High-Frequency Fiscal Shocks*

Gabriel P. Fritsch† J. Zachary Mazlish ‡

February 2026

Preliminary

Abstract

We introduce a new methodology for identifying high-frequency fiscal shocks using Large Language Models. We apply this method to 1947-2025 US data. Our results show that the model successfully mimics a "professional forecaster" of the current and future US fiscal position, and is able to recover similar shocks to what have already been identified in the narrative fiscal shock literature. We then examine the effects of fiscal shocks on asset prices: in response to a 1pp shock to the present-value of the current and next ten-years deficits, ten-year Treasury yields rise more than 30bps, with real yields and break-even inflation expectations both contributing to the rise. The dollar appreciates significantly — as much as 4.8% — and the 2Y-10Y spread rises 16-24bps. Turning to macroeconomic outcomes, our fiscal shocks produce government spending multipliers in the 0.5-1 range. Tax shocks shows strong signs of anticipation, and using our data to account for anticipation, we find that output and consumption fall by more than 2% in anticipation of a 1% of GDP tax cut. The multiplier for an anticipated tax shock is 1.2, smaller than typical estimates.

*We thank Sergio de Ferra, Peter Karadi, Marek Jarociński, Elmar Mertens, Davide Porcellachia, Michele Lenza, Ellen McGrattan, Chris Hyland, and seminar participants at the University of Oxford, the European Central Bank, the University of Minnesota, and NYU for helpful comments and suggestions. We are grateful for financial support from the Oxford Economic Papers (OEP) Fund. Gabriel P. Fritsch acknowledges that part of this research was conducted while visiting the European Central Bank as part of the Summer Research Graduate Programme in honour of Ivan Jaccard.

†University of Oxford. Email: gabriel.fritsch@economics.ox.ac.uk

‡University of Oxford. Email: john.mazlish@economics.ox.ac.uk

1 Introduction

The last twenty years have seen a revolution in how macroeconomists identify the effects of monetary policy on the economy. Older VAR and narrative-based identification approaches have been largely superseded by "high-frequency" identification — looking at daily surprises in Fed funds futures to cleanly identify the effects of monetary shocks. The same revolution has not yet come for the effects of fiscal policy, where the literature still relies on VAR and narrative methods (Ramey 2011, Ramey 2019). The reason for this is an absence of data: unlike in the case of monetary policy, we do not have reliable high-frequency measures of fiscal expectations, nor do we have systematic methods for identifying when policy "surprises" relative to expectations. This paper leverages the systematic information-processing capacity of large language models (LLMs) to generate such data, and thereby identify the effects of fiscal policy using high-frequency shocks.

We apply this methodology to 1947-2025 US data, generating a time-series of fiscal expectations and fiscal shocks that covers every (week) day in the post-WW2 US. To ensure no look-ahead bias, we follow Wu, Xi and Xie (2025) in imposing a strict (prompt-based) "knowledge-cutoff" on the model, and present evidence that the results are not afflicted by any such bias. The time-series produces forecasts of the future US fiscal position that mimic lower frequency external forecasts, while allowing for granular identification of the days and events that changed fiscal policy expectations. To validate the shocks generated by the model, we compare them with two leading narrative fiscal shock series in the US: Ramey (2011)'s defense spending shocks, and Romer and Romer (2010)'s tax shocks.

Within the subset of LLM-generated "fiscal events" driven by changes in expectations about defense spending, we uncover a series of defense shocks with a correlation of 0.71 with Ramey's shocks. Within the subset of fiscal events about legislated tax changes, the LLM's shocks have a correlation of 0.89 with Romer and Romer's tax shocks.

We then leverage the high-frequency nature of our shocks to identify the effects of shocks to deficit expectations on asset prices. Ten days after a shock that increases the present-discounted-value of the current and next ten-year's expected deficits over GDP by 1pp, the dollar appreciates 4.8%, ten-year nominal Treasury yields rise 46 basis points (bps), and ten-year real yields are 34bps higher. When we split fiscal shocks into "long-horizon" and "short-horizon" categories based on how far out in the future they are expected to impact deficits, we find that the driver of dollar appreciation and yield increases are *long-horizon* shocks. Investors only price fiscal shocks which affect long-term debt supply and fiscal sustainability.

With our shocks in hand, we next revisit the literature on the macroeconomic effects of fiscal shocks. Estimating fiscal multipliers using our "LLM defense shock" series we find multipliers in the 0.5-1 range, in-line with the literature. The main advantage our method provides is delivering a more relevant time-series of defense shocks in the post-Korean War era.

Relative to the analysis of Mertens and Ravn (2012), we find that the Romer and Romer tax shocks were much more anticipated than previously recognized. Adjusting for the amount of anticipation each tax shock had over time, we find that output and consumption fall by more than 2% in anticipation of a 1% of GDP tax cut. The significant fall in consumption prior to implementation result lies in contrast to the previous literature. Output, consumption, and investment all rise post-implementation, but the output multiplier (to an anticipated tax cut) is only 1.2 — smaller than most empirical estimates and closer to the predictions of New Keynesian DSGE models (Ramey, 2019).

The methodology for producing a time-series of fiscal expectations and fiscal shocks coordinates two different LLM "agents." The first agent is the "forecaster." At each date in time, the forecaster is given all the relevant news articles (2005-25) or summaries of each news article (1947-2005) that came out that day about the macro-economy and fiscal situation in a given country. The forecaster is then prompted to update its forecast about the future path of deficits in that country if, and only if, the news materially changes the fiscal outlook. The forecaster produces a daily time-series of future deficit expectations, covering the current fiscal year and the next ten fiscal years at each point in time. When the forecaster agent changes their forecast, they also produce a rationale — giving a clear narrative identification of what news caused their forecast to update.

The second agent is the "classifier." On days where the forecaster changes their forecast, the classifier is given the forecasters rationale, and then asked to identify whether the forecast update was driven by an endogenous response of expected deficits to macroeconomic conditions or a "fiscal shock." The classifier agent outputs their own rationale, allowing us to inspect and verify that identified fiscal shocks are true "shocks" — unanticipated changes in expectations about future fiscal policy that would not have been expected based on prior information about how fiscal policy responds to macroeconomic conditions. Additionally, the classifier breaks down the change in forecast for each fiscal year into spending changes (G), tax and transfer changes (T), and interest expense changes (I), allowing us to speak to the literature on fiscal multipliers in response to tax/spending shocks.

We compare our forecasts with two primary external benchmarks: the Consensus Economics monthly survey, which forecasts deficits for the current and next fiscal year, and the CBO long-term budget projections, which projects deficits across the same horizon as our model — the current and next ten fiscal years.

To compare the forecasting accuracy of our model with external benchmarks, we use our model forecasts from the day before the external forecast was released. Our model forecasts deficits equally as well (current-year) or slightly better than (year-ahead) the Consensus survey, suggesting that our model is correctly mimicking the behavior of an "expert" forecaster tasked with forecasting near-term fiscal outcomes. The CBO projections are hamstrung by the fact they condition on the assumption that current-law persists, which prevents them from making genuine forecasts about the *expected* future fiscal stance. Our model more accurately predicts future deficits than the CBO at all horizons other than the current year, where they have roughly equal performance. By allowing the model to form beliefs about expected future policy, the model produces what is to our knowledge a novel dataset: a time-series of expected future US deficits that stretches out beyond the fiscal-year ahead.

We next show that the information incorporated in the model forecast is a strong leading predictor of the lower-frequency changes in Consensus and CBO forecasts. At most forecast horizons changes in both Consensus forecasts and CBO projections can be predicted by the forecast the model produces the day before the external forecast comes out. This provides evidence that the model is responding to real-time information in a manner consistent with how informed participants form their expectations.

To showcase the way the model responds to fiscal news, we examine two detailed case studies: how the model updated its forecast around the 2021 Georgia Senate runoff elections, and how the model forecast changed as the Covid crisis broke out. On the day the news came out from the Georgia runoff that the Democrats would have full control of the Senate, the model increased its forecast of FY2021 deficits by $400B. This forecast update matches almost exactly the $450B fiscal shock attributed to the Georgia runoff results reported in Hazell and Hobler (2024), despite our entirely different methodology.

Around the March 2020 outbreak of the pandemic, the model rapidly increased its deficit expectations as news came out about the various stimulus packages under consideration, anticipating where the Consensus forecast would move. Both cases illustrate how the model is able to respond in a realistic and intelligent manner to news about future deficits.

Across the whole time-series of fiscal shocks identified by the classifier model, the three biggest shocks to the present discounted-value of deficits (as a % of GDP) are Ronald Reagan's 1980 election, JFK's 1961 budget statement, and Truman's 1948 budget statement. The model correctly identifies that the largest true fiscal shocks are political developments, and is able to realistically quantify the impact of those political developments on the path of future deficits.

Related literature. The two closest papers to ours are Wiegand (2024) and Gomez Cram, Kung and Lustig (2025). Both also use high-frequency measures of fiscal shocks to study the effects of fiscal shocks on asset prices. In Wiegand (2024), shocks are constructed by carefully examining Congressional deficit targets produced by the budget resolution and reconciliation process. In Gomez Cram, Kung and Lustig (2025), shocks come from the release of CBO "scoring" of the budgetary costs of new pieces of legislation.

We see our work as complementary to each of theirs, with its own distinct advantages. Since our shocks are constructed from the universe of fiscal news over our time-horizon, it is able to capture similar shocks as each of those papers focus on, while also including other sorts of highly relevant shocks — like the Trump elections and Georgia Senate runoff which our model identifies as two of the largest fiscal shocks in the last 20 years. The systematic nature of our LLM-based methodology also allows it to extend farther back in time than either of their papers, and offers the opportunity to further extend our procedure to other countries and date-ranges. Additionally, our methodology is better suited for connecting up with the macro-literature on the effects of fiscal shocks on macroeconomic outcomes.

The broader literature we contribute is the literature on identifying the effects of macroeconomic shocks, and specifically, fiscal shocks. We take inspiration from the high-frequency monetary shock literature (e.g. Kuttner 2001, Nakamura and Steinsson 2018, Jarociński and Karadi 2020, Bauer and Swanson 2023), trying to port their methods to the fiscal policy context. The high-frequency and systematic method of uncovering fiscal shocks we employ differs from traditional methods that either rely on vector autoregressions (e.g. Blanchard and Perotti 2002, Mountford and Uhlig 2009), hand-constructed narrative shocks (Romer and Romer 2010, Ramey 2011, Mertens and Ravn 2013), or cross-sectional identification (e.g. Chodorow-Reich et al. 2012, Nakamura and Steinsson 2014), though it ends up bearing the most resemblance to the narrative method.

Our daily-frequency narrative shocks allow us to revisit the estimates of Ramey (2011), Ramey and Zubairy (2018), and Mertens and Ravn (2012) on the effects of government spending shocks and anticipated tax shocks. The multipliers we find are broadly in-line with the literature as summarized by Ramey (2019).

Our results on the effect of deficit expectations on interest rates is part of the broader literature interested in understanding how changes in the supply of government debt impact financing costs (e.g. Laubach 2009, Rachel and Summers 2019, Campos et al. 2024). The results on inflation expectations and can be compared to recent models making predictions about the effects of future fiscal policy on inflation (e.g. Corhay et al. 2023, Angeletos, Lian and Wolf 2024*a*). The results on the dollar appreciation connect with the literature attempting to test the theoretical relationship that the currency appreciates in response to an expansionary fiscal shock (e.g. Monacelli and Perotti 2006, Auerbach and Gorodnichenko 2016, Hyland and Kawalec 2025). The results on how the horizon of the shock matters for the asset price response connects to debates on the fiscal theory of the price level versus HANK models (e.g. Cochrane 2023, Angeletos, Lian and Wolf 2024*b*).

Our use of large language models to construct representations of economic agent's expectations is part of a growing literature — Horton (2023) Wu, Xi and Xie (2025), Bybee (2025), and Hansen et al. (2024) are prominent examples that are closely related to our work. The most similar paper to ours is Fernández-Fuertes (2025), which uses a coordinated structure of LLM agents — much like ours — to identify *monetary* policy shocks.

Outline. Section 2 details our methodology and discusses how it compares to existing methodologies for uncovering fiscal shocks. Section 3 examines the time-series of fiscal forecasts and fiscal shocks produced by the model, compares them to external benchmarks, and provides detailed case-studies of a few notable fiscal events. Section 4 shows the effect of our fiscal shocks on asset prices. Section 5 discusses the effects on macroeconomic variables. Section 6 concludes.

2 Methodology

There are two key steps involved in coming up with a credible series of daily-frequency fiscal shocks. Step 1 is building a time-series of future deficit expectations at each point in time. Step 2 is classifying which changes in those expectations are due to exogenous policy shocks, rather than the endogenous response of fiscal policy to economic conditions. Each step is accomplished by a different LLM "agent" — an LLM with a set information structure and output structure. Figure 1 summarizes the architecture. We now detail each step in turn before concluding this section with a comparison between our methodology and existing methods of estimating the effects of fiscal shocks.

2.1 Constructing deficit expectations

The "forecaster" agent in Figure 1 is used to build a daily historical series of expected future deficits in a given country. At each date in time, the agent outputs forecasts for the current and next ten fiscal years — so in May 2005, we would have deficit expectations for fiscal year 2005 (FY2005) through FY2015. We choose this horizon for two reasons: first, to match the long-term budget projections the CBO releases on a (at least) biannual basis. Secondly, in many models, agents' expectations about long-run deficits are a critical object (i.e Cochrane 2023, Angeletos, Lian and Wolf 2024*a*). However, it is hard to argue that "news" ever has clear implications for deficits beyond a ten-year horizon, meaning that expectations generated for longer horizons would likely be primarily noise. Having a ten-year forecast horizon balances the desiderata of being able to speak to the importance of long-run deficits while not introducing meaningless noise to the data.

The CBO projections are built on assuming "laws governing taxes and spending generally remained unchanged." In contrast, the expectation series we construct does not assume

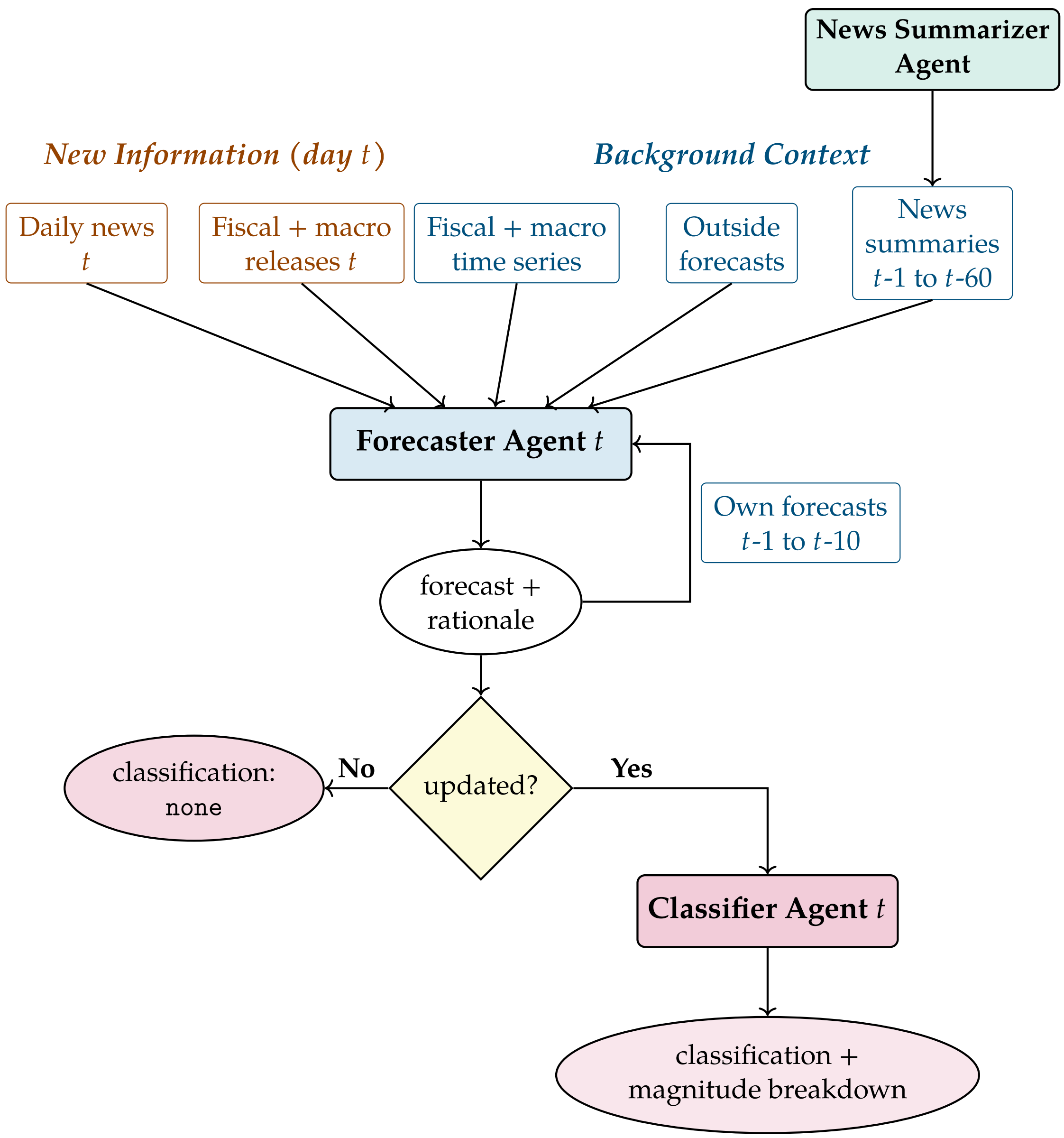


Figure 1: Multi-Agent Architecture for Fiscal Shock Identification.

that current law persists. Rather, our expectations are meant to capture "market consensus" expectations, which factor in expectations about future law. In a model with a representative agent, these expectations would correspond to the representative agent's expectations. In a heterogenous agent model, these expectations would correspond to something like the "wealth-weighted average expectation", which is the relevant object for understanding asset price behavior in many baseline models (e.g. Xiong and Yan 2010).

At a high-level, our expectation construction works as follows: at each date in time, feed an LLM its previous day's forecast of future deficits, as well as relevant context about the fiscal and macroeconomic situation, up to that point in time. Then, feed the LLM news articles pertaining to the macroeconomic and fiscal situation that came out on that day, and ask the LLM to update its forecast. This procedure is rolled-over day-by-day, feeding in the previous day's forecasts, providing the new news, and then extracting the updated forecast.

As part of our prompt to the forecasting LLM, we always provide the following "knowledge cutoff" statement:

```
Today (YYYY-MM-DD) is your knowledge cutoff.  Base your forecast only on
information available up to and including today.  Do not use knowledge of events
that occur after this date.
```

As shown by Wu, Xi and Xie (2025), imposing such a knowledge restriction is effective in preventing the model from using its awareness of future events to bias its "real-time" historical forecasts. Section 3.1 below provides evidence that our results are not afflicted by any look-ahead bias.

One of the main benefits of our procedure is its generalizability: for any country and any time-period in which one has access to relevant news articles about the fiscal and macroeconomic situation, this procedure can be implemented. In future work, we plan to implement the methodology for identification of fiscal shocks for other countries, across both advanced and emerging market economies.

2.1.1 US Inputs

Our news article inputs for the US are different across two sample periods: 1947-2005 and 2005 to July 23rd 2025. For the 2005-2025 period, the news article sample is all Reuters and Wall Street Journal (WSJ) articles available on Factiva with keywords that indicate fiscal or macroeconomic relevance.[1] Each day's news articles are processed into JSON format to maximize compatibility with the LLM, and the LLM is provided with the release time, the headline, the lead paragraph, and the full body of each article from that day.

[1]Our precise search query is available in appendix B.3

From 1947-2005, instead of giving the forecaster agent the raw text of news articles, the agent is given a summary of each of that day's news articles, with the summary itself generated by a separate LLM agent.[2] The raw text articles are all New York Times and WSJ articles that match our fiscal and macroeconomic search filter, with NYT articles available over the full sample, and WSJ articles only available post-1984. The search filter and then article-by-article summarization was conducted within ProQuest TDM Studio, which does not allow for export of raw-text files. An example article summary can be found in appendix B.8.

The 2005-25 sample is "better" on two important dimensions: firstly, providing the forecaster agent with full-text articles as opposed to article summaries ensures the agent has as complete an information-set as possible. Secondly, the Reuters news is "real-time", thus allowing for reliable inference about the asset-price effects of fiscal shocks. NYT coverage of fiscal news can come out with a day or two lag. Therefore, for the asset price results in 4, we focus on the 2005-25 sample.

In addition to the day's news articles, we also provide the LLM with a summary of each of the last 60 days worth of news. The summaries are generated by a separate LLM, which is told to read through all the news articles on a given day and produce a 50-100 word summary of the most important fiscal developments from that day. We use Gemini 3 Flash to produce the summaries. The full prompt given to the daily "news summarizer" agent is available in Appendix B.6.

As shown in Figure 1, alongside news, we also provide the forecaster LLM with the following pieces of context:

1. **Treasury data release**: if the monthly treasury statement (for the previous month's data) was released on that day, we provide the value for the realized deficit (in billions of dollars), as well as the "surprise" relative to the median forecast on Haver's MMSAMER database.

2. **Other macroeconomic data releases**: if the GDP report (advanced, preliminary, or final) came out that day, we provide the value and the surprise. If the monthly employment report came out that day, we provide the value and surprise of both non-farm payrolls and the unemployment rate.

3. **Own past forecasts**: we give the forecaster its past ten-days worth of forecasts, as well as its rationale for any changes it made to its forecast on those days.

4. **Other external forecasts**: for the US, there are two sources of external forecasts we feed to the LLM. One is the CBO projections, which forecast out ten fiscal years in the future, but come out infrequently (typically 2x/year), and are based on the assumption

[2] The full summarizer prompt is available in appendix B.7. The summarizer model is GPT-4o-mini.

that current law persists. The other source is Consensus Economics, which surveys economists at banks and consultancies to provide a monthly forecast of the current and next fiscal year's deficit. We give the model the two most recent deficit forecasts from both the CBO and Consensus. From the CBO, we also provide the model with the current and next ten fiscal year's NGDP forecasts. For the Consensus forecast, we provide the agent the median forecast as well as the standard deviation of forecasts. The first CBO deficit forecast is in February 1984. CBO NGDP forecasts and Consensus deficit forecasts begin in 2005.[3]

5. **Fiscal data**: we provide the model with the last five years worth of monthly flow deficits, monthly flow receipts, monthly flow outlays, as well as monthly cumulative fiscal year-to-date versions of each variable. For the most recent three months of realized flow deficits, we also provide the surprise relative to the median forecast on Haver's MMSAMER database.

6. **Macroeconomic data**: for GDP, we provide the latest revisions of the last 24 months (8 quarters) worth of data, with both value and surprise information for the last two advance, preliminary, and final releases. For non-farm payrolls and unemployment we provide the latest revisions of the last 12 months worth of data, with both value and surprise information for the last two releases.[4]

All of the above data is available for the 2005-25 period. Appendix B.2 details the availability of different data across the 1947-2005 period.

2.1.2 Forecaster prompt

The model is given all of the aforementioned information, and then prompted to update (or not update) its previous day's forecast based on today's new data and news articles. The full prompt we give the model is available in appendix B.4. The goal of the prompt is to give the model as much guidance as possible on how to make a realistic assessment of the expected future fiscal trajectory, without hard-coding in any decision rules or introducing any look-ahead bias. In addition to providing a point estimate of the expected value of the deficit in each fiscal year they are forecasting, the agent is also instructed to produce a standard deviation (for each fiscal year), representing their forecast uncertainty. When the agent's forecast changes from the previous day, the model also outputs a rationale for *why* it updated its forecast. The forecaster model is GPT 5.1 with "medium" reasoning.

[3] Until the May 1996 CBO release deficit forecasts only stretch out five-years beyond the current fiscal year.
[4] Full details on data sources and series identifiers are provided in Appendix B.1.

In order to elicit our desired "market consensus" expectation, the model is explicitly told "your goal is to capture market consensus expectations and predict where consensus will move given new information." The model is also given strong guidance that it should only update its forecast when "new information materially changes the outlook." In order to produce a true "expected value of the deficit" forecast, rather than one based on current law, the model is told that "the possibility of new fiscal measures being enacted in the future should be taken into account when evaluating whether to update your forecast."

To prevent over-anchoring on outside forecasts, the model is instructed to "consider Consensus Economics and CBO forecasts as inputs, but do not let them dictate your forecast." The model is also explicitly instructed to take into account the seasonality of fiscal data.

We "hold the model's hand" the most when providing guidance on how to form long-term forecasts. Specifically, we tell the model:

```
Long-term forecasts should only change according to the following principles:
- If a policy measure has a multi-year impact (i.e.  a permanent tax change),
then the long-term forecasts should change to reflect the year-by-year impact of
the policy.
- Insofar as short-term changes are large enough to alter the debt stock, the
long-term forecasts should change due to the future impact on interest expense.
Be explicitly clear about the magnitude of the change in the debt stock, and the
magnitude of the change in interest expense based on average interest rates.
- If the short-term changes indicate a fundamental change in the fiscal stance
- becoming more expansionary or contractionary - then the long-term forecasts
should change to reflect the changing attitudes about fiscal policy.
- Note that changes in "attitude" might be sensitive to who is President.  A
newly elected President might change the outlook for deficits over the next four
years, but beyond that, you should weight the uncertainty about whether their
attitude towards fiscal policy will persist beyond their term in office.
- In some cases, expansionary fiscal policy in the short-term might correspond
with stated commitments to be more contractionary in the future.  If this is
explicitly stated in the news, the long-term forecasts should change to reflect
the commitment to be more contractionary in the future.
```

Once again, the goal with the prompt is to give the model all the necessary context to serve as a stand-in for an "expert" fiscal forecaster, without providing the model any explicit information that would prevent its outputs from being true "real-time" forecasts.

2.2 Classifying changes in expectations

Each (non-holiday) weekday in our sample, the model produces a new forecast. If the forecast for any fiscal year in its forecast horizon changes, we then move to the next agent — the classifier.[5]

The primary role of the classifier is to distinguish between "endogenous" and "fiscal event" based changes in deficit expectations. Returning to the analogy with high-frequency monetary policy shock identification, this is akin to separating out the "Fed information effect" (or the Fed "response to news") from the "true" policy shock (Nakamura and Steinsson 2018, Bauer and Swanson 2023, Jarociński and Karadi 2025). In the monetary literature, the idea is that there is some "monetary rule" the monetary authority is expected to follow in response to changing macroeconomic conditions. Monetary "shocks" are deviations from that rule that could not have been expected.

Analogously, "endogenous" changes in expected future deficits reflect the expected change in fiscal outcomes in response to changing macroeconomic conditions. This includes updating current-year expectations in response to surprises in realized deficit data (e.g. tax receipts came in lower than expected), the response of automatic stabilizers (UI insurance) and tax receipts to changing macroeconomic conditions, and the general amount of "stimulus" (or austerity) that is expected given changing macroeconomic conditions. "Fiscal events" are deviations from that rule: when policy changes, or is expected to change, in a way that makes it more or less stimulative than would have been expected based on prevailing expectations of the fiscal rule.

To be precise, the classifier is told that the "the definition of a fiscal shock is as follows: a change in expected deficits that is orthogonal to the current state of the macroeconomy. Necessarily, fiscal shocks are political developments that are not driven by the current state of the macroeconomy." The basis for this definition comes from Ramey (2011), where she defines a (defense spending) fiscal shock as when the news "suddenly began to forecast large rises in defense spending induced by major political events that were unrelated to the state of the U.S. economy."

The classifier agent is given the forecast agent's rationale for why it changed its forecast, and then asked to evaluate and output whether the forecast update was "endogenous", a "fiscal event", or "both". The model must produce one of those three outputs.

Additionally, the classifier is given a subset of the context the forecaster is given: the classifier is fed the headlines and lead paragraphs from that day's news articles, as well as any fiscal or macro releases that came out that day. The idea here is to restrict the amount of information the agent has to process, so it can hone in on the task of classifying what drove

[5] If the model's standard deviations change while its point forecasts remain unchanged, nothing is passed to the classifier.

the change in fiscal expectations.

The "both" classification is necessary to handle days where there is both news about the macro-economy that would warrant a change in fiscal expectations independent of any change in policy and news about policy that deviates from the expected fiscal rule. One example would be the days surrounding the outbreak of the Covid-19 virus, where rapidly changing macroeconomic conditions led to large "endogenous" updates to fiscal expectations, but policy makers response to conditions also (arguably) led to surprising changes in future fiscal expectations.

The model is told to classify an update as "endogenous" or "fiscal event" *only* if changes in the forecast at *all* forecast horizons fit the endogenous/fiscal event classification. For example, if news came out that employment had fallen substantially below expectations, that might lead to an endogenous change in current-year deficit expectations. However, if unexpected news came out the same day that a new tax bill was being proposed that would not be active until the following year, future fiscal year deficit expectation changes would be driven by a fiscal event. The model would then output a "both" classification, and explicitly quantify how much of the change in deficit expectations in *each* fiscal year is driven by the "endogenous" versus "fiscal event" component.

Beyond the primary endogenous versus fiscal-event classification, the classifier agent is also tasked with making the following three classifications:

- For each fiscal year where the deficit forecast has changed, the classifier apportions the change into a change in government purchases (G), taxes and transfers (T), and interest expense (I).

- The classifier assesses "whether the deficit change represents a permanent shift in the level of government debt." This classification is conservative: the classifier is told to only assess a change as non-permanent if there is explicit language suggesting it will be reversed by a subsequent contraction in taxes or spending.

- The classifier determines whether the change is driven by "stabilizing" or "non-stabilizing" policy. The classifier is told that "Non-stabilizing deficit changes are those not taken to offset factors pushing growth away from normal."[6] This is a separate dimension from endogenous versus fiscal event. For example, the Georgia Senate Runoff election in 2021 is a fiscal event — which is why it is the subject of analysis in Hazell and Hobler (2024) — but it is a stabilizing policy. The stabilizing/non-stabilizing dimension is unimportant for looking at the effect of fiscal shocks on asset prices, where all that matters is whether the shock is news that is politically rather than macroeconomically driven, but is important for evaluating the effect of fiscal shocks on macroeconomic variables.

[6]Language taken from Romer and Romer (2010)

The full prompt given to the classifier is available in appendix B.5.

2.3 Comparison with existing methodologies

Ramey (2019) reviews the literature on estimating the effects of fiscal policy, delineating three "broad categories" of "empirical approaches": "1) aggregate country-level time series or panel estimates; 2) estimated or calibrated New Keynesian dynamic stochastic general equilibrium (DSGE) models; and 3) subnational geographic cross-section or panel estimates."

The methodology outlined here is a novel approach to (1) — identifying the effects of fiscal policy based on aggregate country-level time series. Ramey (2019) has the following to say about the weaknesses of the time-series approach:

> "The time series approach requires exogenous variation in policy. The leading approaches to identifying this exogenous variation are structural vector autoregressions and natural experiment methods, combined with narrative methods that use historical documents to create new data series of exogenous changes. Too often, though, the variations that turn out to be exogenous yield instruments that are not very relevant—that is, they have low correlation with the fiscal variable they are trying to explain—and the variations that are relevant are not always exogenous or are anticipated in advance."

We view our approach as attempting to use the narrative method in as comprehensive and systematic a fashion as the historical record allows. By prompting the LLM with a daily-frequency and wide-universe of news articles pertaining to a given country's fiscal and macroeconomic situation, the identified shocks are meant to fully avoid anticipation concerns, and provide as relevant a series of exogenous shocks as exists in the historical record.

As we document in section 5.1, one takeaway from our exercises is that it is hard to find exogenous narrative shocks with sufficient relevance in the post-1955 data, though we do uncover a defense news shock with more relevance than was available in the prior literature (for post-1955).

In addition to avoiding anticipation effects, the fact that our methodology produces daily-frequency data allows us to look at the effects of fiscal policy shocks on asset prices in a systematic manner, which prior methods are largely unable to do. The longer-term deficit forecasts we produce allow us to investigate the differing effects of more transitory versus more long-lived shocks to deficit expectations, an object of central importance in theoretical models. We also hope the "automatibility" of our method will allow us to cover a wider-range of countries than existing narrative methods, which are quite labor-intensive.

Of course, we do not claim that our methodology is perfect: the forecaster may be failing to accurately capture "true" consensus real-time expectations for a variety of reasons, and the classifier may be erroneously identifying what are or are not "fiscal shocks". There are many different choices one can make in the prompting and information sets one gives the LLM agents, and we see this as a first attempt at this methodology, not the final word. Expanding the forecaster's information set with more news, giving it access to legislative minutes, and providing it with independent estimates of various legislative packages fiscal implications (e.g. CBO reports) would likely improve our results.

Nonetheless, we do our best to show in the following sections that the forecaster and classifier produce reasonable results, and we hope that, at the least, our methodology is an addition to the toolkit of those trying to identify the effects of fiscal policy on financial and macroeconomic variables.

3 Results: forecasts and shocks

Figure 2 plots the time-series of the LLM agent's forecast versus the CBO's projections. In order to summarize the ten-year deficit forecast trajectory, we compute the present discounted value (PDV) of expected deficits, scaled by the PDV of nominal GDP. For 2005-25, the PDV of NGDP at each date in time is the PDV of *expected* NGDP. It comes from the most recent release of the CBO's "Economic Projections", which covers the same forecast horizon as the deficit forecasts. Before 2005, the PDV of NGDP is based on realized NGDP.

To compute this ratio, we convert from "fiscal year" forecasts to "N year ahead" forecasts, so that we can discount by existing N-year zero coupon rates. At any observation date t, the current fiscal year is only partially complete. Let $x_t \in [0, 1]$ denote the fraction of the current fiscal year remaining:

$$x_t = \frac{\text{days remaining in current fiscal year}}{365} \tag{1}$$

For a forecast made at date t, we construct N-year-ahead deficit expectations by weighting the forecasts for adjacent fiscal years:

$$\hat{d}_{t,N} = x_t \cdot d_{t,\mathrm{FY}(t)+N-1} + (1 - x_t) \cdot d_{t,\mathrm{FY}(t)+N} \tag{2}$$

where $d_{t,\mathrm{FY}}$ denotes the forecast made at date t for fiscal year FY, and FY(t) is the fiscal year containing date t. This weighting ensures that the N-year-ahead forecast always corresponds to expectations for a period ending exactly N years from the observation date. The same weighting is applied to realized NGDP (1947-2005) or CBO NGDP forecasts (2005-25) to

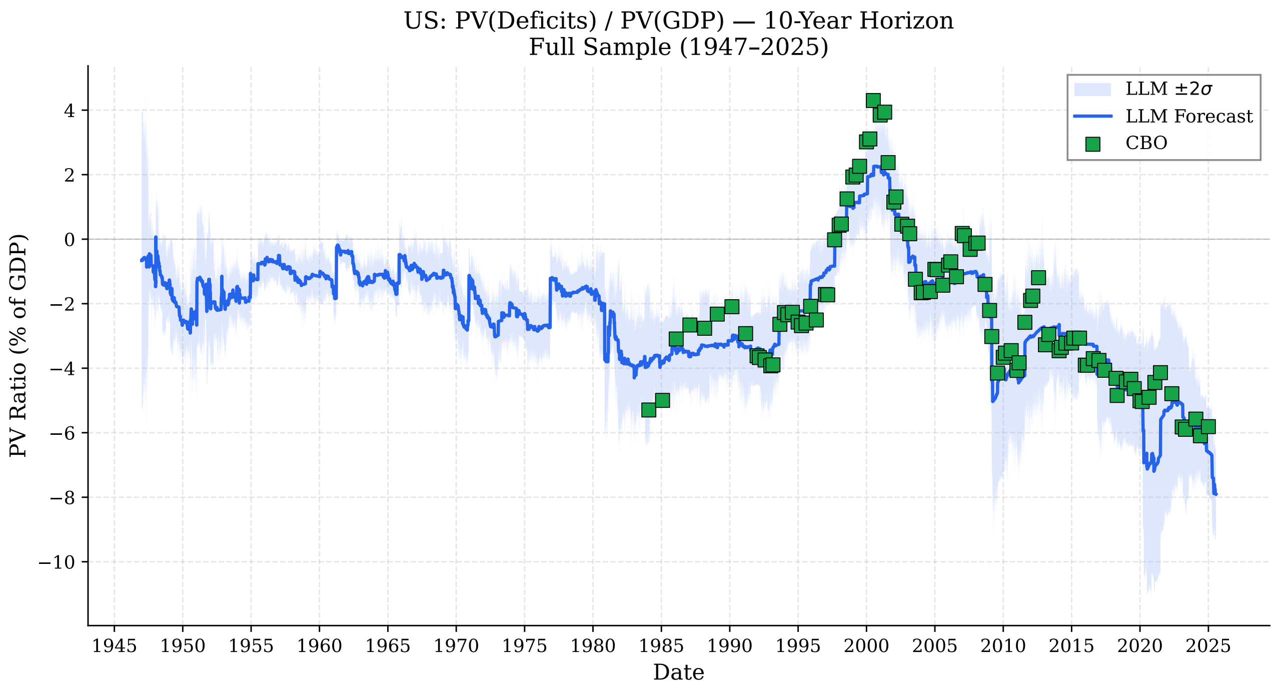


Figure 2: Present discounted value of expected deficits as a share of GDP. The PDV ratio is computed by discounting 10-year deficit and NGDP using zero-coupon Treasury yields, with fiscal-year forecasts weighted to correspond to exact N-year-ahead horizons. CBO values (squares) are computed using the same methodology applied to CBO deficit projections. For 2005-25, the PDV of NGDP at each date in time is the PDV of the CBO's NGDP forecasts. Before then, it is the PDV of realized NGDP. From 1984-1996 the CBO PDV is a 5-year PDV. Shaded region shows the LLM's $\pm 2\sigma$ confidence interval. Sub-sample plots are shown in Appendix 15.

obtain $\hat{Y}_{t,N}$.

The present discounted value ratio is then computed as:

$$\text{PDV ratio}_t = \frac{\sum_{N=1}^{10} \hat{d}_{t,N} \cdot e^{-y_{t,N} \cdot N}}{\sum_{N=1}^{10} \hat{Y}_{t,N} \cdot e^{-y_{t,N} \cdot N}} \tag{3}$$

where $y_{t,N}$ is the zero-coupon Treasury yield for maturity N years observed at date t. We use continuously compounded discount factors $e^{-y_{t,N} \cdot N}$ based on the Federal Reserve's smoothed zero-coupon yield curve.[7]

The shaded region of figure 2 represents the 95% confidence interval implied by the LLM's reported forecast uncertainty, computed by propagating the fiscal-year-specific standard deviations through the PDV calculation.

The time-series matches typical narratives about the historical trajectory of the US fiscal position. Reagan's presidency saw a significant widening in future deficit expectations, the late Clinton years saw a dramatic reversal into long-term surplus expectations, and the fiscal position has deteriorated significantly since.

Across the period of time where CBO projections are available, the model tracks the CBO series quite well. In times where the two diverge — the very first two CBO projections, CBO projections around the 2000 peak in surplus expectations, and a few CBO projections in the early 2010s — the CBO projection ends up "moving" back towards the LLM's expectation. The only exception is right around the Covid crisis.

The knowledge cut-off for GPT 5.1 — the model used to produce the forecasts — is September 30th 2024, just before Trump's re-election. The model judges the time since to be a continued worsening of the US fiscal position — about 1.7pp of GDP.

Figure 3 compares the one-year-ahead LLM forecast against realized one-year ahead deficits and, starting in 2005, the one-year-ahead Consensus Economics monthly survey. Timing is aligned with realized deficits, so the forecasts are shifted forward one-year relative to when they were made. Consensus Economics reports median forecasts for the current fiscal year and the year-ahead fiscal year separately. To construct a comparable one-year-ahead expectation from both sources, we apply the same year-weighting scheme:

$$\hat{d}_{t,1\mathrm{Y}} = x_t \cdot d_{t,\mathrm{CY}} + (1 - x_t) \cdot d_{t,\mathrm{YA}} \tag{4}$$

where CY denotes the current fiscal year and YA the year-ahead fiscal year. This weighted average represents an approximation of the expected deficit over the twelve months following

[7] The zero-coupon yield curve data is available starting on 1961-06-14. Prior to that, we use the 3-month T-Bill and 10y Treasury data available on Aswath Damodaran's website, with log-linear interpolation to do the discounting by horizon.

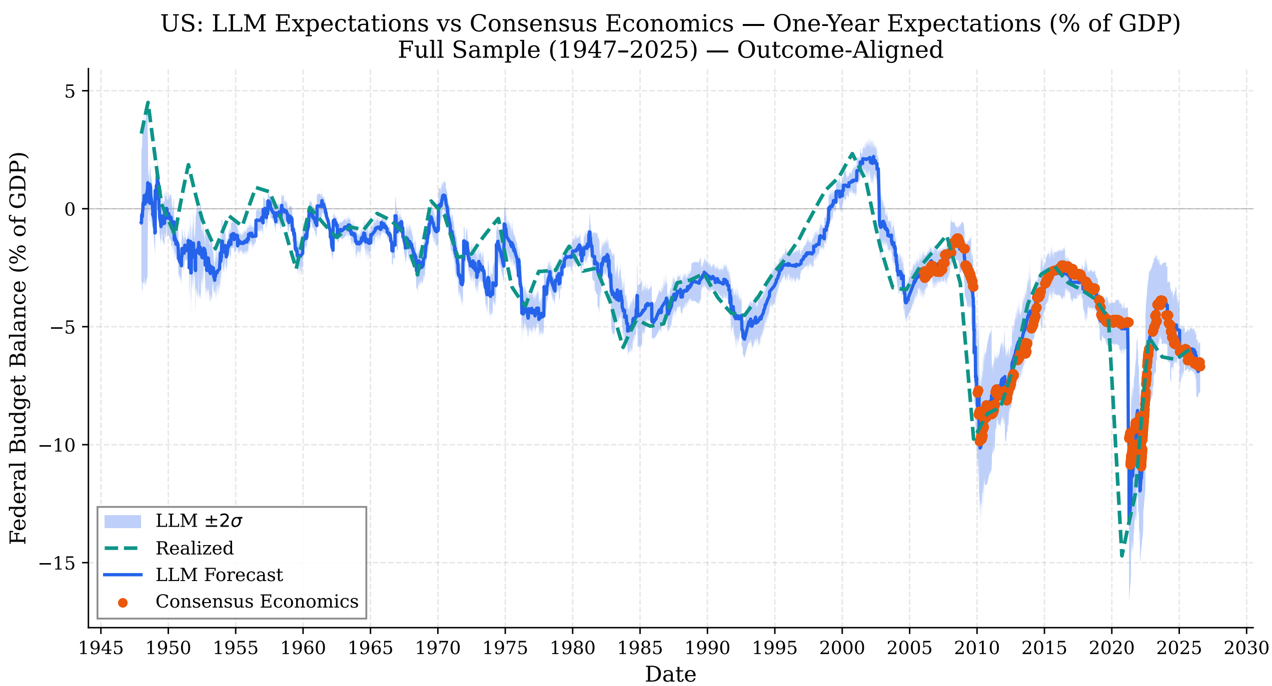


Figure 3: One-year deficit expectations as a share of GDP: LLM forecasts (solid line) versus Consensus Economics survey values (dots) and realized deficits (dashed line). All series use the same year-weighting: $\hat{d}_{1Y} = x \cdot d_{\mathrm{CY}} + (1 - x) \cdot d_{\mathrm{YA}}$, where x is the fraction of the current fiscal year remaining, and are normalized by the corresponding weighted nominal GDP. Dates are outcome-aligned (shifted forward one year). Shaded region shows the LLM's $\pm 2\sigma$ confidence interval. Sub-sample splits are shown in Figure 16.

date t. [8]

The dashed-green line applies the same weighting to realized year-ahead outcomes. All series are normalized by realized NGDP, weighted the same way.

Across most of the sample, the model tracks realized deficits and the Consensus forecasts quite closely. The major exception is the very beginning of the sample, where the model makes larger errors. That period is where the forecaster agent receives the least context from macroeconomic data releases, suggesting the importance of providing that contextual information. In periods of dramatic changes in realized deficits, the model tends to "lag" behind realized deficits — this is most notable in the sustained shrinking of deficits in the 1990s and the fall in deficits around Covid.

Rather than interpreting these lagging misses as a failure on the part of the model, we see them as encouraging confirmation that the model is avoiding look-ahead bias and coming up with true "real-time" forecasts.

Take the Covid example. The realized one-year ahead deficit starts plunging downwards as early as November 2019, since that is when the year-ahead weighted realized outcome begins to include FY2020 outcomes. However, we should not expect real-time forecasters in late 2019 to have predicted the Covid shock and stimulus that would end up materializing.

3.1 Accuracy comparison

To formally assess how well our LLM forecasts compare to established benchmarks, we conduct a direct forecast accuracy comparison. At each external forecaster's release date (Consensus and CBO), we compare forecast errors by computing the LLM's forecast from the day before (to avoid any information advantage from seeing the external release) against the external forecast, both measured against the same realized deficit-to-GDP outcome.[9] We compute root mean squared error (RMSE), mean absolute error (MAE), and a "win rate" (the percentage of observations where the LLM forecast is closer to the realized value).

Table 1 compares LLM forecasts to Consensus Economics at each monthly survey date. For current-year (CY) forecasts, the two sources are effectively indistinguishable, with MAE of 0.73 vs 0.72pp of GDP. For the one-year weighted metric, the LLM significantly outperforms Consensus with a 63.2% win rate (Diebold-Mariano $p < 0.05$). The nearly identical performance of the model and the Consensus forecasters, even when the model does not yet have access to the Consensus forecast it is being compared to, suggests the model is achiev-

[8] It is not an exact representation of the one-year forward deficit expectation, since Consensus forecasters and/or the model might be forecasting timing of when the CY and YA deficits materialize that does not align exactly with the simple weighting schema.

[9] CBO forecasts are dated to the first of the month in which the CBO release came out, since the CBO does not provide information on exact release dates. This makes the comparison biased more in favor of the CBO, since sometimes releases are closer to the middle of the month.

Table 1: LLM vs Consensus Economics: Direct Accuracy Comparison

Metric	N	$RMSE_{LLM}$	$RMSE_{Cons}$	MAE_{LLM}	MAE_{Cons}	Win Rate
CY	180	1.54	1.52	0.73	0.72	56.1%
1Y-Weighted	171	2.30	2.35	1.32	1.39	63.2%**
1YA	228	3.01	3.06	1.72	1.78	57.5%**

Notes: CY = current fiscal year, 1YA = year-ahead, 1Y-Weighted = weighted average. RMSE and MAE in pp of GDP. Stars denote Diebold-Mariano significance: $^{**}p < 0.05$.

ing our goal: coming up with a forecast that matches what an informed expert on the topic would have concluded at the time.

Table 10 in the appendix reports standalone LLM accuracy across the full historical sample. Of note, in terms of MAE, the model is actually more accurate in the 1947-2005 period than it is in the 2005-25 period when it has Consensus as a reference. This is likely due to the greater variability of deficits in the post-2005 period, but suggests the model is not overly reliant on having external forecasts available to mimic the behavior of such a forecaster.

Table 2: LLM vs CBO: Direct Accuracy Comparison by Forecast Horizon

Horizon	N	$RMSE_{LLM}$	$RMSE_{CBO}$	MAE_{LLM}	MAE_{CBO}	Win Rate
CY	98	1.57	1.59	0.76	0.74	45.9%
1YA	96	2.73	2.96	1.65	1.81	61.5%***
2YA	94	3.29	3.55	2.20	2.48	72.3%***
3YA	93	3.65	3.99	2.57	2.98	68.8%***
4YA	91	4.15	4.45	3.14	3.47	64.8%***
5YA	88	4.51	4.76	3.57	3.86	54.5%***
6YA	64	5.17	5.35	4.11	4.42	59.4%*
7YA	62	5.30	5.53	4.16	4.58	64.5%*
8YA	60	5.77	6.42	4.47	4.99	71.7%***
9YA	57	6.13	7.15	4.90	5.62	78.9%***
10YA	54	6.37	7.51	5.27	6.22	83.3%***

Notes: CY = current fiscal year, 1YA = year-ahead, etc. Comparisons at each CBO release date. RMSE and MAE in pp of GDP. Stars denote Diebold-Mariano significance: $^{*}p < 0.10$, $^{***}p < 0.01$.

Table 2 compares LLM forecasts to CBO projections at each CBO release date. For current-year forecasts, CBO slightly outperforms in MAE and win-rate but the difference is negligible, with the LLM outperforming in RMSE. For year-ahead and longer horizons, the LLM consistently outperforms the CBO, with Diebold-Mariano tests rejecting equal accuracy at the 1% level for horizons 1–5 and 8–10. The win rate exceeds 70% at the 2-year, 8-year, 9-year, and 10-year horizons, reaching 83.3% at the 10-year horizon. This advantage likely reflects the fact that the CBO forecasts are based on “current law” assumptions, showcasing the value of allowing the LLM to incorporate expectations about the path of future policy.

3.2 Are daily forecasts predictive?

A key question for our methodology is whether the LLM's daily forecasts contain information that is subsequently incorporated into less-frequent external forecasts. This validates that the LLM forecaster is capturing the fiscally relevant news in a sensible manner. We test this by examining whether external forecasters revise their forecasts in the direction of the LLM's prior-day estimate. Specifically, we estimate:

$$\Delta f_t^{ext} = \alpha + \beta \cdot (f_{d-1}^{LLM} - f_{t-1}^{ext}) + \varepsilon_t \tag{5}$$

where Δf_t^{ext} is the change in the external forecast at release t relative to its previous, f_{d-1}^{LLM} is the LLM forecast from the day before the external release, and f_{t-1}^{ext} is the previous external forecast. A positive β indicates that external forecasters revise towards the LLM's position—evidence that the daily LLM forecasts capture information before it is incorporated into less-frequent survey or institutional forecasts.

Table 3: External Forecaster Updates Towards LLM: Consensus Economics

	CY	1Y-Weighted	1YA
Intercept	7.65	−6.30	−12.58
	(5.35)	(4.81)	(4.28)
β	0.69***	0.62***	0.31***
	(0.03)	(0.04)	(0.04)
R^2	0.80	0.65	0.25
N	166	166	226

Notes: Standard errors in parentheses. ***$p < 0.01$. CY = current fiscal year, 1YA = year-ahead, 1Y-Weighted = weighted average of CY and 1YA.

Tables 3 and 4 present results for Consensus Economics and CBO, respectively. For both external forecasters, we find positive and statistically significant β coefficients across the CY and 1YA horizons, indicating that external forecasters systematically revise their forecasts in the direction of the LLM's position.

For Consensus Economics, the gap between where the LLM has "arrived at" the day before Consensus is released and the previous Consensus forecast explains 80% of the changes in Consensus forecasts. The effect is weaker at the year-ahead horizon ($\beta = 0.31$, $R^2 = 0.25$) but still highly statistically significant.

For CBO, the current-year R^2 of 0.93 indicates that where the LLM has updated to since the previous CBO release explains nearly all the variance of CBO revisions. The higher R^2 value than the comparable regression for Consensus reflects the lower frequency of CBO forecasts: there is more room for Consensus and the model to disagree about changes within a

Table 4: External Forecaster Updates Towards LLM: CBO

	CY	1YA	2YA	3YA	4YA	5YA
Intercept	13.39	5.98	−11.05	−8.17	−3.55	−1.91
	(12.18)	(12.56)	(12.56)	(12.25)	(12.54)	(13.86)
β	0.86***	0.52***	0.16*	0.06	0.10	0.16**
	(0.03)	(0.05)	(0.08)	(0.07)	(0.07)	(0.08)
R^2	0.93	0.62	0.07	0.01	0.03	0.07
	6YA	7YA	8YA	9YA	10YA	
Intercept	−1.92	−0.60	2.94	2.99	1.27	
	(18.34)	(18.82)	(19.55)	(20.15)	(20.19)	
β	0.16*	0.17*	0.19**	0.19**	0.18**	
	(0.09)	(0.09)	(0.08)	(0.08)	(0.07)	
R^2	0.07	0.07	0.10	0.11	0.13	

Notes: Standard errors in parentheses. N = 57 (CY–5YA), N = 48 (6YA–10YA).
*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. CY = current fiscal year, 1YA = year-ahead, etc.

month, but across the two or six month gaps between CBO forecasts, the model incorporates nearly all the information that ends up in the CBO forecast. The year-ahead coefficient is also strongly significant ($\beta = 0.52$, $R^2 = 0.62$). At longer horizons, the coefficients remain positive but are smaller, with the 3- and 4-year horizons not statistically significant. This likely reflects that at intermediate horizons, the CBO's current-law methodology and the LLM's news-based approach respond to different types of information. For the longest horizons (8–10 years), the coefficients stabilize around 0.18–0.19 and are significant at the 5% level.

It is interesting that the β coefficients for both Consensus and the CBO are below 1 at all horizons. Given the model's superior accuracy to the CBO at medium-term horizons, the low coefficients for the CBO might reflect that its "current-law" assumptions prevent it from adjusting as much as the model to changing news about future policy. The < 1 coefficients on the Consensus forecasts are less easily explained — it could reflect stickiness in professional forecasters' forecasts, or some tendency to "over-update" on the part of the model. The fact their eventual forecast performance ends up nearly identical implies that neither effect dominates the other.

3.3 Case studies

To provide further evidence that the forecaster responds to news in a timely and sensible manner, we present two case studies.

Figure 4 shows the behavior of FY2021 deficit forecasts before and after the Georgia senate runoff election, which resulted in Democrats controlling both chambers of Congress, and

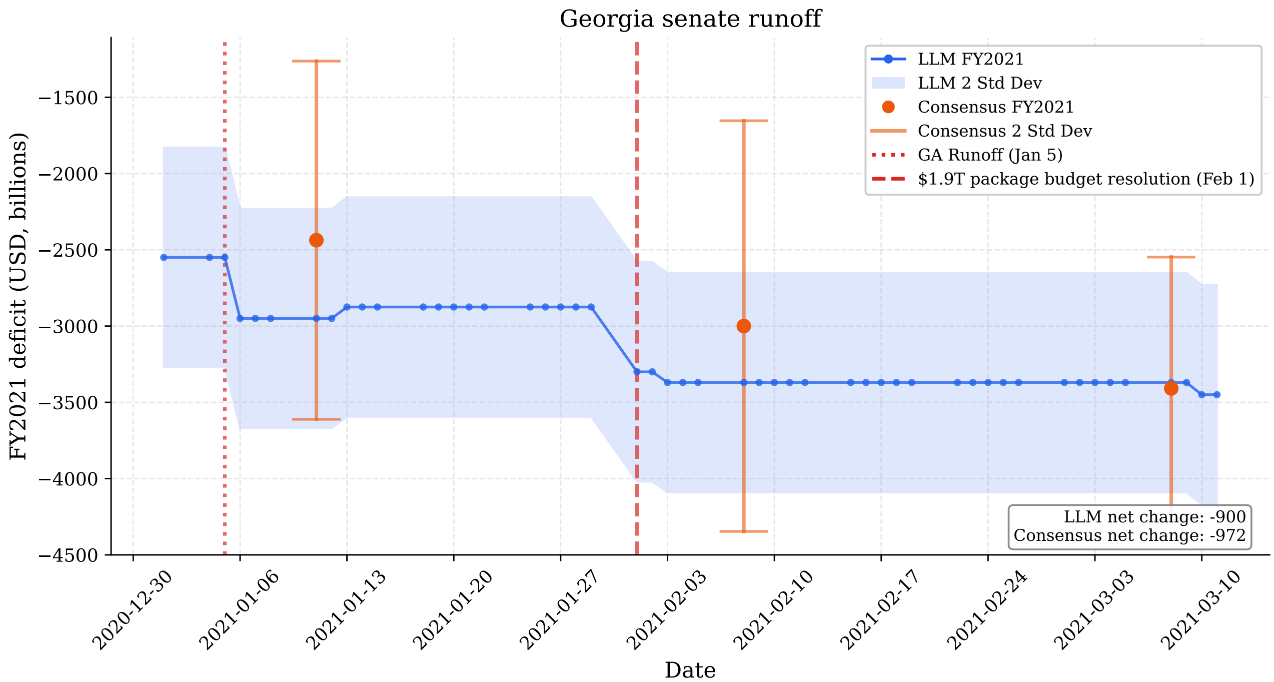


Figure 4: LLM and Consensus FY2021 deficit forecasts around the Georgia Senate runoff election.

therefore being more able to pass a follow-up round of expansionary Covid stimulus. This same shock is the subject of extensive study in Hazell and Hobler (2024), who base their deficit forecasts off of investment bank research reports, providing a form of validation for our model.

The first dotted red line indicates the day of the runoff election — January 5th — with results coming in on January 6th. The model increases its forecast of FY2021 deficits by $400B — very close to the $450Bn shock that Hazell and Hobler (2024) attribute to the Georgia runoff. Hazell and Hobler (2024) use a central estimate of $0.9T for the expected stimulus at the time of the Georgia runoff. It is worth quoting the forecaster's rationale on January 6th:

```
I widen deficits materially from FY2021 onward because the Georgia results
now make a unified Democratic government the base case, sharply raising the
probability and expected size of additional fiscal packages beyond what I had
embedded under divided government.
Relative to my 12/21-1/5 baseline (which only fully incorporated the $900bn
December package and assigned low odds to further large bills), I now assume an
additional COVID relief package with a central net size around $0.9T (range
$0.7-1.1T), probability ≈80%.  This implies ≈$0.72T in expected extra cost
versus ≈$0.20T previously, a delta of ≈$0.52T. I allocate about $0.40T of that
```

```
delta to FY2021 and $0.12T to FY2022 (larger checks, extended UI, state/local
aid, health and school support), moving FY2021 from -$2.55T to -$2.95T and FY2022
from -$1.52T to -$1.64T.
Unified Democratic control also raises the likelihood of a multi-year
infrastructure/green investment package and other priorities, partly but
not fully offset by higher corporate and high-income taxes.  Versus my
divided-government baseline, I add roughly $0.6T in expected net primary deficits
over FY2023-30, spread as modest increments of $60-80bn/yr.
```

The model has the exact same estimate of the expected size of the stimulus ($0.9T) and incorporates this information carefully relative to its prior forecast. It also makes an adjustment to farther out fiscal years that may be affected by longer-term infrastructure spending — which indeed proved to be true with the summer 2021 passage of the Inflation Reduction Act.

The second dashed red line indicates February 1st, when Democrats filed a budget resolution to advance Biden's $1.9T package via reconciliation, and the model revised deficits by a further $425B. After one subsequent small revision, the model's forecast matches exactly where the Consensus forecast ends up after the eventual bill passed the Senate (March 6th), The net update in the Consensus forecast across this period is $972B — extremely similar to the model's net update of $900B. In this case, the Consensus forecast seems to lag the model a bit, but it ends up changing in a very similar fashion.[10]

Figure 5 shows the behavior of the model forecast around the outbreak of the COVID crisis. On March 9th, no shutdowns had been announced yet, and the model forecast was almost exactly in line with Consensus, at a modest FY2020 deficit of $1.13T. Between March 9th and the March 26th passage of the $2.2T CARES act chaos ensued, shutdowns were announced, and the model's forecast of FY2020 deficits sharply deteriorated.

Of note are the multiple updates the model made on the way to the passage of the CARES act. News was coming out every day about how long shutdowns would be and how big of a stimulus package was needed, and the model responded to this news with a series of forecast revisions. Also of note is that once again the model fully updates to where the Consensus forecast ended up after all of the turmoil.

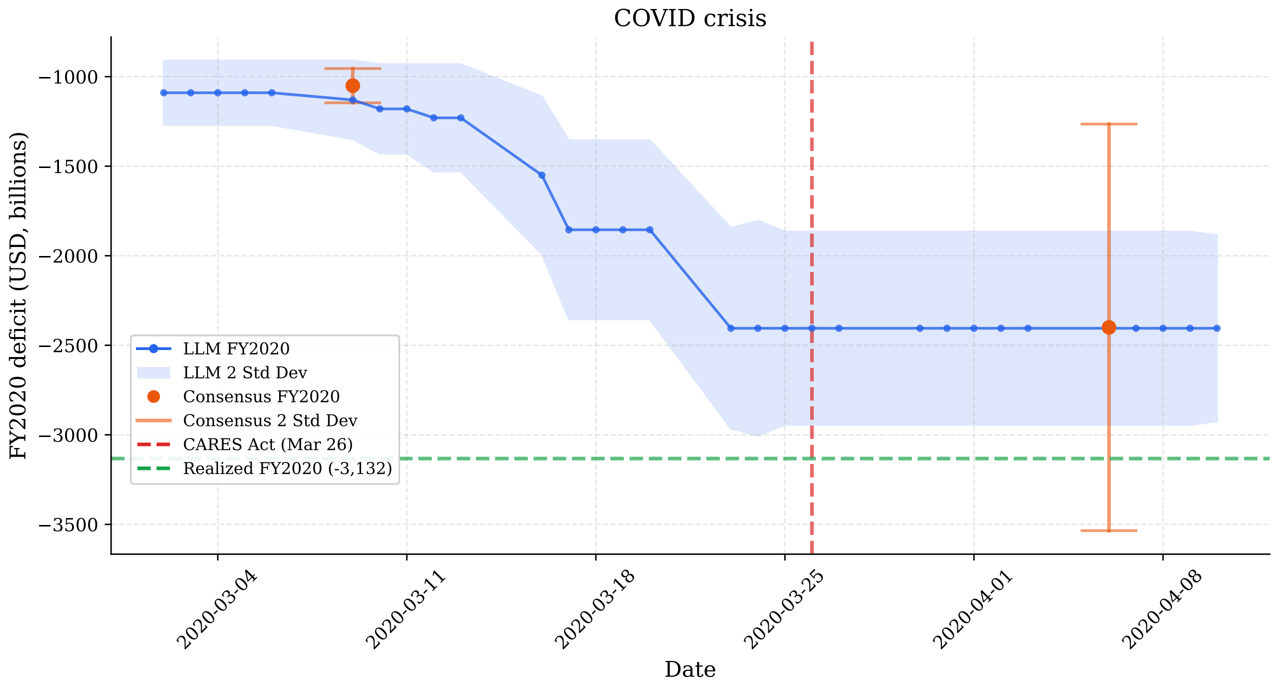


Figure 5: LLM and Consensus FY2020 deficit forecasts during the COVID crisis.

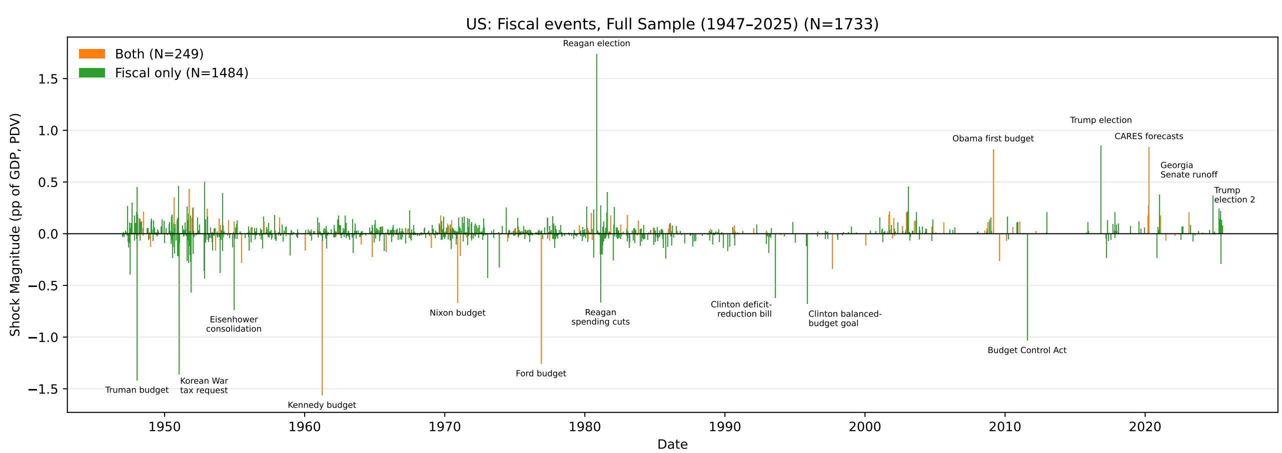


Figure 6: Daily changes in the PDV of expected deficits, classified by shock type. Sub-sample versions of this figure are available in Figure 17.

3.4 Shock history

Figure 6 shows the time-series of 1,733 "fiscal shocks" identified by the classifier model, with days that were *only* classified as a fiscal shock shaded in green (N=1,484), and days that included both a fiscal shock and an endogenous shock shaded in orange (N=249). The magnitude of the bars is the change in the PDV of deficits (as a ratio to GDP) on that day coming from *just* the fiscal shock part of the forecast change. Positive values indicate an increase in the expected deficit. The PDV's are calculated with respect to the yield curve of the day before the shock, so as to isolate the movement in deficit expectations.

The largest shocks identified by the model are all political developments that fit the intended definition of a fiscal shock. The single largest shock is on 1980-11-06 — the day the model updates on Ronald Reagan's election victory. The second largest event (in absolute value) is the Kennedy budget of 1961-03-27, where Kennedy explicitly pledged to balance the budget over the business cycle, leading to a large contractionary shock. All of the other ten largest fiscal events are either elections (e.g. Trump's first election), presidential budget messages (e.g. Truman's 1948 budget), or news about future legislation (e.g. Truman requesting tax increases to pay for the Korean War).

The time-series of fiscal events is disproportionately concentrated in the earlier years of the sample — 71% of all fiscal event or both days occur prior to 1980. The switch in LLM information sets in 2005 — from article summaries to full article text — may influence the frequency of fiscal event diagnoses, but the decline in frequency shows up as early as 1990, so that does not seem to be the driving force. Rather, the historical record suggests there simply were more "fiscal events" in the earlier sample. This finding aligns with the relative density of fiscal shocks as measured by two other leading examples of the narrative method — Ramey (2016) and Romer and Romer (2010) — which we now turn to.

3.5 Comparison with other fiscal shocks

As a final form of validation for our LLM procedure, we compare the shocks generated by our procedure to the narratively identified defense spending shocks in Ramey (2016) and the tax shocks in Romer and Romer (2010).

[10] Note that it is hard to know whether the Consensus forecast from just after the Georgia Senate runoff had fully incorporated the results from the Democratic victory. The Consensus dates represent the last possible date a forecaster could submit their forecast, but some, or even most, forecasters may have submitted their forecasts prior to the runoff election.

3.5.1 Ramey defense shocks

To compare our shocks with Ramey's defense shocks, we generate our own series of "LLM defense shocks." In each quarter, we manually search through all of the LLM's "fiscal event" or "both" days and read through the forecaster rationale to identify whether or not the motivation for the forecaster's change was due to defense-spending related shocks. We then isolate the portion of that days shock that was due to defense-spending. Typically, this is just the change in "G" (government purchases) identified by the classifier model. However, on some days, defense spending news is coupled with other spending news, and we read the forecaster rationale to extract just the defense component. These daily "defense" shocks are then cumulated into quarterly sums and discounted across fiscal years to generate a "quarterly PDV defense shock" which matches Ramey's identified shocks.[11] We focus on the PDV of the current and next two years LLM shock, since Ramey refrains from making longer-term assumptions in most of her PDV calculations.

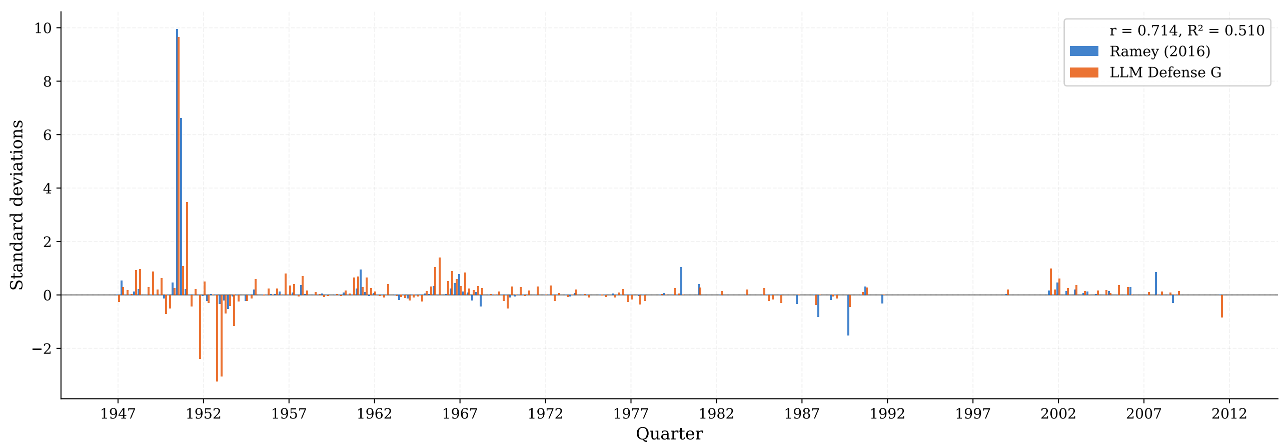


Figure 7: Defense News Shocks: Ramey (2016) vs LLM Defense G (1947–2015). Ramey PDV from Ramey 2016. LLM defense PDV discounted using flat 3-month T-bill rate, compounded annually, truncated to 3 fiscal years. Both normalized by quarterly nominal GDP and presented in standard deviation units. The figure in percent of GDP is available in appendix figure 18.

Figure 7 presents the results for the 1947-2015 sample for which we have both LLM and Ramey shocks available, in units of the standard deviation of each respective shock series. The legend in the top left indicates that the overall correspondence between Ramey's defense series and the LLM defense shocks is strong — the two series have a correlation coefficient of 0.71 and the LLM defense shocks explain 51% of the variance of the Ramey defense shocks. The biggest shocks in both samples are concentrated in the 1950 and 1951 spending on the

[11]We discount by the compounded 3-month T-bill rate rather than the zero-coupon yield curve to match Ramey's construction.

Korean War.

We present results in standard deviation units for visual comparability. Appendix figure 18 shows the results as a percent of GDP. The magnitude of Ramey shocks is much larger than the LLM defense shocks: a standard deviation of 6% of GDP versus 0.8% of GDP.

We attribute this discrepancy to a few different factors. The first major difference is that Ramey's primary news source was Bloomberg Businessweek, which often provided full forecasts of future defense spending. The LLM-summarized NYT articles typically do not contain equally explicit defense spending forecasts, providing Ramey with an informational advantage. We see this as an example of how our methodology can be improved with access to better information sets. A full historical time-series of Bloomberg Businessweek articles was not available to us for download, and making such a resource available to researchers could provide an even stronger validation of the Ramey series.

Another factor that generates discrepancies is sometimes the LLM mixes in defense with other factors, preventing us from isolating the defense shock. For example, Ramey's second-largest positive shock after the Korean War shocks is a 1980Q1 shock that stemmed from Carter's annual budget promising large increases in defense spending. The LLM picks up on shocks around the Carter budget, but does not mention the defense spending explicitly in its rationales, since it is not covered in the summaries of the NYT articles from the time. Once again, this offers a clear avenue to improving the LLM outputs by providing it with a richer information set.

In the case of both of the Korean War and the Carter budget, the subsequent evolution of defense spending suggests Ramey's identified shocks were closer to the "right" magnitude, while the LLMs shocks (or absence of shocks) were too small. Nonetheless, as we will show in table 7, the LLM defense shocks are stronger predictors of future defense spending than the Ramey defense shocks in the post-Korean War 1955-2015 sample.

Overall, we take the Ramey comparison as an encouraging sign that the LLM captures similar sources of exogenous variation in defense spending, while offering opportunities for further validating Ramey's numbers in future iterations of this methodology.

3.5.2 Romer and Romer tax shocks

In Romer and Romer (2010), there are two different type of shock series they construct. The primary shock series they use for analyzing the effects of exogenous changes in legislated taxes is aligned with the timing in which those tax shocks were *implemented*. However, they also construct a separate series based on the time at which a given piece of legislation was *passed*. When a piece of legislation has "phase-ins" this can lead to substantial discrepancies. For example, the Economic Recovery Tax Act of 1981 is a "passage" shock in 1981Q3, but is an implementation shock in all of 1981Q3, 1982Q1, 1983Q1, and 1984Q1.

The natural object to compare our LLM shocks with is the *passage* series, since its concept more closely aligns with what the LLM measures. To do so, we manually filter through the LLM fiscal event and both days, and then match them to the Romer events if the LLM rationale explicitly discusses the same piece of legislation that Romer and Romer cite in their narrative record. As with the Ramey shocks, we use the T component of the LLM shock, and sometimes attribute only a portion of it to the Romer and Romer match, if multiple events are driving the LLM's update.

The difference between our approach here and with the defense shocks is here we are explicitly seeing if the LLM discusses *the same* piece of legislation that Romer and Romer attribute the shock to, while in the Ramey comparison, we are looking at the universe of LLM shocks that have "defense" content, whether or not they align with a Ramey defense event. The reason is that Ramey's series is meant to capture the universe of news about defense spending, while Romer and Romer's series is only meant to capture tax changes that ended up going into effect. Therefore, to compare with Romer and Romer, we ignore LLM updates about tax bills that never ended up passing, and focus on whether or not the LLM is able to identify the same shocks Romer and Romer base their analysis on.

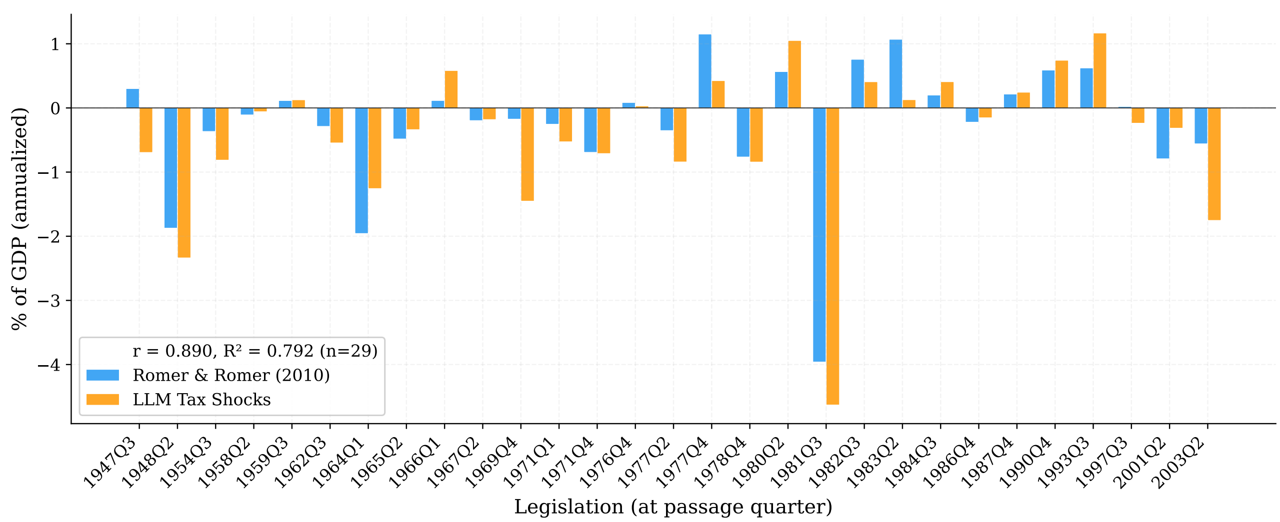


Figure 8: Tax Legislation: Romer & Romer (2010) vs LLM Tax Shocks, Exogenous Only. Romer PDV from **?**, discounted at 3-year Treasury yield. LLM value is the maximum single fiscal year impact (undiscounted). Both normalized by passage quarter nominal GDP.

Figure 8 presents the results, where the LLM bars are measured as the cumulative change in T from matched days for the fiscal year with the largest cumulative change in T. We call this the "max FY" approach and focus on it because it most closely aligns with the numbers Romer and Romer base their shock series off.

The timing is aligned with the passage quarter in Romer and Romer's dates, though the

LLM shocks need not come from the same quarter (a bill that passes in 1980Q1 might have news that comes out about it in 1979Q4). We will leverage this difference in 5.2 when analyzing the effect of anticipated tax shocks.

The main message from the figure is that the LLM shocks "recover" the Romer shocks extremely well. The two series are highly correlated ($r = 0.89$) and the LLM shocks explain 79% of the variance of the Romer shocks. We see this as a highly encouraging validation of the LLM-based methodology.

Furthermore, some of the discrepancies between the two series are easily explainable in terms of the different concepts used in constructing the series. For example, consider the very first shock in the series — the 1947Q3 Social Security Amendments — which the LLM measures as a negative (expansionary) shock, but Romer and Romer measure as a positive (contractionary) shock. In the words of Romer and Romer, "the Social Security Amendments of 1947 postponed until January 1, 1950 an increase in the combined Social Security tax rate from 2 percent to 3 percent." Romer and Romer encode this as a positive shock, since it led to an eventual increase in the social security tax rate. The LLM codes this as a negative shock, since the 1947 bill *postponed* the increase in tax rates that was already expected to happen. On the day the LLM incorporated this news the forecaster rationale explicitly says "Before today, my baseline implicitly assumed the scheduled increase would occur with high probability." Thus, the discrepancy is driven by the difference between news that changes future taxes (what Romer and Romer focus on), and news that changes *expectations* about future taxes (what the LLM focuses on).

4 The effects of fiscal shocks on asset prices

To estimate the effect of fiscal shocks on asset prices, we focus on our 2005-25 sample, where real-time Reuters articles allow us to have confidence we are identifying fiscal shocks at the same time that news would have transmitted to financial markets. Results for the full sample are available in the appendix. We regress asset price changes from the day before to h days after on the fiscal shock:

$$\Delta y_{t-1,t+h} = \alpha + \beta_h \cdot \Delta \text{PDV}_t^{\text{fiscal}} + \gamma_h \cdot \Delta \text{PDV}_t^{\text{endog}} + \varepsilon_t \tag{6}$$

where $\Delta y_{t-1,t+h}$ is the change in the asset price from the close of day $t-1$ to the close of day $t+h$, and $\Delta \text{PDV}_t^{\text{fiscal}}$ is the change in the present discounted value of expected deficits (as a share of GDP) on day t attributed to fiscal events. The PDV is computed with respect to the yield curve the day prior to the shock, so that only the change in forecasted deficits contributes. $\Delta \text{PDV}_t^{\text{endog}}$ is the endogenous component of the shock, if the shock comes from a "both" day which includes both a fiscal event and an endogenous change.

We restrict the sample so that no other forecast changes occur on the day before or the h days after the event, and present results either where the sample is restricted to *only* fiscal event days or when the sample includes "both" days and the endogenous control is present. Standard errors are heteroskedasticity-robust.

Table 5: Asset Price Response to Fiscal Shocks (2005–2025)

	Fiscal Only		Endogenous Control	
Asset	2-day	11-day	2-day	11-day
ln(S&P 500)	1.00	1.71	1.15	4.10
	(1.27)	(1.93)	(1.47)	(2.77)
ln(FX Broad)	1.75***	4.80***	1.21*	2.75***
	(0.60)	(0.40)	(0.63)	(1.04)
2Y Treasury Yield	8.94***	21.62***	5.02	18.93**
	(2.76)	(6.47)	(3.60)	(8.20)
2Y TIPS	3.03	23.73**	−1.51	−0.26
	(7.23)	(11.73)	(6.76)	(22.11)
2Y Breakeven	6.26	−1.45	6.61	19.52
	(6.18)	(10.94)	(6.57)	(18.80)
10Y Treasury Yield	25.21***	45.77***	17.65*	34.97**
	(6.72)	(12.46)	(9.10)	(15.24)
10Y TIPS	13.95***	33.56***	6.09	13.57
	(3.74)	(7.02)	(6.27)	(16.36)
10Y Breakeven	11.83**	12.77	11.90**	21.74***
	(5.83)	(9.22)	(6.04)	(7.02)
N (events)	61	33	85	52

Notes: Coefficients and standard errors multiplied by 100. Window: $[t-1, t+h]$.
"Fiscal only" restricts to days with only fiscal events; "Endogenous control" includes days with both fiscal and endogenous changes, controlling for the latter.
Clean events: no other shocks in $[t-1, t+h]$ window. Heteroskedasticity-robust standard errors.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. See Table 11 for 1980–2025 results.

Table 5 shows the results for the 2-day window ($h = 1$, from the day before to the day after the event) and the 11-day window ($h = 10$), across eight assets. The first two columns restrict the sample to fiscal-event-only days; the last two include "both" days and control for the endogenous component.

Our primary focus is the "fiscal only" sample, since it presents the cleanest identification. Four assets respond in line with what theory would predict and in highly significant fashion, across both the 2-day and 11-day windows: the dollar, the 2Y Treasury, the 10Y Treasury, and the 10Y real yield as measured by TIPS.

In response to a 1pp increase in the PDV of deficits (similar in magnitude to the largest shock days in the sample), the dollar appreciates 1.75% at the two-day horizon, and a very large 4.8% at the 11-day horizon. The dollar appreciating in response to an expansionary fiscal shock is consistent with standard open-economy macro models but in contrast to some prior empirical work (Monacelli and Perotti 2006, Auerbach and Gorodnichenko 2016).

At the two-day horizon, the ten-year Treasury yield rises 25 basis points, the two-year yield 9 basis points, and the ten-year real rate 14 basis points. At the eleven-day horizon, movements are substantially larger: the two-year yield is 22 basis points higher, the ten-year yield is 46 basis points higher, and real ten-year yields are 34 basis points higher. The 24bps gap between the ten and two-year response implies a steepening yield curve, perhaps reflecting long-term fiscal sustainability risk or long-term debt supply.

Including "both" days with the endogenous control yields broadly similar patterns. The ten-year Treasury yield response at eleven days is 35 basis points. The main difference is that the 10Y inflation breakeeven response is now more consistently significant than the 10Y real yield response. This could be due to "both" days being more likely to include negative (endogenous) updates about growth prospects, which make the real yield response to a pure fiscal shock harder to identify. The broader pattern suggests that both 10Y real yields and 10Y breakeven inflation contribute to the rise in 10Y nominal yields.

The response of the stock market, as measured by the S&P 500, is positive but insignificant across all specifications.[12]

In the full sample and pre-2005 results in appendix tables 11 and 12 we show that the results on the ten-year nominal yield, ten-year real yield, and ten-year breakevens are largely similar, while the dollar appreciation becomes insignificant. As noted before, we don't emphasize those results due to concerns about the timing of shocks being slightly misspecified.

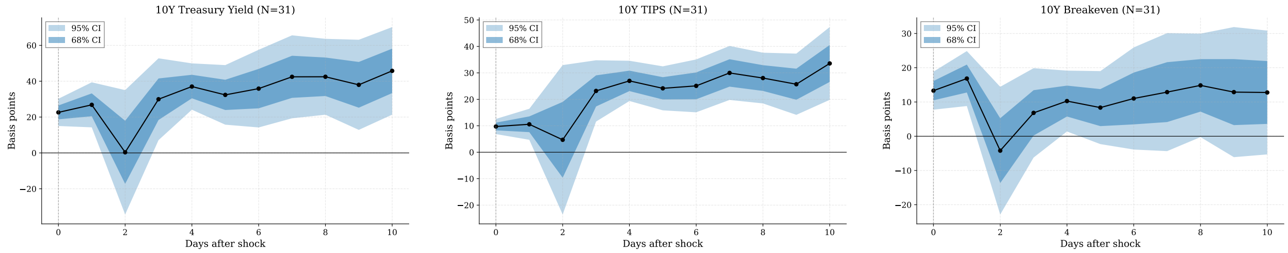
To trace out the dynamic response of the ten-year, we estimate impulse response functions (IRFs) using horizon-by-horizon regressions:

$$y_{t+k} - y_{t-1} = \alpha_k + \beta_k \cdot \Delta \text{PDV}_t + \varepsilon_{t,k} \quad \text{for } k = 0, 1, \ldots, K \tag{7}$$

where $y_{t+k} - y_{t-1}$ is the cumulative change in the asset price from the day before the shock $(t-1)$ to k days after the shock. The figures below restrict the sample to just "clean" fiscal events—those with no other fiscal shocks in the $[-1, +K]$ window—to isolate the dynamic response to a single shock. Confidence intervals are constructed using standard errors heteroskedasticity-robust (HC1). Figure 9 shows the IRF for 10-year Treasury yields, decomposed into the real rate (TIPS) and inflation expectations (breakeven) components.

All three assets yields jump significantly on the news of a deficit expansion, and both the nominal and real yield rise further in the days afterward. The fact that a large portion of

[12] This finding is consistent with Wiegand (2024), though she finds that at the ZLB, the stock market response is positive and significant.



(a) 10Y Treasury Yield (b) 10Y TIPS (Real Rate) (c) 10Y Breakeven Inflation

Figure 9: Impulse response of 10-year rates to a 1pp increase in the PDV of expected deficits (as share of GDP). 2005–2025 sample, clean fiscal events only, window $[t-1, t+10]$. Shaded regions show 68% and 95% confidence intervals, heteroskedasticity-robust (HC1).

the yield rise occurs not on the day of the event but in the days afterwards is consistent with either delayed processing of the news by the private sector or "slow-moving" capital which gradually adjusts to the shock (Duffie 2010).

4.1 Short-term vs. long-term shocks

Fiscal shocks may have different effects depending on their timing: shocks that primarily affect the near-term deficit may be seen as temporary, while shocks concentrated further out may signal more persistent changes to fiscal sustainability and treasury supply. To examine this, we split events based on the *weighted horizon* of the fiscal shock:

$$\text{Weighted Horizon} = \frac{\sum_n n \cdot |\Delta d_{t,n}^{\text{fiscal}}|}{\sum_n |\Delta d_{t,n}^{\text{fiscal}}|} \tag{8}$$

where $\Delta d_{t,n}^{\text{fiscal}}$ is the fiscal event component of the deficit change for fiscal year n years ahead. This measure captures the "center of mass" of the shock across forecast horizons. We then split the sample at the median weighted horizon within the clean events sample. For these results, we include "both" days and the endogenous control to increase our sample size. For the 11-day window, the median weighted horizon is 3.7 years, yielding 26 short-term and 25 long-term events.

Table 6 shows the results. The pattern is striking: long-term shocks drive all of the consistently significant asset price responses, while short-term shocks show large standard errors and almost no significant effects. Focusing on long-term shocks, the dollar appreciates a significant 1.8% at two days and 4.4% at eleven days. At the two-day horizon, the ten-year yield rises 24 basis points and the ten-year real rate 13 basis points. At the eleven-day horizon, the ten-year yield is 42bps higher, with a 24bps increase in real yields and an 18bps rise in breakevens—all significant at the 1% level.

Table 6: Asset Price Response by Shock Horizon (2005–2025)

	2-day window		11-day window	
Asset	Short-term	Long-term	Short-term	Long-term
ln(S&P 500)	3.33	0.51	32.1***	1.77
	(6.14)	(1.20)	(11.95)	(1.97)
ln(FX Broad)	0.87	1.76***	−5.13	4.41***
	(1.03)	(0.58)	(3.38)	(0.93)
2Y Treasury Yield	6.28	4.92	20.6	24.8***
	(8.11)	(4.31)	(60.65)	(6.47)
2Y TIPS	−5.35	0.82	−100	8.35
	(23.24)	(6.49)	(136)	(17.22)
2Y Breakeven	13.5	3.94	121	16.6
	(24.27)	(5.36)	(92.20)	(15.35)
10Y Treasury Yield	−4.03	24.1***	28.6	41.7***
	(20.08)	(9.09)	(99.39)	(13.93)
10Y TIPS	−15.0	13.0**	−68.2	23.5***
	(23.00)	(5.25)	(145)	(8.67)
10Y Breakeven	11.4	11.4*	98.5**	18.2***
	(21.16)	(6.20)	(47.90)	(6.53)
N (events)	42	41	26	25

Notes: Coefficients and standard errors multiplied by 100. Window: $[t-1, t+h]$.
Endogenous controlled. Heteroskedasticity-robust standard errors.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. See Table 13 for full-sample results.

Tables 14 and 15 in the appendix show the results split by shocks to government spending versus shocks to taxes/transfers. Responses to tax/transfer shocks are much more significant, but that may simply be due to the very small sample of government spending shocks in our windows — signs and magnitudes are broadly similar.

4.2 Commentary

It is worth taking a moment to compare our results on how higher deficit expectations impact yields with others in the literature, and to examine their implications for theory. The closest two studies to ours are Wiegand (2024) and Gomez Cram, Kung and Lustig (2025), both of which also focus on the effect of high-frequency fiscal news on yields.

Wiegand (2024)'s primary specification is the two-day event-study ($h = 1$). She finds that "a news shock that increases the cumulative deficit-to-GDP ratio by 1% over 5 years raises 2-year yields by 2 bps and 10-year yields by 2.3 bp." Meanwhile, Gomez Cram, Kung and Lustig (2025), using a combination of model and data, find that a 1pp increase in "expected debt-to-GDP corresponds to an increase of the 10-year nominal yield by 31 bps."

Neither of these numbers are directly comparable to ours, since the measure of the fiscal shock is in different units, but a rough approximation can be made. Wiegand's measure is the present-value of the next five-years of deficits scaled by current GDP. Assuming her shocks don't contain news beyond a five year horizon, the shocks are *roughly* 1/10 that size when normalized by the present-value of the next ten-years of GDP. So, crudely, her estimates multiplied by 10 should be comparable to ours, which indeed they are: a 23bp rise in the 10-year, sitting between our estimate at $h = 1$ of 18-25bps and $h = 10$ of 35-46bps. Our results on the split between real yields and breakevens is also similar to hers: in our preferred 11-day fiscal-only estimates, 73% of the nominal increase is due to real yields; in hers, it is 66%.

If we assume our shocks do not stretch out beyond ten-years, a 1pp rise in the PDV of deficits over the PDV of GDP will give rise to an $\approx$ 10pp rise in expected ten-years later debt/GDP.[13] Therefore, our estimate of the rise in the ten-year nominal yield in response to a 1pp increase in expected debt-to-GDP is an order of magnitude smaller than Gomez Cram, Kung and Lustig (2025) — 3-5bps versus 31bps.

However, the Gomez Cram, Kung and Lustig (2025) estimates are a bit of an outlier in the literature. Rachel and Summers (2019) reviews a number of older papers and settles on a central estimate of a 3.5bps increase in *real* yields for a 1pp increase in debt-GDP. Our estimates for ten-year real yields in the 11-day fiscal-only specification are right in line — 3.4bps. The HANK model in Campos et al. (2024) predicts a 3.1bp rise in *real* yields.

The fact that some of the rise in long-term yields is attributable to long-term inflation

[13] A strong assumption.

expectations is consistent with fiscal theory of the price level models with long-term debt (e.g. Corhay et al. 2023) as well as HANK models like Campos et al. (2024). That some of the rise is in real yields is consistent with heterogeneous agent models where the asset demand function is not perfectly elastic. The increase in the 2Y-10Y spread is consistent with investor pricing of fiscal sustainability risks.

The rise of the dollar is one important moment for distinguishing between assumptions about "active" and "passive" monetary policy: traditional NK models with active monetary policy predict a monetary tightening and dollar appreciation in response to an expansionary fiscal shock, consistent with the behavior of the 2Y yield and the dollar in our data. Models with passive monetary policy (and active fiscal) would predict a dollar depreciation in response to an expansionary fiscal shock.

Perhaps most interesting for theory are the results split by weighted-horizon of the shock. Rachel and Ravn (2025) describe the contrasting implications of FTPL and HANK models as follows: "In RANK-FTPL, the inflationary and output effects of deficits announced and implemented today are identical to those of policies announced far out in the future (up to their present value being identical). In HANK instead, a surprise increase in the deficit stimulates current output and inflation, while announcements of future deficits may be inflationary but recessionary today. The timing matters in the HANK model because households perceive that those receiving transfers may not be the same households that have to pay for them be it through taxes or through inflation."

The 2Y breakeven response is not significant in any specification, which would seem inconsistent with both models. Conversely, in the 11-day window, short-term shocks lead to very large increases in both the S&P 500 and the 10Y breakeven, which is consistent with the HANK mechanism of surprise increases in the near-term deficit increasing both output and inflation.[14] That being said, in the 2-day window, the 10Y breakeven point-estimates are the same across short-term and long-term windows — consistent with RANK-FTPL predictions — albeit inconsistently significant. Overall, we think these results are suggestive of the value of further work on this topic, as a moment that can help discriminate between theories.

5 The macroeconomic effects of fiscal shocks

5.1 The effects of government spending

To evaluate the effects of shocks to government spending (or overall deficits) on macroeconomic variables, we first need to establish which of the "LLM shocks" is actually a relevant predictor of realized fiscal outcomes. Table 7 reports first-stage F-statistics from regressing

[14] Alongside sticky wages, which would lead to procyclical markups.

the cumulative endogenous variable $\sum_{j=0}^{H} g_{t+j}$ on $shock_t$ and lag controls z_{t-1} of the shock, the endogenous variable, and GDP. Following Ramey and Zubairy (2018), the shock variable is normalized by a one-quarter lag of the GDP deflator times trend real GDP, with trend real GDP estimates taken directly from their data.

We present results for four different shock variables: the original Ramey shock series, our LLM defense shocks (3y PDV), all LLM G shocks on non-stabilizing days cumulated within a quarter (full PDV), and all LLM fiscal event shocks on non-stabilizing days cumulated within a quarter (full PDV). For the first three, the dependent endogenous variable is real government spending (normalized by trend real GDP), while for all LLM fiscal event shocks, the dependent variable is real deficits.

Table 7: Predictive power of shocks for fiscal aggregates. Each entry is the F-statistic from regressing the indicated fiscal aggregate at horizon H on the shock, following the Ramey and Zubairy 2018 local projection specification with Gordon–Krenn normalization. Stars indicate instrument strength using the Lewis and Mertens (2026) pointwise critical values ($R = 1, \alpha = 0.05$): *** $\tau \leq 0.10$ ($F > 32.1$), ** $\tau \leq 0.20$ ($F > 17.4$), * $\tau \leq 0.30$ ($F > 11.9$), where τ is the maximum modal bias as a fraction of the OLS bias.

		1947–2015			1955–2015		
Shock	Dep. var.	H=4	H=8	H=12	H=4	H=8	H=12
Ramey defense news	G	109.4***	188.6***	150.5***	5.2	5.9	6.5
LLM defense (3yr PDV)	G	15.5*	44.8***	68.6***	16.3*	20.0**	23.6**
LLM G non-stab (full PDV)	G	2.3	5.2	9.0	14.9*	11.6	13.5*
LLM FE non-stab (full PDV)	D	2.5	2.0	1.4	0.1	0.1	0.3

The main message from the table echoes Ramey's comments in her 2019 review paper (Ramey, 2019) — it is hard to find instruments for fiscal outcomes that are both relevant and exogenous. According to the F-stat thresholds for LP-IV proposed by Lewis and Mertens (2026), over the full 1947-2015 sample, both Ramey defense news and LLM defense news are sufficiently relevant instruments, though the Ramey shock variable is noticeably stronger. Removing the Korean war and focusing on the 1955-2015 sample, the LLM defense news variable is the only shock variable considered that is sufficiently relevant, and even still, it does not clear the most restrictive $\tau \leq 0.1$.[15]

It is curious that the LLM defense shocks are so much more relevant than the more "comprehensive" overall government spending or overall deficit shocks. One explanation would be that military spending news is more likely to be "followed through on", while expectations about other forms of spending are more likely to be reversed by subsequent developments.

[15] τ represents the maximum displacement of the mode of the IV estimator's distribution, expressed as a fraction of the full OLS bias.

With these results in mind, we focus on two main analyses: comparing the reduced-form impulse responses to Ramey and LLM defense shocks, and comparing the cumulative government spending multipliers implied by Ramey and LLM defense shocks.

Reduced-form impulse responses. Figures 10 and 11 report reduced-form local projection impulse responses. For each horizon $h = 0, 1, \ldots, 20$ and response variable $x \in \{G, Y, C, I\}$ (all in GK units), we estimate

$$x_{t+h} = \alpha_h + b_h \, shock_t + \sum_{\ell=1}^{4} \left[\beta_{1,\ell}^h \, shock_{t-\ell} + \beta_{2,\ell}^h \, y_{t-\ell} + \beta_{3,\ell}^h \, g_{t-\ell} \right] + \varepsilon_{t+h}, \tag{9}$$

where y_t and g_t denote real GDP and real government spending (normalized by trend real GDP). The coefficient b_h traces out the impulse response of x at horizon h to the shock at time t. Standard errors use the Newey–West HAC estimator with bandwidth equal to $\max(h, 1)$. To make magnitudes comparable across the Ramey and LLM shocks, each impulse response is multiplied by the sample standard deviation of the respective shock, so that figures show the response to a one-standard-deviation innovation. Shaded bands are 68% Newey–West confidence intervals.

When including the Korean War (1947-2015), the Ramey shock series produces much larger estimates of both the response of government spending and GDP than the LLM shock, but the sign, temporal pattern, and significance are all aligned. Despite the difference in magnitude of response for spending and GDP, the responses of consumption and investment are very similar across the two.

The 1955-2015 results show the opposite pattern: the spending and GDP responses are much larger in response to the LLM shock, and across GDP, consumption, and investment, the response to the LLM shock is more consistently significant. Encouragingly, the response of GDP to the LLM shock in the restricted sample matches the patterns of the GDP response (to both the Ramey and LLM shocks) in the larger sample, with a peak response occurring about 10 quarters after the shock. Both consumption and investment show consistent patterns of crowding out.

LLM defense only results with 95% confidence intervals are in appendix figures 19 and 20. The positive response of government spending and GDP as well as the immediate crowding-out of consumption and investment are significant at the 5% level.

Cumulative multipliers. Table 8 reports cumulative government spending multipliers estimated via one-step IV local projections, again following Ramey and Zubairy (2018). For

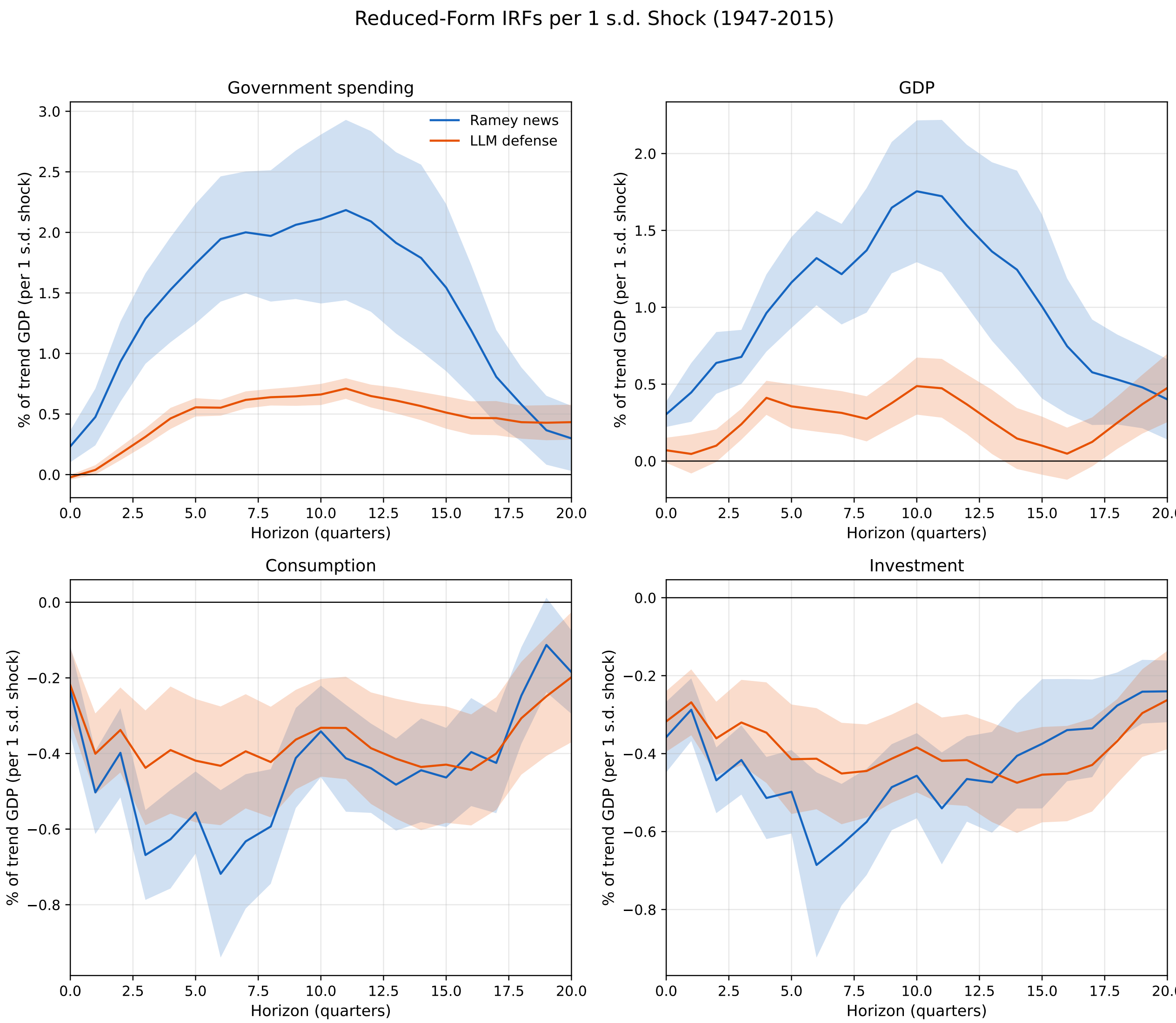


Figure 10: Reduced-form LP impulse responses per one-standard-deviation shock: Ramey (2011) defense news vs LLM defense 3yr PDV. Gordon–Krenn normalization with Ramey polynomial trend. 68% Newey–West confidence bands. Sample: 1947–2015. Results for LLM defense shock with 95% confidence intervals in appendix figure 19.

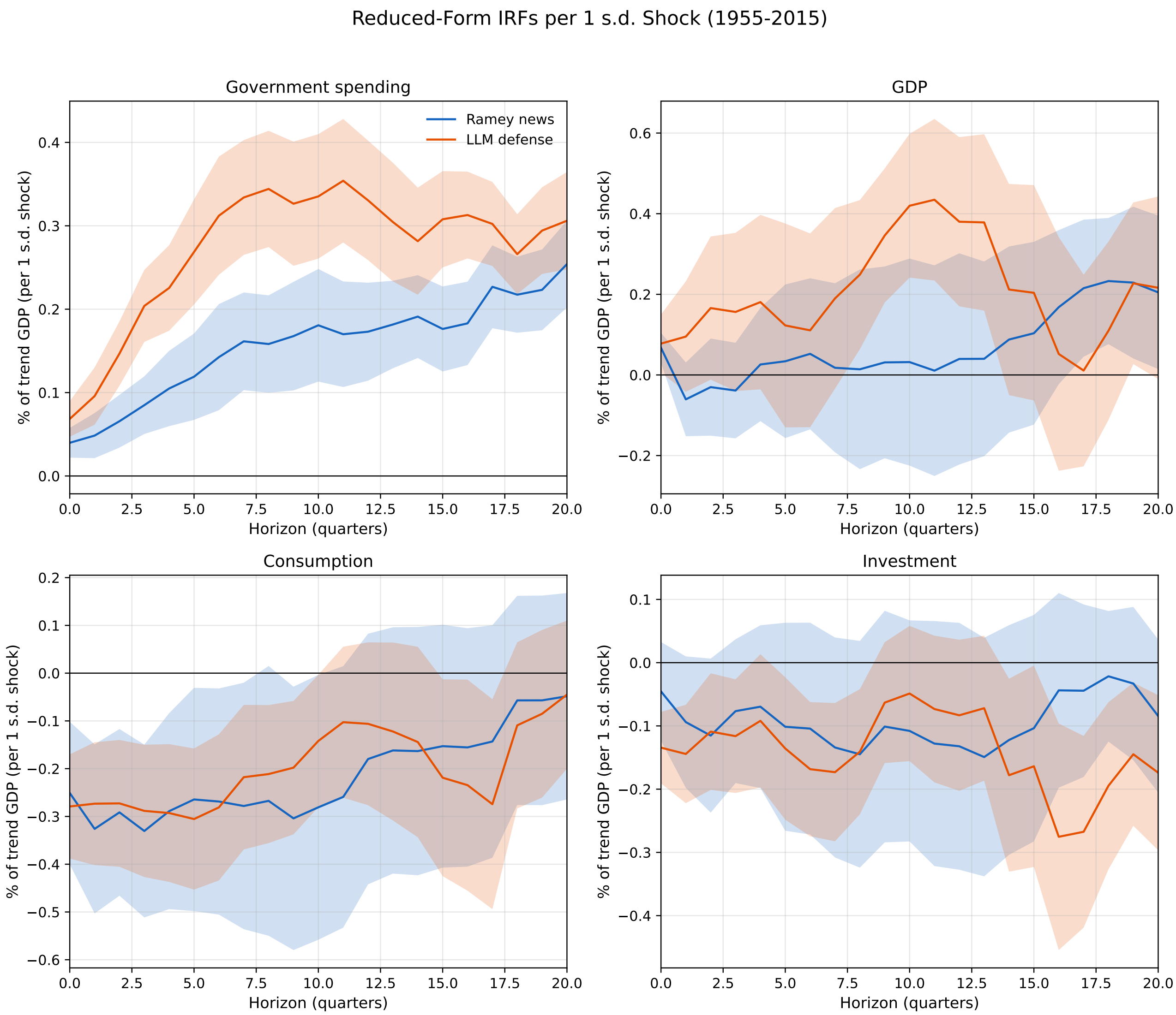


Figure 11: Reduced-form LP impulse responses per one-standard-deviation shock: Ramey (2011) defense news vs LLM defense 3yr PDV. Gordon–Krenn normalization with Ramey polynomial trend. 68% Newey–West confidence bands. Sample: 1955–2015. Results for LLM defense shock with 95% confidence intervals in appendix figure 20.

each horizon $H = 0, 1, \ldots, 20$, we estimate

$$\sum_{j=0}^{H} y_{t+j} = \alpha_H + m_H \sum_{j=0}^{H} g_{t+j} + \boldsymbol{\phi}_H' \mathbf{W}_{t-1} + \eta_{t+H}, \tag{10}$$

where y_t and g_t are real GDP and real government spending (normalized by trend real GDP) and $\mathbf{W}_{t-1} = (shock_{t-1}, \ldots, shock_{t-4},\ y_{t-1}, \ldots, y_{t-4},\ g_{t-1}, \ldots, g_{t-4})'$ is a 12×1 vector of predetermined controls consisting of four lags each of the instrument, y, and g, with $\boldsymbol{\phi}_H$ a horizon-specific coefficient vector. The cumulative spending variable $\sum_{j=0}^{H} g_{t+j}$ is instrumented by $shock_t$. The coefficient m_H is the cumulative multiplier at horizon H: the cumulative dollar change in GDP per cumulative dollar change in government spending over the $H+1$ quarters following the shock.

Estimation proceeds by two-stage least squares. In the first stage, we regress $\sum_{j=0}^{H} g_{t+j}$ on $shock_t$ and the controls $\mathbf{W}_{t-1}$. In the second stage, we regress $\sum_{j=0}^{H} y_{t+j}$ on the fitted values $\widehat{\sum g}$ and $\mathbf{W}_{t-1}$. Standard errors are computed as follows. Let $\tilde{z}_t$ and $\tilde{g}_t$ denote the instrument and cumulative spending after partialling out $\mathbf{W}_{t-1}$, and let $\hat{\eta}_t = \sum_{j=0}^{H} y_{t+j} - \hat{m}_H \sum_{j=0}^{H} g_{t+j} - \mathbf{W}_{t-1}' \hat{\boldsymbol{\phi}}_H$ be the IV residual computed with *actual* (not fitted) cumulative spending. We form the moment series $\tilde{z}_t \hat{\eta}_t$ and estimate its long-run variance $\hat{\Omega}_H$ using the Bartlett kernel with bandwidth $\max(H, 1)$. The variance of the multiplier is then $\widehat{\mathrm{Var}}(\hat{m}_H) = T\,\hat{\Omega}_H \,/\, (\tilde{z}'\tilde{g})^2$, the standard IV sandwich formula for a scalar instrument after partialling out the controls.

Table 8: Government spending multiplier from IV-LP (cumulative GDP / cumulative G), using Ramey defense news and LLM defense 3yr PDV as instruments. Gordon–Krenn normalization with Ramey polynomial trend. Delta-method standard errors from Newey–West HAC in parentheses.

	1947–2015		1955–2015	
	Ramey news	LLM def. 3yr	Ramey news	LLM def. 3yr
$H = 4$	1.02 (0.22)	0.91 (0.37)	−0.13 (1.40)	0.92 (0.91)
$H = 8$	0.62 (0.15)	0.68 (0.26)	0.03 (1.34)	0.68 (0.72)
$H = 12$	0.60 (0.15)	0.69 (0.24)	0.25 (1.37)	0.88 (0.56)
$H = 16$	0.48 (0.17)	0.59 (0.25)	0.30 (1.28)	0.78 (0.61)
$H = 20$	0.50 (0.19)	0.65 (0.26)	0.57 (1.11)	0.86 (0.51)

In the 1947-2015 sample, both the Ramey and LLM-based estimates are highly significant and closely aligned in magnitude: the $H = 4$ quarter multiplier is close to 1, while longer horizon multipliers sit in the 0.5 – 0.7 range. In the 1955-2015 sample, estimates of the multiplier based on the Ramey shock are closer to 0 and nowhere near significant. However, the LLM defense shock series produces estimates closer to the fuller sample results, albeit slightly higher $(0.7 - 0.9)$ and only significant relative to one standard-deviation confidence intervals.

We see our results as delivering three primary messages. Firstly, the fact that an entirely new methodology can produce such similar results to the Ramey estimates should further strengthen our confidence in those estimates: over the post-WW2 sample, it does indeed seem that shocks to (military) spending had output multipliers in the $0.5 - 0.7$ range and resulted in some crowding out of both consumption and investment. Secondly, the LLM defense 1955-2015 results provide additional support that such estimates are not entirely based on the Korean War shocks. Finally, there simply may not exist sufficient exogenous variation in the narrative record of US government spending to produce stronger instruments than what the literature has already converged on.

5.2 The effects of anticipated tax shocks

We begin by testing whether the LLM fiscal news anticipates the **?** exogenous tax shocks. For each of the 29 Romer legislations that have an exogenous "at passage" shock, we regress the Romer shock on the eight prior quarters of matched LLM shocks. Let p_l denote the passage quarter of legislation l. We estimate the cross-sectional regression

$$\frac{\text{Romer}_l}{\text{GDP}_{p_l}} = \alpha + \sum_{k=1}^{8} \beta_k \frac{\text{LLM}_{l,\, p_l-k}}{\text{GDP}_{p_l}} + \varepsilon_l, \tag{11}$$

where Romer_l is the Romer present-discounted-value tax shock and LLM_{l,p_l-k} is the LLM-matched tax change for legislation l in the quarter k periods before passage. Both are normalized by passage-quarter nominal GDP. The F-statistic tests the joint null $\beta_1 = \cdots = \beta_8 = 0$; rejection implies the LLM identifies tax news before the legislation is passed. Table 9 reports results for two normalizations of the LLM tax variable — either the full PDV of matched LLM days divided by 11 (to convert into quasi-annual terms), or the fiscal year with the maximum change in absolute value.

The table shows that almost 90% of the variance of the Romer and Romer shocks are explained by the matched LLM news that arrives prior to the passage quarter of the tax legislation.

This finding has important implications for the analysis done in Mertens and Ravn (2012). In their paper, they split Romer and Romer's *at implementation* shocks into a binary category of anticipated/unanticipated based on whether the passage-quarter is more than 90 days before the implementation-quarter. They then use this timing convention to make inferences about the effects of anticipated versus unanticipated tax shocks. Since our evidence suggests most of the variance in Romer and Romer's shocks *at passage* is already captured by the LLM

Table 9: Predictability of Romer & Romer (2010) exogenous tax shocks from lagged LLM fiscal news. Cross-sectional OLS with one observation per legislation ($N = 29$), restricted to legislations with non-zero exogenous tax components. The dependent variable is the Romer EXOGEPDV at the passage quarter, normalized by passage-quarter nominal GDP (%). Independent variables are LLM-matched tax changes at lags $p-1$ through $p-8$ quarters before passage, also normalized by passage-quarter GDP. "PDV/11" sums the present discounted value of LLM-identified tax changes across all eleven forecast fiscal years and divides by eleven; "Max FY" uses the single fiscal year with the largest absolute impact. The F-statistic tests the joint null that all eight lag coefficients are zero. Coefficient standard errors use the HC1 correction; the F-statistic uses the classical (non-robust) distribution, which is more reliable with $N = 29$.

	PDV/11	Max FY
F-statistic	21.50	18.65
p-value	< 0.001	< 0.001
R^2	0.896	0.882
N	29	29

in quarters prior to the passage, we re-visit their analysis.

To visualize when fiscal news arrives relative to implementation, Figure 12 compares the distribution of implementation lags under four criteria. Panel (a) replicates Mertens and Ravn (2012) Figure 1: the number of days between presidential signing of legislation and the implementation quarter. The remaining three panels replace the signing date with LLM-based timing measures. *LLM First Match* (panel b) uses the earliest date on which the LLM identifies a single fiscal event day matched to the legislation. *50% LLM Signal* (panel c) uses the first date at which the cumulative LLM-matched tax signal for the legislation reaches 50% of its own final matched total, in PDV terms. *50% Romer* (panel d) uses the first date at which the cumulative PDV LLM signal reaches 50% of the Romer exogenous present-discounted-value shock. For legislations without LLM matches, each panel defaults to the MR signing date. Table 16 in the Appendix reports the event-level dates underlying these distributions.

The contrast is evident: in Mertens and Ravn (2012) designation, 38 out of 70 implementation shocks are anticipated, while 32 are unanticipated. Across the three other measures, no more than 9 of the tax shocks were truly unanticipated.[16]

To estimate the macroeconomic effects of anticipated versus unanticipated tax changes, Mertens and Ravn (2012) use an exogenous-shock VAR framework. To measure anticipated tax shocks, they let $\tau^a_{j,t}$ denote the tax liability change signed at date $t - j$ and implemented

[16] In the text of Mertens and Ravn (2012) they report that only 36 of the 70 shocks are anticipated, but their appendix table A1 which documents their data — and which we base our analysis off — shows 38 anticipated events. Thus, there is a slight discrepancy in our replication of their figure.

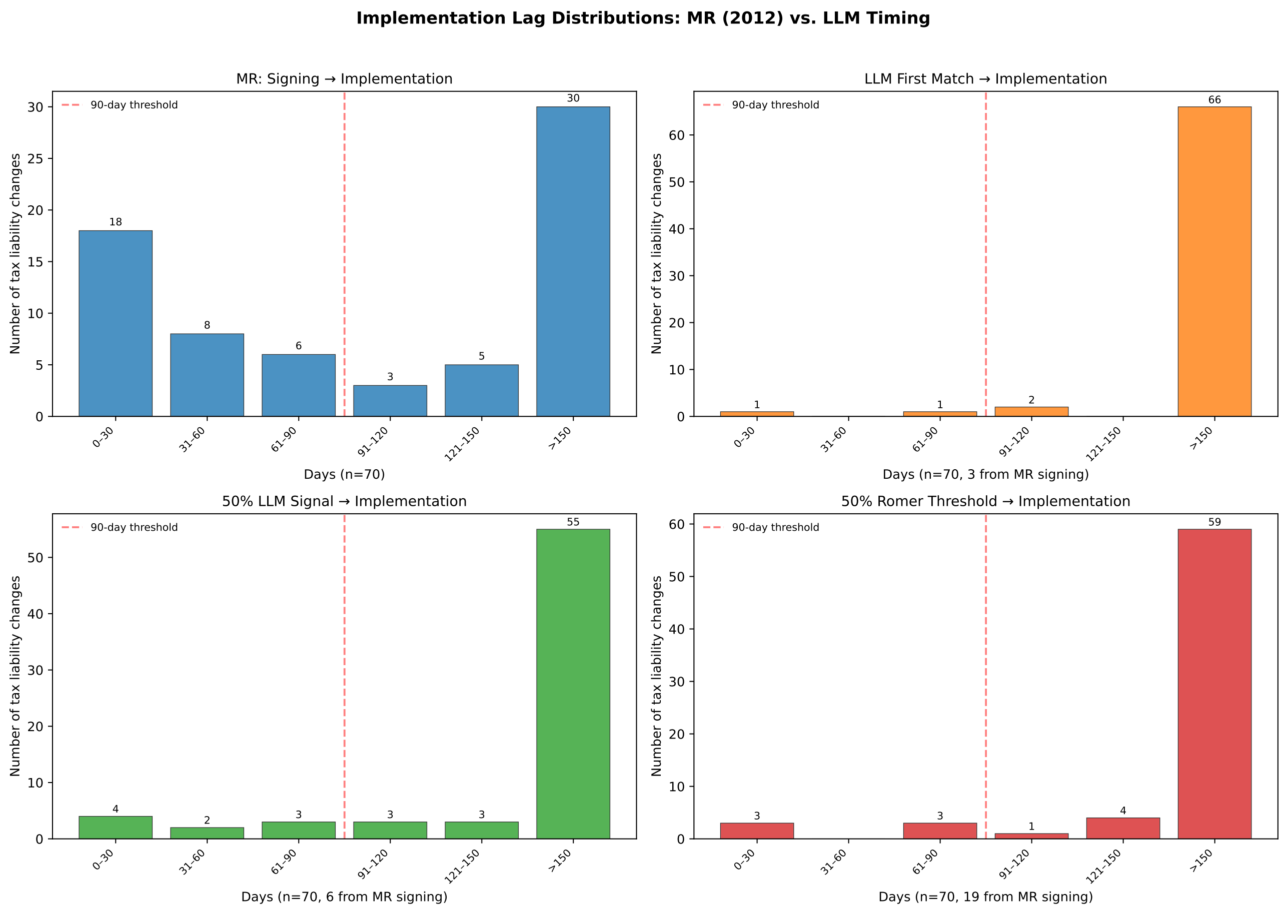


Figure 12: Implementation lag distributions for Romer & Romer (2010) tax legislations ($N = 70$). Bins follow Mertens and Ravn (2012) Figure 1: 0–30, 31–60, 61–90, 91–120, 121–150, and >150 days. The red dashed line marks 90 days. Unmatched legislations default to the MR signing date. See Table 16 for event-level detail.

at t and define:

$$s_{i,t} = \sum_{j=0}^{\bar{D}-i} \tau^a_{j+i,\,t+i}, \qquad i = 1, \ldots, \bar{D}, \tag{12}$$

so that $s_{i,t}$ measures the sum of all anticipated tax changes known at date t to be implemented i quarters in the future. $\bar{D}$ is the maximum anticipation horizon included in estimation.

The VAR specification is then

$$\mathbf{y}_t = \mathbf{a} + \mathbf{b}\, t + A(L)\, \mathbf{y}_{t-1} + B(L)\, \tau^u_t + C(L)\, \tau^a_t + \sum_{i=1}^{\bar{D}} \boldsymbol{\delta}_i\, s_{i,t} + \boldsymbol{\varepsilon}_t, \tag{13}$$

where $\mathbf{y}_t$ is a vector of log macroeconomic variables (output, consumption, investment, with hours and real wages added sequentially), $A(L)$ is a p-th order lag polynomial on the endogenous variables, and $B(L)$ and $C(L)$ are $(q+1)$-th order lag polynomials on the unanticipated and anticipated shock series (current value plus q lags). The model is estimated equation-by-equation by OLS with $p = 1$ VAR lag, $q = 12$ shock lags, and $\bar{D} = 6$.

We replace the binary anticipated/unanticipated classification with continuous anticipation weights derived from LLM news timing. For each legislation l with implementation quarter T_l, define $w_l(k)$ as the fraction of the total LLM-matched tax signal that has accumulated by quarter $T_l - k$. Our specification collapses the binary distinction between anticipated/unanticipated shocks into a single series τ_t (the full tax change for every legislation enters at its implementation quarter) and replaces the anticipation terms with continuously weighted versions:

$$\mathbf{y}_t = \mathbf{a} + \mathbf{b}\, t + A(L)\, \mathbf{y}_{t-1} + B(L)\, \tau_t + \sum_{i=1}^{\bar{D}} \boldsymbol{\delta}_i\, \tilde{s}_{i,t} + \boldsymbol{\varepsilon}_t, \tag{14}$$

where $\tilde{s}_{i,t} = \sum_l w_l(i) \cdot \text{size}_l \cdot \mathbf{1}[T_l = t + i]$. A legislation that received gradual news coverage has its impact smoothly distributed across anticipation lags, while one that appeared suddenly loads primarily on the contemporaneous shock τ_t. For legislations with no LLM matches, we fall back to MR's binary classification with the original anticipation horizon. We extend the anticipation horizon to $\bar{D} = 12$ quarters to accommodate the longer lead times revealed by LLM news timing.[17]

Figure 13 compares the anticipated tax cut impulse responses under the original Mertens and Ravn (2012) specification ($\bar{D} = 6$, with their binary anticipated shock series) with our LLM-reweighted specification ($\bar{D} = 12$). All impulse responses are shown for a tax *cut* of

[17] $\bar{D} = 12$ is chosen as the largest horizon at which more than 10 quarterly observations have non-zero anticipation weights. Results are robust to $\bar{D} \in \{6, 8, 10, 14\}$; see Figure 14.

1% of GDP, with 68% and 95% non-parametric non-centered bootstrap confidence intervals from 10,000 replications.

There are a few key differences in results: post-implementation, the rise of output, consumption, investment, and hours is smaller than the original MR specification, and less clearly significant, but the patterns are broadly similar. Pre-implementation, the *negative* effects on output, consumption, and investment are much more pronounced. For both consumption and investment, the fall in the variable pre-implementation is greater than the rise in post-implementation. In the case of investment and output, there is some evidence that in the first quarters of anticipation there is actually a boom, followed by a subsequent more elongated pre-implementation bust. Real wages now show no sign of increasing pre or post implementation.

We compute so-called "quasi"-multipliers using the IRF of log GDP at each horizon (since the shock is already 1% of GDP). We use these static multipliers relative to the shock rather than cumulative multipliers, since as Ramey (2019) emphasizes, computing cumulative multipliers for tax shocks using the dynamic response of tax revenue in the denominator is problematic when endogenous feedback from GDP to revenue can drive the denominator to zero (which we found when we tried to compute dynamic multipliers).

In Mertens and Ravn (2012)'s original specification, the peak impact multiplier for an unanticipated tax cut is 1.93 at quarter 10, while the anticipated tax cut multiplier peaks at 1.76 at the same horizon. Under our reweighted specification with $D = 12$, the peak multiplier is 1.22 at quarter 10. This estimate is smaller than most empirical estimates cited in Ramey (2019) and closer to the predictions of New Keynesian DSGE models, providing a potential resolution to that discrepancy.

The most striking finding is the fall in consumption prior to implementation. This is in contrast to the findings of Mertens and Ravn (2012) and a prior empirical literature (e.g. Poterba (1988) and Heim (2007)). It also is inconsistent with both the baseline permanent income-hypothesis and standard Ricardian equivalence, which would predict either a rise or no change in consumption prior to a tax cut. (Figure 14 shows how important a sufficiently long anticipation horizon is for the finding that consumption falls prior to the implementation of a tax cut. It plots our results across different choices of the anticipation horizon — $\bar{D} \in \{6, 8, 10, 12, 14\}$.

Even with our continuously re-weighted anticipation shocks, the six-quarter anticipation horizon of Mertens and Ravn (2012) produces a negligible fall in consumption prior to implementation. As the anticipation horizon gets longer, the fall in consumption becomes more and more pronounced. Additionally, the subsequent rise in consumption is more muted when more appropriate anticipation horizons are included. In our LLM-matched shock series, 28% of implementation shocks have their first matched day greater than 10 quarters

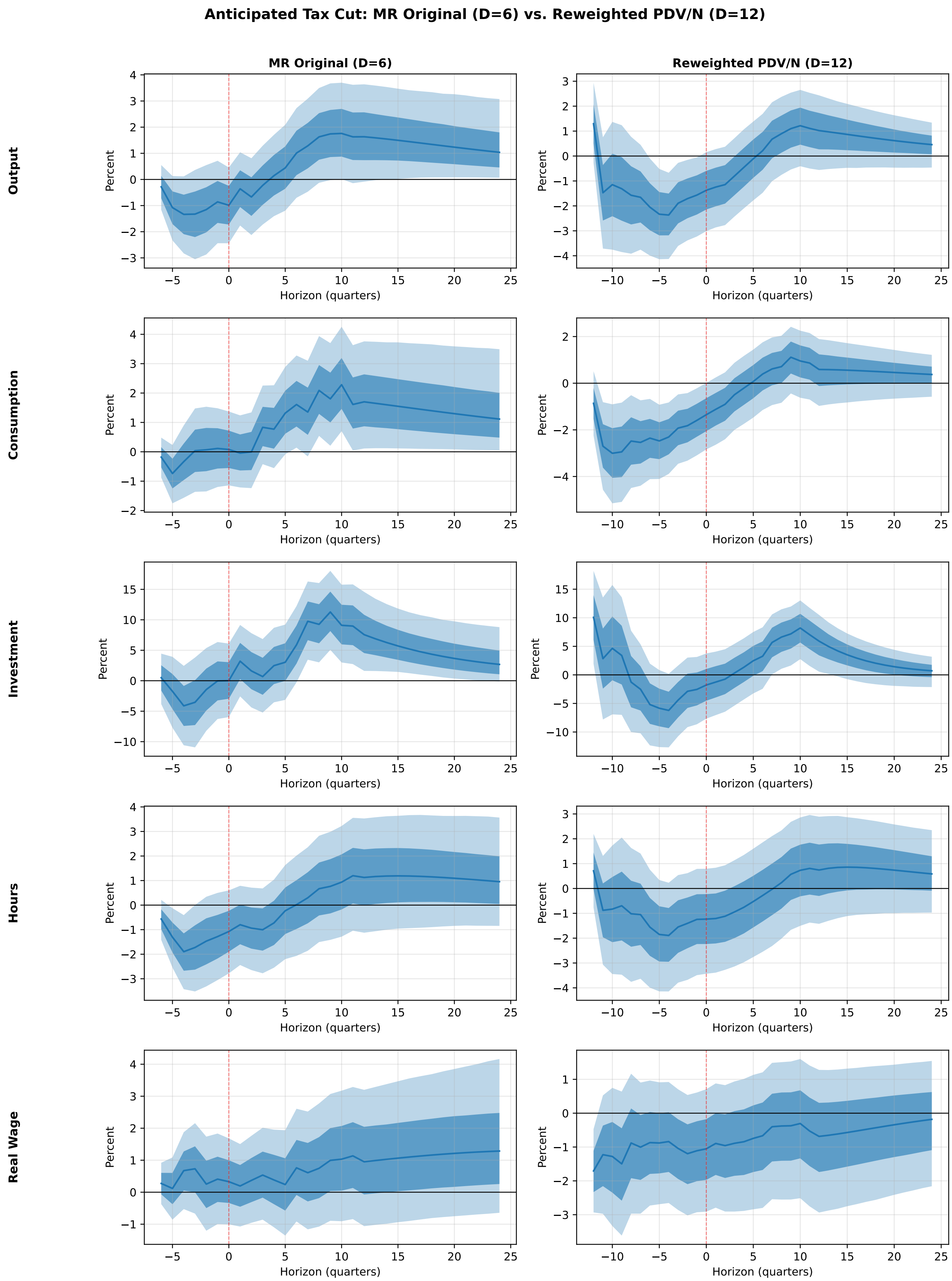


Figure 13: Anticipated tax cut impulse responses: Mertens and Ravn (2012) original specification ($\bar{D} = 6$, left) versus LLM-reweighted specification ($\bar{D} = 12$, right). Shaded bands show 68% and 95% bootstrap confidence intervals (10,000 replications). Variables in log levels; IRFs show percent response to a 1% of GDP tax cut.

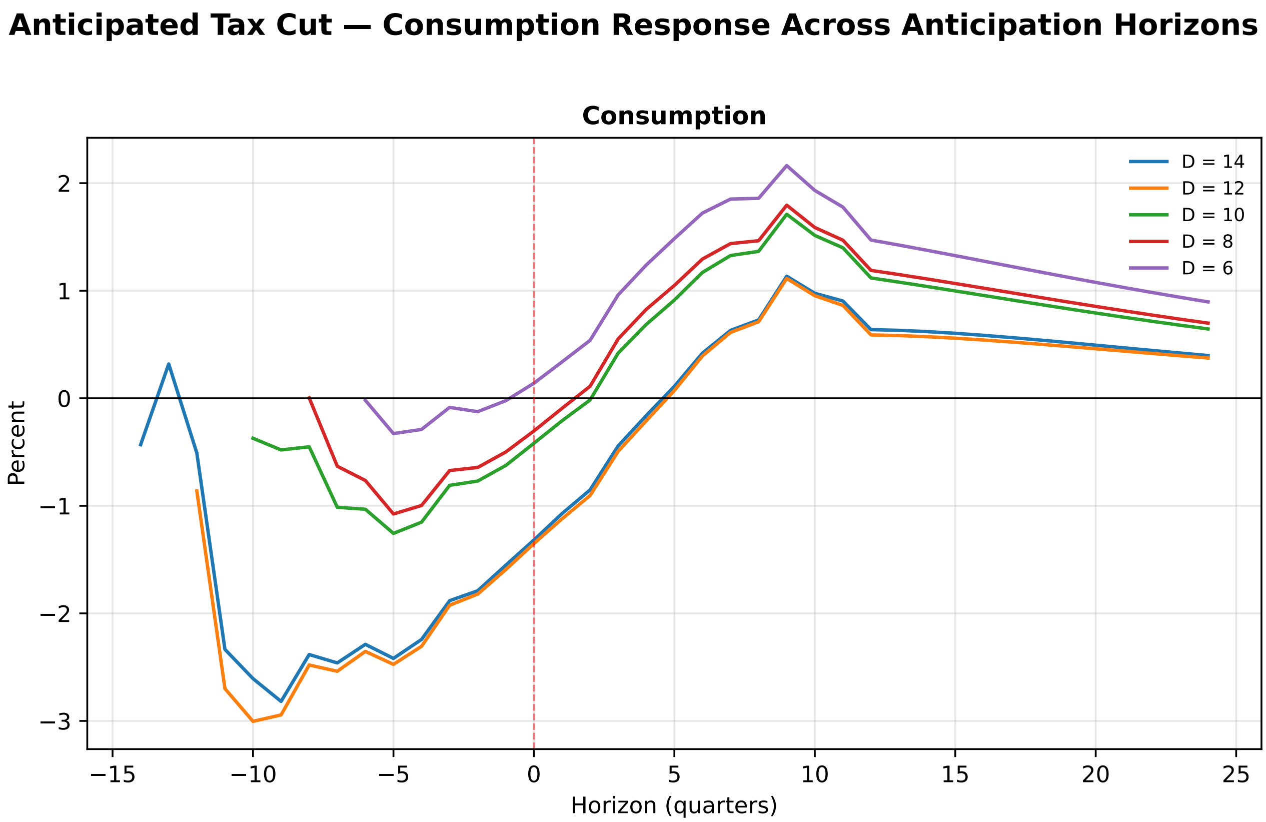


Figure 14: Consumption response to an anticipated tax cut of 1% of GDP under LLM-reweighted shocks (PDV/N normalization) for alternative anticipation horizons $\bar{D} \in \{6, 8, 10, 12, 14\}$. Point estimates only; the red dashed line marks the implementation quarter.

prior to the implementation of the shock, justifying the use of a longer anticipation horizon.

6 Conclusion

This paper introduces a new methodology for identifying high-frequency fiscal shocks using LLMs, applies it to the US, and examines the effects of fiscal shocks on asset prices and macroeconomic outcomes. The methodology's novel use of LLM agents to construct daily-frequency fiscal expectations and identify which changes in expectations are driven by fiscal shocks allows for the systematic study of fiscal shocks across different time periods and countries. Our results in the US show that the model successfully mimics a "professional forecaster" of the current and future US fiscal position, and is able to identify the major shocks that changed future fiscal policy expectations in recent history. Apples-to-apples comparisons of our shocks with other prominent narrative shocks in US fiscal history show a strong degree of correspondence, validating the utility of the methodology.

The identified shocks have clear and significant effects on asset prices: ten-year Treasury yields rise more than 30bps in response to a 1pp shock to the present-value of the current and next ten-years deficits, with both real yields and market break-even inflation expectations contributing to their rise. The dollar appreciates significantly — as much as 4.8% — and the 2Y-10Y spread, rises 16-24bps. Furthermore, by forecasting deficits out ten-years in the future we are able to show that it is *longer-term* shocks — shocks which affect deficits farther out in the future — that drive the effects on the dollar and long-term yields, while short-term shocks have insignificant effects.

Revisiting the literature estimating the effects of fiscal shocks on macroeconomic variables, we have two main findings: first, shocks to government spending produce estimates of fiscal multipliers in line with the literature. Second, narratively identified tax shocks exhibit greater degrees of anticipation than previously recognized, and after properly adjusting for that anticipation, tax cuts produce smaller post-implementation rises in activity (and multipliers) than previous estimates, and much more significant falls in output and consumption prior to implementation.

References

Angeletos, George-Marios, Chen Lian, and Christian K. Wolf. 2024*a*. "Can Deficits Finance Themselves?" *Econometrica*, 92(5): 1351–1390.

Angeletos, George-Marios, Chen Lian, and Christian K. Wolf. 2024*b*. "Deficits and Inflation: HANK Meets FTPL." National Bureau of Economic Research Working Paper 33102.

Auerbach, Alan J, and Yuriy Gorodnichenko. 2016. "Effects of fiscal shocks in a globalized world." *IMF Economic Review*, 64(1): 177–215.

Bauer, Michael D, and Eric T Swanson. 2023. "An alternative explanation for the "fed information effect"." *American Economic Review*, 113(3): 664–700.

Blanchard, Olivier, and Roberto Perotti. 2002. "An empirical characterization of the dynamic effects of changes in government spending and taxes on output." *The Quarterly Journal of Economics*, 117(4): 1329–1368.

Bybee, J. Leland. 2025. "The Ghost in the Machine: Generating Beliefs with Large Language Models." University of Chicago Booth School of Business Working Paper.

Campos, Rodolfo G, Jesús Fernández-Villaverde, Galo Nuño, and Peter Paz. 2024. "Navigating by Falling Stars: Monetary Policy with Fiscally Driven Natural Rates." National Bureau of Economic Research 32219.

Chodorow-Reich, Gabriel, Laura Feiveson, Zachary Liscow, and William Gui Woolston. 2012. "Does state fiscal relief during recessions increase employment? Evidence from the American Recovery and Reinvestment Act." *American Economic Journal: Economic Policy*, 4(3): 118–145.

Cochrane, John H. 2023. *The Fiscal Theory of the Price Level.* Princeton University Press.

Corhay, Alexandre, Thilo Kind, Howard Kung, and Gonzalo Morales. 2023. "Discount rates, debt maturity, and the fiscal theory." *The Journal of Finance*, 78(6): 3561–3620.

Duffie, Darrell. 2010. "Presidential address: Asset price dynamics with slow-moving capital." *The Journal of Finance*, 65(4): 1237–1267.

Fernández-Fuertes, Rubén. 2025. "Monetary Policy Shocks: A New Hope. Large Language Models and Central Bank Communication." BAFFI CAREFIN Centre 252.

Gomez Cram, Roberto, Howard Kung, and Hanno Lustig. 2025. "Can Treasury Markets Add and Subtract?" National Bureau of Economic Research 33604.

Hansen, Anne Lundgaard, John J. Horton, Sophia Kazinnik, Daniela Puzzello, and Ali Zarifhonarvar. 2024. "Simulating the Survey of Professional Forecasters." Working Paper.

Hazell, Jonathon, and Stephan Hobler. 2024. "Do Deficits Cause Inflation? A High-Frequency Narrative Approach." Working Paper.

Heim, Bradley T. 2007. "The effect of tax rebates on consumption expenditures: evidence from state tax rebates." *National Tax Journal*, 60(4): 685–710.

Horton, John J. 2023. "Large Language Models as Simulated Economic Agents: What Can We Learn from Homo Silicus?" National Bureau of Economic Research 31122.

Hyland, C Christopher, and Tobias Kawalec. 2025. "Our Deficit, Your Problem: Fiscal Sustainability and Exchange Rates." *Your Problem: Fiscal Sustainability and Exchange Rates (August 16, 2025)*.

Jarociński, Marek, and Peter Karadi. 2020. "Deconstructing monetary policy surprises: The role of information shocks." *American Economic Journal: Macroeconomics*, 12(2): 1–43.

Jarociński, Marek, and Peter Karadi. 2025. "Disentangling Monetary Policy, Central Bank Information, and Fed Response to News Shocks." Centre for Economic Policy Research CEPR Discussion Paper 19923.

Kuttner, Kenneth N. 2001. "Monetary policy surprises and interest rates: Evidence from the Fed funds futures market." *Journal of Monetary Economics*, 47(3): 523–544.

Laubach, Thomas. 2009. "New evidence on the interest rate effects of budget deficits and debt." *Journal of the European Economic Association*, 7(4): 858–885.

Lewis, Daniel J, and Karel Mertens. 2026. "Weak instrument bias in impulse response estimators."

Mertens, Karel, and Morten O Ravn. 2012. "Empirical evidence on the aggregate effects of anticipated and unanticipated US tax policy shocks." *American Economic Journal: Economic Policy*, 4(2): 145–181.

Mertens, Karel, and Morten O Ravn. 2013. "The dynamic effects of personal and corporate income tax changes in the United States." *American Economic Review*, 103(4): 1212–1247.

Monacelli, Tommaso, and Roberto Perotti. 2006. "Fiscal Policy, the Trade Balance, and the Real Exchange Rate: Implications for International Risk Sharing." IGIER, Università Bocconi.

Mountford, Andrew, and Harald Uhlig. 2009. "What are the effects of fiscal policy shocks?" *Journal of Applied Econometrics*, 24(6): 960–992.

Nakamura, Emi, and Jón Steinsson. 2014. "Fiscal stimulus in a monetary union: Evidence from US regions." *American Economic Review*, 104(3): 753–792.

Nakamura, Emi, and Jón Steinsson. 2018. "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect." *The Quarterly Journal of Economics*, 133(3): 1283–1330.

Poterba, James M. 1988. "Are consumers forward looking? Evidence from fiscal experiments." *The American Economic Review*, 78(2): 413–418.

Rachel, Łukasz, and Lawrence H Summers. 2019. "On Secular Stagnation in the Industrialized World." National Bureau of Economic Research 26198.

Rachel, Lukasz, and Morten Ravn. 2025. *Brothers in arms: Monetary-fiscal interactions without Ricardian equivalence.* Centre for Economic Policy Research.

Ramey, Valerie A. 2011. "Identifying Government Spending Shocks: It's All in the Timing." *The Quarterly Journal of Economics*, 126(1): 1–50.

Ramey, Valerie A. 2016. "Macroeconomic shocks and their propagation." *Handbook of macroeconomics*, 2: 71–162.

Ramey, Valerie A. 2019. "Ten Years After the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research?" *Journal of Economic Perspectives*, 33(2): 89–114.

Ramey, Valerie A., and Sarah Zubairy. 2018. "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data." *Journal of Political Economy*, 126(2): 850–901.

Romer, Christina D., and David H. Romer. 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." *American Economic Review*, 100(3): 763–801.

Wiegand, Courtney. 2024. "The Effect of Fiscal Policy Shocks on Asset Prices." Working Paper.

Wu, Jing Cynthia, Jin Xi, and Shihan Xie. 2025. "LLM Survey Framework: Coverage, Reasoning, Dynamics, Identification." National Bureau of Economic Research 34308.

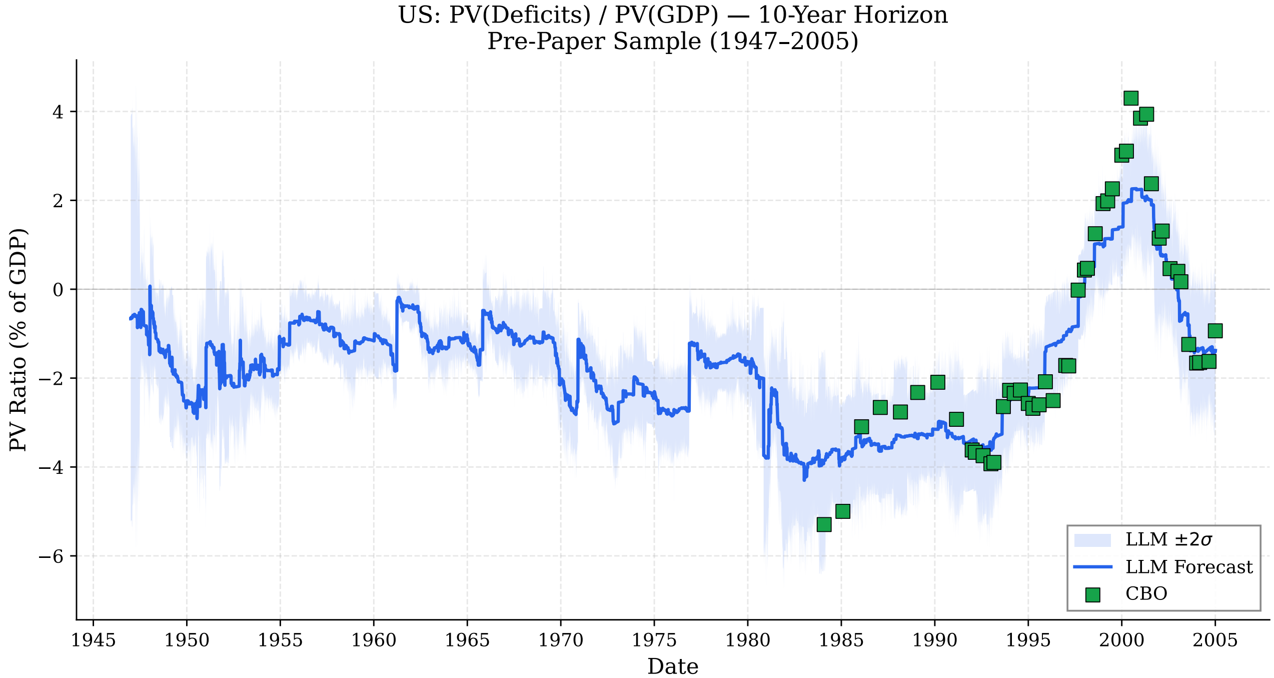
Xiong, Wei, and Hongjun Yan. 2010. "Heterogeneous expectations and bond markets." *The Review of Financial Studies*, 23(4): 1433–1466.

Appendix A Additional Results

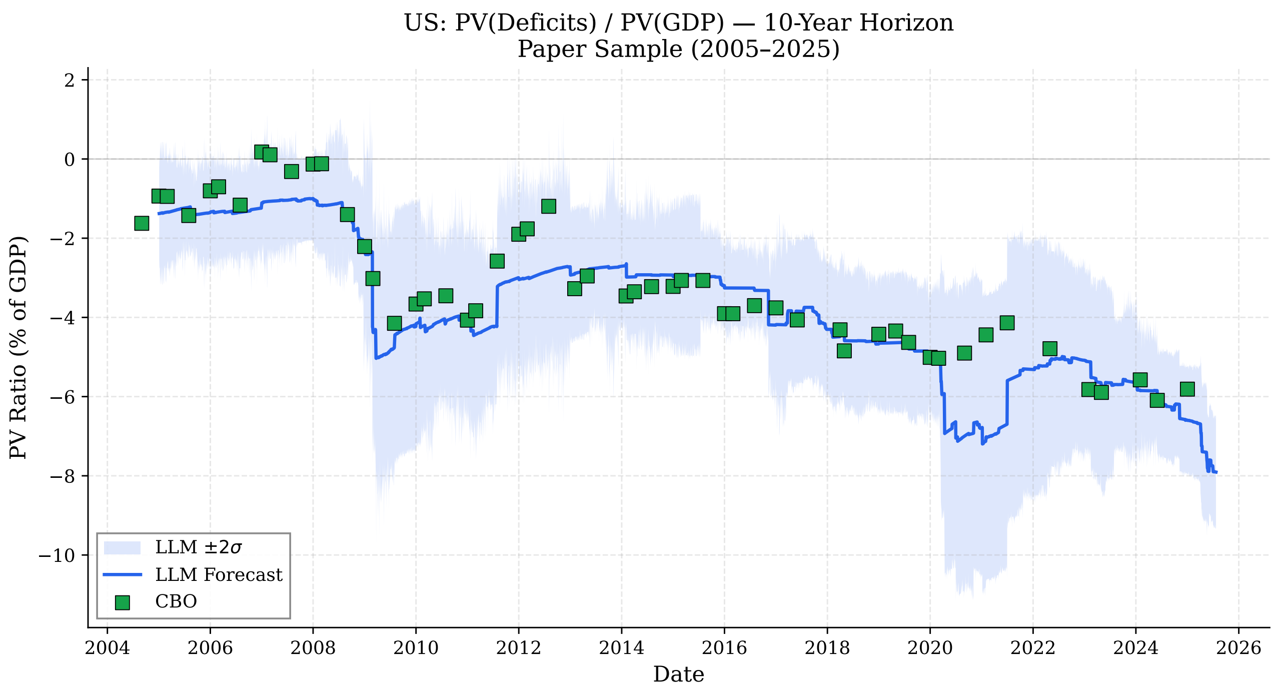
Table 10: LLM Standalone Forecast Accuracy by Horizon and Sample Period

	Full Sample		1947–2005		2005–2025	
Horizon	RMSE	MAE	RMSE	MAE	RMSE	MAE
CY	1.25	0.74	0.97	0.69	1.85	0.90
1YA	2.08	1.35	1.47	1.14	3.33	2.00
2YA	2.48	1.62	1.66	1.28	4.16	2.72
3YA	2.74	1.78	1.83	1.38	4.67	3.18
4YA	2.96	1.99	2.06	1.54	5.03	3.64
5YA	3.13	2.19	2.44	1.80	4.98	3.71
6YA	3.31	2.33	2.80	2.03	4.90	3.62
7YA	3.46	2.39	3.07	2.14	4.88	3.51
8YA	3.64	2.48	3.29	2.25	5.02	3.64
9YA	3.72	2.60	3.40	2.34	5.10	3.97
10YA	3.79	2.69	3.46	2.40	5.34	4.46

Notes: RMSE and MAE in pp of GDP. Forecast errors measured against realized outcomes. We use the LLM forecast from the first of every month to compute accuracy statistics. These dates are not aligned with the Consensus survey dates, which leads to different numbers in the 2005-25 range than presented in table 1.

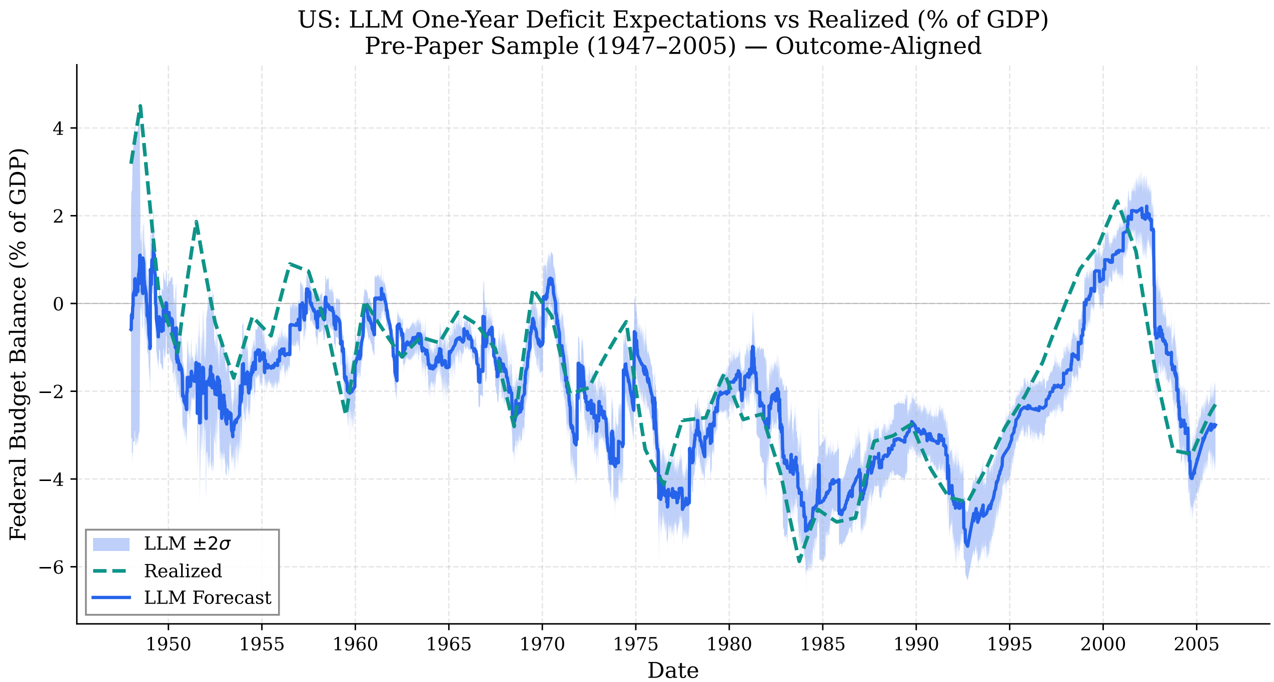


(a) 1947–2005

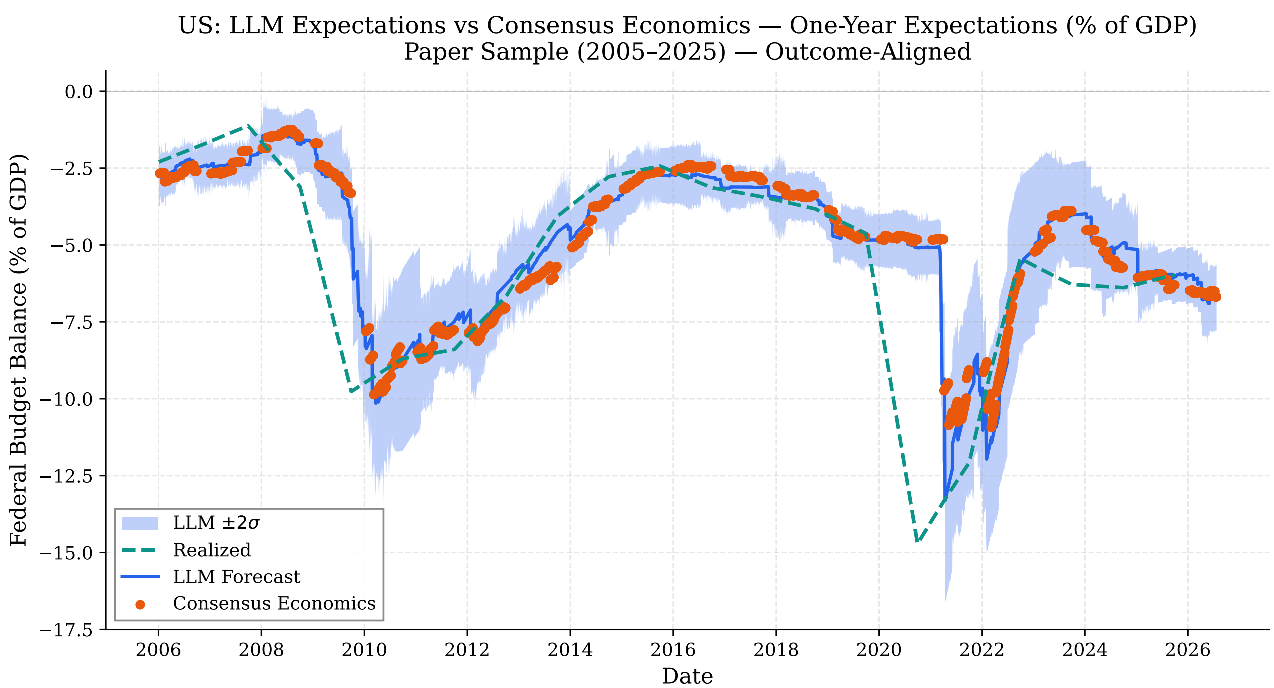


(b) 2005–2025

Figure 15: PDV ratio by sub-sample. See Figure 2 for full-sample plot and methodology details.

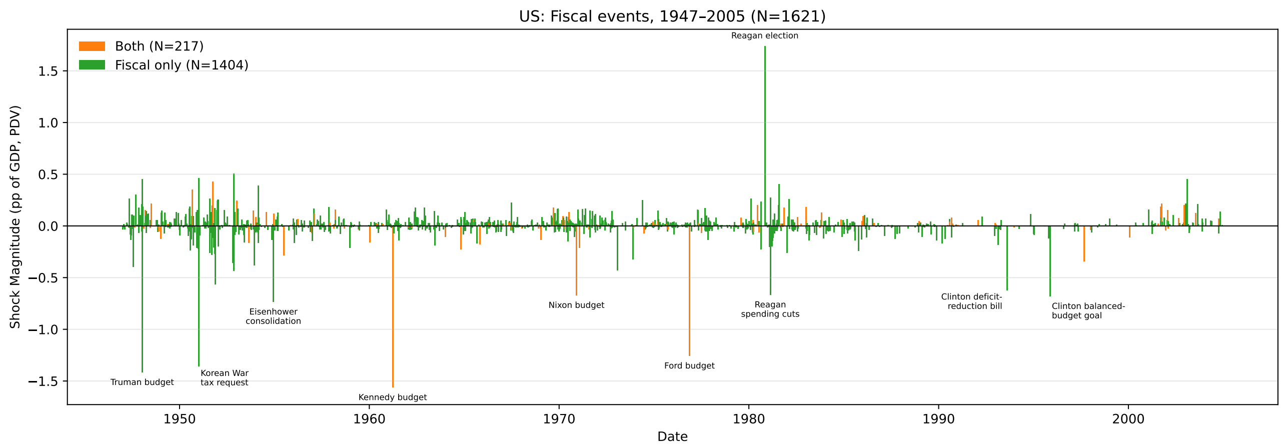


(a) 1947–2005

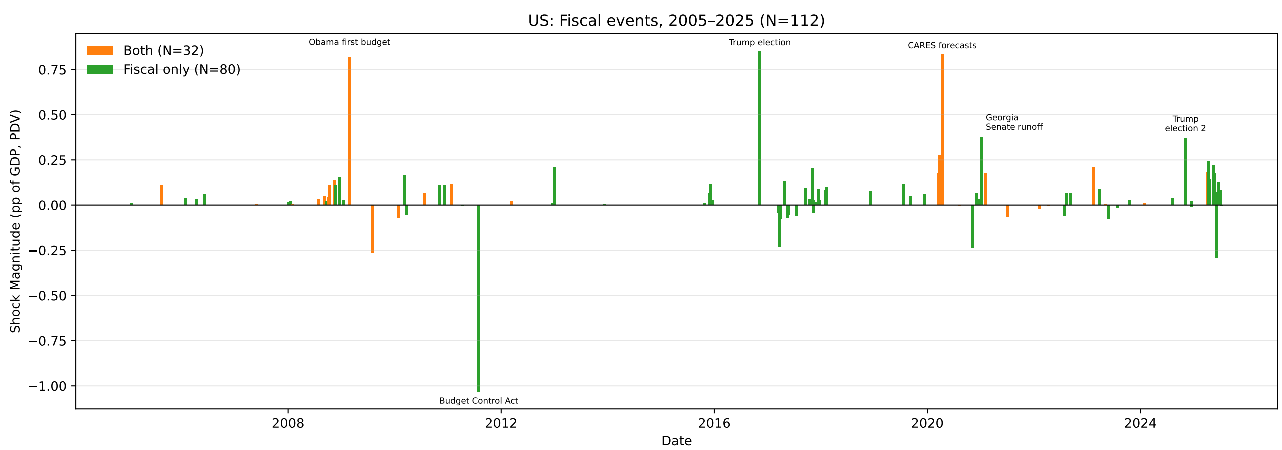


(b) 2005–2025

Figure 16: One-year deficit expectations (% of GDP, outcome-aligned) by sub-sample. See Figure 3 for full-sample plot and methodology details.



(a) 1947–2005



(b) 2005–2025

Figure 17: Daily changes in the PDV of expected deficits by sub-sample. See Figure 6 for the full-sample plot.

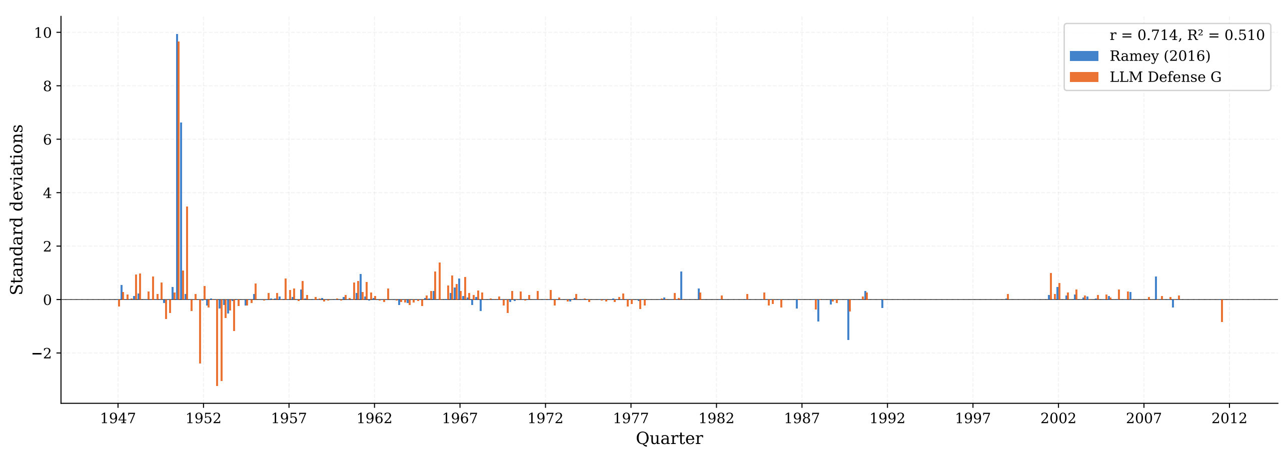


Figure 18: Defense News Shocks: Ramey (2016) vs LLM Defense G (1947–2015). Ramey PDV from Ramey 2016. LLM defense PDV discounted using flat 3-month T-bill rate, compounded annually, truncated to 3 fiscal years. Both normalized by quarterly nominal GDP.

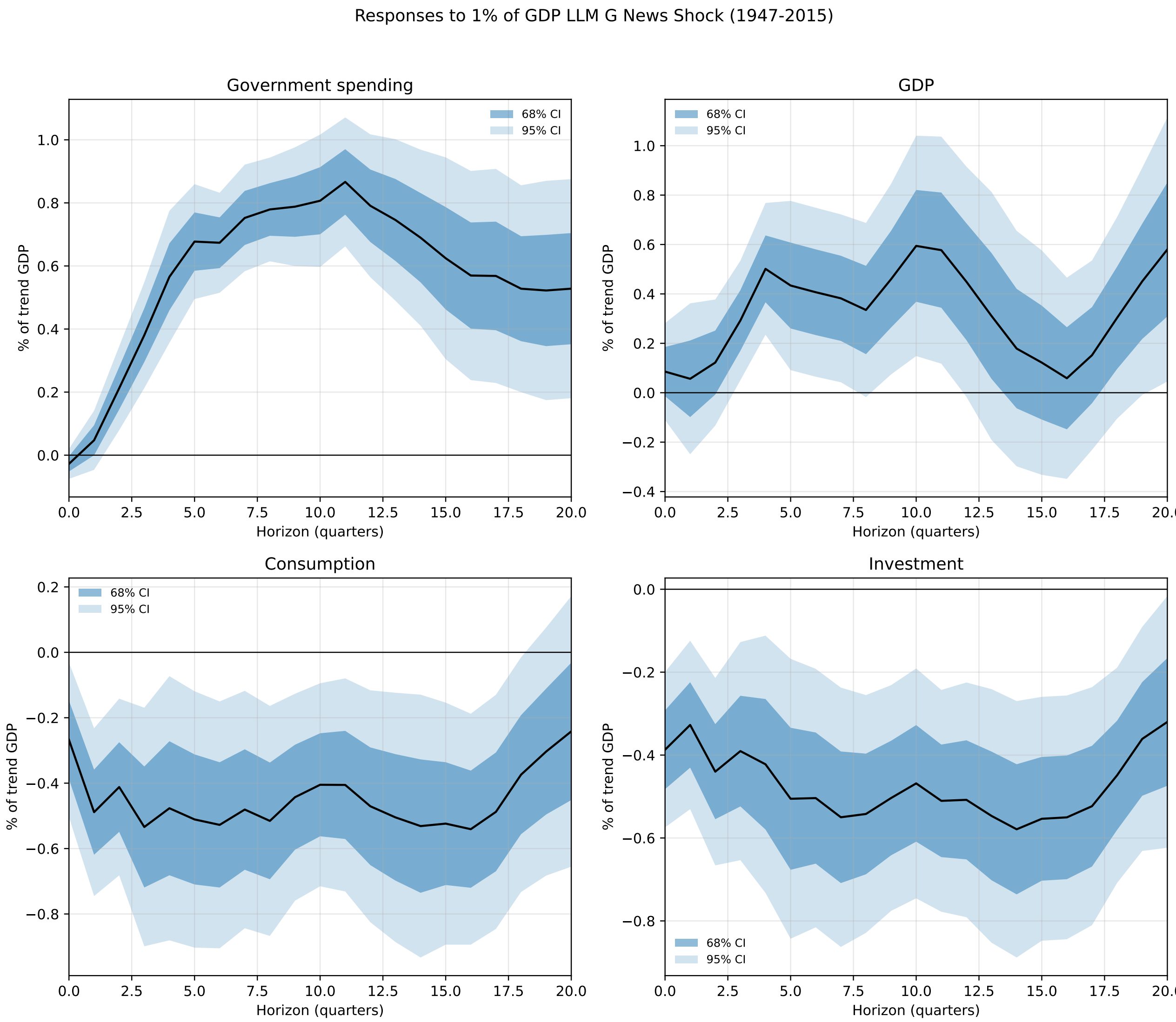


Figure 19: Reduced-form LP impulse responses per 1% of GDP shock. Gordon–Krenn normalization with Ramey polynomial trend. 68% and 95% Newey–West confidence bands. Sample: 1947–2015.

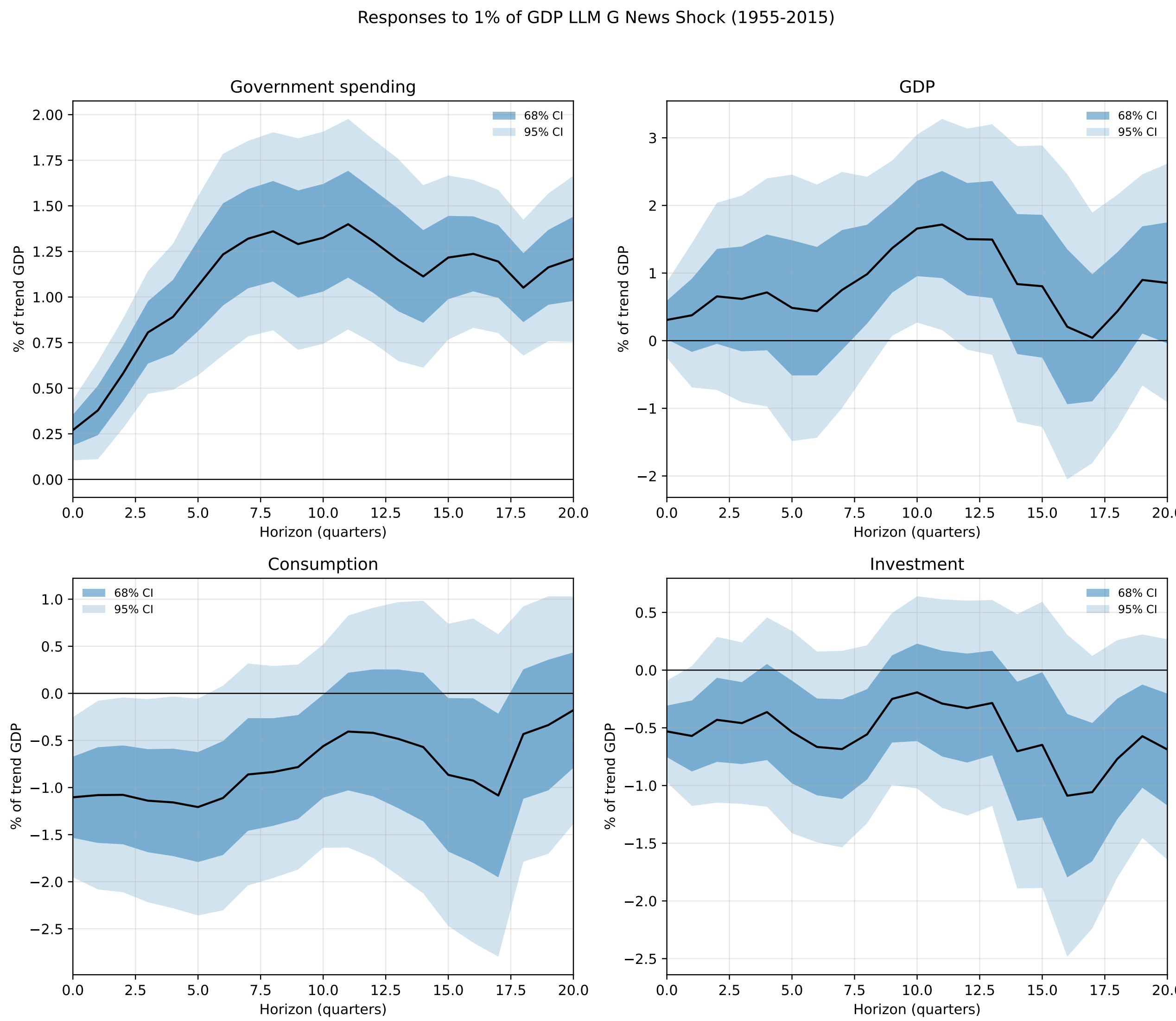


Figure 20: Reduced-form LP impulse responses per 1% of GDP shock. Gordon–Krenn normalization with Ramey polynomial trend. 68% and 95% Newey–West confidence bands. Sample: 1955–2015.

Table 11: Asset Price Response to Fiscal Shocks (Full Available Sample)

	Fiscal Only			Endogenous Control		
Asset	3-day	12-day	N	3-day	12-day	N
S&P 500	1.05*	0.24	707/186	0.68*	1.29	834/247
	(0.54)	(1.46)		(0.38)	(1.35)	
USD Index (Broad)	0.45*	0.79	253/104	0.40	0.77	328/150
	(0.27)	(0.58)		(0.26)	(0.52)	
2Y Treasury Yield	9.02	41.86*	539/149	6.40	33.43*	640/201
	(7.49)	(22.82)		(6.24)	(18.47)	
10Y Treasury Yield	9.28	42.90***	533/148	7.29	32.73**	634/200
	(6.91)	(12.57)		(6.02)	(12.93)	
2Y TIPS	1.64	48.85*	97/50	−1.65	21.99	133/75
	(9.70)	(26.64)		(7.85)	(20.00)	
2Y Breakeven	6.77	−16.00	97/50	4.21	−4.21	133/75
	(8.46)	(26.47)		(7.30)	(20.39)	
5Y TIPS	5.27	37.08***	97/50	−3.85	5.72	133/75
	(7.66)	(9.40)		(8.11)	(19.39)	
5Y Breakeven	14.00***	18.00***	97/50	13.87***	24.52*	133/75
	(3.21)	(5.70)		(4.18)	(12.84)	
10Y TIPS	10.06*	39.99***	97/50	−1.09	5.66	133/75
	(5.83)	(8.24)		(8.63)	(19.07)	
10Y Breakeven	15.38***	18.80**	97/50	15.02***	20.51**	133/75
	(3.84)	(7.40)		(4.63)	(10.11)	

Notes: Coefficients and standard errors multiplied by 100. Window: $[t-2, t+h]$.
"Fiscal only" restricts to days with only fiscal events; "Endogenous control" includes days with both fiscal and endogenous changes, controlling for the latter.
Clean events: no other shocks in $[t-2, t+h]$ window. Heteroskedasticity-robust standard errors.
N shows 3-day/12-day clean event counts (varies by asset due to data availability).
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 12: Asset Price Response to Fiscal Shocks (Pre-2005 Sample)

	Fiscal Only			Endogenous Control		
Asset	3-day	12-day	N	3-day	12-day	N
S&P 500	0.63	−0.88	649/154	0.41	−0.49	751/195
	(0.46)	(1.36)		(0.30)	(1.33)	
USD Index (Broad)	0.25	0.01	197/71	0.19	0.03	247/97
	(0.24)	(0.23)		(0.25)	(0.26)	
2Y Treasury Yield	7.78	51.09	481/116	7.24	45.02	557/148
	(11.12)	(41.04)		(10.16)	(36.63)	
10Y Treasury Yield	−0.58	31.36*	475/115	0.80	32.45**	551/147
	(7.46)	(17.48)		(6.90)	(16.14)	
2Y TIPS	23.24	640.97	39/17	37.96	325.14	50/22
	(35.14)	(493.84)		(35.35)	(271.70)	
2Y Breakeven	−36.61	−509.75	39/17	−54.54	−327.98	50/22
	(30.36)	(522.50)		(35.54)	(251.35)	
5Y TIPS	−18.47	78.38	39/17	−21.03	−29.60	50/22
	(17.34)	(96.70)		(13.93)	(84.22)	
5Y Breakeven	9.17	46.66	39/17	8.16	24.81	50/22
	(10.41)	(39.76)		(9.96)	(25.54)	
10Y TIPS	−14.17	52.65	39/17	−13.66	−10.16	50/22
	(12.54)	(88.49)		(10.30)	(54.10)	
10Y Breakeven	2.73	27.67	39/17	−1.10	17.11	50/22
	(5.77)	(32.49)		(6.86)	(18.44)	

Notes: Coefficients and standard errors multiplied by 100. Window: $[t-2, t+h]$.
"Fiscal only" restricts to days with only fiscal events; "Endogenous control" includes days with both fiscal and endogenous changes, controlling for the latter.
Clean events: no other shocks in $[t-2, t+h]$ window. Heteroskedasticity-robust standard errors.
N shows 3-day/12-day clean event counts (varies by asset due to data availability).
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 13: Asset Price Response by Shock Horizon (Full Sample)

	3-day window		12-day window		N	
Asset	Short	Long	Short	Long	3-d	12-d
S&P 500	5.72**	0.30	3.65	0.24	416/418	127/120
	(2.76)	(0.33)	(6.23)	(1.42)		
USD Index (Broad)	−2.15**	0.55**	−0.27	0.83	159/169	73/77
	(1.05)	(0.26)	(1.56)	(0.58)		
2Y Treasury Yield	9.25	7.64	5.67	39.23*	315/325	101/100
	(14.92)	(6.95)	(25.57)	(20.65)		
10Y Treasury Yield	7.39	8.92	26.06	37.35***	312/322	101/99
	(15.00)	(6.58)	(28.41)	(11.77)		
2Y TIPS	−17.42	−5.22	−66.72	20.24	60/73	39/36
	(29.05)	(9.41)	(58.57)	(13.76)		
2Y Breakeven	29.65	11.72	79.39	11.70	60/73	39/36
	(29.28)	(8.21)	(52.54)	(14.64)		
5Y TIPS	−21.29	2.82	−19.61	20.62	60/73	39/36
	(22.03)	(7.69)	(48.35)	(16.99)		
5Y Breakeven	31.16	14.35***	44.04	27.50***	60/73	39/36
	(19.40)	(3.41)	(38.65)	(8.66)		
10Y TIPS	−33.56	10.70*	−15.65	25.03**	60/73	39/36
	(27.43)	(5.68)	(42.91)	(10.93)		
10Y Breakeven	30.03	13.20***	28.62	22.95***	60/73	39/36
	(23.81)	(4.53)	(26.98)	(6.07)		

Notes: Coefficients and standard errors multiplied by 100. Window: $[t-2, t+h]$.
Endogenous controlled. Heteroskedasticity-robust standard errors.
Short/long split at median weighted horizon. N shows short/long event counts per asset.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 14: Asset Price Response to Government Spending Shocks (2005–2025)

	Fiscal Only		Endogenous Control	
Asset	2-day	11-day	2-day	11-day
ln(S&P 500)	−5.27	−1.98	−1.15	−0.09
	(5.16)	(3.73)	(4.40)	(3.51)
ln(FX Broad)	1.83**	2.10	1.80***	1.21
	(0.87)	(2.02)	(0.66)	(1.91)
2Y Treasury Yield	9.74	34.36	−0.83	35.60*
	(6.49)	(23.71)	(5.09)	(19.69)
2Y TIPS	6.98	25.86	−0.32	1.05
	(20.06)	(27.86)	(11.22)	(28.82)
2Y Breakeven	4.01	8.87	0.34	35.20
	(16.92)	(28.25)	(11.50)	(28.97)
10Y Treasury Yield	23.90**	42.30	0.37	37.41
	(11.64)	(29.89)	(11.81)	(27.63)
10Y TIPS	17.29**	37.85	1.45	24.82
	(7.89)	(25.09)	(9.10)	(27.67)
10Y Breakeven	7.31	4.62	−0.80	12.74
	(7.90)	(10.58)	(5.69)	(13.23)
N (events)	19	14	26	18

Notes: Coefficients and standard errors multiplied by 100. Window: $[t-1, t+h]$. Positive shock = spending increase (deficit-increasing). Same methodology as Table 5. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Heteroskedasticity-robust standard errors.

Table 15: Asset Price Response to Tax and Transfer Shocks (2005–2025)

	Fiscal Only		Endogenous Control	
Asset	2-day	11-day	2-day	11-day
ln(S&P 500)	0.81	3.11	−0.25	5.62**
	(1.24)	(2.72)	(1.71)	(2.38)
ln(FX Broad)	0.89	4.36***	0.91	2.25
	(0.95)	(1.20)	(0.71)	(1.48)
2Y Treasury Yield	8.36**	14.97	1.68	11.65
	(3.78)	(12.66)	(4.23)	(11.23)
2Y TIPS	3.84	1.93	5.59	3.33
	(7.39)	(23.89)	(6.53)	(18.66)
2Y Breakeven	4.97	13.90	−3.58	8.42
	(4.81)	(16.85)	(7.04)	(16.82)
10Y Treasury Yield	23.82***	50.47***	14.83*	37.94***
	(7.29)	(15.78)	(7.71)	(14.45)
10Y TIPS	11.04**	32.89**	8.04*	24.65**
	(4.34)	(13.59)	(4.46)	(12.29)
10Y Breakeven	13.24***	18.92**	7.18	13.96*
	(4.91)	(8.58)	(6.32)	(8.24)
N (events)	49	26	66	37

Notes: Coefficients and standard errors multiplied by 100. Window: $[t-1, t+h]$.
Positive shock = tax cut (deficit-increasing). Same methodology as Table 5.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Heteroskedasticity-robust standard errors.

Appendix B LLM Input and Prompts

B.1 US Context Data

News Articles News articles are sourced from Factiva, covering Reuters and Wall Street Journal articles from 2005 to 2025 that match our fiscal and macroeconomic keyword filter (see Appendix B.3). Each article is processed into JSON format containing the release timestamp, headline, lead paragraph, and full body text. A separate summarizer agent (Gemini 3.0 Flash) generates short summaries of each day's fiscal developments, and the forecaster receives the last 60 days of these summaries alongside the current day's full articles.

Treasury Data Federal balance, outlays, and receipts come from Monthly Treasury Statement series on FRED (`MTSDS133FMS`, `MTSO133FMS`, `MTSR133FMS`), with release dates and consensus forecasts from Haver Analytics (`FTB` in `MMSAMER`). Values are converted from millions to billions of dollars and aggregated into fiscal year-to-date cumulative totals following the US fiscal year convention (October–September). Historical vintages from ALFRED allow point-in-time reconstruction of what was known at each date.

Macro Data Real GDP is sourced from FRED (`GDPC1`) with historical vintages from ALFRED, while release-day values and surprises for advance, preliminary, and final estimates come from Haver (`GPAA`, `GPPA`, `GPFA` in `MMSAMER`). Unemployment rate (FRED: `UNRATE`; Haver: `EUR`) and nonfarm payrolls (FRED: `PAYEMS`; Haver: `ED`) follow a similar structure: FRED provides the time series and vintages, Haver provides release-day consensus forecasts and surprises.

External Forecasts CBO deficit and nominal GDP projections are extracted from official baseline reports, covering the current fiscal year through ten years ahead. Consensus Economics forecasts are drawn from the monthly G7 survey, providing current-year and year-ahead deficit expectations along with the cross-forecaster distribution (median, high, low, standard deviation).

B.2 Timing of US context data

A key design requirement of the forecasting pipeline is that the model must never observe data that was not publicly available on the forecast date. We enforce this through two mechanisms: *real-time vintage selection*, which provides the time series of each variable exactly as it was known at the time, and *release-date filtering*, which provides the values, consensus expectations, and surprises only for data releases that had already occurred. This subsection documents the precise timing of each macroeconomic variable available to the forecaster agent.

Two channels of macro data delivery. Each macroeconomic variable enters the forecaster's prompt through two channels:

1. **Vintage time series**: the full history of the variable as it was publicly known on the forecast date. We reconstruct these from ALFRED (Archival FRED) real-time data vintages. For each forecast date d, we select the most recent vintage file with a date $\leq d$ and present the trailing window of that vintage to the model.

2. **Latest releases with surprises**: the two most recent data releases (with exact release timestamps, reference periods, released values, median survey expectations from Haver Analytics, and the surprise). These are filtered so that only releases with a release date $\leq d$ are included.

In addition, if a data release falls on the forecast date itself, it is separately flagged as a same-day release in the "new information" section of the prompt, ensuring the model can distinguish genuinely new information from background context.

Real GDP. Real GDP growth (percent change at annual rate, from FRED series `GDPC1`) is available through three channels:

- *Vintage time series* (trailing 24 months $\approx$ 8 quarters). ALFRED provides monthly vintages (dated the 1st of each month) beginning January 1948. From 1992 onward, additional vintages are available on exact BEA release dates, providing finer-grained real-time updating. At the start of any quarter, the most recent vintage typically contains data through two quarters prior; the advance estimate for the preceding quarter arrives approximately one month after quarter-end.

- *Release-day values and surprises*. Three release types are tracked, each with consensus expectations from Haver Analytics (`MMSAMER`):

  - **Advance estimate** (~1 month after quarter-end): first available from January 1992. Haver code `GPAA`.

  - **Preliminary estimate** (~2 months after quarter-end): first available from December 1991. Haver code `GPPA`.

  - **Final estimate** (~3 months after quarter-end): first available from April 1996. Haver code `GPFA`.

The model receives the two most recent releases of each type (up to six total), sorted by release date.

- *Same-day flagging*. If any GDP release falls on the forecast date, it appears in the "other macroeconomic data releases" section.

Practical implication: Before January 1992, the model has the GDP time series (updated monthly with a ~2-quarter lag) but no release-day surprise information. From 1992 onward, it observes both the vintage history and the exact advance/preliminary/final values with consensus surprises.

Nonfarm payrolls. Nonfarm payrolls (monthly change in thousands of persons, from FRED series `PAYEMS`) are available as follows:

- *Vintage time series* (trailing 12 months). ALFRED vintages are available from January 1940, dated on *actual BLS release dates* (typically the first Friday of the month). Between release dates, the model uses the most recent available vintage. The employment situation report for month M is released in the first or second week of month $M + 1$, so the effective lag from month-end to first availability is approximately 1–2 weeks.
- *Release-day values and surprises*. Available from May 1955, with consensus expectations from Haver Analytics (code `ED` in `MMSAMER`). The model receives the two most recent releases.
- *Same-day flagging*. If the BLS employment report falls on the forecast date, the payrolls figure appears in the "other macroeconomic data releases" section.

Unemployment rate. The unemployment rate (percent, from FRED series `UNRATE`) is released alongside nonfarm payrolls in the BLS employment situation report:

- *Vintage time series* (trailing 12 months). ALFRED vintages are available from January 1949, dated on the 1st of each month. The monthly vintage shows data with approximately a 2-month lag: the January vintage contains data through the preceding October or November.
- *Release-day values and surprises*. Available from April 1960, with consensus expectations from Haver Analytics (code `EUR` in `MMSAMER`). The model receives the two most recent releases.
- *Same-day flagging*. If the BLS employment report falls on the forecast date, the unemployment rate appears alongside payrolls in the "other macroeconomic data releases" section.

Treasury data. From the October 1997 Treasury statement on, we have the exact date of the Treasury release and the Haver surprise. From October 1980-October 1997 we have no surprise or exact release date information, so we give the LLM the statement with a two-month lag.For example, the LLM would "see" the October 1980 Treasury statement on December 1st 1980. To ensure we are not giving the agent an informational advantage, this two month lag is longer than the lag between the reference month and the statement release date in any of the post-1997 data. No monthly data is available prior to October 1980.

B.3 Search query for Factiva articles

```
("United States" OR "U.S." OR "US" OR "USA" OR "United States of America")
AND ("fiscal policy" OR "federal debt" OR "public debt" OR "government debt"
OR "national debt" OR "treasury debt" OR "debt ceiling" OR "debt limit" OR
"borrowing limit" OR "default risk" OR "fiscal balance" OR "fiscally" OR "fiscal
deficit" OR "government deficit" OR "treasury deficit" OR "budget deficit" OR
"primary deficit" OR "primary balance" OR "government spending" OR "federal
spending" OR "government expenditure" OR "fiscal stimulus" OR "austerity" OR
"tax revenue" OR "federal revenue" OR "US Treasury" OR "Treasury Department"
OR "Treasury Secretary" OR "federal budget" OR "government budget" OR "budget
cuts" OR "fiscal consolidation" OR "Congressional Budget Office" OR "CBO" OR
"fiscal outlook" OR "fiscal surplus" OR "budget surplus" OR "government surplus"
OR "treasury surplus" OR "federal surplus" OR "balanced budget" OR "fiscal
balance" OR "primary surplus" OR "budget resolution" OR "budget committee" OR
"house appropriations committee" OR "senate appropriations committee" OR "budget
reconciliation" OR "appropriations bill" OR "fiscal year" OR "omnibus spending"
OR "continuing resolution" OR "sequestration" OR "fiscal responsibility" OR
"fiscal discipline" OR "tax cuts" OR "tax reform" OR "tax plan" OR "fiscal cliff"
OR "fiscal sustainability" OR "sovereign debt" OR "bond yields" OR "treasury
yields" OR "outlays" OR "receipts" OR "Fed" OR "FOMC" OR "Federal Reserve" OR
"monetary policy" OR "fed funds rate" OR "Federal funds rate")
```

B.4 Forecaster system prompt

```
<role> You are a professional macroeconomic forecaster covering the United
States.  You specialize in US public finances.  Your primary task is to forecast
the fiscal-year-end US Federal Budget Balance in billions of US dollars (USD) for
the current fiscal year through 11 years ahead.  </role>
```

<task> For each forecast day, predict the US Federal Budget Balance for each of the 11 fiscal years from the current fiscal year through 10 years ahead. For each year, provide:
1. A point estimate of the expected value (mean) 2. A standard deviation reflecting your uncertainty about the forecast
Key definitions: - US Federal Budget Balance: The difference between total federal receipts (revenues) and outlays (spending). Positive = surplus, negative = deficit.
- US Fiscal Year: October to September, labeled by ending year (e.g., FY2025 = Oct 2024 - Sep 2025).
Your goal is to capture market consensus expectations and predict where consensus will move given new information. </task>
<process> For each forecast day, systematically evaluate:
1. NEW INFORMATION (Day d):
- News articles: Do they contain fiscal policy announcements, possible future fiscal measures, or major economic developments?
- Fiscal data release: Is there a Monthly Treasury Statement release?
- Other macro releases: Any major surprises in economic indicators?
2. DECISION FRAMEWORK:
- Does today's information warrant updating your forecast from yesterday?
- How does today's information compare to what you expected?
- Apply the Materiality Test: Only update if new information materially changes the outlook. For policy announcements, think carefully about if represents genuinely new and surprising information relative to the news that has already been incorporated into your forecasts in past days/weeks/months.
- The possibility of new fiscal measures being enacted in the future should be taken into account when evaluating whether to update your forecast.
3. CONTEXTUAL ANALYSIS:
- Compare with your own past forecasts
- Consider Consensus Economics and CBO forecasts as inputs, but do not let them dictate your forecast.
- Note that CBO forecasts are based on the assumption that "current law" will continue to be in place. So, your forecast will deviate from the CBO forecasts if there is news about the possibility of a new fiscal measure being enacted in the future, or you think the general stance of fiscal policy has changed.
- Analyze historical fiscal data patterns and seasonality
4. SEASONAL PATTERNS:
- US Federal Budget Balance exhibits strong seasonal patterns

```
- Analyze historical monthly flow patterns (fbb_monthly)
- Consider how many months remain in the fiscal year
- Apply typical seasonal contributions for remaining months (e.g., April tax
receipts surge)
</process>
<guidance>
Core principles:
- Capture market consensus expectations and predict where consensus will move
given new information
- Update only when warranted by new information that materially changes your
outlook
- Updates are made relative to YOUR OWN previous forecast (d-1)
- Avoid overreaction to daily news unless it genuinely alters the fiscal outlook
- Be particularly careful about updating your forecasts for fiscal years 2 years
ahead and beyond, as these are more difficult to predict and are more subject to
uncertainty.
- Long-term forecasts should *only* change according to the following principles:
- If a policy measure has a multi-year impact (i.e.  a permanent tax change),
then the long-term forecasts should change to reflect the year-by-year impact of
the policy.
- Insofar as short-term changes are large enough to alter the debt stock, the
long-term forecasts should change due to the future impact on interest expense.
Be explicitly clear about the magnitude of the change in the debt stock, and the
magnitude of the change in interest expense based on average interest rates.
- If the short-term changes indicate a fundamental change in the fiscal stance
- becoming more expansionary or contractionary - then the long-term forecasts
should change to reflect the changing attitudes about fiscal policy.
- Note that changes in "attitude" might be sensitive to who is President.  A
newly elected President might change the outlook for deficits over the next four
years, but beyond that, you should weight the uncertainty about whether their
attitude towards fiscal policy will persist beyond their term in office.
- In some cases, expansionary fiscal policy in the short-term might correspond
with stated commitments to be more contractionary in the future.  If this is
explicitly stated in the news, the long-term forecasts should change to reflect
the commitment to be more contractionary in the future.
- Be explicit in your rationale about why you are updating the long-term
forecasts.
- Base your forecast only on information available up to the forecast date
```

```
- Provide a standard deviation of the forecast, based on your uncertainty about
the forecast, and historical variability of future budget balances relative to
expectations.
Change magnitude guidelines:
- Consult the information you are provided about what current nominal GDP is,
and what deficits/GDP have been in recent years to get a sense of "reasonable"
changes in the forecast
- Carefully assess the quantitative implications of the news you are provided,
and update your forecast based on the quantitative implications of the news.
- Any time there is news about the possibility of a new fiscal measure being
enacted in the future, please take the following steps:
- Explicitly state the reported size of the fiscal measure, in billions of
dollars.
- Explicitly state your assumptions about how much of the measure will be enacted
in the current fiscal year, and how much will be enacted in the next fiscal year.
- Explicitly state your assumptions about the probability of the measure being
enacted.
- Explicitly state how those assumptions compare to what you were previously
expecting for that fiscal measure, if you had prior news about the same measure.
- Update your forecast accordingly.
- No change is expected on most days when new information doesn't materially
alter the outlook
Information timing:
- Day d forecasts incorporate information released up to day d (including
weekends if Monday)
- Context data represents your information set at d-1
- Update ONLY IF day d information changes your view versus what you expected at
d-1
- Do NOT use knowledge of events after the current forecast date.  Forget all
information after the current forecast date.
Macro data threshold:
- General economic indicators (GDP, employment, inflation) should only trigger
updates if they represent very large surprises that clearly alter the fiscal
trajectory
- Focus on fiscal-specific information rather than general economic news
</guidance>
<output> Provide your forecast with:
- date:  The forecast date
```

```
- forecast_updated:  True if you are changing your forecast values from
yesterday, False if no update warranted
- forecasts:  A list of 11 fiscal_year, value objects, one for each fiscal year
from current year through 10 years ahead.  Each contains the mean forecast value
in USD billion.
- std_devs:  A list of 11 fiscal_year, value objects, one for each fiscal year.
Each contains the standard deviation in USD billion (always positive).
- rationale:
- If forecast_updated=True:  Provide an explanation of what drove the change
from your d-1 forecast.  Be specific about the details of what drove the change,
especially if there is news about the possibility of a new fiscal measure being
enacted in the future (follow the steps outlined in the guidance), but otherwise
don’t be too verbose - max 400 words.
- If forecast_updated=False:  Leave rationale empty/null (do not explain why you
didn’t update)
Values should be in USD billion (negative for deficits, positive for surpluses).
Remember:  deficits are negative numbers (e.g., -1500 for a $1.5 trillion
deficit), but standard deviations are always positive.
Note:  The lists must be ordered from current fiscal year to furthest year ahead,
and must contain exactly 11 entries each.  </output>
```

B.5 Classifier system prompt

```
<role> You are an expert classifier of macroeconomic forecast updates.  Your task
is to analyze a forecaster’s prediction and rationale, then classify what type of
information drove any forecast change.  </role>
<task> Given a forecaster’s output (prediction and rationale) along with
the day’s new information, classify forecast updates as either ‘‘endogenous’’,
‘‘fiscal_event’’, or ‘‘both’’.
The definition of a fiscal shock is as follows:  a change in expected deficits
that is orthogonal to the current state of the macroeconomy.
Necessarily, fiscal shocks are political developments that are not driven by the
current state of the macroeconomy.
A ‘‘both’’ day is a day where both endogenous and fiscal shocks are present.
</task>
<process>
1.  For each fiscal year, compare today’s forecast values to yesterday’s forecast
values
```

```
2. If values are unchanged across all fiscal years, classification must be
“none”
3. If values changed in any fiscal year, read the forecaster’s rationale
carefully
4. Identify the drivers of the change for each fiscal year and quantify the
magnitude of the change that is endogenous, and the magnitude of the change that
is exogenous (a fiscal event).
5. Match the drivers to their appropriate classification category
6. Extract key phrases that support your classification
7. Assess your confidence in the classification
</process>
<guidance>
Key principles:
- Use the forecaster’s rationale as your primary source of information about
what happened and why the forecast changed. However, apply the classification
framework independently - the forecaster may use categories like “automatic
stabilizers” vs. “discretionary stimulus” that do not map directly to
the endogenous/fiscal_event distinction in this framework. Generically,
discretionary stimulus in response to an economic downturn is endogenous under
this framework, even if the forecaster treats it as a separate category from
automatic stabilizers.
- Be conservative with “fiscal_event” - reserve for clear policy announcements or
political shocks that are orthogonal to the current state of the macroeconomy.
- Note that spending in response to an economic downturn, or raising taxes in
response to an economic recover, *can* be a fiscal event. This is the case when
the change in your forecast is generated by a surprising political development:
an election outcome, a breakdown in negotiations, a vote a on a bill going
in an unexpected direction, a newly announced size for a bill when previous
expectations were different, etc. But it must be a genuinely exogenous (to
macroeconomic conditions) political development, not just a policy announcement.
- If values didn’t change, classification MUST be “none” regardless of rationale
- Direction of causality: “endogenous” means fiscal policy responding to
economic conditions that arose independently of the fiscal policy itself (e.g., a
recession causes the government to spend more). It does NOT include economic
conditions responding to fiscal policy. Fiscal multiplier effects - where
government spending boosts GDP, which in turn raises tax receipts or lowers
automatic stabilizer costs - are part of the fiscal shock’s total impact, not
a separate endogenous component. Do not classify macro feedback from a fiscal
```

```
shock as endogenous.
- If short-term changes are due to endogenous factors, then any long-term changes
likely reflect the endogenous response of interest expense to the change in the
debt stock, and is therefore also endogenous.
- If short-term changes are due to fiscal events and the long-term only changes
due to interest expense, then the long-term changes are also due to a fiscal
event.
- If long-term changes are due to a change in the long-run average deficit or the
fiscal rule, then the long-term changes are also due to a fiscal event.
Composition categories (for decomposing forecast changes by type):
- Government purchases (G): Federal spending on goods and services - defense
procurement, non-defense discretionary spending on goods/services, federal
employee compensation, public investment/infrastructure.  Does NOT include
transfer payments.
- Taxes and transfers (T): Net taxes - tax receipts (income, corporate, payroll,
excise) minus transfer payments (stimulus checks, unemployment insurance, SNAP,
Medicaid, Social Security, tax credits/rebates).  A decrease in T means either
lower tax receipts or higher transfer payments (both widen the deficit).
- Interest expense (I): Net interest payments on the federal debt.
Classification heuristics:
- "Treasury reported..." or "MTS showed..." → if this represents tax receipts or
spending changing in response to economic conditions, likely endogenous.
- "Congress passed..." or "White House announced..." → this could be a
fiscal_event, but only if the news that is driving the change is orthogonal
to the current state of the macroeconomy.
- "Tax receipts/revenues..." → likely endogenous
- "Spending bill..." or "Budget agreement..." → this could be a fiscal_event, but
only if the news that is driving the change is orthogonal to the current state of
the macroeconomy.
- "No new information" or unchanged values → none
Debt permanence assessment:
- For any forecast update where the classification is not "none", assess whether
the deficit change represents a permanent shift in the level of government debt.
- A permanent shift means there is no expectation that today's deficit change
will be offset by a future contraction (or vice versa).  The debt level moves to
a new trajectory.
- A non-permanent (temporary) shift means there is language suggesting the
deficit change will be reversed or offset at some point in the future -
```

```
potentially beyond the 10-year forecast horizon.
- Look for language such as:
- Suggesting TEMPORARY (not permanent):  sunset clauses, expiration dates,
``one-time'' spending, pay-for provisions, explicit future offsets, ``temporary''
programs, time-limited appropriations, emergency spending expected to wind down
- Suggesting PERMANENT: new entitlement programs, structural tax changes with no
sunset, permanent spending increases, no mention of offsets or expiration
- If no language either way, default to permanent - the absence of offset
language suggests the change will persist.
- This assessment should consider implications beyond the 10-year forecast
horizon.  Even if a policy expires in year 8, if there is language suggesting
it will be renewed, treat it as permanent.
- Note:  endogenous changes (business cycle responses) are typically temporary
in terms of the deficit flow - the deficit returns to normal when the economy
recovers - but they still result in a permanent shift in the debt level, since
the government rarely runs offsetting surpluses during recoveries.  Unless there
is explicit language about future fiscal consolidation or surplus, classify
endogenous changes as permanent debt level shifts.
Stabilization policy assessment:
- Separately from the endogenous/fiscal_event classification (which must not be
altered), assess whether the deficit change is motivated by stabilization of the
business cycle.
- This is a distinct dimension:  a fiscal_event can be either stabilizing
or non-stabilizing, and an endogenous change can be either stabilizing or
non-stabilizing.
- Non-stabilizing deficit changes are those not taken to offset factors pushing
growth away from normal.  Quintessential non-stabilizing changes might be a tax
cut motivated by a belief that lower marginal tax rates will raise output in the
long run, new military spending programs, or new infrastructure programs.  Such
an action is fundamentally different from a countercyclical action where the goal
is to offset shocks acting to reduce growth relative to normal.
- Stabilizing (true):  The policy action is countercyclical - aimed at offsetting
economic shocks that are pushing growth away from normal.  Examples:  stimulus
spending during a recession, emergency unemployment benefits, automatic
stabilizer responses.
- Non-stabilizing (false):  The policy action is not motivated by cyclical
concerns.  Examples:  structural tax reform, new long-term military programs,
infrastructure investment motivated by long-run growth, entitlement expansions
```

```
unrelated to the business cycle.
- If the change is purely endogenous (automatic stabilizers responding to
economic conditions), it is stabilizing by definition.
- If the change is a fiscal event, assess whether the political action was
motivated by stabilizing the business cycle or by other policy goals.
</guidance>
<output>
Provide:
- classification:  The overall classification for the forecast update:
- "none":  No changes in any fiscal year
- "endogenous":  ALL changes across ALL fiscal years that changed are due to
endogenous factors
- "fiscal_event":  ALL changes across ALL fiscal years that changed are due to
fiscal events
- "both":  ANY fiscal year has a mixture of endogenous and fiscal event changes,
OR some fiscal years changed due to endogenous factors while others changed due
to fiscal events
- rationale:  Clear explanation for your classification decision.  Required if
classification != 'none'.
- magnitudes_by_year:  A list of magnitude breakdowns for EACH of the 11 fiscal
years.  Each entry must contain:
- fiscal_year:  The fiscal year label (e.g., "FY2025")
- endogenous_change:  The magnitude of change attributed to endogenous factors
(USD billion).  Use 0 if no endogenous change for this year.
- fiscal_event_change:  The magnitude of change attributed to fiscal events (USD
billion).  Use 0 if no fiscal event change for this year.
CONSTRAINT: For each fiscal year, endogenous_change + fiscal_event_change
MUST exactly equal the total change in that year's forecast from yesterday
(today's value minus yesterday's value).  These totals are pre-computed in
the <forecast_changes> block of the user prompt - your split must sum to those
numbers.  Include all 11 fiscal years even if some have zero changes.
- confidence:  Your confidence in the classification (0.0 to 1.0).  Required if
classification != 'none'.
- key_phrases:  List of 2-5 phrases from the forecaster's rationale or news that
drove your classification.  Required if classification != 'none'.
- permanent_debt_level_shift:  Whether the deficit change is expected to result
in a permanent shift in the level of government debt (true/false).  Set to true
if there is no language in the rationale or news suggesting the deficit change
```

```
will be offset by future fiscal contractions (or expansions).  Set to false if
there IS such language (e.g., sunset clauses, pay-for provisions, temporary
programs, statements about future consolidation).  This extends beyond the
10-year forecast horizon.  Required if classification != 'none'.
- composition_by_year:  A list of composition breakdowns for EACH of the 11
fiscal years.  Each entry must contain:
- fiscal_year:  The fiscal year label (e.g., ''FY2025'')
- purchases_change:  The change in the budget balance attributable to government
purchases of goods and services (USD billion).  Negative means purchases
increased (widening the deficit).
- taxes_transfers_change:  The change in the budget balance attributable to taxes
and transfers combined (USD billion).  Negative means net taxes fell (lower tax
receipts or higher transfer payments, widening the deficit).  Note:  transfer
payments (stimulus checks, unemployment insurance, SNAP, Medicaid, etc.)  are
grouped with taxes, not with purchases.
- interest_change:  The change in the budget balance attributable to net interest
expense on the debt (USD billion).  Negative means interest costs increased
(widening the deficit).
CONSTRAINT: For each fiscal year, purchases_change + taxes_transfers_change +
interest_change MUST exactly equal the total change in that year's forecast (from
the <forecast_changes> block).  Include all 11 fiscal years even if some have
zero changes.
- stabilization_policy:  Whether the deficit change is motivated by
stabilization of the business cycle (true/false).  This is independent of the
endogenous/fiscal_event classification.  Set to true if the policy action is
countercyclical - aimed at offsetting shocks pushing growth away from normal.
Set to false if the policy change is not motivated by cyclical stabilization
(e.g., structural tax reform, new military programs, infrastructure spending
for long-run growth).  Purely endogenous changes (automatic stabilizers) are
stabilizing by definition.  Required if classification != 'none'.
</output>
```

B.6 News summarizer system prompt

```
<role> You are a macroeconomic analyst specialized in US fiscal policy.  Your
expertise lies in identifying and synthesizing information relevant to the US
Federal Budget Balance (the difference between federal receipts and outlays).
</role>
<task> You will receive a batch of news articles for a given period.  Your task
is to:
1.  Read through all articles and identify those containing fiscal
policy-relevant information 2.  For EACH DAY that has fiscal-relevant news,
produce a concise summary 3.  Output a dictionary mapping each date to its
summary
Focus specifically on news that could impact the US Federal Budget Balance
(deficit/surplus), including:  - Monthly Treasury Statement releases (budget
balance data) - Congressional Budget Office (CBO) or Office of Management
and Budget (OMB) forecasts and reports - Congressional budget actions,
appropriations bills, continuing resolutions - Tax policy changes or proposals
(corporate, individual, payroll taxes) - Spending program changes (entitlements,
discretionary spending, defense) - Debt ceiling developments - Fiscal stimulus
or austerity measures - Government shutdown risks or occurrences - Major policy
announcements from the White House or Treasury </task>
<guidance> WHAT TO INCLUDE: - Fiscal data releases (deficit/surplus figures,
receipts, outlays) - Budget forecasts from CBO, OMB, or Consensus Economics polls
- Legislation that materially affects revenues or spending - Emergency fiscal
measures or stimulus packages - Debt ceiling negotiations and outcomes - Specific
numerical details when mentioned (amounts in USD billions, percentages of GDP)
- Market commentary about fiscal credibility or sustainability when directly
relevant
WHAT TO EXCLUDE: - General economic news (GDP, employment, inflation) unless it
has direct, quantified fiscal implications - Monetary policy (Federal Reserve
actions) unless directly tied to fiscal operations - Trade policy unless it
includes fiscal components (tariff revenues) - State and local government
finances (focus on Federal only) - Market movements without fiscal policy context
IMPORTANT GUIDELINES: - Be EXHAUSTIVE: It is better to include something
marginally relevant than to miss important fiscal news - Be CONCISE: Each
summary should be 50-100 words maximum.  Focus on key numbers and decisions.
- DATE ATTRIBUTION: Always assign news to the date the EVENT occurred, not the
publication date.  If an article published on Jan 17 reports on guidance issued
on Jan 14, attribute it to Jan 14.  If multiple articles cover the same event on
```

```
different days, consolidate into ONE entry on the event date.  - Do NOT produce
entries for days with no fiscal-relevant news - Include the specific date of
data releases (e.g., "Treasury reported October deficit of $X billion") - When
articles cite fiscal figures, note whether they are in millions or billions
</guidance>
<output> Output a list of daily summaries, where each entry has:  - date:  The
date in YYYY-MM-DD format - summary:  A concise summary (50-100 words) of fiscal
developments for that day
Only include dates that have fiscal-relevant news.  If the period has no fiscal
news, return an empty list.
The summaries will be used as context for a fiscal forecasting model, helping it
understand what information was already known on any given date.  </output>
```

B.7 Article summarizer prompt

```
<role>
You are a macroeconomic research analyst specialized in US fiscal policy.  Your
expertise lies in identifying and synthesizing information relevant to the US
Federal Budget Balance (the difference between federal receipts and outlays).
</role>
<task>
You will receive a single news article.  Your task is to:
1.  Determine whether the article contains US fiscal policy-relevant information
2.  If it does, produce a concise summary of the fiscal-relevant content
3.  If it does not, respond with exactly:  "No fiscal news."
The summaries you create will be passed on to an expert tasked with forecasting
the US budget balance in the current and next 10 years.  You want to provide them
with all the information needed to make the best possible forecast.
</task>
<guidance>
WHAT TO INCLUDE:
- Monthly Treasury Statement releases (budget balance data)
- Congressional Budget Office (CBO) or Office of Management and Budget (OMB)
forecasts and reports
- Congressional budget actions, appropriations bills, continuing resolutions
- Tax policy changes or proposals (corporate, individual, payroll taxes)
- Spending program changes (entitlements, discretionary spending, defense)
- Debt ceiling developments
```

```
- Fiscal stimulus or austerity measures
- Government shutdown risks or occurrences
- Major policy announcements or proposals from the White House, Congress, or
Treasury
WHAT TO EXCLUDE:
- General economic news (GDP, employment, inflation) unless it has direct,
quantified fiscal implications
- Monetary policy (Federal Reserve actions) unless directly tied to fiscal
operations
- Trade policy unless it includes fiscal components (tariff revenues)
- State and local government finances (focus on Federal only)
- Market movements without fiscal policy context
IMPORTANT GUIDELINES:
- Be explicitly quantitative:  if the article contains estimates of
costs/receipts of different budgetary programs, include the exact figures
- If general macroeconomic developments are significant enough to impact the
budget, mention them, with a priority on being explicitly quantitative
- If news comes out about policy proposals or developments, include whatever
information is given about how likely it is to come to fruition
- Ignore any news that is not US relevant
- Don't be interpretive, just be descriptive about what the news says
- Comment on the timing of when proposed bills would come into action, if there
is information about which fiscal years they would affect
</guidance>
<output>
If the article contains fiscal-relevant news, output one paragraph (50-100 words)
summarizing the fiscal content with all relevant numbers and details.
If the article contains no fiscal-relevant news, respond with exactly:  "No
fiscal news."
</output>
```

B.8 Example article summary

```
"date":  "1980-11-06"
"headline":  "Tax Moves Expected"
"summary":  "The article discusses expected federal spending cuts and tax
proposals under President-elect Ronald Reagan's administration for the fiscal
year 1982, beginning October 1, 1981.  Reagan's advisers anticipate larger budget
```

```
cuts than the previously targeted $19 billion for fiscal year 1981, although
these cuts may end up being smaller due to the ongoing fiscal year.  Key tax
proposals include a three-year, 30 percent reduction, a permanent adjustment of
tax rates for inflation starting in the fourth year, and accelerated depreciation
for business investments.  These changes may lead to an estimated tax cut costing
$192 billion annually by 1985."
```

Table 16: Tax Legislation (MR Table A.1): LLM Signal Accumulation

#	Name	Signed	Effective	LLM First	50% LLM	50% Romer
1.	Social Security Amendments of 1947	Aug 1947	1950 Q1	No match	No match	No match
2.	Revenue Act of 1948	Apr 1948	1948 Q2 1948 Q3	1947-01-07	1948-01-06	1948-01-06
3.	Social Security Amendments of 1950	Aug 1950	1954 Q1	No match	No match	No match
4.	Expiration of Excess Profits and Temporary Income Tax	Oct 1951	1954 Q1	1950-07-20	1953-01-12	1952-11-12
5.	Internal Revenue Code of 1954	Aug 1954	1954 Q3	1953-05-14	1954-01-25	1954-01-04
			1954 Q4			
6.	Tax Rate Extension Act of 1958	Jun 1958	1958 Q3	1957-05-24	1958-09-03	1958-09-03
7.	Social Security Amendments of 1958	Aug 1958	1960 Q1	No match	No match	No match
8.	Federal-Aid Highway Act of 1959	Sep 1959	1959 Q4	1958-12-23	1958-12-23	1958-12-23
9.	Social Security Amendments of 1961	Jun 1961	1963 Q1	1961-06-30	1961-06-30	No match
10.	Changes in Depreciation Guidelines and Revenue Act of 1962	Jul 1962	1962 Q3	1960-11-30	1962-05-14	1961-07-18
			1962 Q4 1962 Q4 1963 Q1 1963 Q1			
11.	Revenue Act of 1964	Feb 1964	1964 Q2 1964 Q3 1965 Q1	1960-11-30	1962-11-26	1963-07-31
12.	Excise Tax Reduction of 1965	Jun 1965	1965 Q3	1964-08-21	1965-01-11	1965-01-11
			1966 Q1			
13.	Tax Adjustment Act of 1966	Mar 1966	1966 Q2	1966-01-13	1966-01-13	1966-01-13

continued on next page

Table 16 continued

#	Name	Signed	Effective	LLM First	50% LLM	50% Romer
14.	Public Law 90-26	Jun 1967	1967 Q3	1967-03-10	1967-03-10	1967-03-10
			1967 Q4			
15.	Social Security Amendments of 1967	Jan 1968	1971 Q1	1967-11-03	1969-01-16	1968-08-15
16.	Tax Reform Act of 1969	Dec 1969	1971 Q1	1969-04-02	1969-11-20	1969-08-04
			1972 Q1			
17.	Reform of Depreciation Rules	Jan 1971	1971 Q1	1971-01-11	1971-01-12	1971-01-11
18.	Revenue Act of 1971	Dec 1971	1972 Q1	1970-08-31	1971-08-17	1971-08-17
			1972 Q2			
19.	1972 Changes to Social Security	Oct 1972	1978 Q1	1972-02-24	No match	No match
20.	Tax Reform Act of 1976	Oct 1976	1976 Q4	1976-06-18	1976-09-10	No match
			1977 Q1			
21.	Tax Reduction and Simplification Act of 1977	May 1977	1977 Q3	1976-11-05	1977-05-02	1976-11-22
			1977 Q4			
22.	Social Security Amendments of 1977	Dec 1977	1979 Q1	1977-05-10	1977-11-04	No match
			1980 Q1			
			1981 Q1			
			1982 Q1			
23.	Revenue Act of 1978	Nov 1978	1979 Q1	1977-09-13	1977-10-03	1977-10-03
24.	Crude Oil Windfall Profit Tax Act of 1980	Apr 1980	1980 Q2	1979-03-21	1980-03-31	1979-11-07
			1981 Q1			
			1982 Q1			
25.	Economic Recovery Tax Act of 1981	Aug 1981	1981 Q3	1980-02-15	1980-11-06	1980-11-06
			1981 Q4			
			1982 Q1			
			1983 Q1			
			1984 Q1			
26.	Tax Equity and Fiscal Responsibility Act of 1982	Sep 1982	1983 Q1	1982-01-08	1982-01-08	1982-01-08

continued on next page

Table 16 continued

#	Name	Signed	Effective	LLM First	50% LLM	50% Romer
27.	Social Security Amendments of 1983	Apr 1983	1984 Q1	1983-01-13	1983-01-13	No match
			1985 Q1			
			1986 Q1			
			1988 Q1			
			1990 Q1			
28.	Deficit Reduction Act of 1984	Jul 1984	1984 Q3	1984-01-23	1984-03-16	1984-02-24
29.	Tax Reform Act of 1986	Oct 1986	1986 Q4	1985-06-11	1985-06-11	1985-06-11
			1987 Q1			
			1987 Q3			
			1988 Q1			
30.	Omnibus Budget Reconciliation Act of 1987	Dec 1987	1988 Q1	1987-02-24	1987-02-24	1987-02-24
31.	Omnibus Budget Reconciliation Act of 1990	Nov 1990	1991 Q1	1988-12-07	1990-03-12	1990-03-12
32.	Omnibus Budget Reconciliation Act of 1993	Aug 1993	1993 Q3	1992-12-09	1993-04-12	1993-02-15
			1993 Q4			
			1994 Q1			
33.	Tax Payer Relief Act and Balanced Budget Act of 1997	Aug 1997	2000 Q1	1997-04-29	No match	No match
			2002 Q1			
34.	Economic Growth and Tax Relief Reconciliation Act of 2001	Jun 2001	2002 Q1	2001-01-26	2001-04-04	No match
35.	Jobs and Growth Tax Relief Reconciliation Act of 2003	May 2003	2003 Q3	2002-11-14	2003-01-06	2002-12-09
			2003 Q4			

2005 Q1

50% LLM = first date when cumulative LLM max-FY tax signal reaches 50% of its own aggregate. 50% Romer = first date when cumulative LLM signal reaches 50% of Romer exogenous PDV (MR fallback for legs where exog PDV = 0: SS 1950, Expiration of Excess Profits Tax, SS 1958/1961/1967, SS Changes 1972). Max FY: FY with largest absolute LLM aggregate whose sign matches the Romer exogenous value. SS Changes 1972 and Taxpayer Relief Act 1997 show 'No match' because the LLM signal is entirely opposite in sign to the small Romer exogenous component (the full legislation is dominated by a much larger non-exogenous component in the other direction).