Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Tariff Announcements, Implementation and Asset Prices: Evidence from Two Episodes of US Tariffs

Quoc Lap Nguyen

lapnq344@gmail.com

University of Economics and Law

Research Article

Keywords:

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11579670/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Tariff Announcements, Implementation and Asset Prices: Evidence from Two Episodes of US Tariffs

Quoc Lap Nguyen[1,2*]

[1*]University of Economics and Law, Ho Chi Minh City, Vietnam.
[2]Vietnam National University, Ho Chi Minh City, Vietnam.

Corresponding author(s). E-mail(s): lapnq344@gmail.com, lapnqc26602-1-nc@sdh.uel.edu.vn;

Abstract

Policy announcements move asset prices and expectations only to the extent that they are expected to be carried out, yet tariff announcements are often modified, delayed or withdrawn before they apply. This paper measures the gap between announced and applied US tariffs in two episodes, 2018-19 and 2025-26, using a register of 29 events built from primary documents and classified by what happened afterwards, and asks whether equity prices, implied volatility and professional forecasts reflect it. Most determinate announcements took effect in some form, usually within weeks, and the two episodes cannot be told apart at this sample size. Announcements that were later implemented coincided with lower equity prices and higher implied volatility, results that survive a pre-specified multiplicity adjustment although the intervals are wide. Evidence on whether markets foresaw non-implementation, on a second reaction at the effective date and on forecast errors is inconclusive, and minimum detectable effects are reported. A bounding model of announcement and implementation shows that treating announcements as fully implemented overstates the expected implemented fraction of announced tariffs by a factor of about 1.4 to 2.1. The findings argue for treating announced policies as probabilistic in investment, planning and forecasting.

Keywords: Tariff announcements, Policy credibility, Asset prices, Event study

JEL Classification: G14 , F13 , E65 , E37

1 Introduction

Asset prices and forecasts should respond to a policy announcement in proportion to the probability that the policy will be carried out and to the form in which it will be carried out. For tariffs the distance between announcement and implementation is large. Between a presidential statement and the first customs declaration at the new rate lies a sequence of legal steps, negotiations, exemptions and reversals, and a sizeable share of announced measures never completes it. In the trade conflict of 2018-19 and again from 2025, announcements of new duties on steel, on Chinese goods, on Mexican imports or on the whole set of trading partners were followed by delays, partial implementation, suspensions and withdrawals. Investors, forecasters and firms still had to act on the announcement, because the outcome was unknown at the time. The evidence that firms respond to the prospect of tariffs before the tariffs apply is substantial (Handley and Limão 2015; Crowley et al. 2018; Caldara et al. 2020), and it raises a question that has not been placed at the centre: how far apart are what was announced and what was applied, and do prices and forecasts reflect that distance?

A large literature shows that policy uncertainty affects asset prices and real activity (Rodrik 1991; Bloom 2009; Pástor and Veronesi 2012; Baker et al. 2016), and trade policy uncertainty in particular lowers investment, entry and trade (Handley and Limão 2015, 2017; Crowley et al. 2018; Caldara et al. 2020). Event studies of equity markets treat tariff announcements as news about future policy (Egger and Zhu 2020), and recent work on 2025 applies the same logic to the new round of tariffs (Rao et al. 2025; Switzer et al. 2026). The empirical work on the 2018-19 tariffs otherwise concentrates on measures that took effect. It documents near-complete pass-through into US prices and sizeable welfare losses (Amiti et al. 2019; Fajgelbaum et al. 2020; Cavallo et al. 2021), relocation and price effects in individual industries (Flaaen et al. 2020), effects on firm value through trade networks (Huang et al. 2023) and the broader course of the US-China conflict (Bown 2021). Theory explains why announced and applied tariffs can differ, because commitment is imperfect and tariffs are instruments of bargaining over market access (Kydland and Prescott 1977; Bagwell and Staiger 1999; Bown and Crowley 2013). To the best of the author's knowledge, systematic evidence on the distance between announced and implemented tariffs, collected across episodes with a common classification of outcomes, is limited. Studies of implemented tariffs condition on the measure having been carried out, and studies of announcement effects usually take the announcement as the policy change.

This paper treats announcement and implementation as two stages and asks three questions. The first concerns the size and timing of the gap: what share of announced tariffs took effect, in what form and after what delay, and how far collected duties fall below statutory rates. The second concerns pricing: do equity and volatility markets react differently to announcements that are later implemented than to those that are not, do they react again when a measure takes effect, and do rates, inflation compensation and the dollar move in the way a tariff shock would suggest. The third concerns forecasting: if a forecaster treats every announcement as a done deal, by how much does the expected tariff burden become overstated?

The answers rest on a register of 29 tariff events, 17 from 2018-19 and 12 from 2025-26, assembled from primary documents and classified ex post as executed in full,

executed with modification, not executed or de-escalation. Because such a sample is small and the analyst's degrees of freedom are not, the market tests were fixed in a locked design file before they were run on the register: five hypotheses in three families with Holm adjustment, placebo inference that does not rely on asymptotic approximations with few clusters, and a minimum detectable effect reported next to every estimate. Every analysis run after the lock is labelled as such, kept outside the multiplicity adjustment and counted in a ledger of 242 specifications, following the logic of pre-analysis plans and transparent reporting (Olken 2015; Christensen and Miguel 2018) and the evidence on selective reporting in applied economics (Brodeur et al. 2020).

Three findings emerge. First, the gap is large but not uniform. Of 24 announced tariff threats, 17 took effect in some form, usually within a month, and the remaining seven did not; the 2018-19 and 2025-26 episodes cannot be told apart at this sample size. Collected duties, on a simple proxy, also fall short of the statutory rate by roughly a third to a half. Second, announcements that were later implemented coincided with lower equity prices and higher implied volatility over the two-session window. These are the only equity results that pass the pre-specified multiplicity adjustment, although the bootstrap intervals for the raw estimates are wide and touch zero. Treasury yields fell as well, a result that is less robust, and the broad dollar index did not move by more than about a quarter of a percent. Third, the contrasts that would show whether markets anticipated implementation, namely executed versus not executed announcements, resolution without implementation, and a second reaction at the effective date, are too imprecise to support either answer. The minimum detectable effects reported in the paper make this limit explicit, and the absence of a significant contrast is not read as evidence of no difference (Altman and Bland 1995; Gelman and Carlin 2014). A bounding model that links the observed shares to the unobserved behaviour of the announcer shows that treating every announcement as fully implemented overstates the expected implemented fraction of announced tariffs by a factor of roughly 1.4 to 2.1, depending on how much of the announced rate survives in modified measures.

The paper contributes in three ways. The first is a framework and its measurement. It separates the announcement from the implementation, shows what asset prices should do when only the probability of implementation changes, and supplies a documented register of announcements, implementation outcomes, release times and lags across two episodes, together with a competing-risk description of how fast threats are carried out or dropped (Fine and Gray 1999) and a comparison of statutory with collected tariff rates. The second is evidence on how equity prices, implied volatility, yields and the dollar relate to the gap, with an account of what a design-locked event study with 11 independent clusters can and cannot identify, which offers a template for policy episodes in which the number of events can never be large. The third is a bridge to forecasting and scenario analysis. The bounding model translates a gap that is partially identified into a range for the forecast error that follows from assuming full implementation, in the spirit of inference for partially identified parameters (Imbens and Manski 2004), and it is complemented by a benchmark of professional forecast errors. The results are therefore relevant to investors, firms and forecasters who must plan around policies that have been announced but not yet applied.

The study is related to contemporaneous work on financial market reactions to US tariff announcements in 2025 (Rao et al. 2025) and on tariff uncertainty around the USMCA (Switzer et al. 2026). Those papers examine how markets respond to announcements; the present paper conditions on what happened to each announcement afterwards and places the 2018-19 and 2025-26 episodes in one frame. It is also related to research on the strategic reallocation of trade in response to the US-China conflict (Xu et al. 2025), whose results depend on which tariffs were actually applied and when, and more generally to any work that uses announced rates to measure protection or exposure, because such measures can inherit the gap documented here. The paper does not study trade flows, import prices or firm behaviour. Its evidence concerns the measurement of the gap and the pricing of announcements in financial markets, and the link from the gap to real trade outcomes is left to future work.

The remainder of the paper proceeds as follows. Section 2 sets out the conceptual background and the hypotheses. Section 3 describes the data and the event register. Section 4 presents the empirical strategy, the design lock and the bounding model. Section 5 reports the results. Section 6 discusses their interpretation and limits, and Section 7 concludes.

2 Framework

The framework has three parts: why announced and applied tariffs differ, what asset prices should do when the probability of implementation changes, and what forecasters should do. It leads to five testable hypotheses, stated at the end of the section.

2.1 Divergence

Three mechanisms make the distance between announcement and implementation a structural feature of tariff policy and not an accident. The first is commitment. A threat is credible only if it would be carried out when the counterpart does not comply, but carrying it out is costly, and the optimal plan announced in advance is generally not the plan that is optimal once the time comes to act (Kydland and Prescott 1977). An announcer who anticipates this has an incentive to announce measures that are more severe than those intended for implementation, and recipients who anticipate it discount the announcement.

The second mechanism is bargaining. Tariffs are instruments in negotiations over market access, and the theory of trade agreements treats applied tariffs as the outcome of an interaction between governments facing political pressure (Bagwell and Staiger 1999; Maggi and Rodriguez-Clare 1998). Evidence that the use of trade barriers responds to the incentives that govern the self-enforcement of trade agreements (Bown and Crowley 2013) implies that an announcement can be a move in a negotiation whose best outcome for the announcer is that the measure is never applied. The political economy of the 2018-19 conflict, in which retaliation was shaped by domestic political targeting (Fetzer and Schwarz 2021), points in the same direction: tariff measures are the product of strategic interaction and not of a fixed schedule.

The third mechanism is legal and administrative. US tariff measures rest on different statutory authorities, among them safeguard investigations, national-security

findings, trade-practice investigations and emergency economic powers, each with its own procedure, notice requirements and exposure to judicial review (Bown 2021). A measure announced informally can fail for reasons that have nothing to do with the announcer's intentions, and the lag between announcement and effective date depends on the route. The register used below records the authority for each event and the bounding model includes a separate probability of legal veto.

2.2 Asset prices

If announcements are informative about future tariffs, asset prices should respond to the change in the probability and size of implementation. Let X denote the effect on an asset price of full implementation of a measure, let π_{a^-} and π_a denote the market's probability of implementation just before and just after the announcement, and let π_{e^-} denote the probability just before the effective date. In a linear pricing approximation the announcement-day reaction is $(\pi_a - \pi_{a^-})X$ and, for a measure that does take effect, the effective-date reaction is $(1 - \pi_{e^-})X$. Only the unexpected part of a policy move changes prices, as in the response of stock prices to monetary policy surprises (Bernanke and Kuttner 2005), and the timing of news relative to implementation determines the dynamics of the response to fiscal changes (Mertens and Ravn 2012; Leeper et al. 2013). Three implications follow for tariffs. The announcement reaction of measures that are later implemented should have the sign of X, which is negative for equities. If markets can discriminate between measures that will and will not be implemented, the reaction should be larger for the former. And when a measure is withdrawn the probability of implementation falls to zero, so the reaction should have the opposite sign.

Equity prices and implied volatility are the natural outcomes. Volatility indices derived from option prices summarise market-implied uncertainty about equity returns (Whaley 2000), and uncertainty about government policy is priced in both equity returns and volatility (Pástor and Veronesi 2012; Baker et al. 2016; Bloom 2009). Government bond yields and inflation compensation are secondary outcomes. Breakeven inflation extracted from inflation-protected securities measures the compensation investors require for expected inflation but also contains risk and liquidity components (Gürkaynak et al. 2010), so a tariff shock that is read as a growth shock can lower it even though tariffs raise prices mechanically, a possibility that matches the finding that protectionism affects output and prices jointly (Barattieri et al. 2021).

2.3 Forecasts

Professional forecasters revise their projections in response to policy news, but revisions are gradual when information is costly to process (Coibion and Gorodnichenko 2015) and can overshoot when new information is extrapolated (Bordalo et al. 2020). If forecasters assumed that announced tariffs would be implemented in full and implementation proved partial, realised inflation, unemployment and output would have turned out better than forecast. The Survey of Professional Forecasters, which is available as a real-time data set (Croushore and Stark 2001), therefore offers a benchmark: the average signed forecast error across inflation, unemployment and output variables

in tariff episodes can be compared with its average in other periods. The benchmark is descriptive, no test is added after the lock, and the number of surveys in each episode is small.

2.4 Hypotheses

The locked design contains five hypotheses in three families. Family A tests whether announcements were priced and whether markets discriminated between them. Family B tests whether withdrawal and effective-date events carry information. Family C examines secondary outcomes.

H1. Announcements that were later executed, in full or with modification, are followed by a lower S&P 500 and a higher VIX over the two-session window $[t, t+1]$ (family A).

H2. The announcement reaction of executed announcements is more negative for the S&P 500 and more positive for the VIX than that of announcements that were not executed (family A). This contrast is declared informative only if the minimum detectable effect is at most 1.0% for the S&P 500 and 2.0 index points for the VIX.

H3. Resolution without implementation, defined as the union of not-executed announcements and de-escalations, is followed by a higher S&P 500 and a lower VIX (family B).

H4. On pairs whose effective date is not contaminated by other events, the announcement reaction is more negative for the S&P 500 and more positive for the VIX than the reaction at the effective date (family B).

H5. Executed announcements are associated with changes in the 10-year and 2-year Treasury yields, in 5-year breakeven inflation (predicted positive) and in the broad dollar index (family C).

Each family contains four tests and Holm adjustment is applied within families (Holm 1979), with hypotheses that were not run counted as non-rejections. Family A re-estimates results that had been seen before the lock, so its confirmatory status is weaker than that of families B and C, which were new at the lock.

3 Data

3.1 Event register

The register contains 29 events, 17 in the episode 2018-19 (identifiers A01 to A17) and 12 in the episode 2025-26 (E01 to E12). An event is an announcement that names a specific tariff measure or a specific change to an existing measure and can be traced to a dated primary document, such as a presidential statement, proclamation, executive order or agency notice. For each event the register records the announcement date, the statutory authority, the time of the first public release where it could be verified, the effective date of the first implemented measure, the date of resolution for measures that did not take effect, and the source documents. Table 1 lists the register with a short description of each measure.

Table 1: Event register: US tariff threats and measures, 2018-19 and 2025-26

ID	Auth.	Measure	Announced	Outcome	Effective	Lag
Panel A. Episode 2018-19						
A01	S201	Section 201 safeguards: solar (30%) and large washers (20 to 50%)	2018-01-22	Full	2018-02-07	16
A02	S232	Statement: 25% steel and 10% aluminum tariffs	2018-03-01	Modified	2018-03-23	22
A03	S232	Proclamations signed: 25% steel, 10% aluminum	2018-03-08	Modified	2018-03-23	15
A04	S301	Section 301 memorandum proposing tariffs on Chinese imports	2018-03-22	Modified	2018-07-06	106
A05	S301	Section 301 second tranche: 25% on $16bn	2018-08-07	Full	2018-08-23	16
A06	S301	Section 301 third tranche: 10% on $200bn, after Chinese retaliation	2018-09-17	Full	2018-09-24	7
A07	S301	New negotiations with China announced	2018-12-01	De-escalation	n/a	n/a
A08	S301	Post: stage 3 raised to 25%, stage 4 (about $300bn) prepared	2019-05-05	Full	2019-05-10	5
A09	S232	Autos and parts declared a security threat; talks directed	2019-05-17	Not executed	n/a	n/a
A10	S232	Canada and Mexico exempted from steel and aluminum duties	2019-05-17	De-escalation	2019-05-20	n/a
A11	IEEPA	IEEPA threat: 5% on all Mexican goods from June 10, rising to 25%	2019-05-30	Not executed	n/a	n/a
A12	S301	Plan: 10% on about $300bn (stage 4A September 1, 4B December 15)	2019-08-01	Modified	2019-09-01	31
A17	S301	Stage 4B (15% on about $156bn) suspended before taking effect	2019-08-13	Not executed	n/a	n/a
A13	S301	Directive: stages 1 to 3 raised to 30%, stage 4 to 15%	2019-08-23	Modified	2019-09-01	9
A14	S301	Suspension of October increases; phase one deal announced	2019-10-11	De-escalation	n/a	n/a
A15	S232	Announced increase of steel tariffs on Turkey to 50%	2019-10-14	Not executed	n/a	n/a
A16	S232	Announced reinstatement of steel and aluminum tariffs on Argentina and Brazil	2019-12-02	Not executed	n/a	n/a
Panel B. Episode 2025-26						
E01	IEEPA	Executive order: +10% on China	2025-02-01	Full	2025-02-04	3
E02	IEEPA	Executive order: 25% on Canada and Mexico (10% on Canadian energy)	2025-02-01	Modified	2025-03-04	31
E03	S232	Proclamations: 25% on steel and aluminum (all countries)	2025-02-09	Full	2025-03-12	31
E04	IEEPA	Plan announced: reciprocal tariffs from April 2	2025-02-13	Modified	2025-04-02	48
E05	S232	25% on autos and auto parts	2025-03-26	Full	2025-04-03	8
E06	IEEPA	Reciprocal tariffs: 10% baseline and country rates (China 34%, EU 20%)	2025-04-02	Modified	2025-04-05	3

continued on next page

ID	Auth.	Measure	Announced	Outcome	Effective	Lag
E07	IEEPA	Threat: +50% on China unless retaliation reversed by April 8	2025-04-07	Full	2025-04-09	2
E08	IEEPA	US-China joint statement: tariffs cut for 90 days	2025-05-12	De-escalation	2025-05-14	n/a
E09	n/a	50% tariff on EU paused to July 9 (date of the pause)	2025-05-23	Not executed	n/a	n/a
E10	n/a	Letters: 30% on EU and Mexico	2025-07-12	Unclassified	n/a	n/a
E11	S122	Section 122 global tariff of 10% signed after the Supreme Court decision	2026-02-20	Full	2026-02-24	4
E12	S122	Announced raise of Section 122 tariff from 10% to 15%	2026-02-21	Not executed	n/a	n/a

Note: Outcome labels are assigned ex post (Full: executed in full; Modified: executed with modification) and are never used as ex ante information. Lag is the number of calendar days from announcement to the effective date of the first implemented measure. Auth. is the statutory authority (S201, S232, S301: Sections 201, 232 and 301; IEEPA: International Emergency Economic Powers Act; S122: Section 122). The measure column gives a short description and is not a legal citation. Event E10 is unclassified and excluded from the locked tests (post-lock variant in Appendix Table A1). Release times and verification status are in Appendix Table A6.

Each event carries an outcome label assigned after the fact. A measure is *executed in full* if it took effect substantially as announced, *executed with modification* if it took effect with a different rate, scope or date, and *not executed* if it had not taken effect by the end of the sample or was withdrawn. A *de-escalation* is an announcement that suspends, reduces or exempts existing tariffs. One event (E10, letters announcing a 30% tariff on the European Union and Mexico) is left unclassified because the later arrangement with the European Union could be read as a modification or as non-execution. It is excluded from the locked tests and examined as a post-lock variant. The remaining 28 events are called determinate. Of these, 24 are *threats*, meaning announcements of future tariffs that were executed, modified or not executed, and four are de-escalations. Outcome labels use information that was unavailable at the time of the announcement, and they enter the analysis only as grouping variables for the event windows, never as predictors.

Timing matters because announcements outside trading hours are priced in the next session. For each event the register records a release-time bucket (before the open, during the session, after the close, or over a weekend) where a timestamped source could be located. Fifteen of the 29 events have no verified bucket, and in the locked analysis only events verified as released after the close are shifted by one session; events without a verified bucket are treated as released within the session. A later check of release times, completed on 16 October 2026 after the lock, is reported in Appendix Table A6; it does not alter the locked analysis and is used only in the sensitivity variants. Partial verification is a limitation, and its consequences are bounded in Section 5.4 by a variant that places all events with a blank release time after the close.

3.2 Market and tariff data

Daily closing values of the S&P 500, the VIX, the 10-year and 2-year constant-maturity Treasury yields, the 5-year breakeven inflation rate and the nominal broad US dollar index are taken from the Federal Reserve Economic Data (FRED) service of the Federal Reserve Bank of St. Louis. The forecast benchmark uses the Survey of Professional Forecasters of the Federal Reserve Bank of Philadelphia, restricted to surveys from 1990 onward and to targets outside the COVID period, with at least four of seven variables available per survey: consumer prices, core consumer prices, core personal consumption expenditure prices, the GDP deflator, unemployment, real GDP and real consumption. Statutory and realised effective tariff rates by quarter come from the Budget Lab at Yale. Customs receipts are from the Monthly Treasury Statement of the US Department of the Treasury and goods imports from the National Income and Product Accounts of the US Bureau of Economic Analysis. The September 2026 customs receipt and the third-quarter 2026 national accounts were not available when the data were assembled.

4 Methods

4.1 Measuring the gap

The first block is descriptive. Implementation shares are computed for determinate events and for threats, with Wilson confidence intervals that retain coverage with small counts (Brown et al. 2001), and are compared across episodes with the Fisher exact test (Fisher 1922). The lag is the number of calendar days from announcement to the effective date of the first implemented measure; medians and interquartile ranges are reported with bootstrap confidence intervals (Efron 1979), and lags are compared across episodes with the Mann-Whitney test. Because some threats are withdrawn and others are still pending, the time to implementation is treated as a competing-risk problem. Let T denote the time from announcement to the first of two events, implementation or resolution without implementation. The cumulative incidence of cause k is $F_k(t) = \Pr(T \leq t,\ \text{cause} = k)$, estimated nonparametrically at horizons of 15, 30, 60, 90, 180 and 365 days (Fine and Gray 1999), with confidence intervals obtained by bootstrapping over threats. The main variant uses the 21 threats whose outcome date is known; a second variant treats the three remaining threats as censored at the end of the sample.

4.2 Event windows

Let $\Delta y_{j,s}$ denote the change in outcome j on session s: the log return in percent for the S&P 500 and the dollar index, the change in index points for the VIX and the change in basis points for yields and breakeven inflation. Following the standard event-study practice of standardising by recent volatility to account for event-induced variance (Brown and Warner 1985; Boehmer et al. 1991), the demeaned change is divided by the standard deviation over the preceding 60 sessions, lagged by one session so that

the event does not enter its own scale:

$$z_{j,s} = \frac{\Delta y_{j,s} - \overline{\Delta y}_j}{\hat{\sigma}_{j,s-1}}. \tag{1}$$

For event e let $\tau(e)$ denote the first session in which the announcement can be traded. The window outcome is the sum over two sessions,

$$Y_{j,e} = \sum_{h=0}^{1} \Delta y_{j,\tau(e)+h}, \tag{2}$$

and $Y^z_{j,e}$ is defined in the same way from $z_{j,s}$. A two-session window allows for incomplete pricing on the first day and for the release-time uncertainty described above. Events that fall on the same date contribute one observation. Events within five sessions of one another are assigned to the same cluster. For a group G of event days, the estimate is the mean $\bar{Y}_{j,G} = |G|^{-1} \sum_{d \in G} Y_{j,d}$. H2 is the difference between the means for executed and not executed announcements, H3 uses the resolution date in place of the announcement date, and H4 uses the difference $D_i = Y^{\text{ann}}_{j,i} - Y^{\text{eff}}_{j,i}$ for each clean pair i. A pair is clean when the effective date lies more than two sessions from every other announcement or resolution date.

4.3 Inference

Two procedures are combined because each addresses a different weakness of small-sample event studies. Confidence intervals are 95% percentile intervals from a cluster bootstrap with 50,000 draws, resampling clusters of events (Efron 1979). With a small number of clusters, cluster-robust methods can be unreliable and bootstrap refinements are valuable (Cameron et al. 2008; Conley and Taber 2011), and clustering also addresses cross-sectional dependence between events in close succession (Kolari and Pynnönen 2010). The second procedure is a calendar-shift placebo. The observed mean of the standardised outcome is compared with the means obtained when the event dates are moved to other calendar dates by a common shift. Local shifts range from 10 to 250 sessions and global shifts use all feasible shifts. For K shifts, the two-sided p-value is

$$p = \frac{1 + \sum_{k=1}^{K} \mathbf{1}\{|\bar{z}^{(k)}| \geq |\bar{z}^{\text{obs}}|\}}{K + 1}. \tag{3}$$

The placebo compares the event window with the distribution of windows of the same length drawn from non-event periods and therefore reflects fat tails and volatility clustering, whereas the bootstrap reflects sampling variation in the mean across events. The two need not agree in a small sample, and both are reported. Holm adjustment is based on the local p-value.

Power is reported as the minimum detectable effect at 80% power and 5% size, $\text{MDE}_{80} = 2.8\,\hat{\sigma}_{\text{placebo}}$, where $\hat{\sigma}_{\text{placebo}}$ is the standard deviation of the placebo means in raw units. The number 2.8 is the sum of the normal quantiles for the two-sided 5% test and for 80% power. Reporting the MDE next to each estimate serves two purposes.

It prevents a wide interval that contains zero from being read as evidence of no effect (Altman and Bland 1995), and it indicates the size of effect that would be needed for a significant estimate to be informative about magnitude and not only about sign (Gelman and Carlin 2014). The same logic motivates the pre-specified informativeness threshold for H2.

4.4 Design lock

The hypotheses, outcomes, windows, clustering rule, inference procedures and family structure were written to a design file before the locked tests were run, and the tests were executed with that file unchanged. The lock is an internal commitment and not a registration with an external registry, and its credibility therefore rests on the disclosure of what was done before and after it. About 60 specifications were examined in exploratory work before the lock (the exact number cannot be recovered); 12 hypothesis-by-outcome slots make up the locked battery; and 170 specifications were computed after the lock, in sensitivity, model and follow-up analyses. The total of 242 is an upper bound on the specifications that the data have been exposed to, and Appendix Table A4 gives the breakdown. Post-lock variants never enter the Holm families.

4.5 Bounding model

A model with few assumptions links the observed shares of outcomes to the unobserved behaviour of the announcer. It is used as a bounding device and not as a structural estimate. Announcers are of two types. A resolute announcer implements with probability π. A soft announcer uses the threat to bargain; with probability ρ bargaining fails and a modified measure results, and otherwise no measure results. Independently of type, a legal veto with probability λ blocks the measure. The probabilities of the three observable outcomes are then

$$\begin{aligned} \Pr(\text{full}) &= (1-\lambda)\pi, \\ \Pr(\text{modified}) &= (1-\lambda)(1-\pi)\rho, \\ \Pr(\text{not executed}) &= \lambda + (1-\lambda)(1-\pi)(1-\rho). \end{aligned} \tag{4}$$

Three moments, the shares $\hat{p}_{\text{full}}$, $\hat{p}_{\text{mod}}$ and $\hat{p}_{\text{not}}$, are observed and there are three parameters, but they sum to one and the system has only two independent equations. The parameters are therefore not point identified. For any $\lambda \in [0, \hat{p}_{\text{not}}]$ the model implies

$$\pi(\lambda) = \frac{\hat{p}_{\text{full}}}{1-\lambda}, \qquad \rho(\lambda) = \frac{\hat{p}_{\text{mod}}}{(1-\lambda)\left(1-\pi(\lambda)\right)}, \tag{5}$$

so that the identified set is a curve indexed by the legal-veto share. Sampling uncertainty is incorporated by the cluster bootstrap, retaining the draws for which the implied parameters lie in the unit interval (Imbens and Manski 2004).

The forecast implication depends on how much of the announced rate survives in modified measures, a share denoted $\kappa \in [0, 1]$. The expected implemented fraction of the announced tariff is $F(\kappa) = \hat{p}_{\text{full}} + \kappa\,\hat{p}_{\text{mod}}$, and the overstatement from assuming

full implementation is $1/F(\kappa)$. The share κ has not been collected from the legal texts and is varied over an illustrative grid.

5 Results

5.1 The gap

Of the 29 events, 28 have a determinate outcome. Panel B of Table 2 shows nine executed in full, eight executed with modification, seven not executed and four de-escalations. Seventeen of the 28 determinate events (61%, 95% confidence interval 42% to 76%) and 17 of the 24 threats (71%, 51% to 85%) were implemented in some form. The median lag from announcement to the effective date of the first implemented measure was 15 days (interquartile range 5 to 31 days, 17 measures, bootstrap interval for the median 6 to 31 days).

Table 2 The announcement-implementation gap: implementation rates and lags

	All determinate events		Threats only		
Scope	Impl./N	Share % [CI]	Impl./N	Share % [CI]	Lag, median [IQR] (n)
Panel A. Implementation shares and lags					
All events	17/28	61 [42, 76]	17/24	71 [51, 85]	15 [5, 31] (17)
Episode 2018-19	9/17	53 [31, 74]	9/14	64 [39, 84]	16 [9, 22] (9)
Episode 2025-26	8/11	73 [43, 90]	8/10	80 [49, 94]	6 [3, 31] (8)
IEEPA	5/7	71 [36, 92]	5/6	83 [44, 97]	3 [3, 31] (5)
S122	1/2	50 [9, 91]	1/2	50 [9, 91]	4 [4, 4] (1)
S201	1/1	100 [21, 100]	1/1	100 [21, 100]	16 [16, 16] (1)
S232	4/8	50 [22, 78]	4/7	57 [25, 84]	18 [13, 24] (4)
S301	6/9	67 [35, 88]	6/7	86 [49, 97]	12 [8, 27] (6)
Unknown	0/1	0 [0, 79]	0/1	0 [0, 79]	n/a
Panel B. Composition of outcomes (number of events)					
Sample	Full	Modified	Not exec.	De-esc.	Determinate (all)
2018-19	4	5	5	3	17 (17)
2025-26	5	3	2	1	11 (12)
All	9	8	7	4	28 (29)

Note: Scope rows by authority use the statutory authority of the measure (Unknown: authority not recorded). Share % [CI] is the implemented share in percent with a Wilson 95% interval. Lag is in calendar days. Impl./N is the number implemented (executed in full or with modification) over the number of events. Determinate events exclude the unclassified event. Threats are measures announced as future tariffs (executed in full, executed with modification or not executed); de-escalation announcements are determinate events but not threats, so the threat shares are the relevant measure of the gap. Lags run from announcement to the effective date; the number of implemented measures is in parentheses. In Panel B, Not exec. is not executed, De-esc. is de-escalation, and the last column gives determinate events (all events including the unclassified one). Episode comparison (2018-19 versus 2025-26): Fisher exact test of the implemented share, $p = 0.435$ (all determinate events) and 0.653 (threats); Mann-Whitney test of lags, $p = 0.311$. Bootstrap intervals for median lags are in Table 8. With these sample sizes the comparison cannot distinguish the episodes, and authority and episode are confounded.

The two episodes look similar. In 2018-19, 9 of 14 threats were implemented (64%, 39% to 84%), and in 2025-26, 8 of 10 (80%, 49% to 94%). The Fisher test does not reject equality ($p = 0.435$ for all determinate events and 0.653 for threats), nor does the Mann-Whitney test for lags ($p = 0.311$). The comparison has little power with these counts, and the confidence intervals overlap widely; the data cannot show that the episodes differ or that they are the same. Authority and episode are confounded, since the Section 122 route and six of the seven emergency-power events belong to the second episode. Within the table, measures under the emergency authority were implemented with a median lag of three days, against 12 days under Section 301 and 18 days under Section 232, but the cells contain between one and six measures each and the intervals are wide. Figure 1 shows the timeline of each event.

The competing-risk estimates show how quickly threats resolved in either direction. Among the 21 threats with a known outcome date, the cumulative incidence of implementation was 43% after 15 days, 57% after 30 days, 76% after 60 days and 81% after 180 days, while resolution without implementation accounted for 14% by 60 days and 19% at 180 days (Appendix Table A7 and Figure 2). If the three threats without a known outcome date are treated as censored at the end of the sample, the corresponding implementation shares are 38%, 50%, 67% and 71%. The pattern differs in the second episode, where implementation reached 56% after 15 days and 89% after 60 days among nine threats, against 33% and 67% among twelve threats in the first, but the intervals overlap throughout. Implementation, when it happens, is therefore quick: threats that are going to be implemented are mostly implemented within two months. Seven of the 24 threats were not executed, but three of them have no recorded resolution date, which is why the competing-risk estimate of formal resolution (19%) is lower than the share not executed (29%).

5.2 Market reactions

Table 3 reports the locked battery. Announcements that were later executed, in full or with modification, are followed over the two-session window by a fall of 1.39% in the S&P 500 (95% bootstrap interval from −3.12 to 0.08; MDE_{80} 0.93) and a rise of 2.86 points in the VIX (−0.15 to 6.45; MDE_{80} 1.69), using 16 event days in 11 clusters. Both signs match H1. The local placebo p-values are 0.007 and 0.006, the global values 0.004 and 0.004, and the Holm-adjusted p-value is 0.023 for each. The leave-one-cluster-out ranges of the mean (Table 5), from −1.82 to −0.51 for the S&P 500 and from 0.89 to 3.58 for the VIX, exclude zero, so that no single cluster drives the result.

The evidence for H1 is supportive but not decisive. The placebo test and the leave-one-cluster-out ranges favour a non-zero reaction, whereas the bootstrap intervals for the raw means are wide and end at or just below zero. The p-values and the bootstrap address different questions, as noted in Section 4.3, and with 11 clusters neither is definitive. The estimates are best read as a statement on sign and rough magnitude, in the range of one to three percent for the equity index, and not as a precise elasticity. Family A also re-estimates results that were seen before the lock, so the adjusted p-values are less informative than they would be for a new hypothesis.

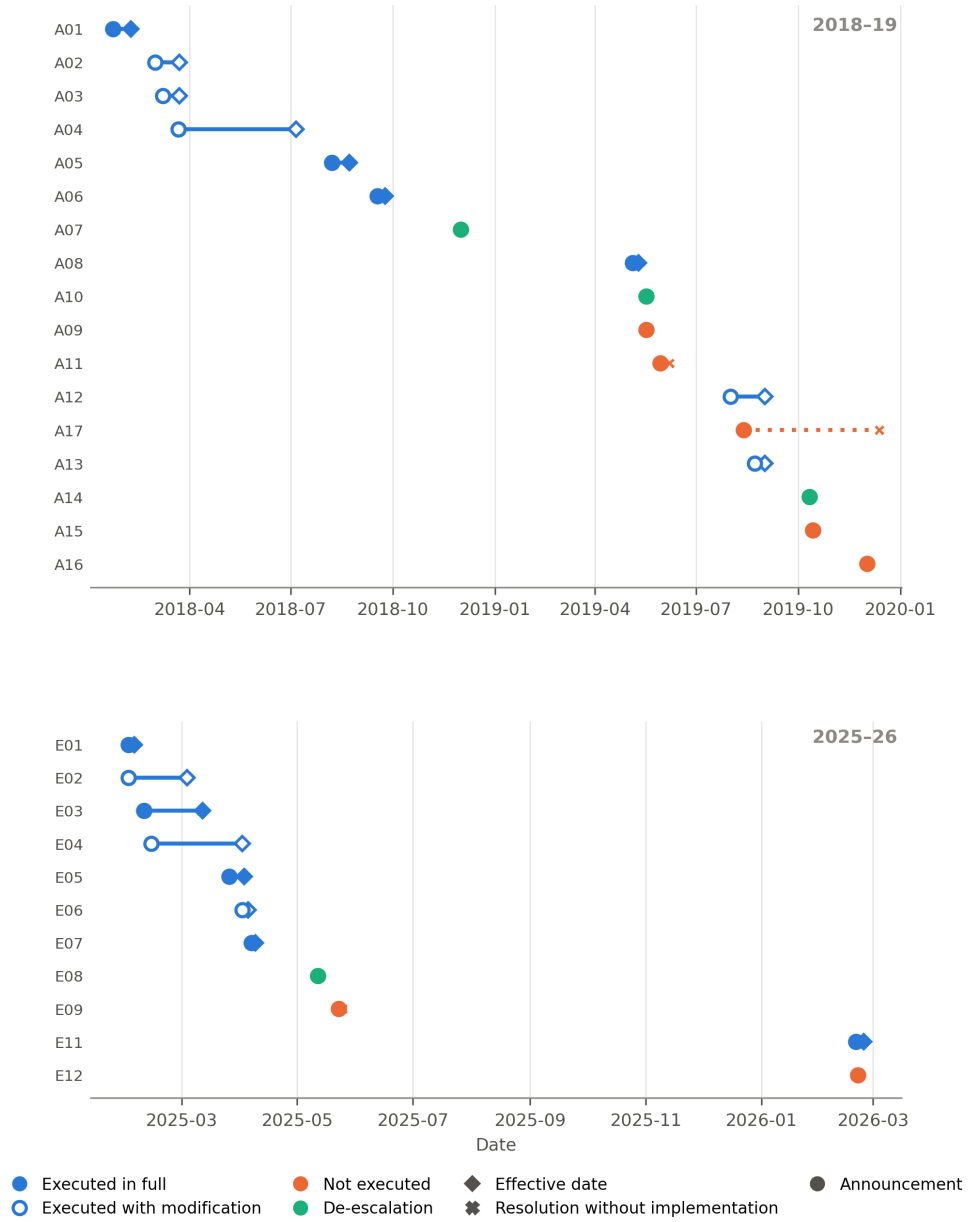

Fig. 1 Timeline of announcements, effective dates and resolutions. Each row is one event, ordered by announcement date within each episode, and colour distinguishes the outcome classes as shown in the legend of the figure. The circle marks the announcement (open for measures executed with modification, filled otherwise), the solid line runs to the effective date (diamond) for executed measures, and the dotted line runs to the resolution date (cross) for measures not executed and for de-escalations. The unclassified event E10 is omitted.

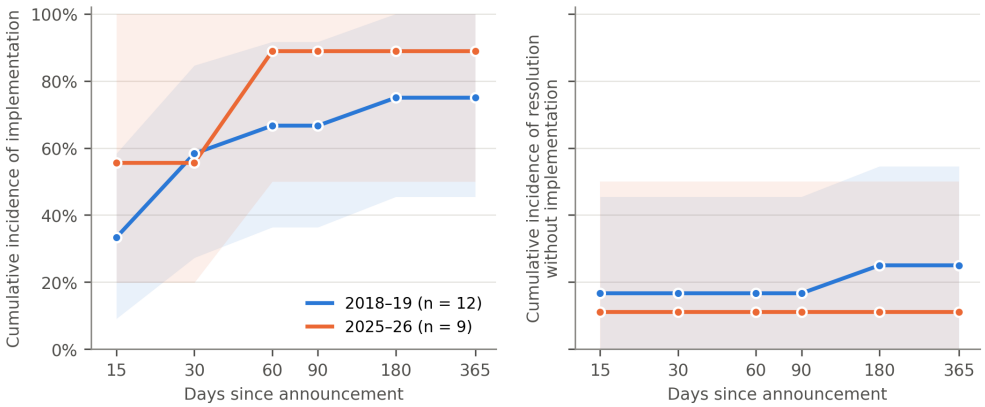

Fig. 2 Cumulative incidence of implementation (left) and of resolution without implementation (right), by episode, for threats with a known outcome date. Lines show the competing-risk estimate at horizons of 15, 30, 60, 90, 180 and 365 days since the announcement (logarithmic horizontal axis), markers the evaluation horizons, and shaded bands the 95% bootstrap intervals. The legend gives the number of threats in each episode.

The contrasts designed to reveal whether markets distinguished measures that were later executed from those that were not do not deliver a verdict. For H2 the difference is −0.83 percentage points for the S&P 500 (−2.80 to 0.98) and 2.36 points for the VIX (−1.14 to 6.23), with signs in the predicted direction, but the minimum detectable effects of 2.19 and 3.76 are more than double the pre-specified thresholds of 1.0% and 2.0 points. By the rule written before the test, H2 is uninformative. The sample contains six event days of announcements that were not executed, and the data neither support nor contradict a market ability to foresee non-execution.

The same limit holds for resolutions. After resolution without implementation the S&P 500 rose by 0.39% (−0.80 to 1.68) and the VIX fell by 0.94 points (−2.25 to 0.52), both with the predicted sign, on eight event days in seven clusters, but the minimum detectable effects of 1.68 and 2.84 exceed the estimates, and the Holm-adjusted p-values are 1.000. The estimates for H3 are in line with the idea that withdrawal relieves part of the burden of an earlier announcement, though they could equally reflect noise.

Among the secondary outcomes, executed announcements coincide with lower Treasury yields: −3.59 basis points for the 10-year yield (−6.81 to −0.71) and −4.50 basis points for the 2-year yield (−7.65 to −1.17), with Holm-adjusted p-values of 0.032 and below 0.001. The intervals exclude zero and the sign is that of a growth-scare or safe-asset response. Breakeven inflation falls by 2.69 basis points (−5.50 to 0.10; Holm-adjusted p-value 0.051), the opposite of the predicted sign, which is compatible with demand and risk-premium components dominating the mechanical price effect of tariffs (Gürkaynak et al. 2010; Barattieri et al. 2021). The broad dollar index changes by −0.05% (−0.25 to 0.15), and the interval excludes moves of more than about a quarter of a percent. The yield results depend on how the placebo reference distribution is chosen: they are significant when local shifts are used (the locked rule), but replacing local with global p-values in the same Holm procedure leaves no adjusted p-value in family C below 0.05, as can be computed from the global column of Table 3.

Table 3 Locked test battery

Test	Days (cl.)	Estimate	95% CI	MDE80	p local	p global	Holm p
Family A (H1 and H2)							
H1 S&P 500	16 (11)	−1.39%	[−3.12, 0.08]	0.93	0.007	0.004	0.023
H1 VIX	16 (11)	2.86 pts	[−0.15, 6.45]	1.69	0.006	0.004	0.023
H2 S&P 500	16 (11); 6 (6)	−0.83%	[−2.80, 0.98]	2.19	0.790	0.748	0.790
H2 VIX	16 (11); 6 (6)	2.36 pts	[−1.14, 6.23]	3.76	0.173	0.138	0.345
Family B (H3 and H4)							
H3 S&P 500	8 (7)	0.39%	[−0.80, 1.68]	1.68	0.896	0.894	1.000
H3 VIX	8 (7)	−0.94 pts	[−2.25, 0.52]	2.84	0.444	0.427	1.000
H4 S&P 500	10 (9)	−0.98%	[−3.43, 1.14]	2.08	0.436	0.400	1.000
H4 VIX	10 (9)	3.72 pts	[0.34, 8.06]	3.63	0.050	0.024	0.200
Family C (H5, secondary outcomes)							
H5 10-year yield	16 (11)	−3.59 bp	[−6.81, −0.71]	4.00	0.011	0.030	0.032
H5 2-year yield	16 (11)	−4.50 bp	[−7.65, −1.17]	4.05	<0.001	0.020	<0.001
H5 5-year breakeven	16 (11)	−2.69 bp	[−5.50, 0.10]	2.95	0.026	0.039	0.051
H5 Broad dollar	16 (11)	−0.05%	[−0.25, 0.15]	0.27	0.819	0.800	0.819

Note: Two-session windows [t, t+1] in raw units (log % for the S&P 500 and the dollar index, index points for the VIX, basis points for yields). Days are distinct event days (cl.: clusters in parentheses); for H2 the two groups are executed and not executed announcements. Confidence intervals are 95% cluster-bootstrap percentile intervals (events within five sessions of one another form one cluster). The p-values are two-sided calendar-shift placebo p-values on the standardised outcome (local: shifts of 10 to 250 sessions; global: all shifts); Holm adjustments use the locked family sizes (m = 4 in each family; hypotheses not run count as p = 1) and are based on the local p-value. MDE80 is 2.8 times the placebo standard deviation (80% power, 5% size) in raw units. H2 is informative only if MDE80 is at most 1.0% (S&P 500) or 2.0 points (VIX). Predicted signs: negative for the S&P 500 and positive for the VIX in H1, H2 and H4, the reverse in H3, positive for breakeven inflation in H5, none for the other H5 outcomes. Family A re-estimates results seen before the lock; families B and C are new. Leave-one-cluster-out ranges are in Table 5. A non-rejection or an interval containing zero is not evidence of no effect.

They are therefore the least robust of the confirmatory results. Figure 3 summarises the estimates together with their minimum detectable effects.

5.3 Effective date

H4 asks whether the announcement is priced more strongly than the effective date. On ten clean pairs in nine clusters, the S&P 500 fell by 1.02% on announcement (−2.26 to −0.05) and by 0.04% at the effective date (−1.50 to 1.62), a difference of −0.98 (−3.45 to 1.13). The VIX rose by 2.23 points at the announcement (0.34 to 4.52) and fell by 1.49 points at the effective date (−4.20 to 0.56), a difference of 3.72 (0.33 to 8.13). Table 4 shows these results. The signs match H4, and the VIX difference has an interval that excludes zero and a global placebo p-value of 0.024, but the local p-value used for adjustment is 0.050 and the Holm-adjusted value 0.200. The minimum detectable effects of the differences are 3.27 percentage points and 5.59 index points in the post-lock run reported in this table, so only differences of that order could have been detected reliably.

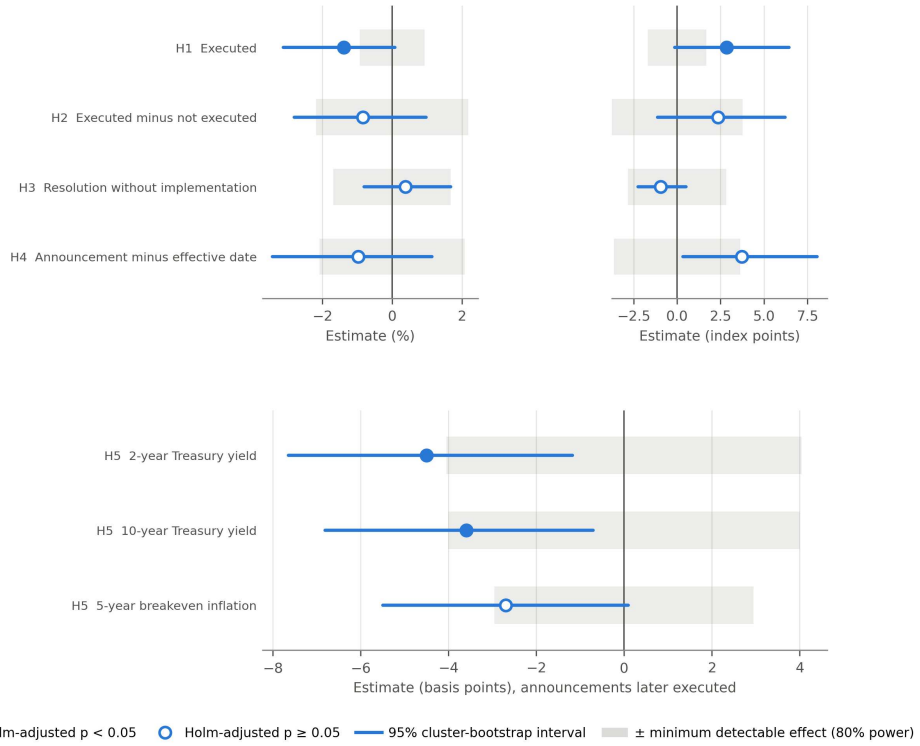

Fig. 3 Locked estimates with 95% cluster-bootstrap intervals. The upper panels show the S&P 500 (two-day log return, percent) and the VIX (two-day change, index points) for H1 to H4. The lower panel shows the two-year and ten-year Treasury yields and five-year breakeven inflation (basis points) for announcements that were later executed (H5). Filled markers indicate a Holm-adjusted p-value below 0.05 and open markers a value of at least 0.05. Grey bands mark plus and minus the minimum detectable effect at 80% power.

Panel B of Table 4 is exploratory and compares the effective-date reaction with what would be expected from mean reversion after a move of the same size on non-event dates. For the VIX the observed effective-date change of −1.82 points compares with a matched-null mean of −0.01, an excess of −1.80 points on nine pairs, and the excess falls to −0.50 points on seven pairs when windows overlapping unlisted policy dates are removed. For the S&P 500 the excess is zero on ten pairs and −0.63 on eight pairs. The pattern is consistent with announcement-day volatility that partly reverses around implementation and not with a fresh shock at the effective date, but no interval accompanies these numbers and the number of pairs is small. Figure 4 shows the pairs one by one.

5.4 Robustness

Table 5 and Figure 5 summarise the sensitivity of H1 to H4 to choices that could not be fixed in advance: the release time of individual events, the length of the window, the label of E10 and the exclusion of windows that overlap policy dates outside the

Table 4 Implementation news: announcement versus effective-date reactions

	S&P 500 (%)	VIX (points)
Quantity	estimate [95% CI]	estimate [95% CI]
Panel A. Reactions in matched pairs		
Clean pairs: 10 pairs in 9 clusters		
Announcement	−1.02 [−2.26, −0.05]	2.23 [0.34, 4.52]
Effective date	−0.04 [−1.50, 1.62]	−1.49 [−4.20, 0.56]
Difference	−0.98 [−3.45, 1.13]	3.72 [0.33, 8.13]
MDE80 (difference)	3.27	5.59
All pairs, including contaminated: 16 days in 11 clusters		
Announcement	−1.39 [−3.11, 0.08]	2.86 [−0.16, 6.48]
Effective date	−0.97 [−1.89, 0.19]	1.03 [−1.27, 3.18]
Difference	−0.42 [−2.12, 1.23]	1.83 [−0.37, 4.14]
MDE80 (difference)	2.42	3.28

Panel B. Effective-date reaction relative to mean reversion

Sample and outcome	Pairs	Obs. ann.	Null ann.	Obs. eff.	Null eff.	Excess
Clean, S&P 500 (%)	10	−1.02	−0.76	−0.04	−0.04	−0.00
Clean, VIX (points)	9	1.77	0.66	−1.82	−0.01	−1.80
Overlap excl., S&P 500 (%)	8	−1.02	−0.77	−0.71	−0.08	−0.63
Overlap excl., VIX (points)	7	1.16	0.30	−0.56	−0.06	−0.50

Note: Units are log % for the S&P 500 and index points for the VIX. Clean pairs are announcement and effective-date pairs whose effective date is more than two sessions from every other announcement or resolution. The difference in Panel A is announcement minus effective date; the H4 estimate in Table 3 is the clean-pair difference. The MDE80 of the difference shown here comes from a post-lock run and differs from the locked value in Table 3. In Panel B, Obs. is the observed reaction, Null the mean over calendar dates whose preceding two-session move matches the observed announcement move, ann. the announcement and eff. the effective date; Excess is the observed effective-date reaction minus its matched-null mean (the null is the mean after a move of the same size); Overlap excl. is the clean sample after dropping windows overlapping unlisted policy dates. Panel B is exploratory and outside the Holm families.

register. For H1 no variant changes the sign. The S&P 500 estimate ranges from −2.03 to −0.66 across nine timing and window variants, with the bootstrap interval excluding zero in one of them, and the VIX estimate ranges from 1.44 to 4.05, with the interval excluding zero in three. The decision variants move the S&P 500 estimate between −1.39 and −0.63 and the VIX estimate between 1.51 and 2.86. The lower end of both ranges comes from the extreme variant that places every event with a blank release time after the close; it is a bound and not a best estimate, and it roughly halves the effects (S&P 500 −0.63, interval −1.46 to 0.19; VIX 1.51, interval −0.17 to 3.17). Removing windows that overlap unlisted policy dates gives −1.35 and 2.58.

For H4 the VIX difference stays between 3.35 and 4.09 across timing and window variants, with the interval excluding zero in all nine, and between 1.52 and 3.72 across decision variants (three of four exclude zero), but it falls to 2.12 with an interval that includes zero once overlapping windows are removed. The S&P 500 difference changes sign in one of four decision variants, the extreme one. The leave-one-cluster-out ranges for H4 do not include zero in either outcome. These results point to a stable direction for H1 and for the volatility part of H4, a sensitivity of magnitude to the unverified

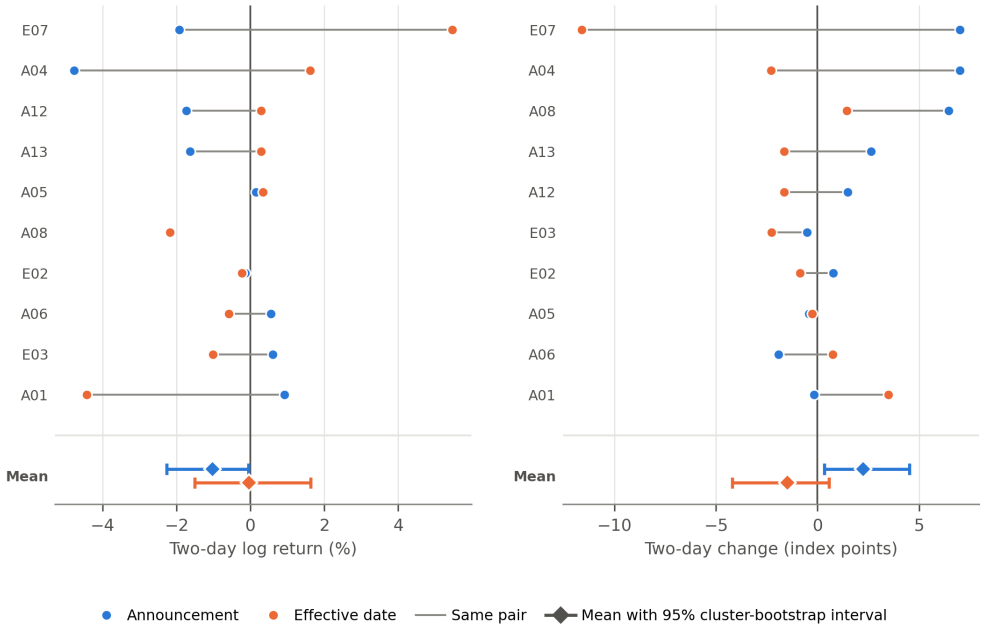

Fig. 4 Announcement and effective-date reactions in clean pairs (H4). Each row is one pair, with a dot at the announcement reaction (blue) and at the effective-date reaction (orange) joined by a grey line, ordered by the size of the difference; the left panel shows the two-day log return of the S&P 500 (percent) and the right panel the two-day change in the VIX (index points). The last row gives the means with 95% cluster-bootstrap intervals.

release times, and no robust finding for the equity part of H4. Full variant tables are in Appendix Tables A1 to A3.

5.5 Forecasts and power

Table 6 presents the forecast benchmark and the power calculation. Outside the tariff episodes the directional index of forecast errors averages −0.117 RMSE units across 125 surveys (interval −0.230 to −0.004), which means that realised outcomes were slightly better than the forecasts on average. In the 2018-19 episode the index is −0.094 on eight surveys (−0.383 to 0.196) and in the 2025-26 episode 0.002 on six surveys (−0.694 to 0.698). The tariff episodes are statistically indistinguishable from the baseline and from zero, and the minimum detectable effects of 0.343 and 0.758 exceed the size of the baseline mean. If forecasters had treated announced tariffs as fully implemented, a negative index would be expected in these periods; the point estimate for 2018-19 has that sign but no inference is possible, and for 2025-26 the interval is too wide to say anything.

Panel B of Table 6 shows why the event tests are limited. A one-sample test of a 1.0% two-day move in the S&P 500 needs about 18 independent clusters at 80% power, and a contrast between two groups needs about 36 per group. For the VIX a two-point move needs 13 clusters, and a contrast 25 per group. The locked battery has

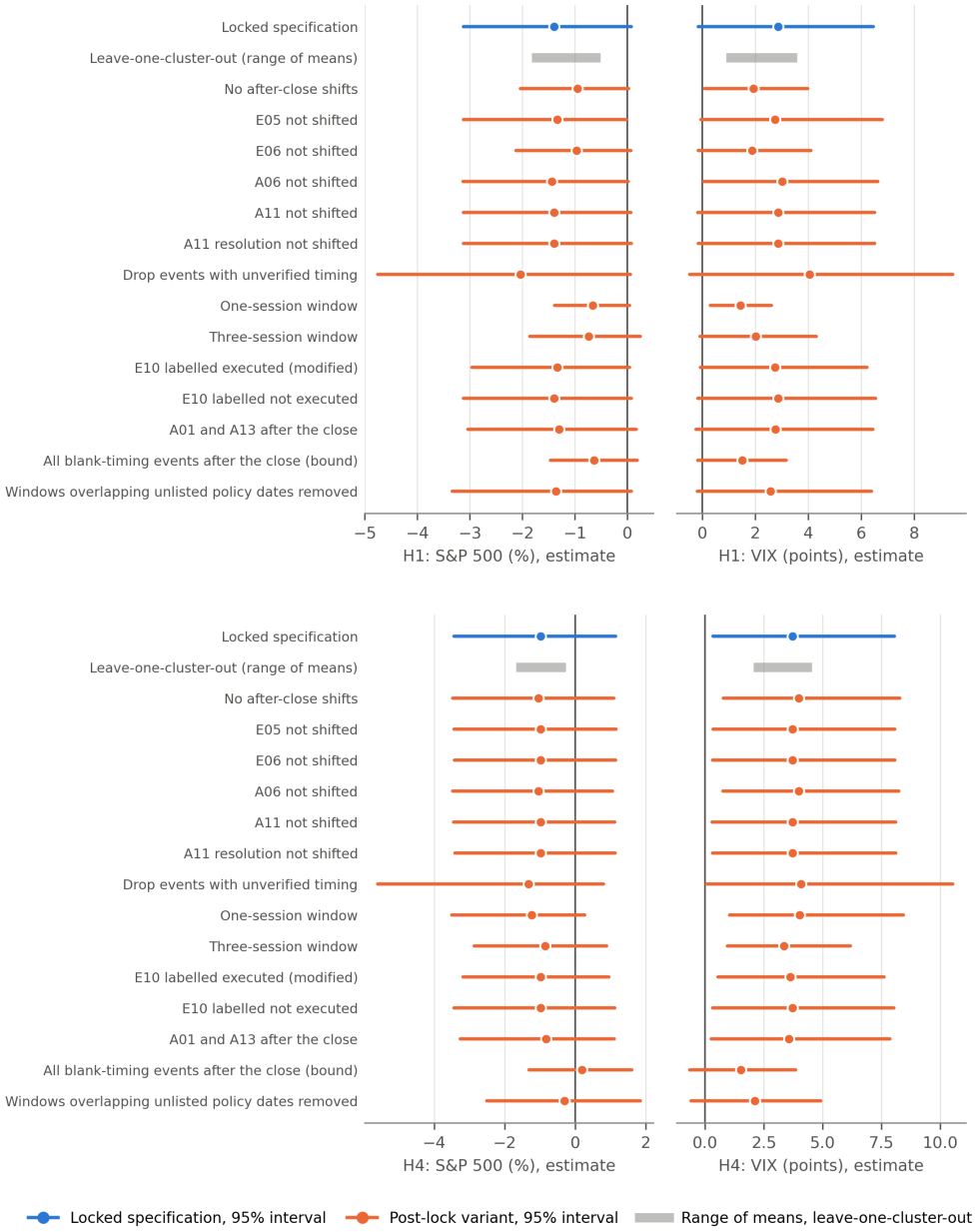

Fig. 5 Robustness of the estimates for H1 and H4 to post-lock variants. Each panel shows the locked estimate (blue) and the post-lock variants (orange) with 95% cluster-bootstrap intervals, for the S&P 500 (percent) and the VIX (index points). The grey bar marks the range of the mean across leave-one-cluster-out re-estimates. Variants cover the release time of single events, the window length, the label of E10, an extreme assumption on blank release times and the exclusion of windows that overlap unlisted policy dates. The variants are outside the Holm families.

Table 5 Robustness of the locked estimates to post-lock variants

Test	Locked estimate [95% CI]	Range	CI excl. 0	Sign flips
Panel A. Leave-one-cluster-out range of the mean				
H1 S&P 500	−1.39% [−3.12, 0.08]	[−1.82, −0.51]	n/a	n/a
H1 VIX	2.86 pts [−0.15, 6.45]	[0.89, 3.58]	n/a	n/a
H3 S&P 500	0.39% [−0.80, 1.68]	[−0.10, 0.77]	n/a	n/a
H3 VIX	−0.94 pts [−2.25, 0.52]	[−1.46, −0.55]	n/a	n/a
H4 S&P 500	−0.98% [−3.43, 1.14]	[−1.68, −0.27]	n/a	n/a
H4 VIX	3.72 pts [0.34, 8.06]	[2.07, 4.54]	n/a	n/a
Panel B. Timing and window variants				
H1 S&P 500	−1.39% [−3.12, 0.08]	−2.03 to −0.66	1/9	0
H1 VIX	2.86 pts [−0.15, 6.45]	1.44 to 4.05	3/9	0
H3 S&P 500	0.39% [−0.80, 1.68]	0.39 to 0.89	1/8	0
H3 VIX	−0.94 pts [−2.25, 0.52]	−1.42 to −0.90	1/8	0
H4 S&P 500	−0.98% [−3.43, 1.14]	−1.32 to −0.86	0/9	0
H4 VIX	3.72 pts [0.34, 8.06]	3.35 to 4.09	9/9	0
Panel C. Decision variants, including extreme timing				
H1 S&P 500	−1.39% [−3.12, 0.08]	−1.39 to −0.63	0/4	0
H1 VIX	2.86 pts [−0.15, 6.45]	1.51 to 2.86	0/4	0
H2 S&P 500	−0.83% [−2.80, 0.98]	−0.86 to 0.12	0/4	1
H2 VIX	2.36 pts [−1.14, 6.23]	0.65 to 2.29	0/4	0
H3 S&P 500	0.39% [−0.80, 1.68]	0.11 to 0.39	0/4	0
H3 VIX	−0.94 pts [−2.25, 0.52]	−0.94 to −0.29	0/4	0
H4 S&P 500	−0.98% [−3.43, 1.14]	−0.98 to 0.19	0/4	1
H4 VIX	3.72 pts [0.34, 8.06]	1.52 to 3.72	3/4	0
Panel D. Windows overlapping unlisted policy dates removed				
H1 S&P 500	−1.39% [−3.12, 0.08]	−1.35	0/1	0
H1 VIX	2.86 pts [−0.15, 6.45]	2.58	0/1	0
H3 S&P 500	0.39% [−0.80, 1.68]	0.39	0/1	0
H3 VIX	−0.94 pts [−2.25, 0.52]	−0.94	0/1	0
H4 S&P 500	−0.98% [−3.43, 1.14]	−0.31	0/1	0
H4 VIX	3.72 pts [0.34, 8.06]	2.12	0/1	0

Note: Panel A reports the range of the mean over leave-one-cluster-out re-estimates. In Panels B to D, Range is the range of point estimates across variants, CI excl. 0 is the number of variants (k/n) whose cluster-bootstrap 95% interval excludes zero, and Sign flips is the number whose sign differs from the locked estimate. Timing and window variants change the release-time bucket of single events or the window length; decision variants relabel event E10 and shift the release time of events with blank or partial timing (the extreme variant places every event with a blank release time after the close); the overlap variant drops windows overlapping policy dates not in the event list. Units are as in Table 3. All variants are post-lock, are counted in the specification ledger (Table A4) and are not part of any Holm family. Full rows are in Tables A1 to A3. Bootstrap draws differ across cells, so intervals for the same specification can differ in the second decimal.

11 clusters for H1 and six for the not-executed group, so H1 is feasible only for effects of the observed size or larger and the contrasts are not. These requirements treat days as independent and are therefore lower bounds when events cluster.

Table 6 Forecast benchmark and sample-size requirements

Period	Surveys	Mean	SE	95% CI	MDE80
Panel A. SPF directional forecast-error index					
Other periods	125	−0.117	0.057	[−0.230, −0.004]	0.160
Tariff episode 2018-19	8	−0.094	0.122	[−0.383, 0.196]	0.343
Tariff episode 2025-26	6	0.002	0.271	[−0.694, 0.698]	0.758

Panel B. Independent clusters required (80% power, 5% size)					
Outcome	Sample	SD (2-day)	Effect	One sample	Contrast, per group
S&P 500 (%)	All days	1.50	0.5	71	142
S&P 500 (%)	All days	1.50	1.0	18	36
S&P 500 (%)	All days	1.50	1.5	8	16
S&P 500 (%)	All days	1.50	2.0	5	9
S&P 500 (%)	ex-COVID	1.34	0.5	57	113
S&P 500 (%)	ex-COVID	1.34	1.0	15	29
S&P 500 (%)	ex-COVID	1.34	1.5	7	13
S&P 500 (%)	ex-COVID	1.34	2.0	4	8
VIX (points)	All days	2.52	1.0	50	100
VIX (points)	All days	2.52	2.0	13	25
VIX (points)	All days	2.52	3.0	6	12
VIX (points)	All days	2.52	4.0	4	7
VIX (points)	ex-COVID	2.26	1.0	41	81
VIX (points)	ex-COVID	2.26	2.0	11	21
VIX (points)	ex-COVID	2.26	3.0	5	9
VIX (points)	ex-COVID	2.26	4.0	3	6

Note: Panel A: the index is in units of each variable's RMSE and is positive when the realised outcome was worse than forecast; other periods are 1990 onward excluding COVID targets. It averages signed forecast errors across at least four variables per survey, each scaled by its own RMSE. It is a benchmark only: no hypothesis tests are added after the lock, and the tariff-episode samples are too small to detect plausible biases (compare MDE80 with the mean). Panel B: Effect is the target effect in the units of the outcome; the columns give the number of independent clusters needed in a one-sample test and per group in a two-group contrast. The calculation treats days as independent, so requirements are lower bounds when events cluster and optimistic because event days are more volatile than average days.

5.6 Tariff rates

The legal rate is an upper bound for the rate that is paid. Table 7 shows that the statutory effective tariff rate of the Budget Lab series rose from 5.30% in the first quarter of 2025 to 15.55% in the third quarter and fell to 12.97% in the first quarter of 2026 and 11.85% in the second, while the realised rate increased from 2.40% to 11.03% by the fourth quarter. The proxy for collected duties relative to the statutory rate was 0.46 in the first quarter of 2025, 0.68 in the fourth quarter, 0.65 in the first quarter of 2026 and 0.57 in the second. A third to a half of the statutory burden was therefore not collected, which can reflect exemptions, trade diversion, timing of payment and refunds. The proxy shares its numerator with the realised-rate series and is not an independent validation, and customs duties from May to August 2026 include unusually large refunds.

Table 7 Legal versus collected tariff rates (percent)

Quarter	Statutory ETR	Realised ETR	Collected proxy	Collected / statutory
2025Q1	5.30	2.40	2.46	0.46
2025Q2	13.96	7.90	7.29	0.52
2025Q3	15.55	10.48	9.78	0.63
2025Q4	15.12	11.03	10.33	0.68
2026Q1	12.97	10.65	8.47	0.65
2026Q2	11.85	n/a	6.81	0.57
2026Q3	9.99	n/a	n/a	n/a
2026Q4	9.36	n/a	n/a	n/a

Note: Statutory and realised effective tariff rates (ETR) are from the Yale Budget Lab series. The collected-duty proxy divides Treasury customs receipts (Monthly Treasury Statement) by NIPA goods imports and is scaled to the Yale realised rate over the overlap quarters; it shares the customs numerator with the Yale realised series and is therefore not independent validation. Cells marked not applicable are quarters without data. Customs duties in May to August 2026 include unusually large refunds; the September 2026 receipt was not available when the data were assembled.

Appendix Table A5 compares the implementation labels for 2025-26 with an objective rule based on a jump in the statutory rate at the effective date. The labels and the rule agree on eight of eleven events. The three disagreements, E07, E08 and E11, occur in windows containing other events, where the statutory rate cannot be attributed to a single measure, and the comparison is treated as exploratory.

5.7 Bounding

Table 8 lists the moments that discipline the model of Section 4.5. Among the 24 threats the shares executed in full, executed with modification and not executed are 0.375, 0.333 and 0.292, so the probability of implementation in some form is 0.708 (95% interval 0.474 to 0.880) in 17 clusters. The shares differ across episodes (0.286, 0.357 and 0.357 in 2018-19, against 0.500, 0.300 and 0.200 in 2025-26), with wide intervals.

Table 8: Data moments used to discipline the bargaining model

Moment	Value	95% CI	n	Clusters
All events				
Executed in full	0.375	[0.176, 0.536]	24	17
Executed with modification	0.333	[0.136, 0.520]	24	17
Not executed	0.292	[0.120, 0.526]	24	17
Implemented in some form	0.708	[0.474, 0.880]	24	17
Episode 2018-19				
Executed in full	0.286	[0.067, 0.538]	14	13
Executed with modification	0.357	[0.077, 0.625]	14	13
Not executed	0.357	[0.133, 0.615]	14	13
Implemented in some form	0.643	[0.385, 0.867]	14	13
Episode 2025-26				
Executed in full	0.500	[0.286, 0.615]	10	4
Executed with modification	0.300	[0.000, 0.462]	10	4
Not executed	0.200	[0.000, 0.667]	10	4
Implemented in some form	0.800	[0.333, 1.000]	10	4
Other moments				

continued on next page

Moment	Value	95% CI	n	Clusters
Median lag, all implemented (days)	15.000	[6.488, 31.000]	17	11
Median lag, 2018-19 (days)	16.000	[7.000, 31.000]	9	8
Median lag, 2025-26 (days)	6.000	[3.000, 31.000]	8	3
Announcement reaction, S&P 500 (%)	−1.390	[−3.124, 0.079]	16	11
Announcement reaction, VIX (points)	2.858	[−0.148, 6.454]	16	11
Collected to statutory rate (latest quarter)	0.649	[0.461, 0.679]	1	1
SPF index, other periods	−0.117	[−0.230, −0.004]	125	125
SPF index, 2018-19	−0.094	[−0.383, 0.196]	8	8
SPF index, 2025-26	0.002	[−0.694, 0.698]	6	6

Note: Ex post labels define the shares (threats only); intervals are cluster-bootstrap percentile intervals (shares) or Wilson and percentile intervals (medians). Units follow Tables 3 and 6. The model is a bounding device, not a structural estimate; the legal-veto share is not point identified and the implemented share of the announced rate in modified measures has not been collected.

Because the legal-veto share cannot exceed the share not executed, the identified set is the curve in Table 9. At $\lambda = 0$ the resolute share is 0.375 and the no-deal share 0.533. At the upper bound $\lambda = 0.292$ the resolute share reaches 0.529 and the no-deal share 1.000, which corresponds to a world in which every soft announcer fails to reach a deal. Across the whole curve the 95% ranges of the resolute share lie between 0.176 and 0.769 and those of the no-deal share between 0.222 and 1.000. The data therefore cannot separate a world dominated by resolute announcers from one dominated by legal vetoes.

Table 9 Identified set along the legal-veto share

λ	π	ρ	π 95% range	ρ 95% range
0.000	0.375	0.533	[0.176, 0.536]	[0.222, 0.786]
0.073	0.404	0.604	[0.200, 0.566]	[0.276, 0.830]
0.146	0.439	0.696	[0.227, 0.609]	[0.364, 0.880]
0.219	0.480	0.821	[0.258, 0.667]	[0.533, 0.936]
0.292	0.529	1.000	[0.300, 0.769]	[1.000, 1.000]

Note: The symbol lambda is the legal-veto share of threats, pi the resolute share (full implementation) and rho the no-deal share. Each row is one admissible world along the identified curve; the legal-veto share cannot exceed the share of non-executed announcements (0.292). Intervals are cluster-bootstrap percentile intervals; all draws were feasible (100%).

The forecast implication does not depend on which point on the curve is true, because it uses only the observed shares and the unobserved κ. Table 10 shows that if modified measures retain 30% of the announced rate, the expected implemented fraction is 0.475 and assuming full implementation overstates it by a factor of 2.11. If they retain 70% the fraction is 0.608 and the factor 1.64, and if they retain the whole announced rate it is 0.708 and the factor 1.41. The lower end of the range, 1.41, is a floor, since it requires modified measures to be as severe as announced. The count-based fraction does not weight measures by the value of trade they cover, and weighting could move the figures in either direction.

Table 10 Implied forecast gap for illustrative values of kappa

κ	Implemented fraction	95% range	Overstatement
0.30	0.475	[0.288, 0.619]	2.11
0.50	0.542	[0.342, 0.684]	1.85
0.70	0.608	[0.400, 0.754]	1.64
1.00	0.708	[0.474, 0.880]	1.41

Note: Kappa is the share of the announced rate that survives in modified measures; its values are illustrative and not estimated. The implemented fraction is the share executed in full plus the modified share times kappa, and the overstatement is its inverse. The measure is a count of events, not weighted by trade value.

6 Discussion

6.1 Interpretation

The results support a reading in which announcement and implementation are linked but distinct events. Most announced tariffs took effect, and mostly quickly, which is why the announcement carried information. A sizeable minority did not, a further third of the threats took effect in a form that differed from the one announced, and collected duties fell short of the statutory rate by a third to a half on a simple proxy. This is the pattern that models of imperfect commitment and of bargaining over market access lead one to expect (Kydland and Prescott 1977; Bagwell and Staiger 1999).

The market evidence is consistent with this mixture. Equity prices fell and implied volatility rose after announcements that proved to be implemented, a reaction that has the sign expected if investors treated the announcement as a signal of a likely cost (Pástor and Veronesi 2012; Baker et al. 2016). The volatility reaction was concentrated at the announcement and not at the effective date in the exploratory comparison of Table 4, which fits a view in which most of the uncertainty about a measure is resolved when it is announced and in which the mean reversion that follows is part of how the shock fades (Bloom 2009). Two secondary outcomes bear on the international side of the question. Treasury yields fell after announcements that were implemented, although these results are the least robust of the confirmatory set. The broad dollar index moved little, and its interval excludes changes of more than about a quarter of a percent over two sessions, so the data give no sign of a sizeable exchange-rate response at the announcement.

What the data do not show is equally important. They do not show that markets could tell ahead of time which announcements would be carried out, since the contrast designed for this purpose is uninformative by the rule written before the test. They do not show that withdrawal is rewarded, since the resolution effects have the predicted sign but are far from distinguishable from zero. And they do not show that forecasters assumed full implementation, since the benchmark of forecast errors cannot separate the tariff episodes from other periods. Each of these is a statement about precision, and the minimum detectable effects reported throughout state what a future sample

would have to be able to detect. A larger and more diverse set of episodes, or intraday data on release times, would speak to these questions more directly than additional analysis of the present 29 events.

The bounding model translates the gap into a range for the forecast error. If modified measures preserved most of the announced rate, assuming full implementation would overstate the implemented fraction by a factor of about 1.4, and if they preserved less than a third of it, by a factor of about 2. The figures are counts of events and are not weighted by trade value, and their purpose is to show that the correction is of first-order size and that its magnitude hinges on a quantity that has not been collected. Collecting the surviving share of the announced rate for each modified measure is the most direct way to narrow the range.

6.2 Implications

Four implications follow for research and practice. For event-study designs on policy announcements, treating every announcement as the treatment mixes measures with different probabilities of implementation, and the interpretation of the average effect depends on the mixture (Brown and Warner 1985). Splitting announcements by their later fate, as is done here, is informative whenever the number of events permits it, and reporting the power of that split is a condition for reading the result. For forecasting and scenario analysis, weighting announced tariffs by an implementation probability close to the observed shares, instead of assuming full implementation or none, is a more cautious rule in the spirit of the evidence on information frictions and overreaction in expectations (Coibion and Gorodnichenko 2015; Bordalo et al. 2020). For the credibility of trade policy, a pattern of announcements that are frequently modified or dropped is what commitment problems predict (Kydland and Prescott 1977). If it persists, the information content of future announcements may fall, a conjecture that the present sample cannot test. And for the measurement of protection, using announced rates as a stand-in for applied rates mixes intentions with barriers, and the statutory rate is itself an upper bound for what is collected. For investors and firms, treating an announcement as a probabilistic event and not as a settled outcome is relevant to scenario planning and risk management, although the paper does not test how firm decisions respond.

6.3 Limitations

Several limitations bear on the interpretation. The paper does not study trade flows, import prices or firm behaviour, so it says nothing about how announced and applied tariffs differ in their effects on trade; its evidence concerns the measurement of the gap and the pricing of announcements in financial markets. The sample is small: the locked tests rest on 16 event days in 11 clusters for H1, six for the not-executed group and eight for resolution. Estimates at this scale are noisy, and statistically significant estimates in underpowered designs tend to exaggerate magnitudes (Gelman and Carlin 2014). Outcome labels are assigned ex post, and the label of one event (E10) is left open. A verified release-time bucket exists for only 14 of the 29 events, and the extreme variant shows that the magnitudes, though not the signs, depend on it. The announcements

are not exogenous: threats are often issued in response to market conditions or to other policy events, and windows of two sessions contain other news, so the estimates describe co-movement around announcements and not causal effects of tariffs. The sample of events is itself a choice; the register is documented, but a different inclusion rule could change the shares.

The design lock is internal, and the number of specifications examined, 242 in all, is large relative to the number of events. The lock and the ledger make the forking paths visible, which is the purpose of pre-specification (Olken 2015; Christensen and Miguel 2018), but they do not remove the possibility of researcher discretion in the choices that preceded the lock, and the p-values of family A are conditional on results seen before it. The yield results of family C depend on the reference set of the placebo. Finally, the legal-veto share and the surviving share of the announced rate are not identified or collected, so the bounding model delivers ranges and not point estimates, and conclusions about the 2025-26 episode rest on events that were still unfolding when the data were assembled.

7 Conclusion

This paper has measured how far US tariff announcements diverge from applied tariffs and asked whether asset prices and forecasts reflect the divergence. In a register of 29 events from 2018-19 and 2025-26, 17 of 24 announced threats took effect in some form, typically within a month, and the share did not differ detectably between episodes. Collected duties fell short of statutory rates by a third to a half on a simple proxy. Announcements that were later implemented coincided with lower equity prices and higher implied volatility, and these two results withstood the pre-specified multiplicity adjustment, although the raw bootstrap intervals are wide. The contrasts that would show whether markets anticipated implementation or reacted again at the effective date are too imprecise to decide, and the forecast benchmark cannot separate the tariff episodes from other periods. A bounding model shows that treating announcements as fully implemented overstates the expected implemented fraction by a factor of about 1.4 to 2.1.

These findings qualify but do not overturn the usual practice of treating announcements as news. They suggest that announced tariffs are better described as probabilistic, and that the probability, once estimated across episodes, matters for event studies, for forecasting and scenario design, for the interpretation of policy credibility and for the measurement of protection. Further work should extend the register to additional episodes and trading partners, weight measures by the value of covered trade, link the register to import prices and trade flows around effective dates, collect the surviving share of the announced rate in modified measures, use intraday data to settle release times, and register the design with an external repository before testing. Only a larger body of events can say whether markets price the probability of implementation, which is the question this study raises and, with 11 clusters, does not settle.

Acknowledgements. This study is funded by University of Economics and Law, Ho Chi Minh City, Vietnam, and Vietnam National University, Ho Chi Minh City, Vietnam.

Declarations

Funding: This study is funded by University of Economics and Law, Ho Chi Minh City, Vietnam, and Vietnam National University, Ho Chi Minh City, Vietnam.

Competing interests: The author declares no competing interests.

Ethics approval and consent to participate: Not applicable. The study uses public market, forecast and tariff data and documents of public record, and involves no human participants, human samples or animals.

Consent for publication: Not applicable.

Data availability: All market, forecast and tariff data are publicly available from Federal Reserve Economic Data (FRED), the Survey of Professional Forecasters, the Budget Lab at Yale, the US Department of the Treasury and the US Bureau of Economic Analysis. The event register, the design file and the derived results are available from the corresponding author upon request.

Materials availability: Not applicable.

Code availability: The code is available from the corresponding author upon request.

Author contribution: Quoc Lap Nguyen was solely responsible for conceptualisation, data collection, methodology, software, formal analysis, visualisation and writing.

Appendix A Supplementary tables

The tables in this appendix document the post-lock variants, the specification ledger, the verification status of the register, the objective check of implementation labels and the competing-risk estimates. All variants are outside the Holm families.

Table A1: Post-lock decision variants: E10 label and release timing of events with blank or partial timing

Test	Days (cl.)	Estimate	95% CI	MDE80
Locked sample (E10 excluded)				
H1 S&P 500 (%)	16 (11)	−1.39%	[−3.11, 0.08]	0.92
H1 VIX (points)	16 (11)	2.86 pts	[−0.16, 6.53]	1.69
H2 S&P 500 (%)	16 (11); 6 (6)	−0.83%	[−2.80, 0.97]	2.21
H2 VIX (points)	16 (11); 6 (6)	2.36 pts	[−1.11, 6.21]	3.78
H3 S&P 500 (%)	8 (7)	0.39%	[−0.81, 1.67]	1.69
H3 VIX (points)	8 (7)	−0.94 pts	[−2.25, 0.53]	2.89
H4 S&P 500 (%)	10 (9)	−0.98%	[−3.41, 1.11]	2.09
H4 VIX (points)	10 (9)	3.72 pts	[0.31, 8.15]	3.60
E10 labelled executed with modification (EU 15% order)				
H1 S&P 500 (%)	17 (12)	−1.33%	[−2.96, 0.05]	0.92
H1 VIX (points)	17 (12)	2.75 pts	[−0.08, 6.21]	1.69
H2 S&P 500 (%)	17 (12); 6 (6)	−0.77%	[−2.68, 0.96]	2.21
H2 VIX (points)	17 (12); 6 (6)	2.25 pts	[−1.08, 5.99]	3.77
H3 S&P 500 (%)	8 (7)	0.39%	[−0.83, 1.67]	1.69
H3 VIX (points)	8 (7)	−0.94 pts	[−2.24, 0.54]	2.89
H4 S&P 500 (%)	11 (10)	−0.97%	[−3.18, 0.96]	1.95

continued on next page

Test	Days (cl.)	Estimate	95% CI	MDE80
H4 VIX (points)	11 (10)	3.62 pts	[0.54, 7.62]	3.38
E10 labelled not executed (resolution 2025-07-31)				
H1 S&P 500 (%)	16 (11)	−1.39%	[−3.12, 0.08]	0.92
H1 VIX (points)	16 (11)	2.86 pts	[−0.16, 6.54]	1.69
H2 S&P 500 (%)	16 (11); 7 (7)	−0.86%	[−2.78, 0.87]	2.02
H2 VIX (points)	16 (11); 7 (7)	2.29 pts	[−1.08, 6.17]	3.57
H3 S&P 500 (%)	9 (8)	0.11%	[−1.06, 1.33]	1.54
H3 VIX (points)	9 (8)	−0.29 pts	[−1.87, 1.60]	2.70
H4 S&P 500 (%)	10 (9)	−0.98%	[−3.43, 1.12]	2.09
H4 VIX (points)	10 (9)	3.72 pts	[0.32, 8.03]	3.60
A01 and A13 announced after the close				
H1 S&P 500 (%)	16 (11)	−1.30%	[−3.03, 0.17]	0.93
H1 VIX (points)	16 (11)	2.76 pts	[−0.24, 6.44]	1.70
H2 S&P 500 (%)	16 (11); 6 (6)	−0.74%	[−2.74, 1.07]	2.18
H2 VIX (points)	16 (11); 6 (6)	2.26 pts	[−1.21, 6.17]	3.75
H3 S&P 500 (%)	8 (7)	0.39%	[−0.78, 1.68]	1.69
H3 VIX (points)	8 (7)	−0.94 pts	[−2.26, 0.54]	2.89
H4 S&P 500 (%)	10 (9)	−0.83%	[−3.26, 1.11]	2.07
H4 VIX (points)	10 (9)	3.57 pts	[0.27, 7.86]	3.63
Extreme timing: blank release times set after the close				
H1 S&P 500 (%)	16 (12)	−0.63%	[−1.46, 0.19]	0.95
H1 VIX (points)	16 (12)	1.51 pts	[−0.17, 3.17]	1.73
H2 S&P 500 (%)	16 (12); 6 (6)	0.12%	[−1.32, 1.53]	2.14
H2 VIX (points)	16 (12); 6 (6)	0.65 pts	[−1.86, 3.17]	3.74
H3 S&P 500 (%)	8 (7)	0.39%	[−0.78, 1.68]	1.69
H3 VIX (points)	8 (7)	−0.94 pts	[−2.27, 0.54]	2.89
H4 S&P 500 (%)	9 (8)	0.19%	[−1.30, 1.61]	2.26
H4 VIX (points)	9 (8)	1.52 pts	[−0.65, 3.86]	3.84

Note: The first block replicates the locked sample (E10 excluded, blank release times treated as not shifted). The extreme variant places every event with a blank release time after the close and is a bound, not a best estimate. Days (cl.): distinct event days with clusters in parentheses; for H2 the two groups are executed and not executed. The p-values are omitted by design: variants are sensitivity checks outside the Holm families. Bootstrap draws differ from those of Table 3, so intervals can differ in the second decimal.

Table A2: Timing and window variants of the announcement-day estimates

Test	Clusters	Estimate	95% CI	MDE80
Locked specification (two-session window)				
H1 S&P 500 (%)	11	−1.39%	[−3.10, 0.08]	0.93
H1 VIX (points)	11	2.86 pts	[−0.15, 6.49]	1.70
H3 S&P 500 (%)	7	0.39%	[−0.77, 1.67]	1.70
H3 VIX (points)	7	−0.94 pts	[−2.25, 0.53]	2.86
H4 S&P 500 (%)	9	−0.98%	[−3.43, 1.16]	2.09
H4 VIX (points)	9	3.72 pts	[0.35, 8.08]	3.64
No after-close shifts				
H1 S&P 500 (%)	11	−0.95%	[−2.04, 0.02]	0.91
H1 VIX (points)	11	1.94 pts	[0.08, 3.98]	1.67
H3 S&P 500 (%)	7	0.52%	[−0.70, 1.85]	1.69
H3 VIX (points)	7	−0.90 pts	[−2.25, 0.60]	2.86
H4 S&P 500 (%)	9	−1.05%	[−3.47, 1.09]	2.09
H4 VIX (points)	9	3.99 pts	[0.79, 8.28]	3.63
E05 not shifted				
H1 S&P 500 (%)	12	−1.34%	[−3.12, −0.01]	0.92
H1 VIX (points)	12	2.75 pts	[−0.06, 6.79]	1.67
H3 S&P 500 (%)	7	0.39%	[−0.80, 1.67]	1.70
H3 VIX (points)	7	−0.94 pts	[−2.26, 0.53]	2.86
H4 S&P 500 (%)	9	−0.98%	[−3.43, 1.16]	2.09
H4 VIX (points)	9	3.72 pts	[0.34, 8.08]	3.64

continued on next page

Test	Clusters	Estimate	95% CI	MDE80
E06 not shifted				
H1 S&P 500 (%)	11	−0.96%	[−2.12, 0.07]	0.92
H1 VIX (points)	11	1.89 pts	[−0.15, 4.10]	1.68
H3 S&P 500 (%)	7	0.39%	[−0.81, 1.67]	1.70
H3 VIX (points)	7	−0.94 pts	[−2.24, 0.54]	2.86
H4 S&P 500 (%)	9	−0.98%	[−3.42, 1.14]	2.09
H4 VIX (points)	9	3.72 pts	[0.32, 8.08]	3.64
A06 not shifted				
H1 S&P 500 (%)	11	−1.43%	[−3.13, 0.01]	0.92
H1 VIX (points)	11	3.02 pts	[0.06, 6.63]	1.70
H3 S&P 500 (%)	7	0.39%	[−0.81, 1.67]	1.70
H3 VIX (points)	7	−0.94 pts	[−2.24, 0.53]	2.86
H4 S&P 500 (%)	9	−1.05%	[−3.47, 1.06]	2.09
H4 VIX (points)	9	3.99 pts	[0.76, 8.25]	3.63
A11 not shifted				
H1 S&P 500 (%)	11	−1.39%	[−3.12, 0.07]	0.93
H1 VIX (points)	11	2.86 pts	[−0.17, 6.51]	1.70
H3 S&P 500 (%)	7	0.39%	[−0.80, 1.67]	1.70
H3 VIX (points)	7	−0.94 pts	[−2.25, 0.54]	2.86
H4 S&P 500 (%)	9	−0.98%	[−3.44, 1.13]	2.09
H4 VIX (points)	9	3.72 pts	[0.30, 8.11]	3.64
A11 resolution not shifted				
H1 S&P 500 (%)	11	−1.39%	[−3.12, 0.08]	0.93
H1 VIX (points)	11	2.86 pts	[−0.15, 6.52]	1.70
H3 S&P 500 (%)	7	0.52%	[−0.70, 1.84]	1.69
H3 VIX (points)	7	−0.90 pts	[−2.25, 0.60]	2.86
H4 S&P 500 (%)	9	−0.98%	[−3.41, 1.13]	2.09
H4 VIX (points)	9	3.72 pts	[0.33, 8.11]	3.64
Events with unverified release time dropped				
H1 S&P 500 (%)	6	−2.03%	[−4.76, 0.06]	1.28
H1 VIX (points)	6	4.05 pts	[−0.47, 9.45]	2.38
H4 S&P 500 (%)	5	−1.32%	[−5.60, 0.80]	2.25
H4 VIX (points)	5	4.09 pts	[0.06, 10.53]	4.06
One-session window				
H1 S&P 500 (%)	11	−0.66%	[−1.38, 0.05]	0.73
H1 VIX (points)	11	1.44 pts	[0.30, 2.61]	1.42
H3 S&P 500 (%)	7	0.89%	[0.17, 1.79]	1.34
H3 VIX (points)	7	−1.42 pts	[−2.45, −0.52]	2.34
H4 S&P 500 (%)	9	−1.24%	[−3.50, 0.27]	1.79
H4 VIX (points)	9	4.02 pts	[1.05, 8.44]	3.05
Three-session window				
H1 S&P 500 (%)	11	−0.74%	[−1.85, 0.25]	1.14
H1 VIX (points)	11	2.01 pts	[−0.08, 4.31]	2.02
H3 S&P 500 (%)	7	0.67%	[−0.62, 1.90]	2.18
H3 VIX (points)	7	−1.17 pts	[−2.33, 0.41]	3.52
H4 S&P 500 (%)	9	−0.86%	[−2.87, 0.89]	2.59
H4 VIX (points)	9	3.35 pts	[0.94, 6.19]	4.23

Note: Variants switch the after-close shift of single events (E05, E06, A06, A11), drop events with unverified release time, or change the window to one or three sessions. The locked specification is the two-session window.

Table A3: Exclusion of windows overlapping policy dates not in the event list

Test	Sample	Clusters	Estimate	95% CI
H1 S&P 500 (%)	Locked	11	−1.39%	[−3.10, 0.07]
H1 S&P 500 (%)	Overlap removed	11	−1.35%	[−3.34, 0.08]
H1 VIX (points)	Locked	11	2.86 pts	[−0.18, 6.49]

continued on next page

Test	Sample	Clusters	Estimate	95% CI
H1 VIX (points)	Overlap removed	11	2.58 pts	[−0.18, 6.40]
H3 S&P 500 (%)	Locked	7	0.39%	[−0.78, 1.67]
H3 S&P 500 (%)	Overlap removed	7	0.39%	[−0.81, 1.66]
H3 VIX (points)	Locked	7	−0.94 pts	[−2.25, 0.54]
H3 VIX (points)	Overlap removed	7	−0.94 pts	[−2.26, 0.54]
H4 S&P 500 (%)	Locked	9	−0.98%	[−3.42, 1.15]
H4 S&P 500 (%)	Overlap removed	8	−0.31%	[−2.51, 1.85]
H4 VIX (points)	Locked	9	3.72 pts	[0.32, 8.01]
H4 VIX (points)	Overlap removed	8	2.12 pts	[−0.60, 4.92]

Note: Unlisted policy dates are used only to exclude overlapping windows and are never added to the event sample. Each unlisted date carries a primary source document.

Table A4: Specification ledger

Stage	n	Status	Content
Pre-lock exploration	60	Approximate; more than 60, exact count not recoverable	Reported as more than 60
Locked battery	12	Exact; 12 of 12 slots run	Holm within families A, B and C
Post-lock sensitivity analyses	91	Exact (counted in code)	H4, base, competing risks, event, implementation, influence, timing, window
Post-lock model analyses	23	Exact (counted in code)	Full implementation, identified set, lag, share of reaction at announcement, shares
Post-lock follow-up analyses	56	Exact (counted in code)	Conditional, decision, extra-date, extreme-day
Total (upper bound)	242	Only the 12 locked slots carry confirmatory Holm adjustment	

Note: Counts include every specification computed after the lock; the pre-lock exploration count is approximate. The total is disclosed in the text.

Table A5: Check of the 2025-26 implementation labels against the Yale statutory rate

ID	Announced	Label	ETR	Jump 60d	Jump eff.	Other	Yale	Label
E01	2025-02-01	Full	4.25	7.84	2.03	4	Yes	Yes
E02	2025-02-01	Modified	4.25	7.84	0.10	4	Yes	Yes
E03	2025-02-09	Full	6.28	13.37	0.50	4	Yes	Yes
E04	2025-02-13	Modified	6.28	5.19	12.78	3	Yes	Yes
E05	2025-03-26	Full	6.88	7.51	7.57	4	Yes	Yes
E06	2025-04-02	Modified	6.88	7.51	2.92	3	Yes	Yes
E07	2025-04-07	Full	19.66	−5.46	0.00	2	No	Yes
E08	2025-05-12	De-esc.	12.35	2.01	2.04	1	Yes	No
E09	2025-05-23	Not exec.	14.39	−0.03	n/a	0	No	No
E11	2026-02-20	Full	13.62	−1.72	−2.78	1	No	Yes
E12	2026-02-21	Not exec.	13.62	−1.72	n/a	0	No	No

Note: ETR is the statutory effective tariff rate (percent) before the announcement; Jump 60d and Jump eff. are changes in percentage points over 60 days and at the effective date; Other is the number of other events within 60 days; Yale and Label (last two columns) say whether the Yale rule and the ex post label, respectively, indicate implementation. A statutory rate can fall at the effective date because another measure was removed or replaced in the same window, so disagreement does not prove a wrong label. Overlapping events cannot be separated; the comparison is exploratory. Label abbreviations: Full, executed in full; Modified, executed with modification; De-esc., de-escalation; Not exec., not executed.

Table A6: Release time and source verification of the event register

ID	Release bucket	Post-lock timing check	Status	Open fields
Panel A. Episode 2018-19				
A01	Not verified	none	Partial	release time (partial)
A02	Intraday	none	Verified	
A03	Not verified	15:39 (intraday) (verified)	Verified	
A04	Not verified	12:45 (intraday) (verified)	Verified	
A05	Not verified	none	Partial	release time (not found)
A06	After close	after close (partial)	Partial	release time (partial)
A07	Weekend	none	Verified	
A08	Weekend	none	Verified	
A09	Not verified	by 09:04, pre-open (partial)	Partial	reason for non-execution (partial), release time (partial)
A10	Not verified	by 14:04, intraday (partial)	Partial	planned effective date (partial), release time (partial)
A11	After close	by 20:07, after close (verified)	Verified	
A12	Not verified	by 14:11, intraday (verified)	Verified	
A17	Not verified	none	Partial	release time (not found)
A13	Not verified	after close (17:00 ET) (verified)	Verified	
A14	Not verified	15:36, intraday (verified)	Verified	
A15	Not verified	none	Open	reason for non-execution (partial), release time (not found)
A16	Pre-open	none	Partial	reason for non-execution (partial)
Panel B. Episode 2025-26				
E01	Weekend	none	Verified	
E02	Weekend	none	Verified	
E03	Weekend	none	Partial	planned effective date (partial)
E04	Not verified	13:35 intraday (verified)	Verified	
E05	After close	17:16 after close (verified)	Verified	
E06	After close	16:06 after close (verified)	Verified	
E07	Not verified	11:14 intraday (verified)	Verified	

continued on next page

ID	Release bucket	Post-lock timing check	Status	Open fields
E08	Not verified	pre-open (03:01 ET) (verified)	Verified	
E09	Not verified	07:43 pre-open (verified)	Verified	
E10	Weekend	weekend (partial)	Partial	planned effective date (partial), release time (partial)
E11	Intraday	10:00 intraday (partial)	Partial	release time (partial)
E12	Weekend	10:59 weekend (partial)	Partial	reason for non-execution (partial), release time (partial)

Note: Release bucket is the verified time bucket of the first public release (Eastern Time); 'Not verified' means that no timestamped source was located (15 of 29 events). In the locked analysis only events verified as released after the close are shifted by one session, and events without a verified bucket are not shifted. The post-lock timing check reports later research on the release time (with confidence) as of 16 October 2026; it is not used in the locked tests and enters only the sensitivity variants (Table 5 and Appendix Table A1). Status counts: 17 verified, 11 partial, 1 open. 'Verified' means the fields used in the analysis were read from a primary or timestamped source; 'partial' means a secondary source or a missing clock time; 'open' means not yet resolved.

Table A7: Competing-risk cumulative incidence by episode and authority (Figure 2 plots the episode strata)

Stratum	Days	Implementation, % [95% CI]	Resolution, % [95% CI]
Panel A. Threats with a known outcome date (main sample)			
All threats (n = 21)	15	43 [20, 65]	14 [3, 35]
	30	57 [32, 80]	14 [3, 35]
	60	76 [50, 93]	14 [3, 35]
	90	76 [50, 93]	14 [3, 35]
	180	81 [56, 96]	19 [4, 44]
	365	81 [56, 96]	19 [4, 44]
Episode 2018-19 (n = 12)	15	33 [9, 58]	17 [0, 45]
	30	58 [27, 85]	17 [0, 45]
	60	67 [36, 92]	17 [0, 45]
	90	67 [36, 92]	17 [0, 45]
	180	75 [45, 100]	25 [0, 55]
	365	75 [45, 100]	25 [0, 55]
Episode 2025-26 (n = 9)	15	56 [20, 100]	11 [0, 50]
	30	56 [20, 100]	11 [0, 50]
	60	89 [50, 100]	11 [0, 50]
	90	89 [50, 100]	11 [0, 50]
	180	89 [50, 100]	11 [0, 50]
	365	89 [50, 100]	11 [0, 50]
IEEPA (n = 6)	15	50 [0, 100]	17 [0, 100]
	30	50 [0, 100]	17 [0, 100]
	60	83 [0, 100]	17 [0, 100]
	90	83 [0, 100]	17 [0, 100]
	180	83 [0, 100]	17 [0, 100]
	365	83 [0, 100]	17 [0, 100]
Section 232 (n = 5)	15	40 [0, 75]	20 [0, 75]
	30	60 [0, 100]	20 [0, 75]
	60	80 [25, 100]	20 [0, 75]
	90	80 [25, 100]	20 [0, 75]
	180	80 [25, 100]	20 [0, 75]
	365	80 [25, 100]	20 [0, 75]
Section 301 (n = 7)	15	43 [14, 71]	0 [0, 0]
	30	57 [29, 100]	0 [0, 0]

continued on next page

Stratum	Days	Implementation, % [95% CI]	Resolution, % [95% CI]
	60	71 [29, 100]	0 [0, 0]
	90	71 [29, 100]	0 [0, 0]
	180	86 [57, 100]	14 [0, 43]
	365	86 [57, 100]	14 [0, 43]
Panel B. Unresolved threats censored at sample end			
All threats (n = 24)	15	38 [17, 59]	12 [0, 29]
	30	50 [27, 73]	12 [0, 29]
	60	67 [40, 85]	12 [0, 29]
	90	67 [40, 85]	12 [0, 29]
	180	71 [45, 88]	17 [4, 37]
	365	71 [45, 88]	17 [4, 37]
Episode 2018-19 (n = 14)	15	29 [8, 54]	14 [0, 36]
	30	50 [21, 79]	14 [0, 36]
	60	57 [29, 82]	14 [0, 36]
	90	57 [29, 82]	14 [0, 36]
	180	64 [36, 88]	21 [0, 46]
	365	64 [36, 88]	21 [0, 46]
Episode 2025-26 (n = 10)	15	50 [20, 90]	10 [0, 50]
	30	50 [20, 90]	10 [0, 50]
	60	80 [33, 100]	10 [0, 50]
	90	80 [33, 100]	10 [0, 50]
	180	80 [33, 100]	10 [0, 50]
	365	80 [33, 100]	10 [0, 50]
IEEPA (n = 6)	15	50 [0, 100]	17 [0, 100]
	30	50 [0, 100]	17 [0, 100]
	60	83 [0, 100]	17 [0, 100]
	90	83 [0, 100]	17 [0, 100]
	180	83 [0, 100]	17 [0, 100]
	365	83 [0, 100]	17 [0, 100]
Section 232 (n = 7)	15	29 [0, 57]	14 [0, 50]
	30	43 [0, 78]	14 [0, 50]
	60	57 [17, 89]	14 [0, 50]
	90	57 [17, 89]	14 [0, 50]
	180	57 [17, 89]	14 [0, 50]
	365	57 [17, 89]	14 [0, 50]
Section 301 (n = 7)	15	43 [14, 86]	0 [0, 0]
	30	57 [14, 86]	0 [0, 0]
	60	71 [29, 100]	0 [0, 0]
	90	71 [29, 100]	0 [0, 0]
	180	86 [57, 100]	14 [0, 43]
	365	86 [57, 100]	14 [0, 43]

Note: Cumulative incidence at the stated number of days since the announcement; Resolution is resolution without implementation, and the two are competing events; intervals are 95% bootstrap intervals over threats. Estimates at long horizons rest on a small number of threats per stratum.

References

Altman DG, Bland JM (1995) Statistics notes: Absence of evidence is not evidence of absence. BMJ 311(7003):485. https://doi.org/10.1136/bmj.311.7003.485

Amiti M, Redding SJ, Weinstein DE (2019) The impact of the 2018 tariffs on prices and welfare. Journal of Economic Perspectives 33(4):187–210. https://doi.org/10.1257/jep.33.4.187

Bagwell K, Staiger RW (1999) An economic theory of GATT. American Economic Review 89(1):215–248. https://doi.org/10.1257/aer.89.1.215

Baker SR, Bloom N, Davis SJ (2016) Measuring economic policy uncertainty. Quarterly Journal of Economics 131(4):1593–1636. https://doi.org/10.1093/qje/qjw024

Barattieri A, Cacciatore M, Ghironi F (2021) Protectionism and the business cycle. Journal of International Economics 129:103417. https://doi.org/10.1016/j.jinteco.2020.103417

Bernanke BS, Kuttner KN (2005) What explains the stock market's reaction to Federal Reserve policy? The Journal of Finance 60(3):1221–1257. https://doi.org/10.1111/j.1540-6261.2005.00760.x

Bloom N (2009) The impact of uncertainty shocks. Econometrica 77(3):623–685. https://doi.org/10.3982/ECTA6248

Boehmer E, Musumeci J, Poulsen AB (1991) Event-study methodology under conditions of event-induced variance. Journal of Financial Economics 30(2):253–272. https://doi.org/10.1016/0304-405X(91)90032-F

Bordalo P, Gennaioli N, Ma Y, et al (2020) Overreaction in macroeconomic expectations. American Economic Review 110(9):2748–2782. https://doi.org/10.1257/aer.20181219

Bown CP (2021) The US–China trade war and phase one agreement. Journal of Policy Modeling 43(4):805–843. https://doi.org/10.1016/j.jpolmod.2021.02.009

Bown CP, Crowley MA (2013) Self-enforcing trade agreements: Evidence from time-varying trade policy. American Economic Review 103(2):1071–1090. https://doi.org/10.1257/aer.103.2.1071

Brodeur A, Cook N, Heyes A (2020) Methods matter: p-hacking and publication bias in causal analysis in economics. American Economic Review 110(11):3634–3660. https://doi.org/10.1257/aer.20190687

Brown LD, Cai TT, DasGupta A (2001) Interval estimation for a binomial proportion. Statistical Science 16(2):101–133. https://doi.org/10.1214/ss/1009213286

Brown SJ, Warner JB (1985) Using daily stock returns: The case of event studies. Journal of Financial Economics 14(1):3–31. https://doi.org/10.1016/0304-405X(85)90042-X

Caldara D, Iacoviello M, Molligo P, et al (2020) The economic effects of trade policy uncertainty. Journal of Monetary Economics 109:38–59. https://doi.org/10.1016/j.jmoneco.2019.11.002

Cameron AC, Gelbach JB, Miller DL (2008) Bootstrap-based improvements for inference with clustered errors. Review of Economics and Statistics 90(3):414–427. https://doi.org/10.1162/rest.90.3.414

Cavallo A, Gopinath G, Neiman B, et al (2021) Tariff pass-through at the border and at the store: Evidence from US trade policy. American Economic Review: Insights 3(1):19–34. https://doi.org/10.1257/aeri.20190536

Christensen G, Miguel E (2018) Transparency, reproducibility, and the credibility of economics research. Journal of Economic Literature 56(3):920–980. https://doi.org/10.1257/jel.20171350

Coibion O, Gorodnichenko Y (2015) Information rigidity and the expectations formation process: A simple framework and new facts. American Economic Review 105(8):2644–2678. https://doi.org/10.1257/aer.20110306

Conley TG, Taber CR (2011) Inference with "difference in differences" with a small number of policy changes. Review of Economics and Statistics 93(1):113–125. https://doi.org/10.1162/REST_a_00049

Croushore D, Stark T (2001) A real-time data set for macroeconomists. Journal of Econometrics 105(1):111–130. https://doi.org/10.1016/S0304-4076(01)00072-0

Crowley M, Meng N, Song H (2018) Tariff scares: Trade policy uncertainty and foreign market entry by Chinese firms. Journal of International Economics 114:96–115. https://doi.org/10.1016/j.jinteco.2018.05.003

Efron B (1979) Bootstrap methods: Another look at the jackknife. The Annals of Statistics 7(1):1–26. https://doi.org/10.1214/aos/1176344552

Egger PH, Zhu J (2020) The US–Chinese trade war: An event study of stock-market responses. Economic Policy 35(103):519–559. https://doi.org/10.1093/epolic/eiaa016

Fajgelbaum PD, Goldberg PK, Kennedy PJ, et al (2020) The return to protectionism. Quarterly Journal of Economics 135(1):1–55. https://doi.org/10.1093/qje/qjz036

Fetzer T, Schwarz C (2021) Tariffs and politics: Evidence from Trump's trade wars. The Economic Journal 131(636):1717–1741. https://doi.org/10.1093/ej/ueaa122

Fine JP, Gray RJ (1999) A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association 94(446):496–509. https://doi.org/10.1080/01621459.1999.10474144

Fisher RA (1922) On the interpretation of χ^2 from contingency tables, and the calculation of P. Journal of the Royal Statistical Society 85(1):87–94. https://doi.org/10.2307/2340521

Flaaen A, Hortçsu A, Tintelnot F (2020) The production relocation and price effects of US trade policy: The case of washing machines. American Economic Review 110(7):2103–2127. https://doi.org/10.1257/aer.20190611

Gelman A, Carlin J (2014) Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. Perspectives on Psychological Science 9(6):641–651. https://doi.org/10.1177/1745691614551642

Gürkaynak RS, Sack B, Wright JH (2010) The TIPS yield curve and inflation compensation. American Economic Journal: Macroeconomics 2(1):70–92. https://doi.org/10.1257/mac.2.1.70

Handley K, Limão N (2015) Trade and investment under policy uncertainty: Theory and firm evidence. American Economic Journal: Economic Policy 7(4):189–222. https://doi.org/10.1257/pol.20140068

Handley K, Limão N (2017) Policy uncertainty, trade, and welfare: Theory and evidence for China and the United States. American Economic Review 107(9):2731–2783. https://doi.org/10.1257/aer.20141419

Holm S (1979) A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6(2):65–70

Huang Y, Lin C, Liu S, et al (2023) Trade networks and firm value: Evidence from the U.S.–China trade war. Journal of International Economics 145:103811. https://doi.org/10.1016/j.jinteco.2023.103811

Imbens GW, Manski CF (2004) Confidence intervals for partially identified parameters. Econometrica 72(6):1845–1857. https://doi.org/10.1111/j.1468-0262.2004.00555.x

Kolari JW, Pynnönen S (2010) Event study testing with cross-sectional correlation of abnormal returns. Review of Financial Studies 23(11):3996–4025. https://doi.org/10.1093/rfs/hhq072

Kydland FE, Prescott EC (1977) Rules rather than discretion: The inconsistency of optimal plans. Journal of Political Economy 85(3):473–491. https://doi.org/10.1086/260580

Leeper EM, Walker TB, Yang SCS (2013) Fiscal foresight and information flows. Econometrica 81(3):1115–1145. https://doi.org/10.3982/ECTA8337

Maggi G, Rodriguez-Clare A (1998) The value of trade agreements in the presence of political pressures. Journal of Political Economy 106(3):574–601. https://doi.org/10.1086/250022

Mertens K, Ravn MO (2012) Empirical evidence on the aggregate effects of anticipated and unanticipated US tax policy shocks. American Economic Journal: Economic Policy 4(2):145–181. https://doi.org/10.1257/pol.4.2.145

Olken BA (2015) Promises and perils of pre-analysis plans. Journal of Economic Perspectives 29(3):61–80. https://doi.org/10.1257/jep.29.3.61

Pástor Ľ, Veronesi P (2012) Uncertainty about government policy and stock prices. The Journal of Finance 67(4):1219–1264. https://doi.org/10.1111/j.1540-6261.2012.01746.x

Rao A, Lucey B, Kumar S (2025) Financial market reactions to U.S. tariff announcements: Evidence from trade-surplus and trade-deficit countries. Finance Research Letters 81:107465. https://doi.org/10.1016/j.frl.2025.107465

Rodrik D (1991) Policy uncertainty and private investment in developing countries. Journal of Development Economics 36(2):229–242. https://doi.org/10.1016/0304-3878(91)90034-S

Switzer LN, El Meslmani N, Bajaj A (2026) From trade agreement to trade war: USMCA, tariff uncertainty, and stock market spillovers. Finance Research Letters 103:110139. https://doi.org/10.1016/j.frl.2026.110139

Whaley RE (2000) The investor fear gauge. The Journal of Portfolio Management 26(3):12–17. https://doi.org/10.3905/jpm.2000.319728

Xu X, Mao L, Zhang M (2025) Geopolitical decoupling and strategic export reallocation: Evidence from China's 'made in China 2025' sectors during the 2018–2019 U.S.–China trade war. Open Economies Review 37(2):377–407. https://doi.org/10.1007/s11079-025-09838-0