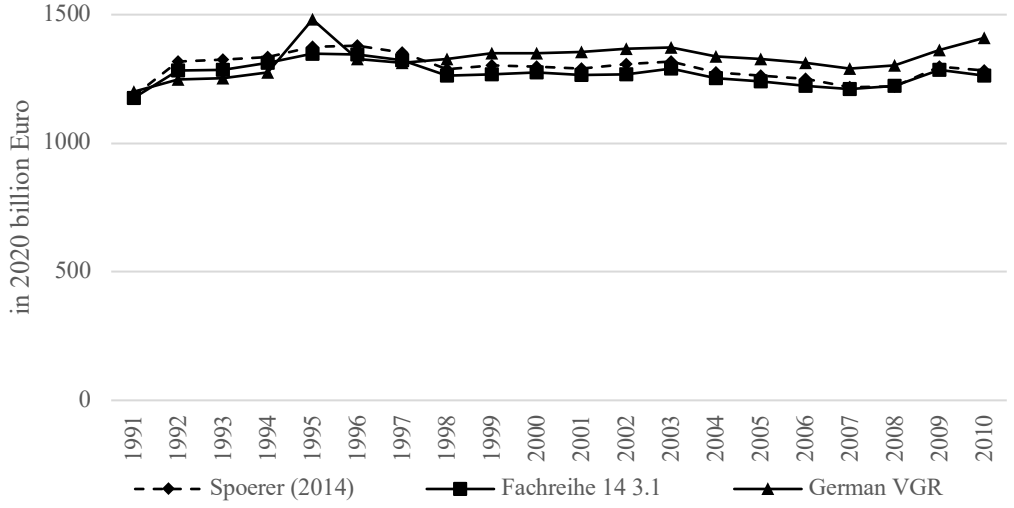
A1: Government Expenditure Dataset

Here, we describe the compilation of the annualised series of German federal government expenditure from 1950 to 2020 according to nine different spending functions (e.g., education, military, social policy, etc.). We combine the following three sources and apply harmonisation techniques to create a dataset that enables us to study trends of fiscal expenditures over seven decades:

1. *Fachreihe 14 3.1* (1950, 1955, 1962–2010): Published by the *Statistisches Bundesamt* (Federal Statistical Office of Germany), this discontinued series documents government expenditure, disaggregated according to eight spending functions over a period of about 50 years.
2. *Spoerer dataset* (1950–2010): Compiled by Mark Spoerer for the book 'Deutschland in Daten', this series records annual government expenditure data across various spending functions for Germany. We rely on this dataset to fill the gaps of the *Fachreihe* between 1950-1962.
3. *German National Accounts* (1991–2020): Published by the *Statistisches Bundesamt*, the national accounts data (German VGR) document government expenditure by issuer and spending function. We mainly use the *German National Accounts* data for years 2010-2020, not covered in the *Fachreihe*.

To ensure comparability across all sources, we begin by harmonising all spending data into billions of euros with 2020 as the base year. Figure A1 presents a validation exercise for the three aggregate expenditure measures for the overlapping years (1991–2010). We find that the discrepancies between the series are always below 2% for every single year, which we consider to be sufficiently accurate.

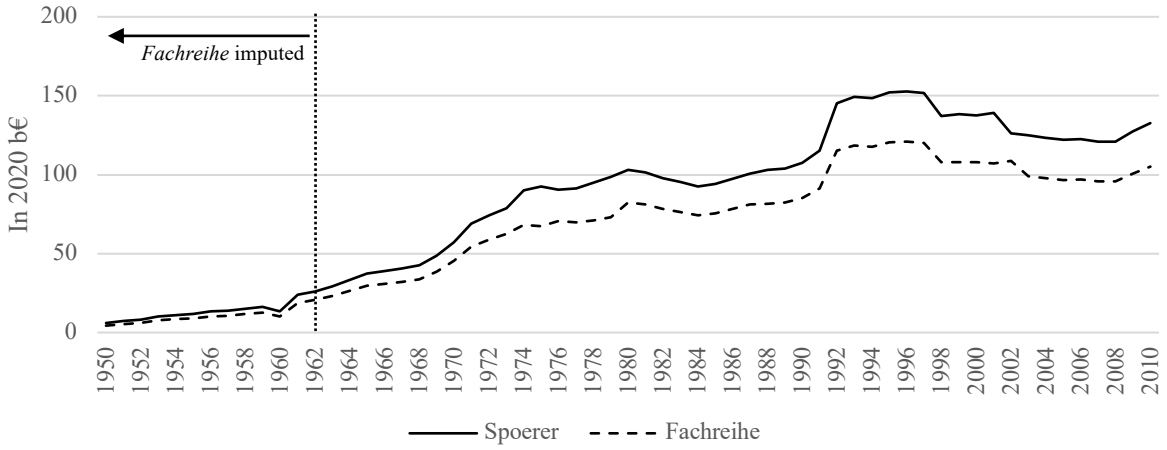
Figure A1: Comparison of alternative government spending series (1991–2010)



Ultimately, our aim is to extend the coverage of the Fachreihe data series to the full period from 1950 to 2020.

Next, we map the spending categories of *Spoerer* (eight categories) to those of *Fachreihe* (15 categories) by semantic and structural mapping. For example, *Spoerer's* expenditure category 'Education' is mapped to the *Fachreihe's* categories of 'General and technical schools', 'Support for pupils and students', 'Universities', and 'Other education'. Figure A2 shows the strong co-movement of education expenditure between the two datasets, with a correlation of more than 99%. We obtain similar results for the remaining spending categories.[27]

Figure A2: Time series of aggregated education expenditure



We then extend the coverage of the *Fachreihe* for the missing years (1950–1962) using imputation techniques. Specifically, using observations from the overlap period (1962–2010), we estimate regression models to predict *Fachreihe* expenditure from the corresponding *Spoerer* expenditure. For example, education expenditure in the *Fachreihe* is predicted using the following regression equation:

$$Education\ (Fachreihe) = 0.79 - 0.41 \times Education\ (Spoerer)$$

Where additional benchmark observations are available, for instance for 1955, we make discretionary adjustments to improve accuracy. The resulting imputed series for education expenditure in the *Fachreihe* is shown in the left-hand side of Figure A2.

Next, we allocate expenditure to the respective issuer for the extended years. Specifically, we filter out expenditure from the federal government only for the years 1950–1962, which is available in the *Fachreihe*, but not in the *Spoerer* Dataset. To achieve this, we rely on the persistent relative distributions between federal government and other issuers (states, local government, municipalities…) within a spending category, as observed between 1962 and 1972. For example, during this period, we find consistently the following allocation of education expenditure: 4% to the federal government, 69% to the states, and 27% to the municipalities. We also apply these shares to earlier years to ensure a meaningful disaggregation of expenditure. As the example of the education category shows, the distinction between total government spending and federal spending is crucial, especially when linking expenditure to parliamentary debates, as it avoids misattributing state spending to parliamentary debates on the federal level.

Finally, we reconcile the harmonised dataset for 1950–2010 with the VGR dataset for 2010–2020, ensuring a consistent methodology over the entire period. The resulting dataset provides

[27] Detailed results are available in the Online Appendix.

a comprehensive, long-term view of German federal government expenditure by spending category. This unified series forms the basis for analysing the correspondence between fiscal policy decisions and parliamentary debates. The descriptive statistics are presented in Table 1.

Table A1: Descriptive statistics of federal government spending (1950–2020, in billion euros)

	Obs.	Mean	St. Dev.	Min	Max
Federal Budget	71	248.74	126.67	34.46	507.29
Education	71	3.67	2.08	0.18	6.99
Healthcare	71	0.90	0.62	0.04	2.03
Housing	71	1.73	0.72	0.72	3.58
National Defence	71	34.90	9.23	13.38	48.89
Public Safety	71	1.81	1.15	0.01	3.97
Regional & Structural Policy	71	14.34	11.08	0.72	58.71
Science & Research	71	6.61	3.54	0.64	14.30
Social Security	71	138.69	83.40	15.61	282.67

Note: Spending functions (Column 1) are defined in Section 2.1.

A2: Unit root tests

Table A2: Augmented Dickey-Fuller tests

Functions	$LnExp_{t-1}$	$LnDebate_{t-1}$	$\Delta LnExp_{t-1}$	$\Delta LnDebate_{t-1}$
Federal Budget	-2.56	-1.55	-3.79*	-5.60**
Education	-2.09	-2.93	-3.52*	-4.85**
Healthcare	-1.88	-3.03	-3.67*	-4.81**
Housing	-1.93	-2.37	-4.70**	-4.49**
National Defence	-3.34	-3.02	-3.89*	-5.10**
Regional & Structural Policy	-2.79	-1.20	-2.91^{+}	-6.90**
Public Safety	-2.25	-2.85	-5.12**	-5.41**
Science & Research	-1.94	-2.56	-4.22**	-5.30**
Social Security	-1.40	-2.02	-4.36**	-4.61**

Note: Critical values for 65 observations at the 5% (*) and 1% (**) levels of significance are -3.48 and -4.10, respectively. Specification: Five lags for the log-level variables and four lags for the variables in first differences, along with a constant and a trend. The '+' symbol indicates rejection of the test at the 1% significance level when using three lags, as recommended by the Akaike Information Criterion.