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. The methodology involves two steps: first, a "forecaster" model is given the daily news that is relevant to fiscal outcomes and updates their forecast for the current and next ten fiscal years of deficits. Second, a "classifier" identifies which changes in the forecast are driven by fiscal shocks — unexpected changes in the stance of fiscal policy. We apply this method to US data, running 2005-2025. Our results 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. We then measure the effects of fiscal shocks on asset prices: in response to a 1pp shock to the present-value of the next eleven-years deficits, ten-year Treasury yields rise more than 30bps, with real yields accounting for 53-60% of the rise and break-even inflation expectations accounting for the remainder. The dollar appreciates significantly — 2.3% — and the term-premium, as measured by the 2Y-10Y spread, rises 14bps. We also show that the response of asset prices is driven by shocks with longer-term implications for deficits, while short-term changes in deficits are not priced by investors. In local projections, fiscal shocks have contractionary effects on realized activity.

*We thank Sergio de Ferra, Peter Karadi, Marek Jarociński, Elmar Mertens, Davide Porcellacchia, Michele Lenza, and seminar participants at the University of Oxford, the European Central Bank, 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 Older VAR and narrative-based identification approaches have been largely superseded by "high-frequency" identification — looking at daily surprises in Fed funds futures to cleanly identify the effects of monetary shocks.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.

Two different LLM "agents" are coordinated to produce a time-series of quantified fiscal shocks. The first agent is the "forecaster." At each date in time, the forecaster is given all the relevant news articles 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.

In virtue of covering the daily universe of fiscal news, the agents are able to produce a time-series of fiscal shocks that is credibly both relevant and exogenous. Since the model forecasts *aggregate* deficits, only changes in expected fiscal policy that materially alter the fiscal stance end up included in the shock series. Since the model responds to daily news, the shocks are true unanticipated changes to expectations about future fiscal policy.

We apply this methodology to the US, covering a period from January 2005 to July 2025. Our primary results are on the effects of fiscal shocks on asset prices, which our daily-frequency shock history allows us to precisely identify. In response to a fiscal shock that increases the

next eleven years present discounted-value of deficits over the present discounted-value of expected GDP by 1 percentage point, the ten-year Treasury yield rises more than 30 basis points (bps). 53-60% of the increase comes in real yields as measured by TIPS and the remainder in inflation breakevens. The significant response of both real yields and market inflation expectations supports models where the asset demand function is not perfectly elastic and where fiscal policy has material implications for inflation. The two-year Treasury yield also responds significantly (18-23bps), but less than the ten-year, giving rise to an increase in the term-premium which is suggestive of attention to fiscal sustainability. Additionally, the dollar depreciates by 2.3% — a very large move — in-line with the directional predictions of standard open-economy models.

In virtue of forecasting deficits out ten-years in the future, we are able to split our sample of shocks by the median "weighted-horizon" of when the shock affects future deficits. Shorter-term shocks (below the median) have no significant effect on asset prices, whereas longer-term shocks have clear and consistent effects. Investors only price fiscal shocks which affect long-term debt supply and fiscal sustainability.

We then use local projections to examine the effects of fiscal shocks on subsequent inflation, GDP, consumption, and investment outcomes. In the short-term, contrary to standard theory, activity contracts in response to fiscal shocks. However, given our relatively short sample that is dominated by two large recessions (2005-25), we see these results as mostly a proof of concept that our methodology can uncover a relevant series of fiscal shocks — as indicated by the strong predictability of future realized deficits from our shock series.

We also explore in detail how our time-series of fiscal expectations compares to lower-frequency external benchmarks, and what our resulting shock series looks like. 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 current and year-ahead deficits equally well as 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 all 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 $450B. This forecast update matches exactly the 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. By early April the model's forecast for FY2020 was very close to realized FY2020 deficits, and significantly closer than the Consensus median expectation. 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 are Trump's 2016 election, Trump's 2024 election, and the Georgia Senate runoff. 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. Furthermore, we show that the time-series of fiscal shocks are unbiased predictors of future realized 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 the shocks 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 the largest fiscal shocks in the last 20 years. The systematic nature of

our LLM-based methodology also allows it to easily extend to different countries and time-horizons.

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 (i.e. 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 (i.e. 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 (i.e. Chodorow-Reich et al. 2012, Nakamura and Steinsson 2014).

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 (i.e. Laubach 2009, Rachel and Summers 2019, Campos et al. 2024). The results on inflation expectations can be compared to recent models making predictions about the effects of future fiscal policy on inflation (i.e. 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 (i.e. Monacelli and Perotti 2006, Auerbach and Gorodnichenko 2016). 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 (i.e. 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 and macroeconomic variables. Section 5 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 fiscal effects.

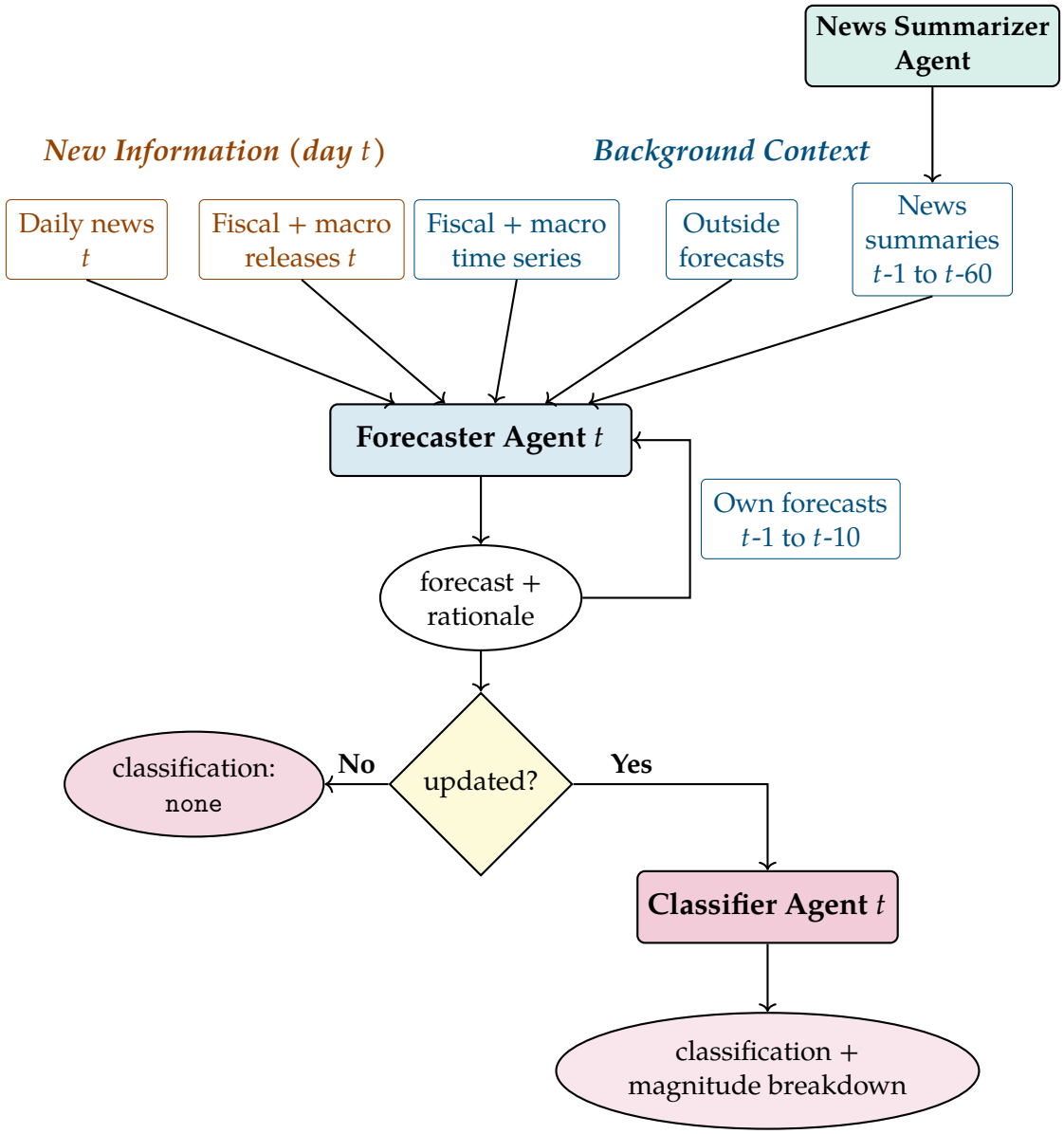


Figure 1: Multi-Agent Architecture for Fiscal Shock Identification.

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 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 (i.e. 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

uk Our news articles for the US are all Reuters and Wall Street Journal (WSJ) articles available on Factiva (from January 3rd 2005 to July 23rd 2025) 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.

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 "news summarizer" agent is available in Appendix B.5.

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

[1] Our precise search query is available in appendix B.2

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.

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.[2]

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.3. 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.

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

[2]Full details on data sources and series identifiers are provided in Appendix B.1.

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.[3]

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

The 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, **?**). 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, we can think of fiscal policy operating according to the following rule:

$$d_t = \alpha + \beta * \tilde{y}_t + \epsilon_t \tag{1}$$

The deficit d_t is equal to the sum of the long-run average deficit α, the expected response β to the output gap $\tilde{y}_t$, and any one-off shocks to fiscal policy ϵ_t. Changes in d_t that are driven by changes to y_t represent the endogenous response of fiscal policy to economic conditions. Any movement in α, β or ϵ represents a shock — a "fiscal event". Note that movements in α and β are *changes* to the fiscal rule — such as changes in tax policy, or a change in how stimulative the government is expected to be in recessions — while one-off policy shocks (i.e. one-time transfers) that are not expected to affect future fiscal policy are captured by ϵ_t.

In prompting the classifier agent, we explicitly give the agent this equation, and tell it to "classify the forecast update according to the above framework". The 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.

In order to make sure the agent understands the concept of a fiscal rule, we explicitly tell it the following:

```
Historically in the US a  1pp rise in the output gap leads to a  1pp rise in
the deficit/GDP ratio.  This is *not* all due to automatic stabilizers - a large
chunk of it is due to the *predictable* pattern whereby discretionary fiscal
spending rises in response to deteriorating economic conditions.  Even if this
fiscal policy is ‘‘discretionary’’, it is part of the fiscal rule, since it is part
of the expected response of the government to economic downturns.  When deciding
whether a policy announcement is endogenous or fiscal news, think explicitly
about what the current economic news implies about future output gaps, and
whether the announced discretionary fiscal policy is more or less expansionary
than the historical norm.  Only if it deviates from the historical norm should it
be considered a fiscal event.
```

The coefficient cited in the prompt is estimated from historical data.[4]

Additionally, the classifier is given a subset of the context the forecaster is given: the classifier is only 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 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) surprised relative to pre-existing expectations of the fiscal rule.

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. The full prompt given to the classifier is available in appendix B.4.

[4]We regress realized deficit/GDP on the annual average of the CBO's quarterly measure of the output gap over the 20-year period 1984-2004. The exact coefficient is 1.05. The period is chosen to represent what would have been the agent's information set at the start of our sample (January 2005).

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 address both of these concerns. As we will document below, many of the variations in deficit forecasts identified by our classifier as "fiscal events" are strongly relevant predictors of future fiscal policy, and are also relevant predictors of future realizations of less-frequent survey or forecast measures of expected future policy. Furthermore, 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 unlikely to have been anticipated in advance.

In addition to helping with relevance and 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. 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 expected nominal GDP. The PDV of expected GDP at each date in time comes from the most recent release of the CBO's "Economic Projections", which covers the same forecast horizon as the deficit forecasts.

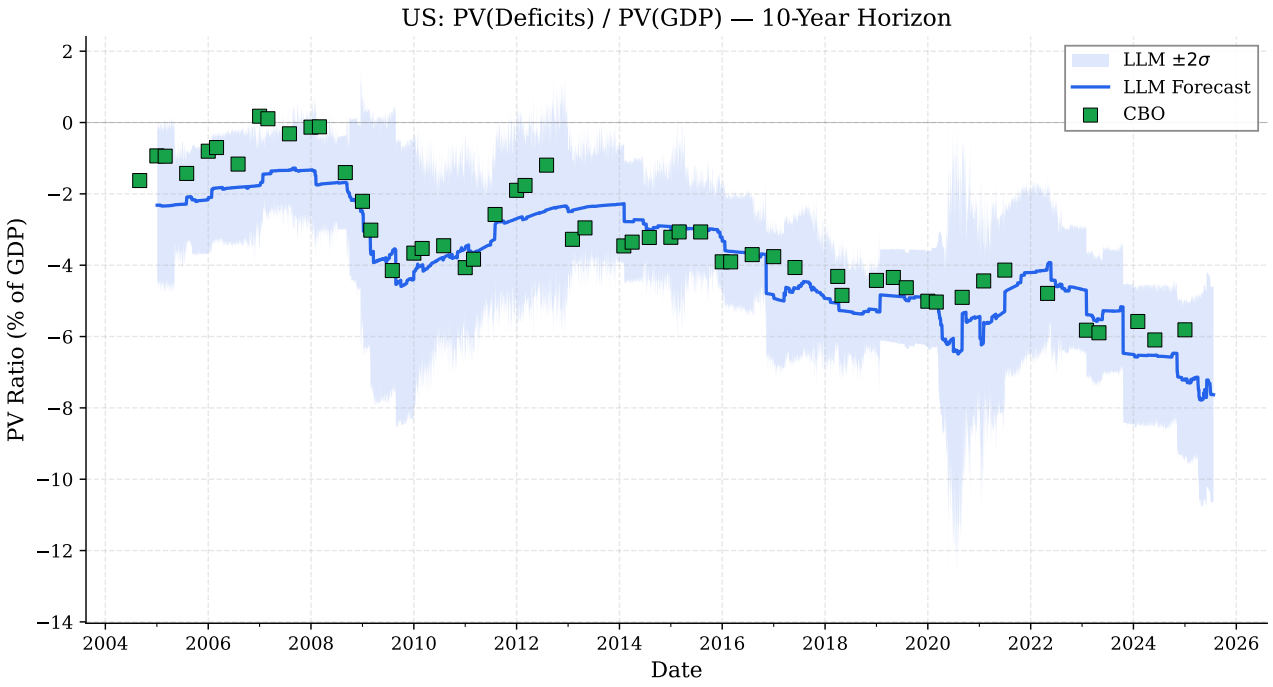


Figure 2: Present discounted value of expected deficits as a share of expected GDP. The PDV ratio is computed by discounting 10-year deficit and GDP forecasts using zero-coupon Treasury yields, with fiscal-year forecasts weighted to correspond to exact N-year-ahead horizons. CBO values (squares) are computed using the same methodology applied to CBO deficit projections. Shaded region shows the LLM's $\pm 2\sigma$ confidence interval.

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:

$$x_t = \frac{\text{days remaining in current fiscal year}}{365} \tag{2}$$

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

$$\hat{d}_{t,N} = x_t \cdot d_{t,\text{FY}(t)+N-1} + (1 - x_t) \cdot d_{t,\text{FY}(t)+N} \tag{3}$$

where $d_{t,\text{FY}}$ 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 GDP forecasts to obtain $\hat{g}_{t,N}$.

The present discounted value ratio is then computed as:

$$\text{PDV ratio}_t = \frac{\sum_{N=1}^{10} \hat{d}_{t,N} \cdot e^{-y_{t,N} \cdot N}}{\sum_{N=1}^{10} \hat{y}_{t,N} \cdot e^{-y_{t,N} \cdot N}} \tag{4}$$

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.

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.

Aside from a few observations in 2007 and 2008, all CBO forecasts lie within the LLM's confidence interval, and the two series generally track each other well. Across the full sample, the PDV of deficits has decreased substantially — from around -2% of GDP to almost -8%. Interestingly, the GFC shows up as a larger deterioration in the present value of the US fiscal position than the Covid crisis. Since the start of 2022, the present discounted value of deficits/GDP has almost doubled. 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 1pp of GDP.

Figure 3 compares the one-year-ahead LLM forecast against the one-year-ahead Consensus Economics monthly survey. Consensus Economics reports median forecasts for the current fiscal year and the year-ahead fiscal year separately. To construct a comparable one-year-

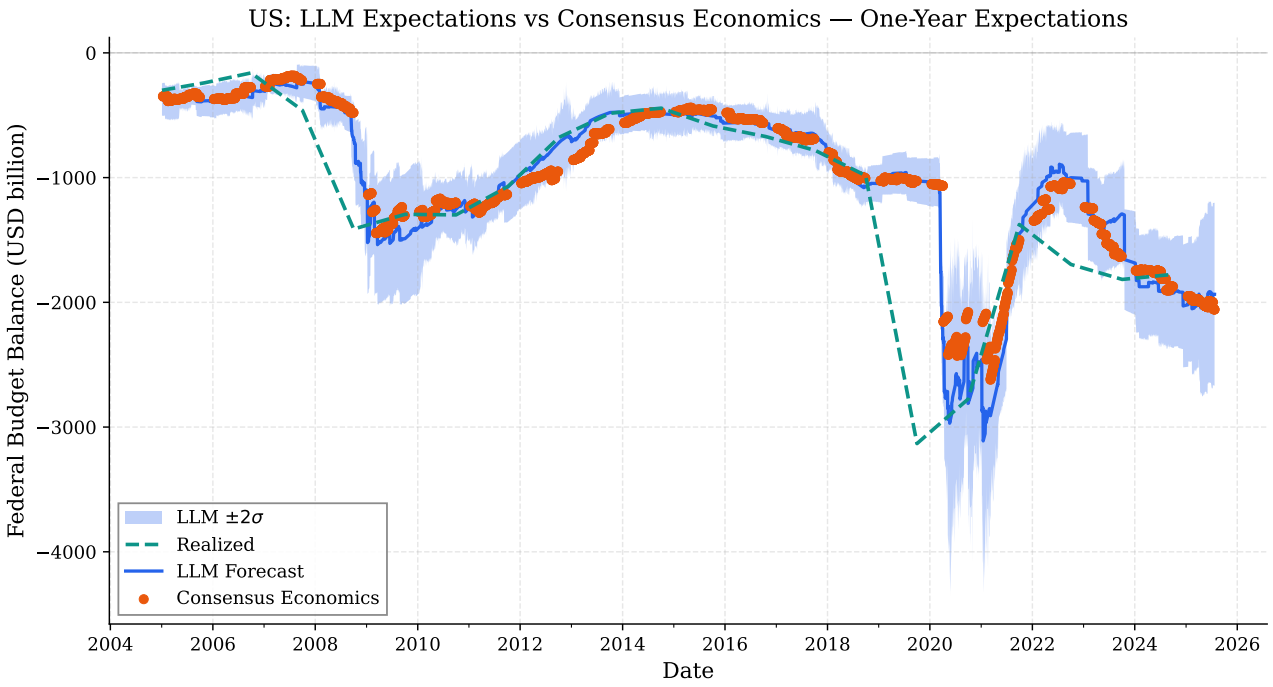


Figure 3: One-year deficit expectations: LLM forecasts (solid line) versus Consensus Economics survey values (dots) and realized deficits (dashed line). All series use the same year-weighting: $\hat{d}_{1Y} = x \cdot d_{\mathrm{CY}} + (1 - x) \cdot d_{\mathrm{YA}}$, where x is the fraction of the current fiscal year remaining. Shaded region shows the LLM's $\pm 2\sigma$ confidence interval.

ahead expectation from both sources, we apply the same year-weighting scheme:

$$\hat{d}_{t,1Y} = x_t \cdot d_{t,CY} + (1 - x_t) \cdot d_{t,YA} \tag{5}$$

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 date t.[5]

The dashed-green line applies the same weighting to realized year-ahead outcomes. Across most of the sample, the model tracks Consensus forecasts quite closely, and both forecasts correspond well with realized deficits. Both 2012-2014 and 2020 show up as periods where the model tracks realized deficits more closely than Consensus.

There are three noticeable "misses" for both the model and the Consensus forecast: the beginning of the GFC, the beginning of the Covid crisis, and a period in 2022. Rather than interpreting these 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 outcome.[6] 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, Consensus outperforms the LLM with an RMSE of $318.5B vs $368.8B. However, for year-ahead forecasts (1YA), the two are statistically indistinguish-

[5] 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.

[6] 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	187	368.8	318.5	172.1	144.4	47.1%
1Y-Weighted	180	479.7	465.4	262.0	261.0	50.6%
1YA	240	609.1	597.1	326.0	330.6	52.5%

Notes: CY = current fiscal year, 1YA = year-ahead, 1Y-Weighted = weighted average. RMSE and MAE in $B.

able, with the LLM winning 52.5% of head-to-head comparisons. 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 achieving our goal: coming up with a forecast that matches what an informed expert on the topic would have concluded at the time.

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

Horizon	N	$RMSE_{LLM}$	$RMSE_{CBO}$	MAE_{LLM}	MAE_{CBO}	Win Rate
CY	52	482.0	434.9	217.3	178.3	51.9%
1YA	51	707.4	739.5	395.8	426.1	62.7%
2YA	49	751.7	802.9	492.3	535.7	67.3%
3YA	47	775.7	859.9	553.5	624.1	70.2%
4YA	46	899.6	969.7	651.4	724.2	69.6%
5YA	44	963.3	1009.6	685.2	755.3	63.6%
6YA	41	981.8	1010.8	679.6	765.0	65.9%
7YA	38	941.2	940.3	650.7	734.4	60.5%
8YA	36	1034.4	1089.9	719.0	798.2	55.6%
9YA	34	1147.2	1203.9	850.2	930.4	52.9%
10YA	31	1209.8	1234.5	955.1	1055.5	51.6%

Notes: CY = current fiscal year, 1YA = year-ahead, etc. Comparisons at each CBO release date. RMSE and MAE in $B.

Table 2 compares LLM forecasts to CBO projections at each CBO release date. For current-year forecasts, CBO outperforms the LLM in RMSE and MAE, tough not in win rate. Some degree of CBO out-performance should perhaps be expected given that the CBO has direct access to administrative fiscal data. However, for year-ahead and longer horizons, the LLM consistently outperforms the CBO, winning greater than 60% of head-to-head comparisons at 2–7 year horizons. 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. The model continues "beating" the CBO at the 8-10 year horizon, but only by a very small amount. Once again, we take this as comforting evidence that the model is outputting forecasts that respect its knowledge cut-off at each point in time. Forecasting far out deficits is a hard task, and it would be more concerning if

the model did "too well" on these forecasts.

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. We test this by examining whether external forecasters revise their forecasts in the direction of the LLM's prior-day estimate. Specifically, we estimate:

$$\Delta f_t^{ext} = \alpha + \beta \cdot (f_{d-1}^{LLM} - f_{t-1}^{ext}) + \varepsilon_t \tag{6}$$

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	9.15	−6.16	−10.41
	(5.69)	(5.33)	(4.47)
β	0.51***	0.45***	0.20***
	(0.02)	(0.03)	(0.03)
R^2	0.77	0.58	0.18
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 all horizons, indicating that external forecasters systematically revise their forecasts in the direction of the LLM's position.

For Consensus Economics, the effect is strongest for current-year forecasts, with the gap between where the LLM has "arrived at" the day before Consensus is released and the previous Consensus forecast explaining 77% of the changes in Consensus forecasts. The effect is weaker at the year-ahead horizon but still highly statistically significant.

For CBO, the current-year R^2 of 0.96 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:

Table 4: External Forecaster Updates Towards LLM: CBO

	CY	1YA	2YA	3YA	4YA	5YA
Intercept	0.45	7.94	3.30	14.35	18.72	9.47
	(16.65)	(19.44)	(19.91)	(18.40)	(16.79)	(18.09)
β	0.71***	0.47***	0.37***	0.31***	0.35***	0.33***
	(0.03)	(0.06)	(0.13)	(0.11)	(0.10)	(0.10)
R^2	0.96	0.71	0.22	0.20	0.29	0.29
	6YA	7YA	8YA	9YA	10YA	
Intercept	13.55	10.97	9.97	4.90	4.03	
	(21.03)	(21.68)	(22.55)	(23.20)	(23.11)	
β	0.35***	0.28***	0.26***	0.23***	0.24***	
	(0.10)	(0.09)	(0.09)	(0.09)	(0.08)	
R^2	0.28	0.24	0.23	0.20	0.26	

Notes: Standard errors in parentheses. N = 31 for all horizons. $^{***}p < 0.01$.
CY = current fiscal year, 1YA = year-ahead, etc.

there is more room for Consensus and the model to disagree about changes within a month, but across the two or six month gaps between CBO forecasts, the model incorporates all the information that ends up in the CBO forecast. The coefficients remain positive and significant across all horizons out to 10 years, ranging from 0.23 to 0.35 for longer-term forecasts. This suggests that even the LLM's long-term forecasts are able to respond to current news in a manner consistent with the news's expected effect on deficits.

It is interesting that the β coefficients for both Consensus and the CBO are < 1 at all horizons. Given the model's superior accuracy to the CBO at medium-term horizons, the low coefficients might reflect that the CBO's "current-law" assumptions are preventing 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 therefore being more able to pass a follow-up round of expansionary Covid stimulus. This

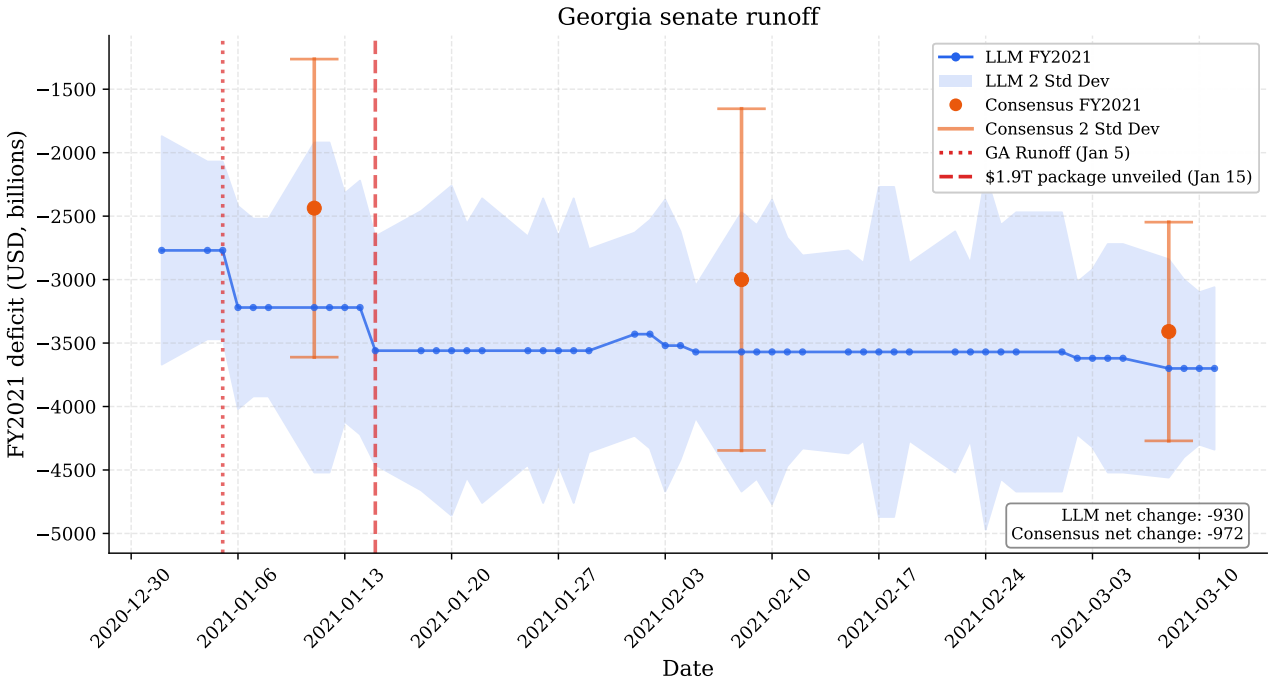


Figure 4: LLM and Consensus FY2021 deficit forecasts around the Georgia Senate runoff election.

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 $450Bn — the exact magnitude of the shock in Hazell and Hobler (2024)'s central estimate. Ten days later, on January 15th, Biden proposed that the stimulus package would be $1.9T, significantly larger than previous expectations. 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 in full how the forecaster explains their update on the news of the $1.9T proposal:

```
On 1/6 I assumed, conditional on the Georgia results, an expected 0.8-0.9T of
additional near-term COVID relief beyond the December bill, plus a separate
medium-term investment/green package (net ≈ 0.7-0.8T over FY2022-31).
Yesterday's news (referenced in today's FX and muni articles) that Biden has put
forward a ≈ 0.9T near-term package shifts the distribution notably.  My updated
scenario set for near-term relief above the December bill is:
- 1.8T package:  25%
- 1.4T compromise:  40%
```

```
- 0.9T scaled package:  25%
- 0.4T minimal add-on:  10%
    ⟹     Probability-weighted expected size ≈ 0.3T, versus ≈ 0.85T previously
(increment ≈ 0.45T).
Given the design (checks, UI, state/local aid, health/school spending), I assume
≈ 80% of this expected package hits FY2021 and 20% hits FY2022.  Relative to
my 1/6 baseline this adds roughly 0.34T to FY2021 and 0.09T to FY2022 primary
deficits.
```

The model correctly anchors on a previous expected stimulus of around \$0.9T, understands the quantitative implications of the new larger proposal, and adjusts its forecast accordingly.

Across the time from before the runoff election to after the eventual bill passed the Senate (March 6th), the Consensus median forecast sits above the LLM forecast. However, the net update in the Consensus forecast across this period is \$972B — extremely similar to the model's net update of \$930B. In this case, the Consensus forecast seems to lag the model a bit, but it ends up changing in a very similar fashion.[7]

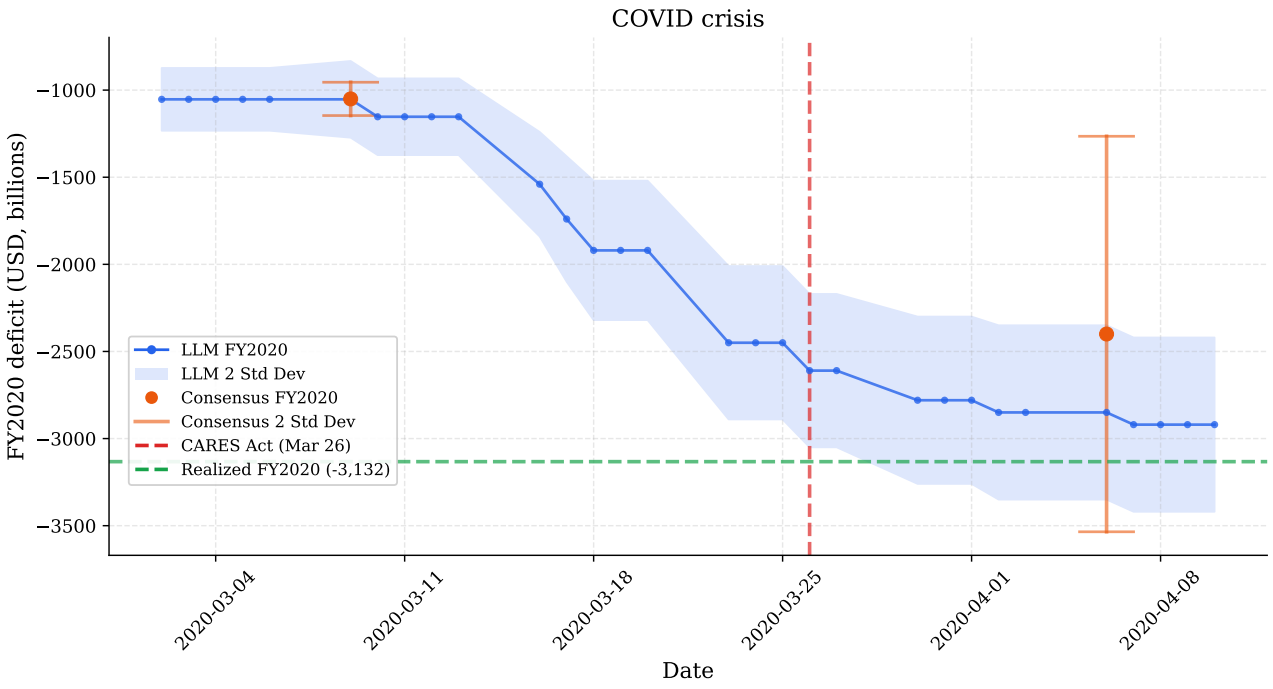


Figure 5: LLM and Consensus FY2020 deficit forecasts during the COVID crisis.

[7]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.

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 exactly in line with Consensus, at a modest FY2020 deficit of $1.05T. 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 by April 6th the model's forecast of FY2020 deficits was $450B lower than the Consensus median and much closer to the eventual realized FY2020 deficit of $3.1T.

3.4 Shock history

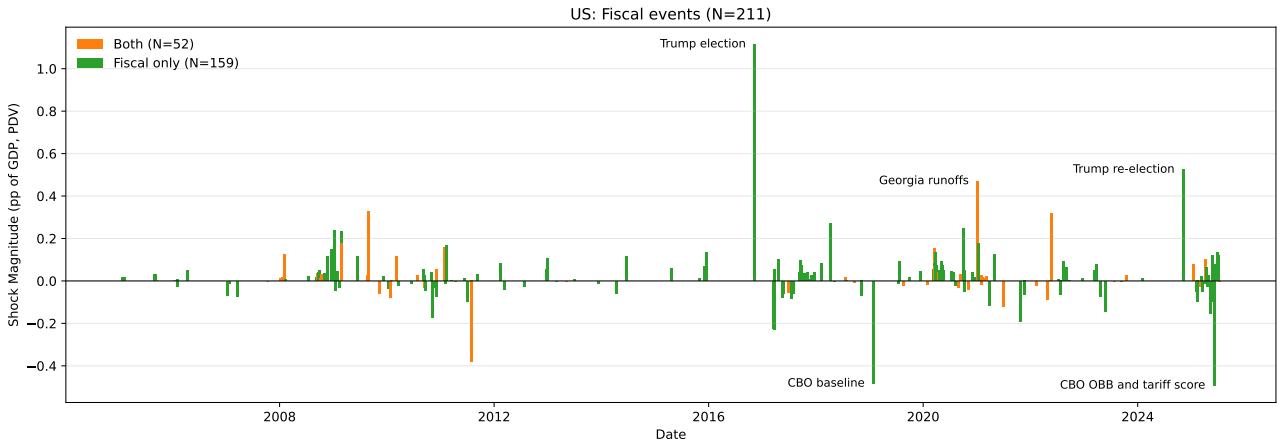


Figure 6: Daily changes in the PDV of expected deficits, classified by shock type.

Figure 6 shows the time-series of 211 "fiscal shocks" identified by the classifier model, with days that were *only* classified as a fiscal shock shaded in green (N=159), and days that included both a fiscal shock and an endogenous shock shaded in green (N=52). 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. 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 three largest fiscal shocks in the sample are, in order, Trump's 2016 election, Trump's 2024 re-election, and the January 5th 2021 Georgia Senate runoff which resulted in Democratic control of both chambers of Congress. All three of these events match well the intended classification of a fiscal shock: they changed expectations about the path of future policy in

a manner independent of business-cycle conditions.

It is of note that these explicitly political developments are classified as larger fiscal shocks (in PDV terms) than any of the news about stimulus packages surrounding the GFC or the Covid outbreak. There are a few reasons for this: the biggest reason is that a large portion of the forecast updates in these periods were classified as endogenous changes, rather than fiscal shocks. This makes sense given the historical fiscal "rule-of-thumb" (from 1985-2005) that 1pp increases in the output gap corresponded to 1pp higher deficits/GDP. Other reasons are that both periods saw a series of shocks across time, rather than having one isolated big "fiscal event", and that much of the news about deficits coming out in those periods was about the next year or two of deficits, while the Trump elections and Senate results each had longer-term fiscal implications.

The two biggest downward revisions in deficit forecasts (revising towards less expansionary expectations) come from the January 2019 release of new CBO forecasts and the June 2025 release of two separate CBO reports — one providing scoring for Trump's "One Big Beautiful Bill" (OBBB), one providing scoring for Trump's tariff policy at the time. As is argued by Gomez Cram, Kung and Lustig 2025, it makes sense for private-sector forecasts to respond to CBO news, given that the CBO has an informational advantage in scoring bills. Deficit forecasts becoming *less* expansionary in response to OBBB and tariff scoring may seem surprising, but the forecaster output explicitly notes that the scoring of OBBB was significantly lower ($1.4T) than a previous CBO score, based on intervening legislative developments, and that the tariff scoring implied more deficit reduction than previous private sector estimates ($0.8T).

3.5 Relevance of fiscal shocks for fiscal realizations

As a final validation test, we demonstrate that the classifier model's identified fiscal shocks — not just overall forecast changes — are significant predictors of future realized deficits. To test this, we regress realized 1-year ahead deficits on the cumulative change in the model forecast (of 1-year ahead deficits) attributable to fiscal shocks from the start to end of each month. That is, we cumulate just the change in the model forecast due to fiscal shocks, not the change due to endogenous factors, in order to test whether the shock series is a relevant predictor of future fiscal outcomes.

Table 5 shows the results. The β coefficient is highly significant and almost exactly 1, implying that fiscal shocks pass, on average, one-for-one into future realized deficits. The low R^2 value is due to the fact we are only looking at one-month of changes and excluding endogenous changes: most of the variance in one-year-ahead deficits is not due to the fiscal shocks within a given month, but those fiscal shocks are highly significant predictors of

Table 5: Predictive Regression of Realized Deficits on LLM Fiscal Shocks

Horizon	Intercept	β	R^2	N
1-Year Ahead	−166.40	1.02***	0.039	237
	(30.91)	(0.33)		

Notes: Standard errors in parentheses. $^{***}p < 0.01$.

future deficits.

4 Results: the effects of fiscal shocks

We now turn to examining the effects of fiscal shocks. First, their effect on asset prices and then their effect on macroeconomic variables.

4.1 Asset prices

To estimate the effect of fiscal shocks on asset prices, we regress asset price changes from the day before to h days after on the fiscal shock:

$$\Delta y_{t-1,t+h} = \alpha + \beta_h \cdot \Delta \text{PDV}_t^{\text{fiscal}} + \gamma_h \cdot \Delta \text{PDV}_t^{\text{endog}} + \varepsilon_t \tag{7}$$

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 clustered by event.

Table 6 shows the results for $h = 1$ (from the day-before to the day-after the event) and $h = 10$, across eight different assets. The first two columns restrict the sample to only fiscal event days, the third column includes both days and the endogenous control.

Focusing first on the fiscal event only sample, at both the $h = 1$ and $h = 10$ horizons the S&P is about 1% higher than the day before the event — an economically significant move — but the estimated rise is not statistically significant. The Fed's broad US dollar index appreciates in both an economically and statistically significant fashion at both horizons — a 1.3% appreciation a day after the shock, and a 2.3% appreciation 10 days after the shock.

Table 6: Asset Price Response to Fiscal Shocks (PDV)

Asset	$\beta_{h=1}$ (2-day)	$\beta_{h=10}$ (11-day)	$\beta_{h=10}$ (11-day, endog. ctrl.)
ln(S&P 500)	1.34	0.87	1.22
	(0.83)	(1.30)	(1.20)
ln(FX Broad)	1.27***	2.34**	2.27**
	(0.45)	(1.18)	(1.15)
2Y Treasury Yield	4.58**	17.88***	21.71***
	(2.13)	(5.08)	(5.17)
2Y TIPS	1.01	15.37*	16.24*
	(5.22)	(8.94)	(8.83)
2Y Breakeven	3.64	3.04	5.84
	(4.63)	(6.10)	(7.67)
10Y Treasury Yield	13.04*	31.93***	36.69***
	(6.73)	(9.72)	(7.71)
10Y TIPS	5.98*	18.66***	19.76***
	(3.31)	(6.77)	(6.08)
10Y Breakeven	7.23	13.52***	17.13***
	(4.90)	(5.18)	(4.73)
N (events)	116	48	66

Notes: Coefficients and standard errors multiplied by 100.
Standard errors clustered by event in parentheses.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

An appreciation in response to a fiscal expansion is consistent with standard open-economy macro models, but in contrast to some previous empirical work (Monacelli and Perotti 2006, Auerbach and Gorodnichenko 2016).

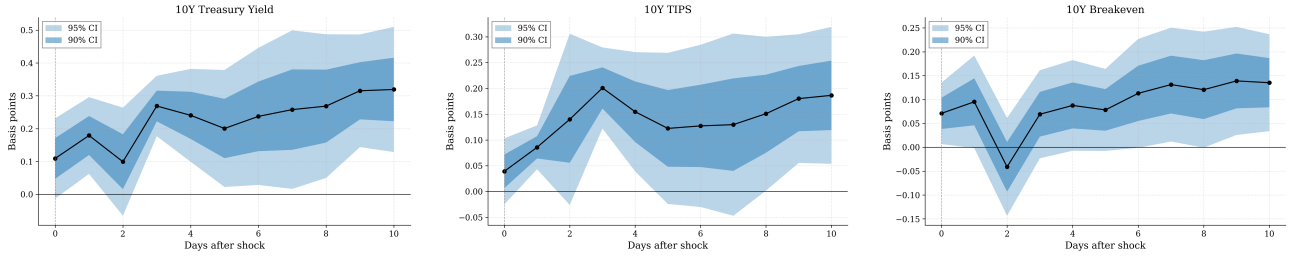
At the $h = 1$ horizon, all of the ten and two-year treasury yield, real rate, and break-even inflation rate rise, with the movement in the two-year yield, the ten-year yield, and the ten-year real rate showing varying levels of significance. At the eleven-day horizon movements are significantly larger and more significant: the two-year yield is 18 basis points higher, primarily due to a rise in real two-year yields, while the ten-year yield is 32 basis points higher, with 58% of the movement coming in real yields and the remainder in break-even inflation. The gap between the ten and two year response implies an increase of 14bps in the term-premium.

The results with "both" days included are very similar: the main difference is that both the ten-year and two-year Treasury responses are slightly larger — 37bps and 22bps respectively.

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

$$y_{t+k} - y_{t-1} = \alpha_k + \beta_k \cdot \Delta \text{PDV}_t + \varepsilon_{t,k} \quad \text{for } k = 0, 1, \ldots, K \tag{8}$$

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 clustered by event. Figure 7 shows the IRF for 10-year Treasury yields, decomposed into the real rate (TIPS) and inflation expectations (breakeven) components.



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

Figure 7: Impulse response of 10-year rates to a 1pp increase in the PDV of expected deficits (as share of GDP). Shaded regions show 68% and 95% confidence intervals. Sample restricted to clean fiscal events with no other shocks in the estimation window.

All three assets yields jump on the news of a deficit expansion, and then rise further in the days afterward. The fact that a large portion of the yield rise occurs not on the day of the event but in the next few 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).

4.1.1 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. To examine this, we split events based on the *weighted horizon* of the fiscal shock:

$$\text{Weighted Horizon} = \frac{\sum_n n \cdot |\Delta d_{t,n}^{\text{fiscal}}|}{\sum_n |\Delta d_{t,n}^{\text{fiscal}}|} \tag{9}$$

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 is 3.7 years, yielding 32 short-term and 32 long-term events.

Table 7: Asset Price Response by Shock Horizon

	2-day window		11-day window	
Asset	Short-term	Long-term	Short-term	Long-term
ln(S&P 500)	1.55	2.33***	0.14	1.21
	(4.49)	(0.87)	(11.83)	(1.04)
ln(FX Broad)	0.37	1.18***	−0.61	2.36**
	(1.06)	(0.46)	(4.01)	(1.16)
2Y Treasury Yield	−0.22	6.06***	−32.17	22.70***
	(7.43)	(2.09)	(35.90)	(4.49)
2Y TIPS	4.78	−4.19	60.82	16.04*
	(25.47)	(5.80)	(95.74)	(8.58)
2Y Breakeven	−4.14	10.26*	−94.73	7.21
	(29.48)	(5.44)	(117.63)	(7.04)
10Y Treasury Yield	2.66	18.56***	−43.73	37.86***
	(19.73)	(4.58)	(47.47)	(7.98)
10Y TIPS	2.00	7.71***	−53.23	22.38***
	(14.40)	(2.56)	(38.73)	(5.17)
10Y Breakeven	1.09	11.24***	8.89	15.72***
	(17.67)	(2.88)	(33.80)	(5.38)
N (events)	76	76	32	32

Notes: Coefficients and standard errors multiplied by 100. Endogenous controlled. Standard errors clustered by event in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 7 shows the results. The pattern is striking: long-term shocks drive all of the significant asset price responses, while short-term shocks show large standard errors and no significant effects. Relative to the fiscal only event sample not split by weighted horizon, two-day asset price responses are far more significant when looking only at longer-horizon shocks: the S&P jumps a significant 2.3%, and the response of all the different ten-year assets are highly significant. At the eleven-day horizon, the ten-year is 38bps higher, with a 22bps increase in real yields and a 16bps rise in breakevens.

4.1.2 Commentary

It is worth taking a moment to compare our results on how deficit expansions 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 13-19bps and $h = 10$ of 32-38bps. Our results on the split between real yields and breakevens is also similar to hers: in our preferred estimates, 53-60% 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 PDV(deficit)/PDV(GDP) will give rise to an $\approx$ 10pp rise in expected ten-years later debt/GDP.[8] 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) — 3-4bps 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 real yields (at the ten-year horizon) are a little more than half of that — 1.9-2.2bps. 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 expectations is consistent with fiscal theory of the price level models with long-term debt (i.e. 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.

Most important for theory are the results split by weighted-horizon of the shock. Shocks that are only expected to affect near-term deficits are not priced by investors, while long-term shocks are. Taken at face value, we interpret these results as more consistent with the predictions of FTPL-style models than HANK models. However, exactly how HANK models respond to deficit shocks of different horizons is an area of ongoing research, so we hope these results provide impetus for more clarification of this issue (i.e. Angeletos, Lian and Wolf 2024*b*).

[8] A strong assumption.

4.2 Macroeconomic variables

To examine the effects of fiscal expectation shocks on macroeconomic outcomes, we estimate monthly and quarterly local projections following Jordà (2005). For each outcome variable y, we estimate:

$$y_{t+h} - y_{t-1} = \alpha_h + \beta_h \cdot \text{shock}_t + \sum_{\ell=1}^{p} \gamma_{h,\ell} \cdot \text{shock}_{t-\ell} + \mathbf{X}_t' \boldsymbol{\delta}_h + \varepsilon_{t+h} \tag{10}$$

where shock_t is the cumulative fiscal event component of the expectation shock (within the month or quarter) —constructed using smooth forecast weights that blend current-year and year-ahead forecasts based on the fiscal calendar. We work with year-ahead deficit expectations rather than the full PDV shock since they are a stronger instrument for realized deficits, which is more important in a local projection analysis. The control vector $\mathbf{X}_t$ includes the cumulative endogenous component of the shock (with contemporaneous value and lags 1 through p), dummy variables for the 2008-2009 global financial crisis and the 2020 COVID-19 period, and seasonal (quarterly or monthly) fixed effects. We use HAC-robust standard errors with lag length equal to the horizon h. For monthly outcomes, we use $p = 12$ lags; for quarterly outcomes, $p = 4$ lags.

Due to our relatively short-sample (2005-25) and the confounding presence of both the GFC and the Covid crisis in that sample, we do not want to over-interpret the results of our local projections. We mostly see it as a proof of concept for what this methodology can do, but extending our sample farther back in time — which we plan to do — will be critical for reliable inference.

Instrument validity. A key concern in using expectation shocks as instruments is whether they have sufficient predictive power for realized fiscal outcomes. One way of demonstrating the *potential* of our methodology for estimating the effects of fiscal shocks is showing that it creates relevant instruments. Table 8 reports first-stage F-statistics from regressions of cumulative realized deficit changes (from $t-1$ to $t+h$, as % of GDP) on the expectation shock and its lags. The relevant horizon is $h = 4$ quarters for the quarterly specification (aligning with the 1-year forecast horizon) and $h = 12$ months for the monthly specification.

The full sample quarterly F-statistic of 5.35 falls below the conventional weak instrument threshold of 10, raising concerns about instrument strength. However, this weakness is driven almost entirely by the timing of outlays in the COVID-19 period. The Q2 2020 deficit reached 37.6% of GDP (annualized) while the cumulative fiscal shock (excluding the endogenous shock) was only about 2 pp. This disconnect is due in large part to the lumpy timing of the Covid stimulus, which was almost entirely concentrated in Q2, while our forecasts

Table 8: First-Stage F-Statistics: Do Expectation Shocks Predict Realized Deficits?

	Quarterly ($h = 4$)		Monthly ($h = 12$)	
	F-stat	R^2	F-stat	R^2
Full sample (2005–2025)	5.35	0.31	18.83	0.39
Excluding 2020–2021	20.58	0.63	14.13	0.22

Notes: F-statistics test joint significance of the expectation shock and its lags in predicting cumulative realized deficit changes from $t - 1$ to $t + h$ (as % of GDP). The weak instrument threshold is typically 10. Sample starts January 2005 when LLM forecast data begins.

are at the fiscal-year horizon. When we exclude 2020–2021, the quarterly F-statistic rises to 20.58, well above the weak instrument threshold. The F-stats for the monthly local projection are encouragingly high in both the full sample and the restricted sample.

Full sample results. Figure 8 shows the response of core CPI, headline CPI, and the University of Michigan Consumer Survey measure of 1-year inflation expectations to a 1 percentage point increase in year-ahead expected deficits (as a share of GDP). Both core and headline CPI are measured in logs.

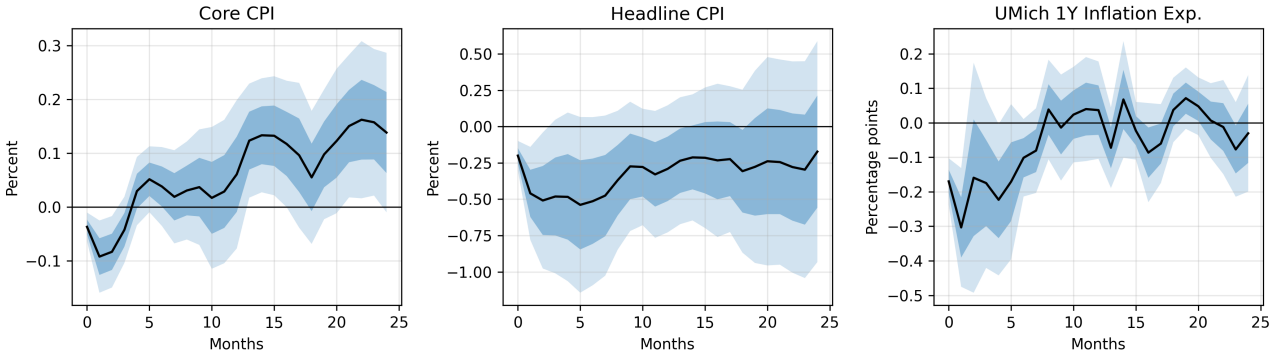


Figure 8: Response of inflation measures to a 1pp increase in expected deficits (as % of GDP). Left: Core CPI (log change × 100). Center: Headline CPI (log change × 100). Right: UMich 1-year inflation expectations (percentage points). Shaded regions show 68% and 95% confidence intervals using HAC standard errors. Specification controls for endogenous shock component and includes crisis dummies for 2008-2009 and 2020.

All three measures fall in response to the shock. Headline inflation remains persistently low, core inflation rises above trend after about five months, and inflation expectations recover to their previous level after about seven months.

Figure 9 presents the response of quarterly real activity measures.

Once again, all three measures fall below trend in response to the shock, with consumption and GDP recovering about eight quarters later, but never moving significantly above trend. Investment remains below trend out to 15 quarters later.

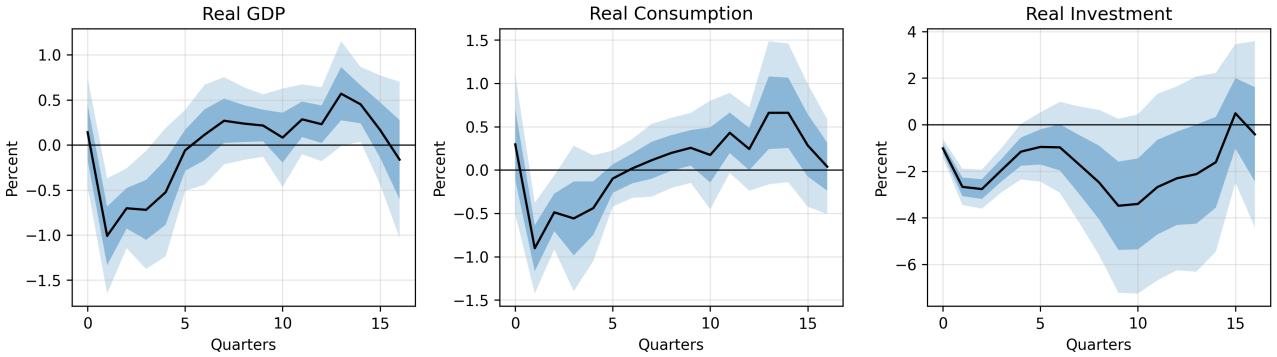


Figure 9: Response of real activity to a 1pp increase in expected deficits (as % of GDP). Left: Real GDP. Center: Real consumption. Right: Real investment. All outcomes are in percent (log change × 100). Shaded regions show 68% and 95% confidence intervals using HAC standard errors. Specification controls for endogenous shock component and includes crisis dummies for 2008-2009 and 2020.

As mentioned, the large declines in output and inflation in the GFC and 2020 dominate the variation in our small sample, likely contributing to the counter-intuitive result that fiscal expansions lead to declines in activity and inflation.

Figures 12 and 13 in the appendix show that if we simply remove both periods from our shock sample, the response of all variables is highly positive and significant. However, since this concentrates our sample on the recovery periods and removes the recession periods, we do not put too much weight on these results.

A slightly more "fair" test is removing the large outlier that was the Covid crisis. To do so, we exclude all observations in our sample where either the shock variable *or* the outcome variable was measured in 2020 or 2021. Figures 10 and 11 present the results.

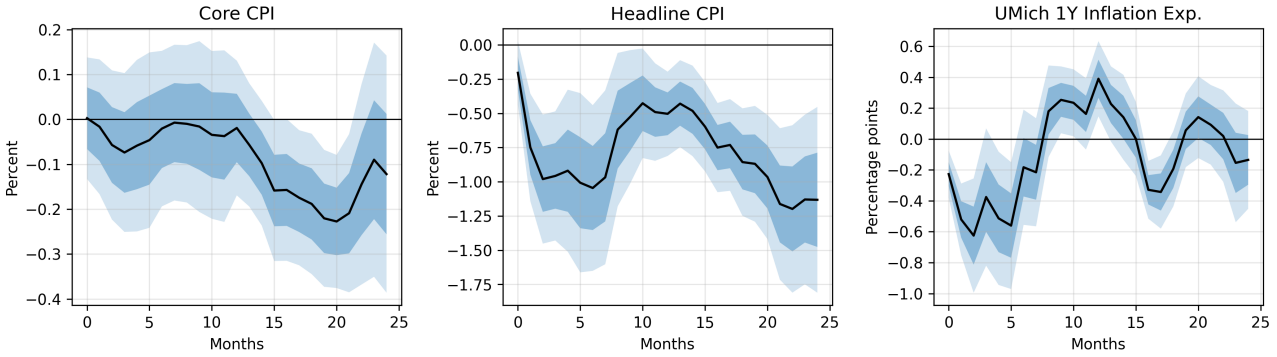


Figure 10: Response of inflation measures to a 1pp increase in expected deficits, excluding 2020–2021. Left: Core CPI. Center: Headline CPI. Right: UMich 1-year inflation expectations. Specification controls for endogenous shock component and includes crisis dummies for 2008–2009. F-statistic at $h = 12$: 14.13.

Even with Covid out of the sample, the responses are not what conventional theory would

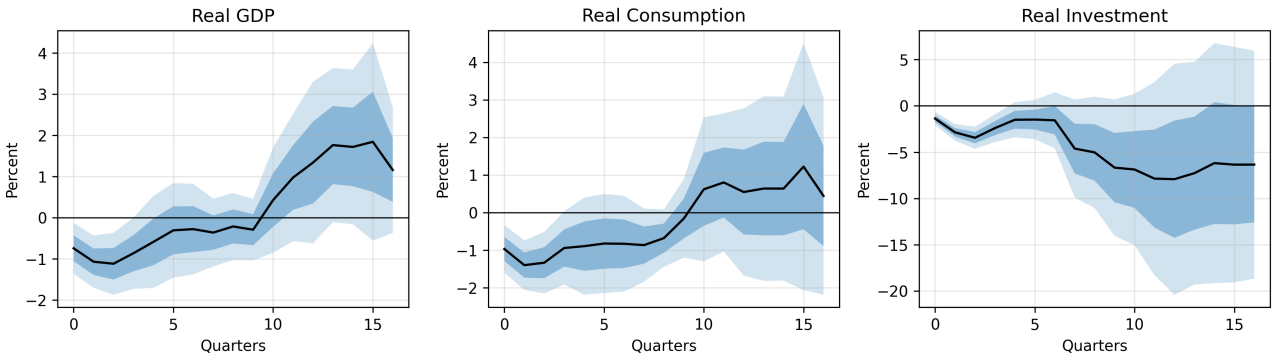


Figure 11: Response of real activity to a 1pp increase in expected deficits, excluding 2020–2021. Left: Real GDP. Center: Real consumption. Right: Real investment. Specification controls for endogenous shock component and includes crisis dummies for 2008–2009. F-statistic at $h = 4$: 20.58.

predict: inflation falls below trend (immediately for headline, with a delay for core), and real activity sits below trend for about two years after the shock (the full four years for investment). It is fair to worry that the GFC is dominating the sample, but the controls for endogenous shocks should somewhat account for negative activity responses to the GFC, so while we do not put too much weight on these results, we do find them suggestive. We leave a more thorough exploration of the effects of fiscal shocks on activity to future work.

5 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. We are working on extending our results to other countries and time-periods. 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.

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 next eleven-years deficits, with real yields accounting for 53-60% of the rise and market break-even inflation expectations accounting for the remainder. The dollar appreciates significantly — 2.3% — and the term-premium, as measured by the 2Y-10Y spread, rises 14-15bps. 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 are priced by investors, while short-term shocks have insignificant effects. In our sample, fiscal shocks have contractionary effects on realized activity — we are working on our extending our sample to cover a longer time horizon.

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.

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

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

Jordà, Òscar. 2005. "Estimation and Inference of Impulse Responses by Local Projections." *American Economic Review*, 95(1): 161–182.

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.

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.

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

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. 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.

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

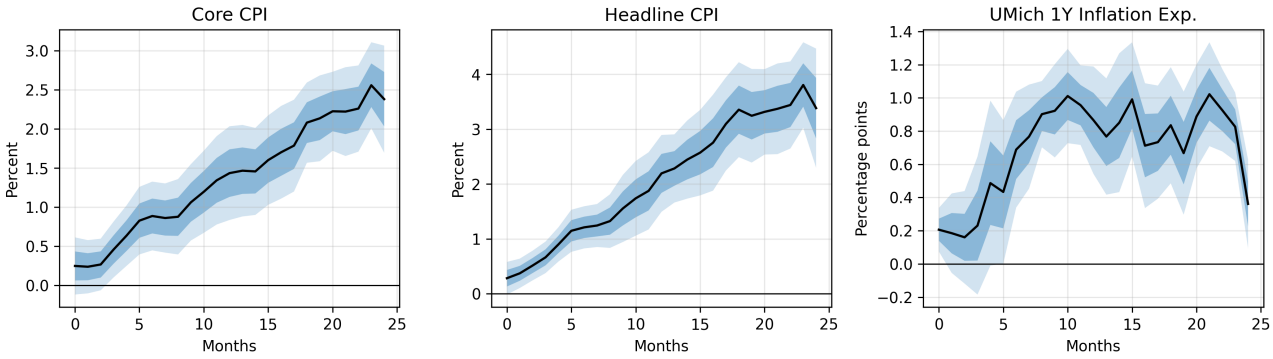


Figure 12: Response of inflation measures to a 1pp increase in expected deficits, excluding GFC (2008–2009) and 2020. Left: Core CPI. Center: Headline CPI. Right: UMich 1-year inflation expectations. Specification controls for endogenous shock component. F-statistic at $h = 12$: 17.18.

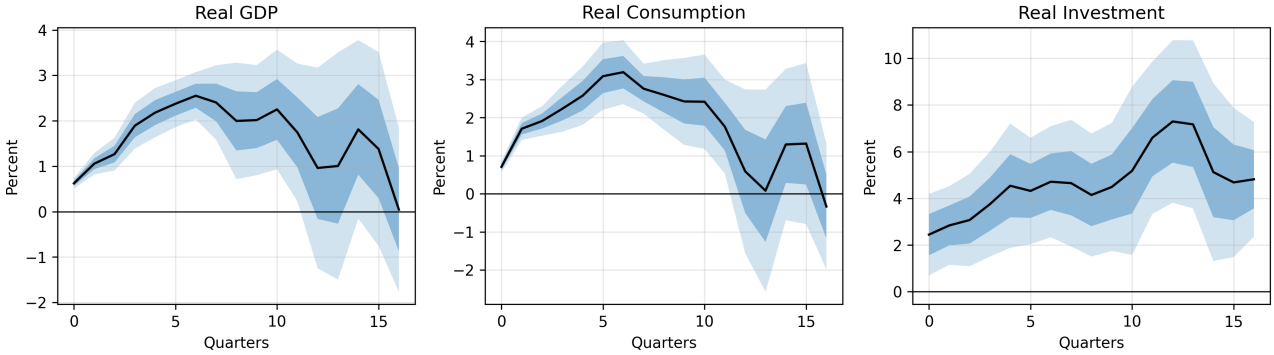


Figure 13: Response of real activity to a 1pp increase in expected deficits, excluding GFC (2008–2009) and 2020. Left: Real GDP. Center: Real consumption. Right: Real investment. Specification controls for endogenous shock component. F-statistic at $h = 4$: 65.39.

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.2). 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 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.3 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.4 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 the forecast update based on the following framework:
The expected deficit in a given fiscal year can be modeled as follows:
d = a + b*o_gap + epsilon
where:
- d is the deficit
- a is the intercept, which represents current expectations about the long-run
average deficit.
- b is the slope which represents the "fiscal rule" of the government - how the
government responds to economic conditions.
- o_gap is a measure of something like the output gap or the unemployment rate
gap.  The idea is that when o_gap is negative, output is below the natural rate,
and the government will need to run a deficit to stimulate the economy.
- epsilon is the error term which represents a "fiscal shock" - a deviation
of expected spending from what would typically be expected based on the fiscal
response to economic conditions.
Whenever the forecast of future fiscal deficits has changed for a given fiscal
year, classify the forecast update according to the above framework.  Any of
the four RHS variables can change, or a combination of them can change.  If a
combination of variables changes, quantify the magnitude of the change that is
endogenous, and the magnitude of the change that is a fiscal event.  For example,
```

if the change in forecast is 20B, split that into XB that is endogenous and YB that is a fiscal event, where X + Y = 20.
Do this separately for each fiscal year.
- If the long-run average deficit has changed, classify as a "fiscal event".
- If the fiscal rule has changed, classify as a "fiscal event".
- If the change in future deficits is a predictable response to a changing output gap, classify as "endogenous". This represents a change in the economic conditions that the government is responding to, not a shock to fiscal policy.
- If the fiscal shock has changed, classify as a "fiscal event".
- Please be explicit about the nature of the change - both whether it is a fiscal event or an endogenous change, and if it is an endogenous change, whether it is a change in the long-run average deficit, a change in the fiscal rule, or a pure fiscal shock.
</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 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:
- Focus on the forecaster's stated reasoning, not your own interpretation of the news
- Be conservative with "fiscal_event" - reserve for clear policy announcements or political shocks that deviate from what would typically be expected based on the fiscal rule.
- Remember that even if something is a policy announcement, it may not be a fiscal event if it is not a surprise relative to the policy that would be expected based on economic conditions (the fiscal rule).

- If values didn't change, classification MUST be "none" regardless of rationale
- When an event is classified as "both", think carefully about to what extent fiscal stimulus (or contraction) is simply responding to economic conditions, versus acting in an unusually expansionary or contractionary manner. Do not just default to a 50/50 split.
- VERY IMPORTANT: Historically in the US a 1pp rise in the output gap leads to a 1pp rise in the deficit/GDP ratio. This is *not* all due to automatic stabilizers - a large chunk of it is due to the *predictable* pattern whereby discretionary fiscal spending rises in response to deterioriating economic conditions. Even if this fiscal policy is ''discretionary'', it is part of the fiscal rule, since it is part of the expected response of the government to economic downturns. When deciding whether a policy announcement is endogenous or fiscal news, think explicitly about what the current economic news implies about future output gaps, and whether the announced discretionary fiscal policy is more or less expansionary than the historical norm. Only if it deviates from the historical norm should it be considered a fiscal event.
- 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.
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 response is greater or less than what would typically be expected based on the fiscal rule.
- "Tax receipts/revenues..." → likely endogenous
- "Spending bill..." or "Budget agreement..." → this could be a fiscal_event, but only if the response is greater or less than what would typically be expected based on the fiscal rule.
- "No new information" or unchanged values → none
</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, including the
sub-classification if applicable (long-run average deficit, fiscal rule, output
gap, or fiscal shock).  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.
Note:  For each fiscal year, endogenous_change + fiscal_event_change should equal
the total change in that year's forecast from yesterday.  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'.
</output>
```

B.5 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>