A Causal Approach to Forecasting Central Bank Decisions

Matthieu Garcin* Martino Grasselli † Mathis Guenet‡

January 7, 2026

We propose a causal framework to assist market participants in forecasting cumulative adjustments to the Federal Open Market Committee (FOMC) target rate by integrating market-implied expectations from the CME FedWatch tool with a broad set of U.S. macroeconomic indicators. In this paper, we make three contributions. First, we show the existence of a regime shift: at short horizons, market-implied expectations fully capture the available information, making macroeconomic variables redundant, while at longer horizons, macroeconomic variables provide substantial additional predictive value beyond market expectations. Second, we introduce a causal discovery framework that identifies horizon-specific direct drivers reflecting the underlying mechanisms of FOMC monetary policy decisions. This approach yields parsimonious and interpretable forecasting models, mitigates overfitting, and enhances out-of-sample performance. Finally, we identify and quantify horizon-specific direct and total causal effects of inflation and market expectations on FOMC decisions, showing that standard regressions conflate these effects and are prone to confounding and mediation biases, while our causal framework delivers well-defined and economically interpretable causal estimates.

JEL classification: C51, C53, C54, E43, E44, E47, E52, E58, G10.

Keywords: Causal inference; Causal discovery; Monetary policy; FOMC decisions; Interest rate forecasting; Market-based expectations; FedWatch tool; Macroeconomic indicators; Stochastic Discontinuities, Target Rate, SOFR.

1 Introduction and Motivation

The Federal Reserve[1] (Fed) is the central bank of the United States and is responsible for the formulation and implementation of monetary policy. Policy decisions are taken by the Federal Open Market Committee[2] (FOMC) and rank among the most closely monitored events in global financial markets. The FOMC meets eight times per year[3] to determine the target range for the federal funds rate, which constitutes the primary instrument of U.S. monetary policy. This target is implemented operationally through the Interest on Reserve Balances (IORB), the overnight rate paid on reserve balances held by commercial banks at the Federal Reserve. In practice, the Effective Federal Funds Rate (EFFR), shown in Figure 1, and the Secured Overnight Financing Rate (SOFR), shown in Figure 2, are the two most widely used measures of U.S. short-term interest rates and typically fluctuate around the IORB. The EFFR reflects unsecured overnight lending between depository institutions and therefore captures conditions in the interbank market, while the SOFR represents the secured overnight borrowing rate in the Treasury repurchase agreement market. Together, these rates provide key indicators of short-term funding conditions and overall market liquidity.

In the financial literature, the dynamics of SOFR and the EFFR are typically modeled as mean-reverting processes fluctuating around the Interest on Reserve Balances (IORB), with a stochastic

*De Vinci Higher Education, De Vinci Research Center, Paris (France). Email: matthieu.garcin@devinci.fr

†Department of Mathematics, University of Padova (Italy) and De Vinci Higher Education, De Vinci Research Center, Paris (France). Email: grassell@math.unipd.it

‡De Vinci Higher Education, De Vinci Research Center, Paris (France) and Scalnyx, Paris (France). Email: mathis.guenet@devinci.fr

[1]https://www.federalreserve.gov/

[2]https://www.federalreserve.gov/monetarypolicy/fomc.htm

[3]For the FOMC calendar see e.g. https://www.federalreserve.gov/monetarypolicy/fomccalendars.htm

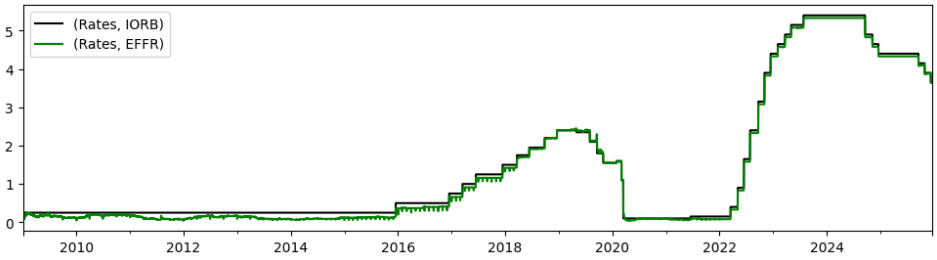


Figure 1: Effective Federal Funds Rate (EFFR) vs Interest on Reserve Balances (IORB) over time (Source: Bloomberg).

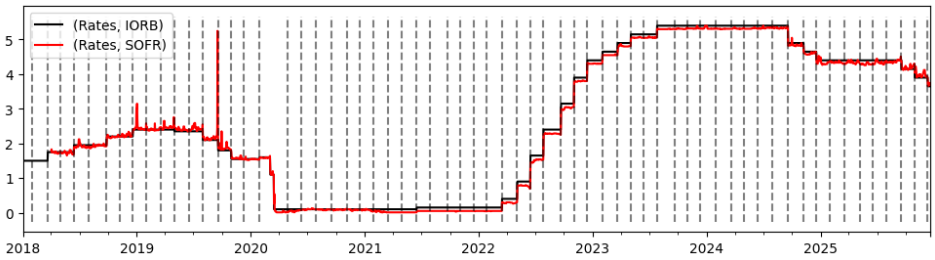


Figure 2: Secured Overnight Financing Rate (SOFR) vs Interest on Reserve Balances (IORB) over time. Grey vertical lines represent dates of Federal Open Market Committee (FOMC) scheduled meetings (Source: Bloomberg).

spread that is often assumed to be independent and Gaussian. By contrast, the IORB itself is commonly represented as a piecewise constant process featuring scheduled jumps corresponding to policy decisions. This structural asymmetry gives rise to the so-called problem of stochastic discontinuities, namely the presence of stochastic processes that exhibit jumps at predetermined dates, which significantly complicates the mathematical treatment of jump dynamics. This issue has long been studied in Gaussian settings. For instance, [Piazzesi, 2005], [Piazzesi, 2010] provides an early attempt to model the Federal Funds target rate accounting for the FOMC meeting calendar. In her model the target rate follows a pure jump process driven by Poisson processes, with jumps during FOMC meetings triggered by elevated jump intensities. [Kim and Wright, 2014] were the first to introduce a model where rates can jump at deterministic times and the jump size is random. More recently, multiple theoretical papers have also studied the modeling of stochastic discontinuities in relation to term structure modeling see e.g. [Keller-Ressel et al., 2019], [Fontana et al., 2020], [Schlögl et al., 2024] and [Fontana et al., 2024].

When adjusting the target range for the federal funds rate, the FOMC operates under a dual mandate of promoting maximum employment and ensuring price stability. More specifically, its objective is to maintain inflation at a low and stable level—articulated by the Federal Reserve as a 2 percent longer-run goal—while fostering conditions consistent with full employment. In practice, the FOMC evaluates a wide array of macroeconomic indicators and financial conditions to inform its policy decisions. These include measures of inflation and inflation expectations, labor market conditions, economic activity and growth, as well as broader financial conditions.

Table 1 reports the correlations between FOMC decisions and selected macroeconomic and financial indicators at different horizons. Consistent with the dual mandate, inflation and the unemployment rate exhibit the strongest correlations with policy decisions. At short horizons, these variables, together with Federal Reserve Bank total assets, show significant correlations with FOMC actions. Other indicators, such as GDP growth, year-on-year inflation measures, and the ISM Manufacturing PMI, while not significantly correlated at short horizons, become increasingly correlated with FOMC decisions as the forecasting horizon lengthens.

Within this data-dependent and forward-looking policy framework, financial markets continuously incorporate incoming macroeconomic releases and evolving financial conditions into asset prices, thereby forming real-time expectations about future FOMC actions. As a consequence, short-term monetary policy expectations have become increasingly transparent, largely owing to the widespread

Variable	1 day	1 month	3 months	6 months
Inflation US	0.57***	0.57***	0.62***	0.64***
Unemployment Rate US	-0.20**	-0.20**	-0.21**	-0.24***
Bloomberg Financial Condition	-0.12	-0.10	-0.10	-0.06
FED Reserve Bank Total Assets	0.32***	0.32***	0.36***	0.40***
GDP YoY	0.07	0.11	0.14*	0.21***
ISM Manufacturing PMI	0.13	0.14*	0.20**	0.24***

Table 1: Correlation between the sum of the FOMC adjustment decision and some examples of macroeconomic and financial indicators at different time horizons from 2009 to 2025. (* denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Source: Bloomberg)

adoption of market-based forecasting tools. Among the most prominent of these are the Chicago Mercantile Exchange (CME) FedWatch Tool [CME-Group, 2025] and the Bloomberg World Interest Rate Probability (WIRP) function [Bloomberg Finance L.P., 2023], both of which are widely used to anticipate FOMC policy rate decisions.

These methodologies rely on the assumption that the price of a 30-day Federal Funds futures contract for a given calendar month reflects the market's consensus forecast of the average effective federal funds rate over that month. By jointly exploiting information from multiple futures contracts—including those corresponding to months without scheduled FOMC meetings—the FedWatch tool and the WIRP function infer implied pre-meeting and post-meeting target rates. This procedure yields a discrete probability distribution over possible policy outcomes, such as rate hikes, cuts, or unchanged rates, for upcoming FOMC meetings. Moreover, because the methodology provides an implied rate adjustment for each forthcoming meeting, it naturally allows for the computation of the cumulative implied rate adjustment as the sum of the implied adjustments from the current date up to the next scheduled meeting. The Atlanta Fed's Market Probability Tracker (MPT) [Atlanta-Fed, 2023] constitutes another market-based forecasting tool, relying on option prices traded on the CME that reference the three-month compounded average Secured Overnight Financing Rate (SOFR). However, because each option contract aggregates market expectations over a three-month horizon rather than being tied to specific FOMC meeting dates, the MPT—while informative about the general policy stance over a quarterly window—does not allow for a precise estimation of post-meeting implied rate adjustments or cumulative adjustment probabilities at individual FOMC meetings. In addition, SOFR-based derivatives of this type have only been available since 2023, which limits the historical depth of the data and prevents the methodology from being applied consistently over longer time periods.

There also exist alternative approaches to forecasting FOMC policy rate decisions that do not rely on market-based expectations, but instead exploit large language models (LLMs) and natural language processing (NLP) techniques applied to the textual analysis of FOMC minutes and statements. Examples include:

- [Lee et al., 2021], which develops an interpretable multi-component NLP framework to decode Federal Reserve communications, producing sentiment indicators, document summaries, and predictions of movements in the federal funds rate;
- [Seok et al., 2024], which employs a multi-agent LLM architecture that facilitates interactions among simulated agents (e.g. voting members and economists) to generate projections of the federal funds rate at specific FOMC meeting dates, including longer maturities.

While informative from a communication and sentiment-analysis perspective, these approaches do not focus on market-based expectations or realised macroeconomic indicators. Given our objective of combining market-implied information with macroeconomic fundamentals to predict cumulative FOMC adjustments at specific horizons and to produce meeting-specific forecasts—while maintaining a sufficiently large historical sample of meetings—we therefore adopt a futures-based expectations framework in the spirit of the FedWatch and WIRP methodologies in the subsequent analysis.

Table 2 reports the correlation between the cumulative implied rate adjustments produced by the FedWatch tool and the corresponding cumulative FOMC adjustment decisions at different horizons.

The correlations are particularly high at short horizons, indicating that the FedWatch tool captures a substantial share of the information relevant for near-term policy decisions. This evidence suggests that FedWatch is capable of delivering high predictive accuracy in the short run. Consistent with this observation, recent empirical work shows that the FedWatch tool can correctly anticipate FOMC rate adjustments with accuracy levels of up to 88% within one month of a meeting date [Bonini et al., 2025]. In addition, Table 3 indicates that the corresponding regression framework achieves an R^2 as high as 0.93 within 30 days of a meeting date.

	1 day	**1 month**	**3 months**	**6 months**	**9 months**
Implied change rate	0.99***	0.95***	0.84***	0.62***	0.31***
Cumulative implied change rate	0.99***	0.95***	0.92***	0.84***	0.75***

Table 2: Correlation between the FedWatch Tool (cumulative) implied rate change and the (cumulative) realised FOMC adjustment decisions at different horizons from 2009 to 2025. (* denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level. Source: Bloomberg)

	1 day	**1 month**	**3 months**	**6 months**	**9 months**
Implied change rate	0.99	0.91	0.65	0.33	0.07
Cumulative implied change rate	0.98	0.91	0.83	0.69	0.50

Table 3: R^2 obtained by using the FedWatch Tool (cumulative) implied change rate to predict the (cumulative) realised FOMC adjustment decision at different horizons from 2009 to 2025. (Source: Bloomberg)

Furthermore, Table 2 also indicates that the correlation between the FedWatch-implied and realised cumulative FOMC adjustment decisions remains significant at longer horizons, albeit at progressively lower levels. As the forecast horizon extends, the strength of this correlation declines. A similar pattern emerges in Table 3, where the predictive power of the FedWatch tool—while exceptionally high at short horizons—gradually diminishes as the horizon increases.

The high short-term predictive accuracy of cumulative FOMC adjustment decisions indicates that near-term policy actions are largely anticipated by markets and therefore contain little element of surprise. In practice, a broad set of information—including macroeconomic releases, forward-looking market signals, and FOMC communications—appears to be efficiently reflected in the futures prices that underpin the FedWatch tool and the Bloomberg WIRP function. As the forecast horizon extends, however, uncertainty surrounding future policy decisions naturally increases, and the cumulative implied rate adjustments inferred from FedWatch progressively lose informational content.

This naturally raises the question of whether macroeconomic variables provide incremental predictive power beyond that already embedded in market-based expectations, particularly at longer horizons and for more distant FOMC meetings. Accordingly, the objective of this study is to identify the forecast horizons at which macroeconomic indicators add meaningful value relative to market-implied expectations derived from tools such as FedWatch and the Bloomberg WIRP function.

To address these challenges, we adopt a causal framework aimed at clarifying the underlying mechanisms driving FOMC adjustment decisions while mitigating the limitations inherent in macroeconomic datasets. Such datasets are intrinsically high-dimensional, yet their effective sample size is small—particularly in the context of FOMC decisions, where the number of observations is constrained by the frequency of scheduled meetings. This setting necessitates remaining within a linear modeling framework in order to limit the risk of overfitting.

In addition, macroeconomic variables, market-based indicators such as the FedWatch tool, and FOMC decisions exhibit strong dependencies and interdependencies, as evidenced by the high correlations reported in Figure 7 for different forecast horizons. These relationships often give rise to redundant information and increase the likelihood of identifying spurious correlations rather than genuine causal effects. A causal approach is therefore essential to disentangle these interactions and to isolate the variables that exert a direct and economically meaningful influence on policy decisions.

As discussed in [Pearl, 2009, Angrist and Pischke, 2009, Pearl and Mackenzie, 2018, Neal, 2020]

and [MacKinnon and Lamp, 2021], empirical settings of this type are particularly susceptible to several well-known sources of bias, including confounder bias [Rubin, 1974, Rosenbaum and Rubin, 1983, Imbens and Rubin, 2015, Pearl, 2022a, Hernán et al., 2011], mediator bias [Pearl, 2012, Richiardi et al., 2013, Pearl, 2022b, Rijnhart et al., 2021], and collider bias [Cole et al., 2010, Lu et al., 2022, Munafò et al., 2018, Hernán and Monge, 2023]. These biases typically arise when the regression model is misspecified, an issue that has recently been characterised in quantitative finance and econometrics as undercontrolled or overcontrolled regression [Lopez de Prado, 2023, Lopez de Prado et al., 2024, Lopez de Prado et al., 2025]. The specific mechanisms underlying these biases are discussed in detail in the next section.

As a consequence, the resulting linear regression coefficients may yield incorrect estimates of causal effects, erroneously attributing predictive power to variables that merely reflect correlation structures present in the training sample rather than stable causal relationships with FOMC decisions. When the joint distribution of macroeconomic indicators shifts over time or across economic regimes, such spurious associations are unlikely to persist, leading to unstable predictions, degraded out-of-sample performance, and potentially misleading economic interpretations.

For these reasons, this study adopts a graphical causal framework based on causal discovery methods to first identify the underlying causal structure linking macroeconomic indicators, the FedWatch tool indicator, and FOMC policy decisions. The inferred causal graph allows us to characterise the full set of conditional independencies and causal relationships among the variables. Building on this structure, we then derive the most parsimonious set of genuine causal drivers—that is, the minimal yet most informative subset of macroeconomic indicators that is free from spurious correlations. This approach enables the construction of predictive models that deliver more coherent causal interpretations and improved out-of-sample performance in forecasting FOMC decisions across different horizons.

The paper is organised as follows. Section 2 introduces the causal framework and the associated terminology, clarifying its key features and distinguishing it from the concept of Granger causality, which may otherwise lead to misunderstandings. In particular, Granger causality captures a form of statistical association akin to correlation, whereas we explain how a proper definition of causality allows one to move beyond mere associations. We present the tools used to recover the causal graph, with a particular focus on two approaches: the Peter–Clark (PC) algorithm and the Linear Non-Gaussian Acyclic Model (LiNGAM). Section 3 presents the empirical application of the proposed causal framework and the associated results, while Section 4 explains why a causal approach is essential for interpreting these findings, showing how causal graphs both mitigate overfitting through parsimonious variable selection and provide a principled distinction between total and direct causal effects that is robust to confounding, mediation, and collider biases. Section 5 concludes the paper. The appendix reviews selected foundational concepts in causal inference that may be helpful for a general understanding of the proposed approach.

2 The Causal Approach

Statistical inference seeks to draw conclusions about the properties of an underlying probability distribution based on observed samples assumed to be representative of that distribution. In applied settings, such as econometrics, this often involves analysing joint distributions to assess whether variables are statistically associated—namely, whether observing the value of one variable conveys information about another. On the basis of such associations, predictive models may then be constructed to forecast the value of a variable conditional on others with which it is correlated.

Causal inference, by contrast, constitutes a distinct branch of statistical inference that aims to identify and quantify causal relationships between variables, that is, to determine whether changes in one variable generate changes in another. While predictive accuracy is important, econometric analysis often places greater emphasis on uncovering cause–effect relationships and estimating causal effects, as these are essential for evaluating the impact of policy interventions and for understanding how changes in one variable propagate to others.

Within this perspective, causal inference—and in particular the graphical causal framework—provides a principled means of identifying the most parsimonious yet informative subset of variables for predicting an outcome Y. By construction, this subset is free from spurious correlations and redundant information. As a result, researchers can construct predictive models that are both parsimonious and maximally informative, grounded in genuine causal relationships rather than misleading statistical associations.

2.1 Correlation vs Causality

In empirical analysis, it is important to distinguish clearly between correlation and causality. Correlation describes a statistical association between two variables, indicating that they tend to move together in a systematic way. Such an association, however, does not by itself provide information about the underlying mechanism generating the observed dependence. Correlated variables may be linked through a direct relationship, driven by a common underlying factor, or connected indirectly through more complex interactions. Causality, by contrast, refers to a directional cause–effect relationship, whereby changes in one variable produce changes in another as a consequence of an underlying mechanism. Establishing causality therefore requires going beyond the observation of statistical association and addressing questions related to directionality, intervention, and counterfactual outcomes. While correlation is a necessary condition for causality, it is not sufficient: causal relationships cannot be inferred from association alone without additional structural assumptions or methodological tools designed to disentangle genuine cause–effect links from spurious dependencies, see e.g. [Holland, 1986, Angrist and Pischke, 2009, Pearl and Mackenzie, 2018].

2.2 Granger Causality as Temporal Association

Granger causality provides one of the earliest and most widely used notions of causality in time series analysis. Introduced by [Granger, 1969, Granger, 1980], it defines a variable X as "causing" another variable Y if past values of X contain incremental predictive information about Y beyond that contained in past values of Y alone. This concept has been extensively applied in econometrics and finance, particularly within linear vector autoregressive frameworks, and remains a standard tool for assessing predictive relationships in time series data. However, as emphasized by [Pearl, 2009], despite its terminology, Granger causality does not correspond to causality in the structural or interventionist sense. Rather, it captures a form of temporal associativity or predictive dependence: it identifies whether one time series helps forecast another, not whether changes in one variable would generate changes in the other under an intervention. As such, [Kilian, 2006, Peters et al., 2017] emphasize that Granger causality is sensitive to omitted variables and thus cannot reliably distinguish direct causal effects from indirect or spurious relationships; in addition, its sensitivity to model specification, lag length, and contemporaneous correlations further limits causal interpretation. Consequently, while Granger causality is valuable for detecting predictive temporal structure, it should be interpreted as a measure of temporal association rather than evidence of genuine cause–effect relationships.

2.3 Regression Bias

In regression analysis, parameter estimates can be systematically biased when the model fails to account properly for the underlying causal structure, particularly in the presence of mediators, colliders or unobserved and omitted factors. Confounder bias arises when two variables appear statistically related because they are both influenced by a third unobserved or omitted variable. In this case, this variable is called a confounder and the regression incorrectly attributes the spurious association induced by this confounder to the variable of interest (see Figure 9 in Appendix). Mediation bias occurs when a third variable lies on the causal pathway between an explanatory variable and the outcome: conditioning on this mediator partially or fully blocks the causal effect, removing relevant information and causing the estimated coefficients to no longer reflect the true causal relationship (see Figure 8 in Appendix). Collider bias, by contrast, emerges when two otherwise independent variables become statistically associated as a result of conditioning on a collider, that is a third variable that is jointly affected by both, thereby inducing spurious correlations driven purely by selection effects (see Figure 10 in Appendix). In all these cases, standard regression techniques may yield misleading parameter estimates, as they conflate genuine causal effects with artefacts generated by inappropriate conditioning or omitted variable structures, see e.g. [Cinelli et al., 2019, Cinelli and Hazlett, 2020, Chernozhukov et al., 2022, Lopez de Prado et al., 2024, Lopez de Prado et al., 2025, Lopez de Prado, 2023].

2.4 Causal Graphs

The causal graphical framework, originally introduced by Pearl in [Pearl, 1993, Pearl, 1995] and fully developed in [Pearl, 2009], provides a graphical alternative to the potential outcomes framework

[Rubin, 1974]. It offers a formal representation of cause–and–effect relationships by means of a causal graph and an associated structural causal model, which together encode the data-generating process, namely the causal mechanisms through which the observed data arise.

Causal graphs [Pearl, 2009] constitute a fundamental component of this framework, as they provide a graphical representation of the causal relationships among variables. A causal graph is represented by a Directed Acyclic Graph (DAG) $G = (V, E)$, where V denotes the set of nodes and E the set of edges, with each edge connecting a pair of nodes. Each node represents a random variable, while a path is defined as a sequence of edges connecting two nodes. A directed path is a path in which all edges are oriented in the same direction.

If a directed edge exists from node X to node Y in a DAG G, X is referred to as a parent of Y, and Y as a child of X. More generally, if there exists a directed path from node X to node Y, then X is said to be an ancestor of Y, and Y a descendant of X [Pearl, 2009, Koller and Friedman, 2009]. For a given node X, its sets of parents, children, descendants, and ancestors are denoted by $Pa(X)$, $Ch(X)$, $Desc(X)$, and $Anc(X)$, respectively. A causal graph is thus a DAG in which each parent is assumed to exert a direct causal influence on its children, and each ancestor is assumed to be a cause—either direct or indirect—of its descendants.

While the causal graph specifies which variables exert direct influences on others, it does not by itself describe the underlying causal mechanisms that generate these influences. To move from a purely qualitative representation to a quantitative model, the causal graph is augmented with explicit causal mechanisms. A causal mechanism is defined as a (possibly stochastic) function that determines the value of a variable X as a function of the values of its direct causes $Pa(X)$.

The first formalization of such causal mechanisms was introduced through *causal Bayesian networks* (CBNs) [Pearl, 1993, Pearl, 1995] and subsequently generalised within the framework of *Structural Causal Models* (SCMs) [Pearl, 2009]. In this setting, each variable X_i in the causal graph is associated with a corresponding causal mechanism $P(X_i \mid \mathrm{Pa}(X_i))$, which specifies how the variable is generated from its direct causes.

In a Structural Causal Model (SCM), each variable's causal mechanism is formalised through a structural equation. Specifically, for each variable X_i, one writes

$$X_i := f_i(\mathrm{Pa}(X_i), U_i),$$

where $\mathrm{Pa}(X_i)$ denotes the set of direct causes of X_i in the causal graph, U_i is an exogenous noise term capturing all unobserved factors that influence X_i, and f_i is a deterministic function, which may be specified parametrically or non-parametrically. The collection of these structural equations fully characterises the data-generating process. An SCM can therefore be defined as a triple $M = \langle U, V, F \rangle$ where U is the set of exogenous variables, V is the set of endogenous variables, and F is the set of structural equations. In linear SCM, the structural equations are linear functions, such that:

$$X_i = \sum_{X_j \in \mathrm{Pa}(X_i)} \beta_{ij} X_j + U_i, \tag{1}$$

where β_{ij} are called the structural coefficients. Through this construction, SCMs provide both a graphical representation, via the induced causal graph, and a functional representation, via the system of structural equations.

2.4.1 Causal Markov Assumption

The causal Markov assumption asserts that each variable in a causal graph is conditionally independent of its non-descendants, given its direct causes (i.e. its parents). Formally, for every variable $X_i \in V$,

$$\forall X_i \in V, \quad X_i \perp\!\!\!\perp_P \mathrm{NonDesc}(X_i) \setminus \mathrm{Pa}(X_i) \mid \mathrm{Pa}(X_i).$$

This assumption is a cornerstone of the causal graphical framework, as it establishes the fundamental link between the qualitative structure of a causal graph and the associated probability distribution.

Under the causal Markov assumption, the joint distribution of the variables admits a factorisation into a product of lower-dimensional conditional distributions, each depending only on the corresponding causal mechanism encoded in the graph. This representation is known as the Bayesian network factorisation [Pearl, 2009] and is formally equivalent to the causal Markov assumption [Koller and Friedman, 2009].

Specifically, given a joint distribution $P(X_1, \ldots, X_n)$ and a direct acyclic graph G, the distribution can be expressed in terms of the conditional probabilities $P(X_i \mid \mathrm{Pa}(X_i))$, where each term represents the causal mechanism generating variable X_i from its direct causes. Consequently, the overall data-generating process decomposes into a product of local causal mechanisms, one for each node in the graph:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \mathrm{Pa}(X_i)).$$

The Causal Markov condition also gives rise to a rich set of graphical criteria for reading off conditional independencies directly from the causal graph. Key concepts such as blocked paths, d-separation, flow of causation, as well as confounding, mediation, and collider bias are detailed in Appendix A.2.

2.4.2 Variable Selection

A direct implication of the causal Markov assumption is that the direct causes (parents) of a variable fully characterise its causal mechanism, which is assumed to be independent of the mechanisms generating the other variables in the structural causal model. Furthermore, this implies that the direct causes d-separate all other variables from the target variable. In other words, once all direct causes of a target variable are fixed, variations in any of its indirect causes have no effect on its value. Formally,

$$\forall Z \in \mathrm{Anc}(Y) \setminus \mathrm{Pa}(Y), \quad P(Y \mid \mathrm{Pa}(Y), Z) = P(Y \mid \mathrm{Pa}(Y)).$$

This implies that indirect causes Z that are not direct parents of Y do not enter the structural equation of Y, and conditioning on them does not alter the distribution of Y once its parents are already controlled for. Consequently, such variables are redundant for predicting Y given its direct causes, and the parent set $\mathrm{Pa}(Y)$ constitutes a sufficient set for prediction.

By contrast, including any descendant of Y disrupts its causal mechanism and break the d-separation criterion. Although descendant variables may appear to enhance predictive performance, their inclusion can introduce spurious associations by unblocking backdoor paths through collider structures, thereby inducing collider bias. Similarly, omitting any direct cause of Y breaks the causal mechanism and the d-separation criterion, resulting in information loss or the emergence of spurious associations due to confounding bias, as backdoor paths from the omitted direct causes become unblocked.

Taken together, these considerations imply that the direct causes $Pa(Y)$ of Y form the minimal set of variables required to predict Y in a maximally informative way while remaining free from spurious correlations. Thus, $Pa(Y)$ represents the most parsimonious and informative subset of predictors for Y, by construction devoid of redundant information and misleading statistical associations.

2.5 Causal Discovery

Causal discovery refers to the problem of inferring the underlying causal graph from observational data by exploiting statistical regularities in their joint distribution. Several methodological paradigms have been developed for this purpose, which can be broadly grouped into three main classes: constraint-based, score-based, and functional causal model (FCM) approaches.

Constraint-based methods identify conditional independence relationships among variables in the data and use these relationships to impose constraints on the presence or absence of edges, thereby recovering the structure of the causal graph. Score-based approaches, by contrast, perform a search over the space of admissible causal graphs, assigning each candidate graph a score that reflects its fit to the data, and then select the graph that optimises this criterion. Functional causal model approaches explicitly represent each variable as a function of its direct causes plus an independent noise term, and leverage independence properties of these noise terms to identify the causal structure.

Comprehensive reviews of causal discovery methodologies can be found in [Peters et al., 2017, Glymour et al., 2019, Spirtes and Zhang, 2016, Guo et al., 2020, Hasan et al., 2023, Niu et al., 2024, Zanga et al., 2022]. These approaches typically rely on a common set of assumptions, including causal sufficiency, the Markov condition, faithfulness, and acyclicity.

2.5.1 Assumptions

Acyclicity assumes that the data-generating process can be represented by a Directed Acyclic Graph (DAG), implying the absence of feedback loops in which a variable could causally influence itself through a sequence of other variables. The connection between this graphical structure and observed statistical independencies is established through the Markov condition, which states that two variables X and Y are conditionally independent given a set of variables $\mathcal{S}$ in the joint distribution P if they are d-separated by $\mathcal{S}$ in the DAG G.

The causal faithfulness assumption posits the converse implication: if two variables X and Y are conditionally independent given $\mathcal{S}$ in the joint distribution P, then they must be d-separated by $\mathcal{S}$ in the DAG G. In other words, faithfulness ensures that observed statistical independencies are not accidental cancellations of effects, but instead reflect genuine features of the underlying causal structure. This assumption is critical, since violations of faithfulness may lead to inferred graphs that fail to capture the true conditional independencies and, consequently, the true causal relationships. Under the joint assumptions of the Markov condition and faithfulness, the following equivalence holds:

$$X \perp\!\!\!\perp_P Y \mid \mathcal{S} \iff X \perp\!\!\!\perp_G Y \mid \mathcal{S}.$$

Causal sufficiency further assumes that all relevant confounders affecting the variables under analysis are observed. This rules out the presence of latent common causes, which would otherwise invalidate conditional independence tests and lead to causal graphs that misrepresent the data-generating process.

Even when acyclicity, the Markov condition, faithfulness, and causal sufficiency all hold, the causal graph is typically only partially identifiable from observational data. In particular, a single DAG cannot usually be uniquely recovered, since multiple DAGs may encode the same set of d-separation relations and therefore imply identical conditional independence structures Such statistically indistinguishable graphs form what is known as a Markov Equivalence Class (MEC). Two DAGs are Markov equivalent if and only if they share the same skeleton (i.e. the same underlying undirected graph) and the same set of immoralities, see e.g [Verma and Pearl, 2022, Andersson et al., 1997].

A Markov Equivalence Class can be compactly represented by a Completed Partially Directed Acyclic Graph (CPDAG). In a CPDAG, an edge is directed if and only if it has the same orientation across all DAGs in the equivalence class; otherwise, it remains undirected. Directed edges thus reflect causal relationships that are identifiable from observational data alone, while undirected edges indicate structural ambiguity that cannot be resolved without additional assumptions or interventional data. The CPDAG is therefore the most informative causal structure recoverable under the standard assumptions from purely observational data.

2.5.2 The Peter-Clark (PC) Algorithm

The PC algorithm [Spirtes et al., 2000], named after its original developers Peter Spirtes and Clark Glymour, is a flexible constraint-based procedure that allows the practitioner to incorporate any suitable statistical test for assessing conditional independence. Under the assumptions of acyclicity, causal faithfulness, and causal sufficiency, the algorithm is designed to recover the CPDAG of the Markov equivalence class to which the underlying causal DAG belongs.

The PC algorithm proceeds through three main stages. First, in the initialisation step, it begins with a fully connected undirected graph in which every variable is connected to every other variable. Second, during skeleton discovery, the algorithm systematically removes edges by performing conditional independence tests[4]: an undirected edge between two variables X and Y is removed if X is found to be conditionally independent of Y given some conditioning set $\mathcal{S}$. This step is iterative, starting with the empty conditioning set $\mathcal{S} = \{\}$ and progressively increasing the size of $\mathcal{S}$ by one at each iteration. Finally, in the edge orientation phase, the algorithm first identifies and orients all unshielded triples (Rule 0) and then applies a sequence of orientation rules (Rules 1–4) to direct as many remaining undirected edges as possible, while ensuring that no new v-structures or directed cycles are introduced.

The skeleton discovery phase relies on conditional independence tests to construct decision rules

[4]In the Appendix A.3, we revisit the Neyman-Pearson framework and we provide more details on the notion of Pearson partial correlation (ParCorr) that is employed to test conditional independence.

that determine which edges should be removed from the graph:

$$\text{Decision rule:} \quad \begin{cases} \text{keep edge } X_i - X_j & \text{if } H_0 \text{ is rejected,} \\ \text{remove edge } X_i - X_j & \text{if } H_0 \text{ fails to be rejected.} \end{cases}$$

If the p-value is smaller than α, or equivalently if the test statistic t exceeds the critical value t_α, the null hypothesis H_0 is rejected in favour of the alternative H_α. Within the PC algorithm, this outcome implies that the edge between X_i and X_j is retained, as the data provide sufficient evidence of dependence given the conditioning set $\mathcal{S}$.

Conversely, if H_0 is not rejected, this does not constitute acceptance of conditional independence, but rather indicates that H_0 remains a plausible explanation for the observed data. In the PC algorithm, this leads to the removal of the edge between X_i and X_j, suggesting conditional independence given $\mathcal{S}$. This behavior reflects the conservative nature of constraint-based methods such as the PC algorithm, where the absence of an edge represents uncertainty about dependence and ensures that only relationships that are strongly supported by the data are retained.

The primary orientation rule in the PC algorithm concerns the identification and orientation of unshielded triples as immoralities, as illustrated in Figure 3. This rule, together with three additional orientation rules, has been shown to be both sound and complete for recovering the Markov equivalence class, even in the presence of background knowledge [Meek, 2013, Andersson et al., 1997]. Soundness ensures that all causal orientations inferred by these rules are correct, in the sense that reversing any of them would violate either the Markov equivalence class or the acyclicity assumption. Completeness, on the other hand, guarantees that the resulting Completed Partially Directed Acyclic Graph (CPDAG) is maximally informative, meaning that all causal relations that are identifiable within the Markov equivalence class are recovered by the algorithm. In particular, every compelled edge—those edges that are oriented in the same direction across all DAGs within the equivalence class—is represented as directed in the CPDAG.

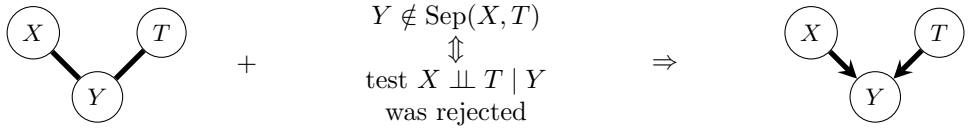


Figure 3: Unshielded Triples Orientation: For every unshielded triple $X - Y - T$, where X and T are not adjacent (directly connected), orient it as a v-structure $X \to Y \leftarrow T$ if Y is not in the d-separation set of X and T, that is Y was not in the set conditioning on which X and T became independent. Such orientation is possible because Y can only be a collider in this configuration. Three additional orientation rules, described in [Meek, 2013, Andersson et al., 1997], guarantee that the resulting Partial Directed Acyclic Graph (PDAG) is the most complete PDAG.

2.5.3 The DirectLiNGAM Algorithm

The Linear Non-Gaussian Acyclic Model (LiNGAM) [Shimizu et al., 2006] constitutes a particular class of Functional Causal Models (FCMs) in which causal relationships among variables are assumed to be linear, while the disturbance terms are statistically independent and non-Gaussian.

Formally, under the assumptions of acyclicity and causal sufficiency, for a set of random variables $V = \{X_1, \ldots, X_p\}$, the model is defined as follows.

$$X_i = \sum_{j \in \text{PA}(X_i)} \beta_{ij} X_j + e_i, \quad e_i \perp\!\!\!\perp \text{PA}(X_i), \tag{2}$$

where $\text{PA}(X_i)$ denotes the set of direct causes of X_i in the causal graph, and e_i are mutually independent non-Gaussian disturbance terms. The coefficient matrix β encodes the structural dependencies among the variables.

In contrast to constraint-based methods such as the PC algorithm, which can identify only the Markov equivalence class of admissible DAGs, LiNGAM achieves full identifiability of both the causal ordering and the associated structural coefficients. As a result, it enables recovery of the complete

causal graph rather than an equivalence class. This identifiability stems from the non-Gaussianity assumption on the disturbance terms, which permits the application of Independent Component Analysis (ICA) to uniquely determine causal directions among variables [Comon, 1994, Hyvärinen and Oja, 2000, Hyvärinen, 2013].

The first estimation procedure for LiNGAM models was introduced in [Shimizu et al., 2006] and relies on Independent Component Analysis (ICA). To enhance robustness and computational efficiency, the DIRECTLiNGAM algorithm [Shimizu et al., 2011] was subsequently proposed, avoiding iterative ICA estimation by identifying the causal ordering in a recursive manner. At each step, the algorithm selects the most exogenous variable, defined as the variable that is most statistically independent of the residuals obtained when the remaining variables are regressed on it. Once identified, this variable is removed from the system, and the procedure is repeated on the reduced set until a complete causal ordering is obtained. The measure of independence employed in this process is based on mutual information.

Moreover, once the causal ordering has been identified, each variable can be regressed on its predecessors in this ordering to estimate the corresponding structural coefficients in equation 2. To this end, regularised regression techniques such as the LASSO [Tibshirani, 1996] can be employed, which are particularly well suited to high-dimensional settings and promote sparsity in the estimated relationships. Formally, for each variable X_i, the vector of structural coefficients β_i associated with its predecessors in the causal ordering is obtained by solving the following optimisation problem:

$$\hat{\beta}_i = \underset{\beta \in \mathbb{R}^{|\{j : X_j \prec_K X_i\}|}}{\arg\min} \|X_i - X_{\{j : X_j \prec_K X_i\}} \beta\|_2^2 + \lambda \|\beta\|_1,$$

where $X_{\{j : X_j \prec_K X_i\}}$ denotes the set of variables X_j that are predecessors of X_i in the causal ordering K, and $\lambda > 0$ is the regularization parameter. This step ensures that only statistically significant causal connections are retained, resulting in a sparse, strictly lower-triangular matrix $\hat{B}$ that is consistent with the identified causal ordering. The nonzero entries b_{ij} of $\hat{B}$ then correspond directly to directed edges $X_j \to X_i$ in the inferred causal graph.

3 Empirical Application

3.1 Methodology

The objective of this empirical study is to apply the proposed causal framework to assess whether macroeconomic variables provide incremental predictive power beyond that already embedded in market-based expectations, such as those derived from the FedWatch tool and the Bloomberg WIRP function. In doing so, the analysis aims to identify the forecast horizons at which macroeconomic indicators, in addition to market-implied information, become relevant for predicting FOMC adjustment decisions.

Motivated by this perspective, we focus on a single prediction target: the cumulative interest rate adjustment decision, defined as the sum of FOMC rate adjustments up to a given meeting. As reported in Table 3, the coefficient of determination R^2 associated with the cumulative implied rate adjustment is systematically higher than that obtained when forecasting the adjustment at a specific meeting, supporting the choice of the cumulative target.

This result is intuitive, as at longer horizons it becomes increasingly difficult to forecast the precise timing of individual policy actions, whereas the overall magnitude of the expected policy rate path tends to be more stable and thus easier to predict. Aggregating rate adjustments across multiple meetings helps to smooth meeting-specific noise and reduce idiosyncratic uncertainty. Although both single-meeting adjustments and cumulative adjustments can be mapped into an implied post-meeting target rate, the cumulative measure provides a more robust and informative indicator of the expected policy stance at medium and long horizons. Consequently, for longer-term forecasts, it is more appropriate to focus on cumulative FOMC rate adjustments rather than on individual meeting decisions.

In line with the methodologies employed by the FedWatch tool and the Bloomberg WIRP function, FOMC adjustment decisions may be modeled either as continuous outcomes, representing cumulative rate changes, or as discrete categorical outcomes with associated probability distributions in 25 basis point increments. However, given the limited sample size and the emphasis on cumulative adjustments, discretising the target variable—even under the assumption that policy moves occur in discrete 25 basis

point steps—would lead to an excessive number of categories and a substantial loss of predictive power. For this reason, we adopt a regression-based modeling framework.

In addition, we do not adopt a time-series modeling approach and therefore do not include lagged variables in our analysis. This choice is motivated by both conceptual and practical considerations. From a conceptual perspective, our objective is not to model the dynamic evolution of macroeconomic variables or market expectations per se, but rather to identify the contemporaneous causal mechanisms that govern FOMC adjustment decisions at given forecast horizons. Introducing lagged variables would shift the focus toward temporal dependence and predictive dynamics, potentially confounding causal interpretation with autoregressive effects.

From a practical standpoint, the available sample size is inherently limited by the frequency of FOMC meetings, which constrains the number of observations. Incorporating lagged variables would further reduce the effective sample size and exacerbate the risk of overfitting, especially in a high-dimensional macroeconomic setting. Moreover, many of the macroeconomic indicators considered are already constructed to reflect economic conditions over specific reporting periods, such that lagged versions may add little incremental information while increasing model complexity. For these reasons, we restrict attention to contemporaneous variables and adopt a cross-sectional causal framework tailored to horizon-specific predictions, rather than a traditional time-series specification.

Furthermore, to mitigate the risk of unobserved confounding, we adopt a general-to-specific modelling strategy. Our baseline specification incorporates a broad set of macroeconomic indicators that are widely used in the monetary policy and interest rate literature. In addition, the dataset is augmented with forward-looking measures of expectations as well as proxies for financial conditions.

While this comprehensive approach is intended to capture the full range of potential determinants of policy decisions, it also poses important challenges. Macroeconomic datasets are inherently high-dimensional, yet the available sample size is limited—particularly in the context of FOMC decisions, where the number of observations is constrained by the frequency of scheduled meetings. To address these limitations and to limit the risk of overfitting, we confine the analysis to linear specifications.

From this broad information set, we construct horizon-specific candidates for the direct causal drivers of FOMC adjustment decisions using two complementary causal discovery algorithms. First, we estimate a partially directed acyclic graph using the Peter–Clark (PC) algorithm, employing partial correlation tests as the basis for (conditional) independence assessment. Second, we estimate a fully directed acyclic graph (DAG) using the DirectLiNGAM algorithm.

Under the assumption of causal sufficiency, the resulting graphs are then used to infer the underlying causal mechanisms governing FOMC adjustment decisions at each forecast horizon. These mechanisms are characterised by the set of direct causes of the policy decision, which forms the most parsimonious yet informative subset of variables. By construction, this subset is free from spurious correlations and redundant information, and represents the minimal conditioning set required to explain and predict policy adjustments within the proposed linear structural framework.

Consequently, we translate these causal hypotheses into predictive models by estimating, for each forecast horizon, two linear regressions, each restricted to the specific set of direct causes identified by the causal discovery algorithms. The resulting models are then assessed on the basis of their out-of-sample predictive performance.

In addition, we benchmark the predictive accuracy of these causal-based models against three alternative specifications: (i) the FedWatch indicator alone, represented by the cumulative implied rate adjustment; (ii) a linear regression that includes the full set of macroeconomic indicators together with the FedWatch indicator; and (iii) a linear regression that relies exclusively on the full set of macroeconomic indicators.

To evaluate out-of-sample generalisability, we employ a stratified k-fold cross-validation procedure, using the coefficient of determination (R^2) as the performance metric. The dataset is partitioned into k stratified folds, such that each fold preserves the empirical distribution of the true cumulative interest rate adjustment. The model is then iteratively trained on $k-1$ folds and evaluated on the remaining fold, with performance metrics averaged across all iterations.

Because variability in the estimated performance can arise both from random partitioning and from the limited sample size, we repeat the k-fold cross-validation procedure n times. In the empirical analysis, we implement repeated stratified k-fold cross-validation with $k = 3$ folds and $n = 10$ repetitions. Relative to a single train–test split, this approach yields a more robust and reliable assessment of model stability and predictive performance, particularly in small and potentially imbalanced datasets.

In general, macroeconomic variables are often endogenous with respect to interest rate decisions, as they are jointly determined with monetary policy actions over time. This joint determination gives rise to feedback relationships, whereby macroeconomic conditions influence interest rates, and, in turn, interest rate changes affect macroeconomic outcomes. However, the scope of our analysis is deliberately narrower. We focus exclusively on FOMC policy decisions at scheduled meeting dates—specifically, the decision to raise, lower, or maintain interest rates—rather than on the subsequent economic effects of these decisions.

Within this framework, interest rate decisions are not treated as causal drivers of macroeconomic variables. Our objective is to model the decision-making process itself, not the post-decision dynamics through which monetary policy affects the economy. Moreover, we restrict attention to macroeconomic indicators whose values are available at the time of each FOMC meeting, thereby imposing a clear temporal ordering. This ordering precludes the use of interest rate decisions to explain variables that are already observed prior to the policy decision. Consistent with this perspective, we impose an explicit temporal constraint in both the PC and DirectLiNGAM algorithms, enforcing a causal direction that runs exclusively from macroeconomic variables to interest rate adjustment decisions, and not vice versa.

3.2 Data

Our baseline specification incorporates a broad set of macroeconomic indicators that are widely used in the monetary policy and interest rate literature [Piazzesi, 2010, Kim and Wright, 2014], including measures of inflation, the unemployment rate, and GDP growth. The dataset is further enriched with forward-looking indicators of expectations and proxies for financial conditions, such as financial stress and financial conditions indices, the level of the risk-free rate, and the size of the Federal Reserve's balance sheet.

The empirical analysis is conducted on data spanning the period from January 2009 to December 2025. This timeframe encompasses several distinct monetary policy regimes. It includes the post–global financial crisis period, the European sovereign debt crisis, and the COVID-19 pandemic, all characterised by prolonged episodes of low interest rates, large-scale quantitative easing, and significant balance sheet expansion. The sample then extends into the post-pandemic phase, marked by a pronounced tightening cycle driven by elevated inflation and successive interest rate hikes, and finally covers the most recent period, characterised by a transition toward monetary easing following the peak of the tightening cycle, with the implementation of initial policy rate cuts.

All FOMC meetings in the sample are scheduled meetings, with the exception of two unscheduled emergency meetings held during the COVID-19 pandemic. These emergency meetings are excluded from the analysis, as they reflect exceptional decision-making processes that are not representative of regular, predictable policy adjustments under typical macroeconomic conditions. After this exclusion, the final sample comprises 135 FOMC meetings. Furthermore, the FedWatch tool and the Bloomberg WIRP function provide cumulative implied rate adjustments and the associated probability distributions only from 2015 onward. To maximise the sample size and ensure the robustness of the empirical analysis, we therefore replicate the underlying methodology in order to construct cumulative implied adjustment rates starting from 2009.

To assess the consistency and reliability of our implementation, we compare the cumulative implied adjustment rates produced by our replication with those obtained directly from the Bloomberg Terminal over the overlapping period from 2015 to the present. The resulting correlation is approximately 0.99, and the predictive performance metrics for FOMC adjustment decisions are nearly identical across the two implementations. Moreover, the metrics derived from our replication closely match those reported by the original methodology (see Table 4), providing strong evidence in support of the accuracy and consistency of our approach.

While the set of macroeconomic variables considered is not exhaustive, it is designed to capture the main dimensions of the information set relevant for monetary policy decisions. In addition to market-based expectations derived from the FedWatch tool, the empirical analysis relies on a broad set of standard U.S. macroeconomic indicators, measures of expectations, and proxies for financial and monetary conditions commonly used in the monetary policy literature. Inflation captures realised price pressures faced by the economy, while the five-year forward inflation expectation reflects longer-term inflation expectations monitored by the Federal Reserve. Real economic activity is proxied by year-on-year GDP growth, the unemployment rate, the ISM Manufacturing PMI, and the Federal

	1 day	1 month	3 months	6 months	9 months
From Bloomberg Terminal	0.990	0.925	0.833	0.674	0.478
From our replication	0.988	0.922	0.832	0.675	0.480

Table 4: R^2 values for the prediction of cumulative FOMC interest rate adjustment decisions across different forecast horizons. using the data from the Bloomberg Terminal and our replication.

Reserve Weekly Economic Index. Financial conditions are measured using the St. Louis Fed Financial Stress Index and the Bloomberg Financial Conditions Index, capturing stress in funding markets and broader financial tightness. Finally, monetary policy implementation and balance sheet conditions are represented by the level of the Interest on Reserve Balances (IORB) and total Federal Reserve assets, which proxy respectively for the policy stance and the scale of balance sheet operations. All variables are sourced from the Bloomberg Terminal and the FRED[5].

Table 5 summarises the out-of-sample predictive performance of several benchmark and proposed models for forecasting cumulative FOMC target rate adjustments across horizons, ranging from one day to nine months. As a baseline, we report the predictive performance of our replication of the FedWatch cumulative implied adjustment rate.

We then compare this benchmark against a set of linear regression specifications based on: (i) the replicated FedWatch cumulative implied adjustment rate alone, (ii) the selected macroeconomic variables, and (iii) the replicated FedWatch cumulative implied adjustment rate together with the selected macroeconomic variables. Finally, we report the performance of the proposed causal approach, which relies on a parsimonious set of predictors identified through causal graphs inferred using the DirectLiNGAM and PC algorithms.

	1 day	1 month	3 months	4 months	6 months	9 months
Cumulative implied change rate	0.98	0.91	0.83	0.81	0.69	0.50
Linear Regression (FedWatch)	0.98	0.90	0.80	0.80	0.69	0.46
Linear Regression (Macro)	0.23	0.20	0.40	0.46	0.59	0.59
Linear Regression (Macro + FedWatch)	0.98	0.88	0.81	0.82	0.81	0.75
PC (ParCorr) (Macro + FedWatch)	0.98	0.90	0.83	0.83	0.81	0.73
DirectLiNGAM (Macro + FedWatch)	0.98	0.90	0.82	0.84	0.84	0.81

Table 5: R^2 values for the prediction of cumulative FOMC interest rate adjustment decisions across different forecast horizons. A separate model is estimated for each horizon, and all results are evaluated out of sample using stratified three-fold cross-validation with ten repetitions.

	1 month	3 months	6 months
DirectLiNGAM	0.38	0.36	0.45
PC (ParCorr)	0.01	0.16	0.32

Table 6: R^2 values for predicting the cumulative implied rate change derived from the FedWatch tool across different forecast horizons, using the sets of direct causal drivers identified by the DirectLiNGAM and PC (partial correlation) algorithms.

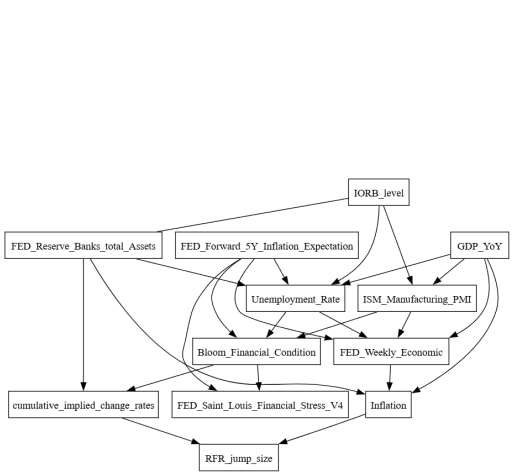
3.3 Predicting cumulative FOMC interest rate adjustment decisions over short-term horizons

At short forecast horizons, the FedWatch indicator—measured by the cumulative implied rate adjustment—exhibits substantially stronger predictive performance for cumulative FOMC adjustment decisions than a linear regression based solely on the selected macroeconomic indicators.

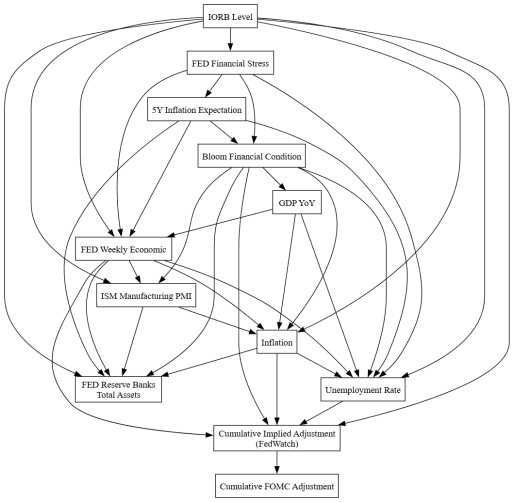
[5]https://fred.stlouisfed.org/

To further assess the informational content embedded in the FedWatch indicator, we examine the extent to which it can be explained by its direct causal drivers as inferred from the causal graphs, shown in Figure 4. The corresponding R^2 values for predicting cumulative implied rate changes are notably low. Under the PC algorithm, only a single direct cause of the FedWatch indicator is identified, resulting in an R^2 of 0.01. The DirectLiNGAM algorithm identifies a larger set of direct causes, improving the fit to an R^2 of 0.38, which nevertheless remains modest. These results suggest that, while the FedWatch indicator reflects information related to the selected macroeconomic variables, it also incorporates additional sources of information beyond those captured by the macroeconomic dataset considered in this analysis.

This finding is consistent with the causal graph identified at the one-month horizon, shown in Figure 4. The DirectLiNGAM algorithm identifies the FedWatch indicator as the sole direct causal driver of the FOMC adjustment decision, implying—via the d-separation criterion—that the remaining macroeconomic indicators are conditionally independent of the policy decision. The PC algorithm, by contrast, identifies inflation alongside the FedWatch indicator as a direct cause. While this suggests that some residual association between inflation and the FOMC decision may persist after conditioning on FedWatch, this additional link does not translate into meaningful gains in predictive performance.



(a) Causal graph for the one-month forecast horizon estimated using the PC algorithm with partial correlation tests at a significance level of $\alpha = 0.05$. The FedWatch indicator (cumulative implied rate adjustment), together with inflation, is identified as the set of direct causes of the FOMC adjustment decision, thereby d-separating the remaining macroeconomic indicators and rendering them conditionally independent of the policy decision.



(b) Causal graph for the one-month forecast horizon inferred using the DirectLiNGAM algorithm. The FedWatch indicator (cumulative implied rate adjustment) is identified as the sole direct cause of the FOMC adjustment decision, thereby d-separating the remaining macroeconomic indicators and rendering them conditionally independent of the policy decision.

Figure 4: Causal graphs for the 1 month forecast horizon obtained by the PC algorithm (a) and the DirectLiNGAM algorithm (b).

Taken together, these findings indicate that at short forecast horizons, although macroeconomic indicators are indeed related to FOMC adjustment decisions, the FedWatch indicator already encapsulates most of their relevant informational content, rendering them largely redundant for forecasting purposes. Moreover, the sizeable unexplained component in the FedWatch measure suggests that it incorporates additional information beyond that contained in standard macroeconomic fundamentals.

More precisely, at short horizons the cumulative implied rate adjustments derived from the FedWatch tool appear to subsume the predictive content of the selected macroeconomic indicators, such that these variables do not provide incremental forecasting power once FedWatch is included. The presence of substantial residual variation further indicates that FedWatch captures information not directly reflected in the macroeconomic dataset considered.

A plausible economic interpretation is that, over short horizons, the most recent macroeconomic releases may already be outdated and thus insufficient to reflect prevailing economic conditions. As a result, short-term FOMC decisions are largely driven by market-implied expectations embedded in

futures prices underlying the FedWatch tool, which incorporate forward-looking assessments of upcoming macroeconomic developments. In this setting, macroeconomic variables add little incremental information once FedWatch is taken into account. In addition, FedWatch likely reflects dimensions not explicitly captured by the selected macro indicators, such as market liquidity conditions and geopolitical risks.

3.4 Predicting cumulative FOMC interest rate adjustment decisions over medium and long-term horizons

Table 5 shows that up to the four-month forecast horizon, all specifications that include the FedWatch indicator—the FedWatch measure itself, the macro-augmented linear regression, and the regressions based on causal graphs inferred via the DirectLiNGAM and PC (partial correlation) algorithms—exhibit very similar predictive performance, with R^2 values around 0.83. Beyond this horizon, however, the predictive power of the FedWatch indicator alone deteriorates more markedly than that of models that combine market-based expectations with macroeconomic information. In particular, the FedWatch-only specification declines to an R^2 of 0.65 at six months and 0.45 at nine months, whereas the macro-augmented regression ($R^2 = 0.81$ and 0.75), the DirectLiNGAM-based regression ($R^2 = 0.84$ and 0.81), and the PC (ParCorr)-based regression ($R^2 = 0.81$ and 0.73) retain substantially higher predictive accuracy at the corresponding horizons.

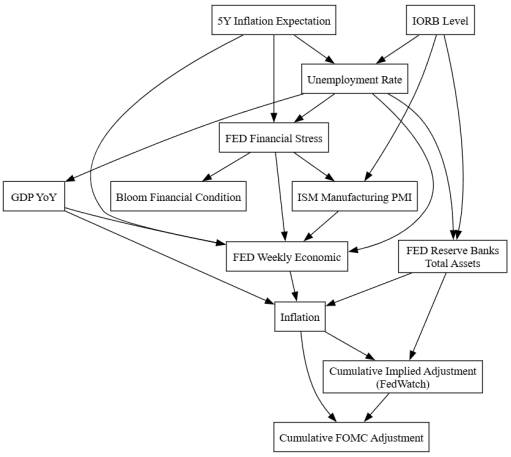
By contrast, the linear regression based exclusively on the selected macroeconomic indicators displays a different pattern: its predictive power increases with the forecast horizon, although it remains below that of models incorporating the FedWatch indicator. At the nine-month horizon, this macro-only specification reaches R^2 values of approximately 0.59, highlighting the growing relevance of macroeconomic fundamentals at longer horizons, while also underscoring the continued importance of market-based expectations.

Furthermore, Table 6 shows that predicting the cumulative implied rate changes of the FedWatch indicator using the direct causal drivers inferred from the causal graphs at the three- and six-month horizons—reported in Figures 5 and 6, respectively—results in substantially lower residual errors than those observed at the one-month horizon. In particular, the explanatory power increases to an R^2 of 0.45 under DirectLiNGAM and 0.32 under the PC approach at the six-month horizon. This pattern indicates that macroeconomic variables account for an increasing share of the variation in FedWatch-implied rate changes as the forecast horizon lengthens.

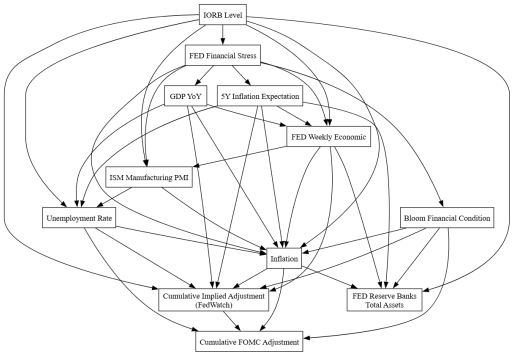
Taken together, these findings suggest that as the forecast horizon increases, macroeconomic indicators become progressively more relevant for predicting FOMC decisions, while the informational content of the FedWatch indicator alone diminishes. In contrast to the short-term horizon, medium- and long-term forecasting requires the explicit inclusion of macroeconomic fundamentals, as they provide incremental predictive power beyond market-based expectations. This conclusion is supported by the superior performance of models that combine FedWatch with macroeconomic variables relative to FedWatch-only specifications, by the rising predictive accuracy of models based solely on macroeconomic indicators, and by the growing ability of macroeconomic variables to explain the dynamics of the FedWatch indicator itself at longer horizons.

This evidence is consistent with the causal graphs inferred using the DirectLiNGAM and PC (partial correlation) algorithms at the three- and six-month horizons, shown in Figures 5 and 6, respectively. At both horizons, and under both causal discovery approaches, the FedWatch indicator is no longer identified as the sole direct causal driver of the FOMC adjustment decision. Instead, multiple macroeconomic indicators emerge as additional direct causes as the forecast horizon increases. As a result, the FedWatch indicator no longer d-separates the remaining macroeconomic variables from the policy decision, implying that these variables remain conditionally dependent on the FOMC adjustment. This finding underscores that macroeconomic indicators are not redundant at medium and long horizons, but instead contribute incremental predictive information beyond that embedded in the FedWatch indicator.

In contrast to the short-term horizon, these findings indicate that at medium and long horizons the explicit incorporation of market-based expectations of macroeconomic variables becomes less essential. Models relying on observed macroeconomic indicators alone are sufficient to deliver additional predictive power beyond that provided by the FedWatch indicator and to account more effectively for the dynamics of FedWatch itself. This pattern points to a gradual transition from a regime dominated

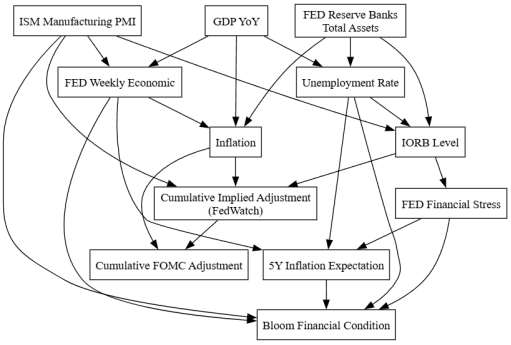


(a) Causal graph for the three-month forecast horizon estimated using the PC algorithm with partial correlation tests at a significance level of $\alpha = 0.05$. The FedWatch indicator (cumulative implied rate adjustment), together with inflation, is identified as the set of direct causes of the FOMC adjustment decision, thereby d-separating the remaining macroeconomic indicators and rendering them conditionally independent of the policy decision.

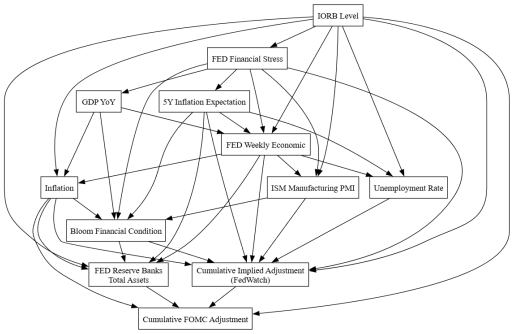


(b) Causal graph for the three-month forecast horizon inferred using the DirectLiNGAM algorithm. The FedWatch indicator (cumulative implied rate adjustment), together with the unemployment rate, inflation, and the Bloomberg Financial Conditions Index, is identified as the set of direct causes of the FOMC adjustment decision, thereby d-separating the remaining macroeconomic indicators and rendering them conditionally independent of the policy decision.

Figure 5: Causal graphs for the 3 months forecast horizon obtained by the PC algorithm (a) and the DirectLiNGAM algorithm (b).



(a) Causal graph for the six-month forecast horizon estimated using the PC algorithm with partial correlation tests at a significance level of $\alpha = 0.05$. The FedWatch indicator (cumulative implied rate adjustment), together with inflation, is identified as the set of direct causes of the FOMC adjustment decision, thereby d-separating the remaining macroeconomic indicators and rendering them conditionally independent of the policy decision.



(b) Causal graph for the six-month forecast horizon inferred using the DirectLiNGAM algorithm. The FedWatch indicator (cumulative implied rate adjustment), together with the unemployment rate, inflation, the Bloomberg Financial Conditions Index, the IORB level, and the Federal Reserve's total assets, is identified as the set of direct causes of the FOMC adjustment decision. These variables d-separate the remaining macroeconomic indicators, rendering them conditionally independent of the policy decision.

Figure 6: Causal graphs for the 6 months forecast horizon obtained by the PC algorithm (a) and the DirectLiNGAM algorithm (b).

by market-expectation-driven dynamics in the short run to one in which underlying macroeconomic fundamentals assume an increasingly important causal and predictive role in shaping monetary policy decisions at longer horizons.

4 Discussion

The causal graphs allow us to characterize the conditional independence relationships between the FOMC policy adjustment decision and its underlying determinants through the d-separation criterion. This framework provides a structural justification, developed in Section 3, for the observed transition from a regime in which the FedWatch indicator dominates at short horizons to one in which macroeconomic indicators play an increasingly prominent role at medium and long horizons.

Beyond providing qualitative insight, the causal graph identifies the most parsimonious set of genuine causal drivers—namely, the minimal yet maximally informative subset of macroeconomic indicators that is free from spurious correlations and redundancy. As a result, this framework supports the construction of predictive models that are less susceptible to overfitting and that exhibit superior generalization and improved out-of-sample performance. In particular, Table 7 reports in-sample and out-of-sample R^2 statistics across forecast horizons for the unrestricted linear regression that includes macroeconomic indicators together with the FedWatch indicator, as well as for the two causal-based regressions defined by the direct causal drivers identified in the causal graphs inferred via the PC (ParCorr) and DirectLiNGAM algorithms. The corresponding causal graphs are shown in Figure 4 for the one-month horizon, Figure 5 for the three-month horizon, and Figure 6 for the six-month horizon.

All three model specifications exhibit high and broadly comparable in-sample performance across forecast horizons. While the unrestricted regression continues to perform well in-sample at longer horizons, its out-of-sample performance deteriorates more substantially, resulting in an increasing gap between in-sample and out-of-sample fit that is indicative of overfitting. In contrast, the DirectLiNGAM-based regression displays a markedly smaller divergence between in-sample and out-of-sample performance at medium and long horizons, and consistently delivers superior out-of-sample performance relative to the unrestricted specification from the four-month forecast horizon onward. The PC (ParCorr) specification exhibits lower in-sample performance at longer horizons relative to both the unrestricted and the DirectLiNGAM-based regressions, reflecting its more conservative identification of a limited set of direct causal drivers. Nevertheless, despite its inability to capture the full complexity of the FOMC policy adjustment process, the divergence between its in-sample and out-of-sample performance remains smaller than that observed for the unrestricted regression. As a result, the PC model attains out-of-sample performance that is comparable to the unrestricted specification from the four-month forecast horizon onward.

These findings indicate that restricting the model to the set of direct causal drivers identified by the causal discovery algorithms substantially mitigates overfitting. By concentrating on the underlying causal mechanisms and excluding redundant variables and spurious associations, the causal specifications exhibit a markedly smaller divergence between in-sample and out-of-sample performance, thereby enabling more accurate forecasting, particularly at medium and long horizons. Consequently, the proposed causal variable selection framework yields more stable and robust predictive models, in which a modest reduction in in-sample fit is more than compensated by improved out-of-sample generalization.

Linear Regression	**Sample**	**1 day**	**1 month**	**3 months**	**4 months**	**6 months**	**9 months**
Unrestricted (Macro + FedWatch)	In-sample	0.99	0.92	0.90	0.89	0.88	0.87
	Out-of-sample	0.98	0.88	0.81	0.82	0.81	0.75
Causal (PC - ParCorr)	In-sample	0.99	0.92	0.87	0.88	0.84	0.78
	Out-of-sample	0.98	0.90	0.83	0.83	0.81	0.73
Causal (DirectLiNGAM)	In-sample	0.99	0.92	0.89	0.88	0.88	0.87
	Out-of-sample	0.98	0.90	0.82	0.84	0.84	0.81

Table 7: In-sample and out-of-sample R^2 measurements for predicting cumulative FOMC interest rate adjustment decisions over different forecast horizons. One model is fitted for each horizon. The unrestricted linear regression includes the macroeconomic indicators alongside the FedWatch indicator, while the causal regressions are restricted to the direct causes identified by the PC (ParCorr) and DirectLiNGAM algorithms. Out-of-sample results are obtained using a stratified 3-fold cross-validation with 10 repeats.

Throughout this paper, it is important to distinguish between the total causal effect and the direct

causal effect of X on Y[6]. These two quantities correspond to distinct causal estimands and generally differ in economic settings where indirect transmission channels play a non-negligible role.

In the absence of explicit knowledge of the data-generating process encoded in the causal graph, a natural—but ultimately naïve—approach to estimating the total causal effect of X on Y is to rely on an unrestricted linear regression and interpret the estimated coefficient on X causally. Such an approach, however, remains purely associational: it is vulnerable to confounding, mediation, and collider biases, and is therefore incapable of distinguishing between total and direct causal effects.

Table 8 illustrates these limitations in the context of estimating the effect of inflation on the cumulative FOMC policy adjustment across different forecast horizons. Methods (i)–(iii) report estimates from conventional regression specifications that differ solely in their conditioning sets: (i) a linear regression including inflation alone; (ii) a linear regression including both macroeconomic indicators and the FedWatch indicator; and (iii) a linear regression including macroeconomic indicators only. By contrast, specifications (iv)–(vi) are grounded in the proposed causal-graph-based framework. Specification (iv) restricts the conditioning set to the direct causal drivers of the FOMC policy decision identified by the DirectLiNGAM algorithm and therefore targets the direct causal effect of inflation, as defined in Equation 4. Specifications (v) and (vi) implement identification strategies consistent with the backdoor criterion implied by the estimated causal graph obtained via DirectLiNGAM, and thus recover the total causal effect of inflation. Specifically, the total effect is identified either through a backdoor-adjusted linear regression as in Equation 3, or via the path-based decomposition in Equation 5. Under the assumption of no unobserved confounders and given the assumed causal graphs, specifications (v) and (vi) therefore yield valid estimates of the total causal effect of inflation on the cumulative FOMC policy adjustment, whereas the coefficient on inflation in specification (iv) captures only its direct causal effect.

	Estimand	**Methods**	**1 month**	**3 months**	**6 months**	**9 months**
(i)	Association β	LR (Inflation)	0.051	0.131	0.201	0.290
(ii)	Association β	LR (Macro + FedWatch)	-0.002	0.041	0.101	0.180
(iii)	Association β	LR (Macro)	0.060	0.130	0.215	0.281
(iv)	Direct causal effect β_{Direct}	LR in Equation 4	0.000	0.043	0.094	0.166
(v)	Total Causal Effect β_{Total}	LR in Equation 3	0.071	0.159	0.277	0.409
(vi)	Total Causal Effect β_{Total}	Equation 5	0.069	0.153	0.275	0.396

Table 8: Comparison of associational and causal estimates of the effect of inflation on the cumulative FOMC policy adjustment across forecast horizons. Specifications (i)–(iii) report coefficients from conventional linear regressions with different conditioning sets and therefore capture purely associational effects. Specifications (iv)–(vi) exploit the estimated causal graph obtained via the DirectLiNGAM algorithm (see Figures 4b, 5b, and 6b) to distinguish between direct and total causal effects. In particular, specification (iv) identifies the direct causal effect through conditioning on direct causal parents, while specifications (v) and (vi) recover the total causal effect using, respectively, a backdoor-adjusted regression and a path-based decomposition.

We observe that the coefficients reported in Table 8 for the traditional regression specifications (i)–(iii) systematically underestimate the causal effect relative to the backdoor-adjusted approaches (v) and (vi). This discrepancy arises because the causal graphs in Figures 4b, 5b, and 6b indicate the presence of confounders that jointly influence inflation and the cumulative FOMC policy adjustment at all forecast horizons, thereby inducing backdoor paths and confounding bias. In addition, the graphs reveal the presence of mediating variables along the causal pathways from inflation to the FOMC decision at all horizons, so that conditioning on these mediators leads to further bias through overcontrol (mediation bias).

For instance, estimating specification (i), which includes inflation as the sole regressor, avoids conditioning on mediating variables and therefore mitigates mediation bias. However, this specification remains vulnerable to confounding bias, since common determinants of inflation and the cumulative FOMC policy adjustment are omitted from the regression.

A seemingly more coherent strategy is therefore to condition on all available covariates, leading to regression specifications such as (ii) and (iii). While this approach aims to mitigate confounding bias,

[6]In Appendix A.4 we recall the notion of total and direct causal effect of X on Y.

it does so at the cost of introducing mediation bias. According to the causal graphs, specification (ii) conditions on the FedWatch indicator and the unemployment rate, whereas specification (iii) conditions on the unemployment rate alone; both variables act as mediators along the causal pathways from inflation to the FOMC policy adjustment. As a consequence, these specifications block indirect causal channels and conflate direct and total causal effects. Consistent with this interpretation, the estimated coefficients reported in Table 8 systematically underestimate the total causal effect of inflation across all forecast horizons, and in the case of specification (ii), yield an effect close to zero at short horizons. This does not imply that specifications (ii) and (iii) lack informational value; rather, their limitation lies in the fact that they identify neither the total causal effect nor isolate it cleanly. Under the assumed causal graph, specification (ii) effectively recovers a direct causal effect comparable to that obtained in specification (iv), while specification (iii) provides a coherent estimate of the direct causal effect in settings where the FedWatch indicator is excluded.

Although the estimated causal graphs do not feature collider variables—since the analysis does not condition on the consequences of the FOMC policy adjustment—it is important to emphasize that, in more general unrestricted regression settings, including a collider as a control variable can introduce bias rather than mitigate it. Care must therefore be exercised to avoid such pitfalls in broader empirical applications.

Taken together, these results underscore the limitations of traditional regression approaches for causal interpretation. Depending on the choice of conditioning set, the estimated coefficient on inflation may be affected by confounding bias, mediation bias, or both. By contrast, the proposed causal-graph-based framework offers a principled means of distinguishing between distinct causal estimands: the direct causal effect, captured by regression specification (iv), and the total causal effect, recovered through either the backdoor-adjusted regression in specification (v) or the path-based decomposition in specification (vi). This ability to explicitly target well-defined causal effects yields estimates that are more coherent across forecast horizons, more aligned with economic intuition, and less sensitive to arbitrary modeling choices.

Furthermore, the estimated direct and total causal effects obtained from specifications (iv)–(vi) are fully consistent with our empirical findings and reinforce the interpretation of a regime shift from a market-expectation-driven mechanism at short horizons to a macroeconomic-fundamentals-driven mechanism at medium and long horizons. As the forecast horizon lengthens, inflation displays an increasingly large and economically meaningful positive total causal effect on the cumulative FOMC policy adjustment. At the same time, its direct causal effect—once mediating variables such as the FedWatch indicator are appropriately controlled for—also becomes progressively more positive and statistically significant. Taken together, these results underscore the increasingly prominent role of inflation, and more broadly of macroeconomic indicators, as central components of the causal mechanism underlying cumulative FOMC policy adjustments.

A revealing contrast emerges when the same analysis is applied to the cumulative implied adjustment of the FedWatch indicator, as reported in Table 9. Both its total causal effect, as identified by specifications (v) and (vi), and its direct causal effect, as captured by specification (iv), on realized FOMC policy adjustments remain positive but become progressively weaker and less statistically significant as the forecast horizon increases. This pattern indicates that, while market expectations continue to co-move with policy outcomes, their causal influence diminishes at longer horizons.

	Estimand	**Methods**	**1 month**	**3 months**	**6 months**	**9 months**
(iv)	Direct causal effect β_{Direct}	LR in Equation 4	1.076	0.924	0.913	0.817
(v)	Total Causal Effect β_{Total}	LR in Equation 3	1.077	0.994	0.913	0.817
(vi)	Total Causal Effect β_{Total}	Equation 5	1.077	0.994	0.913	0.817

Table 9: Comparison of associational and causal estimates of the effect of the cumulative implied adjustment decision of the FedWatch indicator on the cumulative realised FOMC adjustment decision across forecast horizons. Methods (iv)-(vi) leverage the estimated causal graph obtained by DirectLiNGAM (see Figure 5b, 5b, and 6b) to distinguish between the direct and total causal effects, using either direct-cause adjustment (iv) for direct causal effect, backdoor-adjusted regression (v) for total causal effect, or path-based decomposition (vi) for total causal effect.

Consequently, the proposed causal-graph-based framework fulfills a dual role. By restricting atten-

tion to a parsimonious set of genuine causal drivers, it mitigates overfitting and enhances out-of-sample generalization. At the same time, it provides a coherent explanation for the empirical findings by offering a principled distinction between total and direct causal effects. This approach yields estimates that are explicitly targeted to well-defined causal estimands, robust to confounding, mediation, and collider biases, and therefore more interpretable and more closely aligned with economic intuition.

5 Conclusion

In this paper, we introduce a novel causal framework for predicting FOMC target rate adjustment decisions. By employing causal discovery techniques—specifically the PC algorithm and DirectLiNGAM—we systematically extract, from a comprehensive set of macroeconomic indicators together with the market-based CME FedWatch tool, horizon-specific causal graphs that characterise the underlying mechanisms driving FOMC decisions across different forecast horizons.

This causal perspective provides a coherent interpretation of the empirical results. At short horizons, up to one month, virtually all predictive power for FOMC rate adjustments is captured by the FedWatch tool, which emerges as the sole direct causal driver under both causal discovery algorithms. This finding suggests that market-based expectations already incorporate forward-looking assessments of anticipated macroeconomic developments, liquidity conditions, and geopolitical risks. At medium and long horizons, the causal analysis reveals an evolving causal structure in which macroeconomic variables emerge as additional direct causes of FOMC decisions, thus increasingly adding predictive information beyond what is contained in market-implied expectations. Moreover, the causal approach based on the DirectLiNGAM algorithm delivers parsimonious, interpretable, and asymptotically unbiased predictive models that consistently outperform traditional linear regressions in terms of out-of-sample forecasting accuracy and causal effect estimation across horizons. These improvements further enable the construction of new estimates of the cumulative implied rate adjustment and the corresponding future price curves, providing a more coherent and economically grounded representation of market expectations.

References

[Andersson et al., 1997] Andersson, S. A., Madigan, D., and Perlman, M. D. (1997). A characterization of Markov equivalence classes for acyclic digraphs. *The Annals of Statistics*, 25(2):505–541.

[Angrist and Pischke, 2009] Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

[Atlanta-Fed, 2023] Atlanta-Fed (2023). Atlanta fed's market probability tracker (mpt). https://www.atlantafed.org/markettprobtracker.

[Baba et al., 2004] Baba, K., Shibata, R., and Sibuya, M. (2004). Partial correlation and conditional correlation as measures of conditional independence. *Australian & New Zealand Journal of Statistics*, 46(4):657–664.

[Bloomberg Finance L.P., 2023] Bloomberg Finance L.P. (2023). Explore predicted global interest rates with the bloomberg terminal - wirp. https://www.bloomberg.com/professional/insights/markets/bloomberg-pro-tips-explore-predicted-global-interest-rates-with-the-bloomberg-terminal/.

[Bonini et al., 2025] Bonini, S., Huang, S., and Simaan, M. (2025). Watching the FedWatch. *Available at SSRN 5093703*.

[Chernozhukov et al., 2022] Chernozhukov, V., Cinelli, C., Newey, W., Sharma, A., and Syrgkanis, V. (2022). Long story short: Omitted variable bias in causal machine learning. Technical report, National Bureau of Economic Research.

[Cinelli and Hazlett, 2020] Cinelli, C. and Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(1):39–67.

[Cinelli et al., 2019] Cinelli, C., Kumor, D., Chen, B., Pearl, J., and Bareinboim, E. (2019). Sensitivity analysis of linear structural causal models. In *International conference on machine learning*, pages 1252–1261. PMLR.

[CME-Group, 2025] CME-Group (2025). Understanding the CME Group FedWatch Tool Methodology. https://www.cmegroup.com/articles/2023/understanding-the-cme-group-fedwatch-tool-methodology.html. Accessed: 2024-06-13.

[Cole et al., 2010] Cole, S. R., Platt, R. W., Schisterman, E. F., Chu, H., Westreich, D., Richardson, D., and Poole, C. (2010). Illustrating bias due to conditioning on a collider. *International journal of epidemiology*, 39(2):417–420.

[Comon, 1994] Comon, P. (1994). Independent component analysis, a new concept? *Signal processing*, 36(3):287–314.

[Fontana et al., 2020] Fontana, C., Grbac, Z., Gümbel, S., and Schmidt, T. (2020). Term structure modelling for multiple curves with stochastic discontinuities. *Finance and Stochastics*, 24(2):465–511.

[Fontana et al., 2024] Fontana, C., Grbac, Z., and Schmidt, T. (2024). Term structure modeling with overnight rates beyond stochastic continuity. *Mathematical Finance*, 34(1):151–189.

[Glymour et al., 2019] Glymour, C., Zhang, K., and Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Frontiers in genetics*, 10:1–15.

[Granger, 1969] Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438.

[Granger, 1980] Granger, C. (1980). Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, 2:329–352.

[Guo et al., 2020] Guo, R., Cheng, L., Li, J., Hahn, P. R., and Liu, H. (2020). A survey of learning causality with data: Problems and methods. *ACM Computing Surveys (CSUR)*, 53(4):1–37.

[Hasan et al., 2023] Hasan, U., Hossain, E., and Gani, M. O. (2023). A survey on causal discovery methods for iid and time series data. *arXiv preprint arXiv:2303.15027*.

[Hernán et al., 2011] Hernán, M. A., Clayton, D., and Keiding, N. (2011). The simpson's paradox unraveled. *International journal of epidemiology*, 40(3):780–785.

[Hernán and Monge, 2023] Hernán, M. A. and Monge, S. (2023). Selection bias due to conditioning on a collider. *Bmj*, 381.

[Holland, 1986] Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960.

[Hyvärinen, 2013] Hyvärinen, A. (2013). Independent component analysis: recent advances. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984):20110534.

[Hyvärinen and Oja, 2000] Hyvärinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430.

[Imbens and Rubin, 2015] Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge university press.

[Keller-Ressel et al., 2019] Keller-Ressel, M., Schmidt, T., and Wardenga, R. (2019). Affine processes beyond stochastic continuity. *The Annals of Applied Probability*, 29(6):3387–3437.

[Kilian, 2006] Kilian, L. (2006). New introduction to multiple time series analysis. *Econometric theory*, 22(5):961–967.

[Kim and Wright, 2014] Kim, D. H. and Wright, J. H. (2014). *Jumps in bond yields at known times*. Number w20711. National Bureau of Economic Research Cambridge.

[Kim, 2015] Kim, S. (2015). ppcor: an R package for a fast calculation to semi-partial correlation coefficients. *Communications for statistical applications and methods*, 22(6):665.

[Koller and Friedman, 2009] Koller, D. and Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

[Lee et al., 2021] Lee, J., Youn, H. L., Stevens, N., Poon, J., and Han, S. C. (2021). Fednlp: an interpretable nlp system to decode federal reserve communications. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 2560–2564.

[Lopez de Prado et al., 2024] Lopez de Prado, M., Lipton, A., and Zoonekynd, V. (2024). The case for causal factor investing. *Available at SSRN 4774522*.

[Lopez de Prado et al., 2025] Lopez de Prado, M., Lipton, A., and Zoonekynd, V. (2025). Causal factor analysis is a necessary condition for investment efficiency. *Available at SSRN*.

[Lopez de Prado, 2023] Lopez de Prado, M. M. (2023). *Causal factor investing: can factor investing become scientific?* Cambridge University Press.

[Lu et al., 2022] Lu, H., Cole, S. R., Howe, C. J., and Westreich, D. (2022). Toward a clearer definition of selection bias when estimating causal effects. *Epidemiology*, 33(5):699–706.

[MacKinnon and Lamp, 2021] MacKinnon, D. P. and Lamp, S. J. (2021). A unification of mediator, confounder, and collider effects. *Prevention Science*, 22(8):1185–1193.

[Meek, 2013] Meek, C. (2013). Causal inference and causal explanation with background knowledge. *arXiv preprint arXiv:1302.4972*.

[Munafò et al., 2018] Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M., and Davey Smith, G. (2018). Collider scope: when selection bias can substantially influence observed associations. *International journal of epidemiology*, 47(1):226–235.

[Neal, 2020] Neal, B. (2020). Introduction to causal inference. *Course lecture notes (draft)*, 132.

[Niu et al., 2024] Niu, W., Gao, Z., Song, L., and Li, L. (2024). Comprehensive review and empirical evaluation of causal discovery algorithms for numerical data. *arXiv preprint arXiv:2407.13054*.

[Pearl, 1993] Pearl, J. (1993). From Bayesian networks to causal networks. In *Mathematical models for handling partial knowledge in artificial intelligence*, pages 157–182. Springer.

[Pearl, 1995] Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4):669–688.

[Pearl, 2009] Pearl, J. (2009). *Causality*. Cambridge university press.

[Pearl, 2012] Pearl, J. (2012). The causal mediation formula—a guide to the assessment of pathways and mechanisms. *Prevention science*, 13(4):426–436.

[Pearl, 2022a] Pearl, J. (2022a). Comment: understanding simpson's paradox. In *Probabilistic and causal inference: The works of judea Pearl*, pages 399–412.

[Pearl, 2022b] Pearl, J. (2022b). Direct and indirect effects. In *Probabilistic and causal inference: the works of Judea Pearl*, pages 373–392.

[Pearl and Mackenzie, 2018] Pearl, J. and Mackenzie, D. (2018). *The book of why: the new science of cause and effect*. Basic books.

[Peters et al., 2017] Peters, J., Janzing, D., and Schölkopf, B. (2017). *Elements of causal inference: foundations and learning algorithms*. The MIT Press.

[Piazzesi, 2005] Piazzesi, M. (2005). Bond yields and the federal reserve. *Journal of Political Economy*, 113(2):311–344.

[Piazzesi, 2010] Piazzesi, M. (2010). Affine term structure models. In *Handbook of financial econometrics: Tools and Techniques*, pages 691–766. Elsevier.

[Richiardi et al., 2013] Richiardi, L., Bellocco, R., and Zugna, D. (2013). Mediation analysis in epidemiology: methods, interpretation and bias. *International journal of epidemiology*, 42(5):1511–1519.

[Rijnhart et al., 2021] Rijnhart, J. J., Lamp, S. J., Valente, M. J., MacKinnon, D. P., Twisk, J. W., and Heymans, M. W. (2021). Mediation analysis methods used in observational research: a scoping review and recommendations. *BMC medical research methodology*, 21(1):226.

[Rosenbaum and Rubin, 1983] Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

[Rubin, 1974] Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

[Schlögl et al., 2024] Schlögl, E., Skov, J. B., and Skovmand, D. (2024). Term structure modeling of sofr: Evaluating the importance of scheduled jumps. *International Journal of Theoretical and Applied Finance*, 27(02):2450009.

[Seok et al., 2024] Seok, S., Wen, S., Yang, Q., Feng, J., and Yang, W. (2024). Minifed: Integrating llm-based agentic-workflow for simulating fomc meeting. *arXiv preprint arXiv:2410.18012*.

[Shimizu et al., 2006] Shimizu, S., Hoyer, P. O., Hyvarinen, A., and Kerminen, A. (2006). A Linear Non-Gaussian Acyclic Model for Causal Discovery. *Journal of Machine Learning Research 7*.

[Shimizu et al., 2011] Shimizu, S., Inazumi, T., Sogawa, Y., Hyvarinen, A., Kawahara, Y., Washio, T., Hoyer, P. O., Bollen, K., and Hoyer, P. (2011). Directlingam: A direct method for learning a linear non-gaussian structural equation model. *Journal of Machine Learning Research-JMLR*, 12(Apr):1225–1248.

[Spirtes et al., 2000] Spirtes, P., Glymour, C. N., and Scheines, R. (2000). *Causation, prediction, and search*. MIT press.

[Spirtes and Zhang, 2016] Spirtes, P. and Zhang, K. (2016). Causal discovery and inference: concepts and recent methodological advances. In *Applied informatics*, volume 3, pages 1–28. Springer.

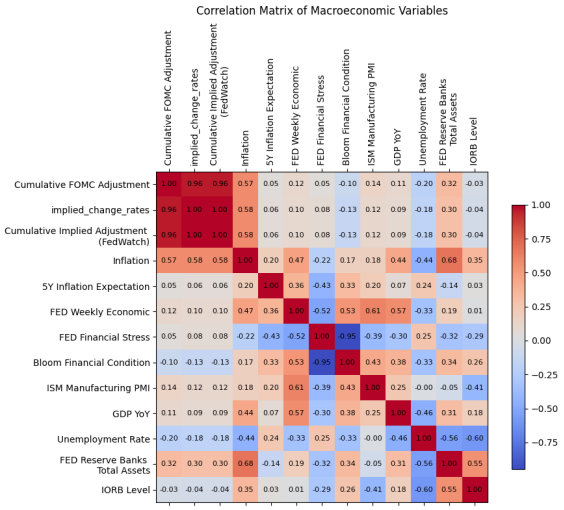
[Tibshirani, 1996] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288.

[Verma and Pearl, 2022] Verma, T. S. and Pearl, J. (2022). Equivalence and synthesis of causal models. In *Probabilistic and causal inference: The works of Judea Pearl*, pages 221–236.

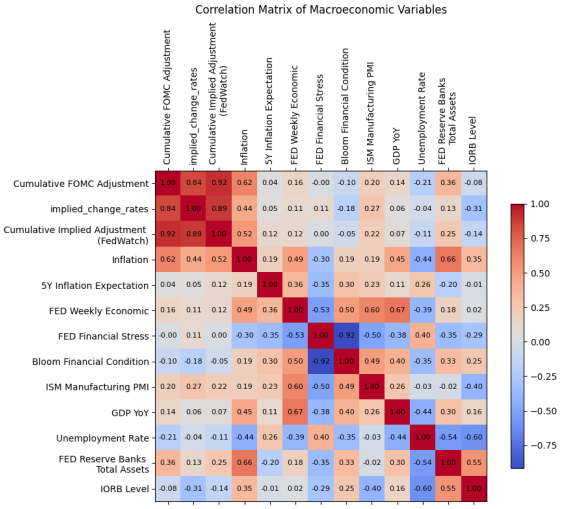
[Zanga et al., 2022] Zanga, A., Ozkirimli, E., and Stella, F. (2022). A survey on causal discovery: theory and practice. *International Journal of Approximate Reasoning*, 151:101–129.

A Appendix

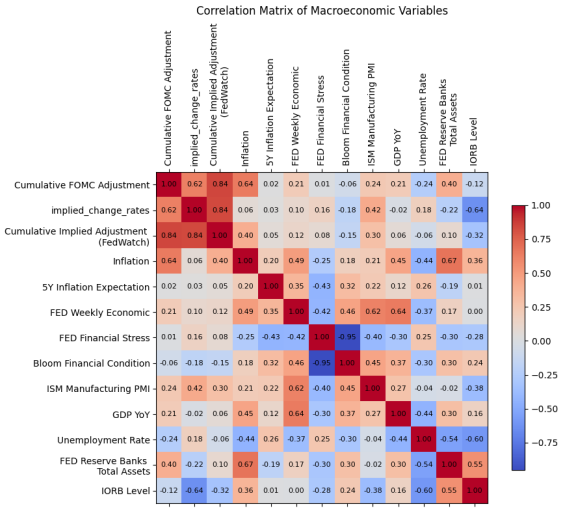
A.1 Additional figures



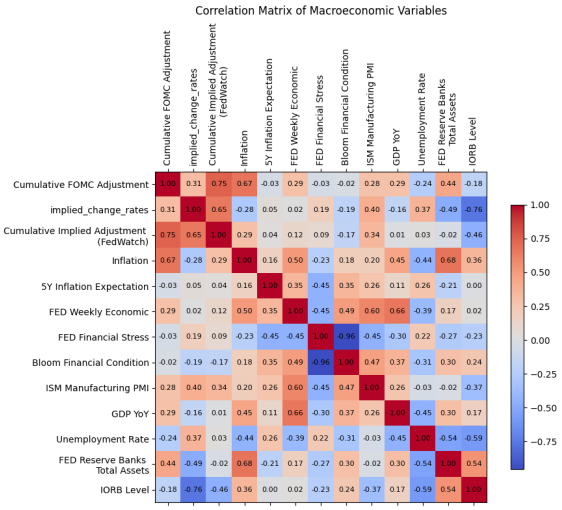
(a) Correlation matrix for the one-month forecast horizon.



(b) Correlation matrix for the three-month forecast horizon.



(c) Correlation matrix for the six-month forecast horizon.



(d) Correlation matrix for the nine-month forecast horizon.

Figure 7: Empirical Pearson correlation matrices between the main variables used in our analysis for the 1-month, 3-month, 6-month, and 9-month forecast horizons. These matrices assess the linear dependence between macroeconomic indicators, market-based indicators with the (cumulative) implied FOMC rate adjustment (from the FedWatch Tool), and the (cumulative) realised FOMC rate adjustment. The sample covers 135 FOMC meetings from 2009 to 2025. Notably, correlations between market-based indicators and rate adjustments tend to decrease with the forecast horizon, while those between macroeconomic indicators and realised rate adjustments tend to increase. (Source: Bloomberg, FRED)

A.2 Additional content on the causal approach

A.2.1 Blocked path, flow of association and flow of causation

The causal Markov assumption and the bayesian network factorization tell us about the flow of association, that is whether any two variables in the causal graph are associated (statistically dependent) or not associated (statistically independent). In causal graphs, we can define the flow of association using the minimal building blocks of graphs: chain, fork and immorality. Figure 8 is an example of a chain, where there exist an association between the indirect cause X and the target variable Y which is blocked by conditioning on the mediator M. According to the bayesian network factorization, the joint distribution of a chain can be factorized as $P(X, Y, M) = P(X)P(M \mid X)P(Y \mid M)$. Without conditioning on M, X and Y remains associated through mediator M because:

$$P(X, Y) = \sum_{M} P(X)P(M \mid X)P(Y \mid M) \neq P(X)P(Y).$$

However, when we condition on M, X and Y are no longer associated because:

$$P(X, Y \mid M) = P(X \mid M)P(Y \mid M) \implies X \perp\!\!\!\perp Y \mid M.$$

(a) Not conditioning on the mediator M let the causal association between X and Y flow through M, as the directed path (blue arrow), also called causal path, remains unblocked.

(b) Conditioning on M blocks the (causal) association between X and Y, rendering X and Y conditionally independent given M, i.e., $X \perp\!\!\!\perp Y \mid M$ (mediator bias).

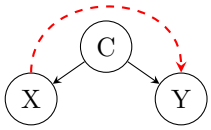
Figure 8: Illustration of a chain structure $X \to M \to Y$, where M is a mediator transmitting the causal association of X on Y. Panel (a) shows that X and Y remains associated when M is not conditioned on, while panel (b) shows that conditioning on M blocks the association between X and Y and renders X and Y conditionally independent given M (mediator bias).

Figure 9 is an example of a fork, where a variable X and the target variable Y are associated through a common cause C, also called confounder. According to the bayesian network factorization, the joint distribution of a fork can be factorized as $P(X, Y, C) = P(C)\,P(X \mid C)\,P(Y \mid C)$. Without conditioning on C, X and Y remain associated because:

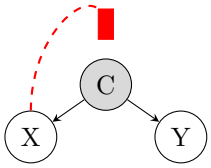
$$P(X, Y) = \sum_{C} P(C)\,P(X \mid C)\,P(Y \mid C) \neq P(X)\,P(Y).$$

However, when we condition on C, X and Y are no longer associated because:

$$P(X, Y \mid C) = P(X \mid C)\,P(Y \mid C) \implies X \perp\!\!\!\perp Y \mid C.$$



(a) The two variables X and Y are spuriously associated through the confounder C as the non-directed path (red arrow), also called backdoor path or non-causal path, remains unblocked when C is not conditioned on (confounder bias).



(b) Conditioning on C blocks the backdoor path and renders X and Y conditionally independent given C, i.e., $X \perp\!\!\!\perp Y \mid C$, thus removing the spurious association between X and Y.

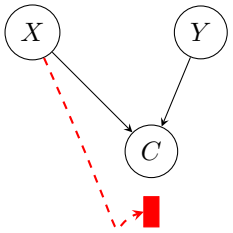
Figure 9: Illustration of a fork structure $X \leftarrow C \rightarrow Y$, where C is a confounder (common cause) inducing a spurious association between the two independent variables X and Y. Panel (a) shows that X and Y remain spuriously associated when C is not conditioned on (confounder bias), while panel (b) shows that conditioning on C removes the spurious association between X and Y and correctly renders X and Y conditionally independent given C.

Figure 10 is an example of a immorality structure, where a variable X and the target variable Y are independent but associated when conditioning on the collider Z. According to the bayesian network factorization, the joint distribution of a collider can be factorized as $P(X, Y, Z) = P(X)\,P(Y)\,P(Z \mid X, Y)$. Without conditioning on C, X and Y are independent because:

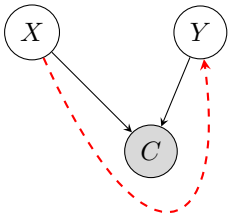
$$P(X,Y) = P(X)\,P(Y)\sum_{C} P(C \mid X,Y) = P(X)\,P(Y) \implies X \perp\!\!\!\perp Y.$$

However, when we condition on C, X and Y becomes associated because:

$$P(X,Y \mid C) = \frac{P(X,Y,C)}{P(C)} = \frac{P(X)\,P(Y)\,P(C \mid X,Y)}{P(C)} \implies X \not\perp\!\!\!\perp Y \mid C.$$



(a) Not conditioning on the collider$C$ let the non-directed path (red arrow), also called backdoor path or non-causal path, unblocked and X and Y remain independent



(b) Conditioning on the collider C opens the a backdoor path and induces a spurious association between X and Y, rendering them conditionally dependent given C, i.e., $X \not\perp\!\!\!\perp Y \mid C$ (collider bias).

Figure 10: Illustration of an immorality structure $X \rightarrow C \leftarrow Y$, where C is a collider (common effect) inducing a spurious association between the two independent variables X and Y when conditioning on. Panel (a) shows that X and Y remains independent when C is not conditioned on, while panel (b) shows that conditioning on C induces a spurious association between X and Y (collider bias).

More formally, a path between X and Y is blocked (i.e, rendered non-associational) if either the path traverses a collider, and the collider and its descendants have not been conditioned on (see figure 10), or a variable in the path between X and Y, which is not a collider, has been conditioned on (see figure 8 and 9). Thus, associations flow along unblocked paths such that any two variables are associated if and only if there exists an unblocked path between them.

Furthermore, under the causal graph assumption that an edge from X to Y implies that X is a cause of Y, we can distinguish between causal associations and spurious associations. Causal associations flows along unblocked directed paths, also called causal paths, where the direction of the path follows the direction of causality from one variable to another (see figure 8). In contrast, spurious associations flows along unblocked non-causal paths, also called backdoor paths (see figure 9 and 10).

A.2.2 The notion of d-separation and global Markov condition

The connection between the causal graph and the observed statistical independencies is established through the global Markov condition. This assumption states that the conditional independence relationships implied by the graph are reflected in the joint distribution of the variables. The graphical criterion that determines these independencies is known as d-separation. Intuitively, d-separation characterizes whether information can flow between two variables along the paths of a causal graph once a set of variables is conditioned upon.

Formally, two variables X and Y are d-separated by a set of variables $\mathcal{S}$ in a DAG G if every path between X and Y is blocked by $\mathcal{S}$; otherwise, they are d-connected. Specifically, under the global Markov assumption, if X and Y are d-separated by $\mathcal{S}$, then they are conditionally independent in the joint distribution P:

$$X \perp\!\!\!\perp_G Y \mid \mathcal{S} \implies X \perp\!\!\!\perp_P Y \mid \mathcal{S}$$

In other words, if $\mathcal{S}$ d-separates X and Y, then once all variables in $\mathcal{S}$ has been conditioned on, knowledge of X does not provide any additional information about Y.

A.3 Additional content on conditional independence tests

A.3.1 Neyman-Pearson framework

To assess (conditional) dependence between two variables X_i and X_j given a conditioning set $\mathcal{S}$, we formulate the null hypothesis that X_i and X_j are conditionally independent given $\mathcal{S}$, against the alternative that they are conditionally dependent given:

$$H_0 : X_i \perp\!\!\!\perp X_j \mid \mathcal{S}$$
$$H_\alpha : X_i \not\perp\!\!\!\perp X_j \mid \mathcal{S}$$

Statistical hypothesis testing proceeds by computing a test statistic T from the data and comparing its realised value t to a rejection region determined by a chosen significance level α. Within the Neyman–Pearson framework, the rejection region is defined such that the probability of observing a value of T at least as extreme as t under the null hypothesis is sufficiently small. Specifically, the null hypothesis is rejected whenever the corresponding p-value falls below α, reflecting confidence at level $1 - \alpha$ that the observed dependence is incompatible with conditional independence.

Formally, in the one-sided case, the null hypothesis is rejected if $P(T > t|H_0) < \alpha$, while in the two-sided case rejection occurs if $P(|T| > |t| \ |H_0) < \alpha$. Equivalently, critical values t_α (one-sided) or $t_{\alpha/2}$ (two-sided) are chosen such that observations exceeding these thresholds lie in the rejection region:

$$t > t_\alpha \implies \text{Reject } H_0, \quad |t| > t_{\alpha/2} \implies \text{Reject } H_0.$$

In this study, we rely on Pearson-based (partial) correlation tests to construct the test statistic T. This choice is motivated by several considerations. First, Pearson tests are well suited to continuous variables and linear dependence structures, which are natural assumptions in the macroeconomic and financial settings considered here. Second, under mild regularity conditions, Pearson correlation admits a well-understood sampling distribution, enabling reliable inference even in moderate sample sizes. Third, Pearson-based tests integrate naturally into constraint-based causal discovery algorithms, where repeated conditional independence testing is required and computational efficiency is essential. Alternative nonparametric or information-theoretic tests, while more general, typically suffer from reduced power in small samples and higher computational cost. Given the limited number of observations available at the frequency of FOMC meetings, Pearson tests therefore offer a principled and pragmatic balance between statistical power, interpretability, and computational tractability.

A.3.2 Pearson partial correlation

In settings where the data are assumed to be Gaussian, it is standard practice to employ the Pearson partial correlation coefficient $\rho_{ij,\mathcal{S}}$ as the test statistic for assessing conditional independence. This quantity provides a measure of conditional association and is defined as the correlation between two variables X_i and X_j in their conditional distribution given a set of variables $\mathcal{S}$ [Baba et al., 2004]. Equivalently, it captures the residual linear dependence between X_i and X_j after removing the linear effect of the conditioning variables set $\mathcal{S}$.

An efficient estimator $\hat{\rho}_{ij,\mathcal{S}}$ of the partial correlation coefficient can be obtained through inversion of the empirical covariance matrix [Kim, 2015]. Specifically,

$$\hat{\rho}_{ij,\mathcal{S}} = -\frac{\omega_{ij}}{\sqrt{\omega_{ii}, \omega_{jj}}},$$

where ω_{ij} denotes the (i,j)-th entry of the precision matrix $\Omega = \Sigma^{-1}$, the inverse of the covariance matrix Σ. The hypothesis test is then formulated by taking the null hypothesis that the partial correlation between X_i and X_j given S is zero, against the alternative that it is non-zero:

$$H_0 : \rho_{ij,\mathcal{S}} = 0$$
$$H_\alpha : \rho_{ij,\mathcal{S}} \neq 0$$

To compute the corresponding p-value, the Fisher z-transformation is applied:

$$z(\hat{\rho}_{ij,\mathcal{S}}) = \frac{1}{2} \ln \left(\frac{1 + \hat{\rho}_{ij,\mathcal{S}}}{1 - \hat{\rho}_{ij,\mathcal{S}}} \right).$$

Under the null hypothesis $\rho_{ij,\mathcal{S}} = 0$, the transformed statistic satisfies the approximation $\sqrt{N - \mid S \mid -3} z(\rho_{ij,\mathcal{S}}) \sim \mathcal{N}(0,1)$, where N denotes the number of observations and $\mid \mathcal{S} \mid$ the cardinality of the conditioning set. Within the Neyman–Pearson framework for a two-sided test, the null hypothesis is rejected if the associated p-value is smaller than the significance level α, that is,

$$2(1 - \Phi(\mid \sqrt{N - \mid \mathcal{S} \mid -3} z(\hat{\rho}_{ij,\mathcal{S}}) \mid)) < \alpha \implies \text{Reject } H_0,$$

where Φ denotes the cumulative distribution function of the standard normal distribution. Equivalently, rejection can be based on the critical value criterion

$$\sqrt{N - \mid \mathcal{S} \mid -3} \mid z(\hat{\rho}_{ij,\mathcal{S}}) \mid > \Phi^{-1} \left(1 - \frac{\alpha}{2}\right) \implies \text{Reject } H_0,$$

with Φ^{-1} denoting the inverse standard normal distribution function.

In the special case of unconditional independence testing, the partial correlation coefficient $\rho_{ij,\mathcal{S}}$ reduces to the ordinary Pearson correlation coefficient ρ_{ij}. The same Fisher z-transformation applies, with the asymptotic result $\sqrt{N-3} \mid z(\rho_{ij}) \mid \sim \mathcal{N}(0,1)$ under the null hypothesis $\rho_{ij} = 0$.

A.4 Total and direct causal effect of X on Y

The total causal effect captures the overall influence of X on Y through all causal pathways, including both direct paths and indirect causal paths that operate through mediating variables. To estimate the total causal effect of a variable X on an outcome Y, we seek a conditioning set that blocks all spurious (non-causal) associations between X and Y while leaving all causal associations from X to Y unblocked. In particular, mediators from X to Y must be excluded to preserve the total causal association flowing from the causal path. A conditioning set $\mathcal{S}$ satisfying these requirements is called a valid adjustment set and must fulfill the backdoor criterion: it blocks all backdoor paths from X to Y and does not contain any descendants (i.e., effects) of X. Therefore, under the assumption of no unobserved confounders between X and Y, the total causal effect of X on Y can be consistently estimated by regressing Y on X while controlling for $\mathcal{S}$. Formally, it can be estimated as the coefficient $\beta_{\text{Total}} \in \mathbb{R}$ in the linear regression model:

$$Y = \beta_{\text{Total}} X + \Gamma \mathcal{S} + \epsilon \tag{3}$$

where $\Gamma \in \mathbb{R}^q$ is the vector of coefficients for the adjustment set $\mathcal{S} \in \mathbb{R}^q$ satisfying the backdoor criterion, with $q = |\mathcal{S}|$ denoting the size of the adjustment set.

By contrast, the direct causal effect measures the influence of X on Y that is not transmitted through any mediators. This effect corresponds exclusively to the causal association associated with the direct edge $X \to Y$ in the causal graph. To isolate this effect, we seek to condition on the direct causes of Y other than X, denoted as $\mathcal{D} = \text{Pa}(Y) \setminus \{X\}$. This conditioning set blocks both indirect causal paths and backdoor paths, thus preserving the direct causal link from free from spurious associations. Accordingly, under the same assumption of no unobserved confounders assumptions, the direct causal effect as the coefficient $\beta_{\text{Direct}} \in \mathbb{R}$ in the linear regression model:

$$Y = \beta_{\text{Direct}} X + \Gamma \mathcal{D} + \epsilon \tag{4}$$

where $\Gamma \in \mathbb{R}^d$ is the vector of coefficients for the adjustment set $\mathcal{D} \in \mathbb{R}^d$ with $d = |\mathcal{D}|$ denoting the size of the adjustment set.

Importantly, in a linear SCM as defined in Equation 1, the direct causal effect β_{Direct} corresponds exactly to the structural coefficient β_{ij} of the edge $X_j \to X_i$ in the graph. This follows from the fact that conditioning on all other parents of Y isolates the direct contribution of X, which is precisely what the structural equation encodes. Consequently, estimating the direct causal effect provides a consistent estimator of the underlying structural parameter.

Alternatively, the total causal effect of X on Y in a linear SCM, as defined in Equation 1, can be estimated by summing, over all causal paths from X to Y, the product of the structural coefficients along each path. Those structural coefficients can be estimated as the direct causal effect. More formally, if $\mathcal{P}_{X \to Y}$ denotes the set of all causal paths from X to Y in the causal graph, and if each edge $(V_j \to V_i)$ along a path $\mathcal{P}$ has an associated structural coefficient β_{ij}, then:

$$\beta_{\text{Total}} = \sum_{\mathcal{P} \in \mathcal{P}_{X \to Y}} \left(\prod_{(V_j \to V_i) \in \mathcal{P}} \beta_{ij} \right) \tag{5}$$