

Filtering Through the Storm

Comparing Volatility-Treatment Methods for DSGE Filtration and Estimation in Small Open Economies

Óscar I. Ávila-Montealegre* Fredy A. Castañeda-Valdés[†] Oscar Jaulín-Méndez[‡]
Juan A. Rincón-Galvis[§]

Macroeconomic Modeling Department[¶]
Banco de la República

September 1, 2026

Abstract

COVID-19 exposed a gap in standard DSGE practice: the constant-variance, mean-reverting shock processes these models rely on have no principled way to absorb a handful of quarters an order of magnitude more volatile than anything in the estimation sample, and practitioners need a general protocol for extreme-volatility episodes, not an ad-hoc pandemic patch. We compare four such protocols, no treatment, missing real-side data, heteroskedastic AR(1) shocks, and heteroskedastic ARMA(1,1) shocks, on CUMBIA, Banco de la República's estimated small-open-economy DSGE model for Colombia, across Bayesian parameter estimates, impulse responses, historical shock decompositions, and out-of-sample forecast accuracy. CUMBIA's transmission mechanism proves robust to the choice of treatment, but the treatments disagree sharply on what a crisis quarter's variance represents and how long its effects should persist. The heteroskedastic AR(1) treatment is the most accurate post-COVID forecaster on every variable and horizon we evaluate, while the ARMA(1,1) treatment is easier to implement without error and easier to communicate to a policy committee. We recommend the ARMA(1,1) specification as a production default, with the AR(1) treatment reported alongside it as a supplementary, scarring-robust check in the years following any future extreme-volatility episode.

Keywords: DSGE; Kalman Filter; Heteroskedastic Shocks; COVID-19; Bayesian Estimation; Forecast Evaluation

*Head of Model Development Section, Macroeconomic Modeling Department, Banco de la República, e-mail: oav-ilamo@banrep.gov.co

[†]Specialized Economist, Macroeconomic Modeling Department, Banco de la República, e-mail: fcastava@banrep.gov.co

[‡]Junior Research Economist, Macroeconomic Modeling Department, Banco de la República, e-mail: ojaulime@banrep.gov.co

[§]Specialized Economist, Macroeconomic Modeling Department, Banco de la República, e-mail: jrincoga@banrep.gov.co

[¶]A special thanks to Johannes Pfeifer for his valuable feedback as an external reviewer on the CUMBIA project. We also thank Franz Hamann, and José Pulido, for their valuable comments. Finally, we appreciate the contributions of our junior economists, César A. Vargas, Natalia Rodríguez and Sergio E. Avilán, for their superb research assistance. All remaining errors are ours.

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are built on a simplifying premise that is rarely stated out loud because it is rarely tested: shocks are drawn, period after period, from the same distribution. Persistence is allowed to vary shock by shock, but variance is not; a demand shock in a tranquil quarter and a demand shock in a crisis quarter are treated by the Kalman filter as draws from the identical Gaussian. COVID-19 made that premise untenable in a way no ordinary recession had. Colombian GDP, like output nearly everywhere, fell and rebounded within a handful of quarters by an order of magnitude more than anything in the pre-pandemic estimation sample, and a model that insists on explaining a move that large with an ordinary-variance shock is not describing an extreme event, it is quietly reinterpreting the entire shock distribution to accommodate one. Left unaddressed, this does not just distort the fit of a few pandemic quarters: because a DSGE model's shocks are structural, an outlier absorbed by the wrong process, or forced through a homoskedastic, mean-reverting channel, can bias the estimated parameters governing the model's propagation mechanism for the whole sample, contaminating exactly the object, the structural model, that motivated using a DSGE framework in the first place.

COVID-19 is this paper's motivating case, but it is not the only one, and treating it as a one-off problem requiring a one-off patch would be a mistake. Colombia's own recent history offers two other episodes of comparably extreme, if less globally synchronized, macroeconomic volatility: the 2008–09 global financial crisis and the 2014–16 collapse in oil prices, which hit a commodity-exporting small open economy's terms of trade, fiscal position, and exchange rate simultaneously. Any protocol a modeling team adopts for handling COVID-19 will be asked, sooner or later, to handle the next such episode, whatever form it takes. The research question this paper asks is therefore general by design: how do the volatility-treatment methods available to a DSGE practitioner affect filtration and estimation outcomes, not only during COVID-19 specifically, but across extreme episodes of different origins, and which of them should a modeling team adopt as its default going forward?

Figure 1.1 shows that this is not a Colombian idiosyncrasy. For a panel of OECD economies plus Colombia (coverage varies slightly by series; see the caption), it plots each country's own standard deviation of quarterly GDP growth, CPI inflation, and the change in a short-term interest rate during τ_{Covid} against that same country's standard deviation in normal quarters, excluding τ_{GFC} , τ_{Oil} , and τ_{Covid} itself. For real activity the pattern is close to universal: every country in the panel was more volatile in GDP growth during the pandemic than in its own typical quarter, several of them by several multiples. Inflation and the policy rate tell a more selective story: only a minority of countries were more volatile in these two series during τ_{Covid} itself, since most central banks held policy rates near the effective lower bound throughout the window and the inflation surge that eventually followed the pandemic arrived, for most economies, only after it ended. That asymmetry across variables is not incidental to this paper's argument, it is close to its center: it is precisely the real-activity variables whose 2020–21 behavior an ordinary, constant-variance shock cannot explain, while the policy-relevant nominal variables a filter still needs to condition on remained comparatively well-behaved in real time, a pattern one of the treatments below, which withholds exactly these real-activity variables during a high-volatility window, is built to exploit directly. Appendix C.6 develops the same cross-country evidence with two quantile-quantile comparisons and shows that, where volatility did spike, most dramatically in GDP growth, it did so as a small number of extreme, fat-tailed quarters rather than as a uniform widening of the whole distribution, exactly the kind of departure from the constant-variance Gaussian shock that Section 4 is built to address.

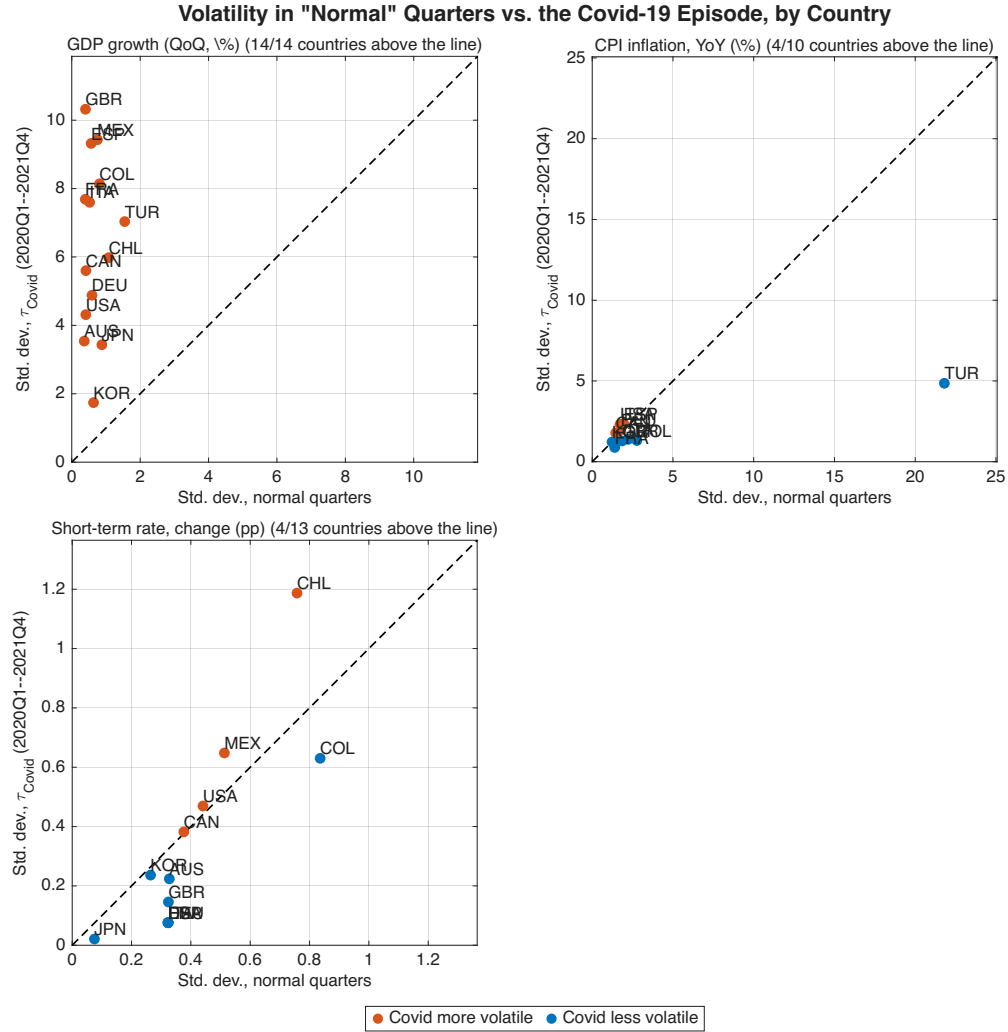


Figure 1.1: Standard deviation of quarterly GDP growth, CPI inflation, and the change in a short-term interest rate during normal quarters (x-axis) against the standard deviation of the same series during τ_{Covid} (y-axis), for a panel of OECD economies plus Colombia. Points above the 45-degree line were more volatile during the pandemic than in the country's own typical quarter; the count above the line is reported in each panel's title, and countries are labeled by ISO3 code. Data availability varies across the three series (14 countries for GDP growth, 10 for CPI inflation, 13 for the interest rate) since not every country reports every series at the required frequency; see Appendix C.6 for construction details.

We compare four such methods, described in full in Section 4. Treatment 1 is the do-nothing benchmark: the model is estimated with its ordinary, constant-variance shock processes, and every episode is left to be absorbed however the likelihood sees fit. Treatment 2 withholds the real-side observables, GDP, consumption, investment, and trade flows, during a high-volatility window, letting the Kalman filter's cross-equation restrictions infer a model-consistent counterfactual path for the real economy from the observables that remain, policy rates and inflation, rather than forcing those restrictions to rationalize an extreme, and possibly misleading, realized value. Treatments 3 and 4 instead modify the shock processes directly, using Dynare's `heteroskedastic_shocks` block to allow selected shocks' variances to rise during a designated episode window: Treatment 3

rescales the variance of the shock's own, ordinary AR(1) innovation, so that an extreme realization decays at the shock's normal persistence once the episode ends, while Treatment 4 routes the added variance through a dedicated, transitory auxiliary shock constructed, following Smets and Wouters (2007)'s precedent, to self-cancel within one to two quarters regardless of that shock's ordinary persistence. These four choices are not exhaustive of everything the literature has proposed, Section 2 surveys structural-break-robust filtering, dedicated Covid-shock designs, and stochastic-volatility alternatives we do not implement directly, but they represent the tools most readily available inside a standard Dynare-based estimation workflow, which is precisely why a systematic comparison between them is of practical, not just academic, interest.

We conduct that comparison on CUMBIA, Banco de la República's small-open-economy New Keynesian DSGE model for Colombia, estimated on Colombian data (Section 3). CUMBIA is a natural testbed for three reasons. First, it is estimated, actively used, and periodically re-estimated for policy work at a central bank, so the question of which volatility treatment to adopt is not hypothetical for its maintainers; it is a decision the modeling team has had to make, and will have to make again. Second, Colombia's own macroeconomic history conveniently supplies all three of the extreme-volatility episodes motivating this paper, the 2008–09 financial crisis, the 2014–16 oil-price collapse, and COVID-19, in a single, consistently constructed dataset (Table 5.1), letting us stress-test every treatment against more than one kind of shock. Third, CUMBIA's small-open-economy structure, with its debt-elastic risk premium and direct exposure to foreign demand, foreign prices, and remittances, gives the comparison real stakes: several of the channels most likely to be contaminated by an untreated extreme episode, the preferences shock, the risk premium, foreign demand, are also channels central to how the model transmits external shocks to the domestic economy, so getting their estimation wrong is not a narrow technical concern.

This paper's contribution is threefold. First, it provides, to our knowledge, the first side-by-side comparison of these four volatility-treatment methods implemented on one estimated model and evaluated against one common dataset, rather than each method being validated in isolation on a different model, country, or benchmark, as has been the norm in the literature surveyed in Section 2. Second, because the comparison spans every dimension a modeling team would actually need to weigh before adopting a treatment, impulse responses, posterior parameter estimates, historical shock decompositions, and out-of-sample forecast accuracy (Section 6), it yields concrete, actionable guidance rather than a purely theoretical ranking, guidance we make explicit in our recommendation (Section 7). Third, the Dynare implementation underlying Treatments 2–4, including the practical pitfalls we ran into and resolved along the way (Appendix B.1), is designed to be reusable by any DSGE modeling team facing its own extreme-volatility episode, not just by CUMBIA's maintainers.

The results, previewed here and developed fully in Sections 6 and 7, converge on a picture that is more nuanced than either “always treat volatility” or “never bother.” CUMBIA's qualitative transmission mechanism turns out to be robust to the choice of treatment: across fifteen real, nominal, and financial variables and five shocks spanning both Covid-treated and untreated processes, every treatment produces impulse responses of the same sign, shape, and approximate timing, so this choice does not put the model's basic economics on the table. What it does change is more consequential than it first appears. Left untreated, the shocks best positioned to absorb an extreme episode, preferences, foreign demand, wages, have their estimated variances measurably contaminated, four to seven Treatment-1 standard errors in some cases; every alternative treatment corrects this in the same direction, but Treatments 2 and 4 share an implicit assumption that Treatment 3 does not, that the episode leaves no lasting mark on the estimated law of motion for real activity, one through an information gap and one through construction, while only Treatment

3 allows a crisis-quarter disturbance to echo forward at the shock's own ordinary persistence. This distinction is not merely conceptual: Treatment 3 is the more accurate forecaster of Colombian macroeconomic aggregates in the years following COVID-19 on every variable and horizon we evaluate, a result that survives a robustness check across alternative episode-scale calibrations, yet Treatment 4 remains considerably easier to implement without error and easier to defend to a policy committee under time pressure. Our recommendation, developed in Section 7, is accordingly conditional rather than a single winner: Treatment 4 as the production default, with Treatment 3 reported alongside it as a supplementary, scarring-robust check, particularly in the years immediately following any future extreme-volatility episode.

The remainder of the paper proceeds as follows. Section 2 situates our four treatments within the broader literature on structural breaks, missing data, and time-varying volatility in state-space and DSGE models. Section 3 summarizes the CUMBIA model. Section 4 details the four volatility treatments and their Dynare implementation. Section 5 describes the data and the three high-volatility episodes used to identify each treatment's window. Section 6 reports results across a missing-data counterfactual, Bayesian estimation, impulse responses, shock decomposition, and forecast evaluation. Section 7 weighs the resulting trade-offs and states our practical recommendation, and Section 8 concludes.

2 Literature Review

Structural breaks and outliers in DSGE estimation State-space representations have long provided the natural language for handling structural breaks and outlying observations in time series. Harvey (1989) and Durbin and Koopman (2012) show that once a model is cast in state-space form, extreme or missing observations can be accommodated directly within the Kalman filter and smoother, either by allowing time-varying disturbance variances or by simply omitting the update step for a given observation, without altering the underlying transition equations. DSGE models inherit this machinery through their linearized state-space representation, but they add a complication that pure statistical models do not face: the parameters governing the transition equations are structural, so an outlier that is absorbed into the wrong shock, or forced through a homoskedastic, mean-reverting AR(1) process, does not just distort a reduced-form fit; it can bias the estimated structural parameters themselves and contaminate the model's implied propagation mechanisms. Canova (2014) makes a related point about the risks of imposing an overly rigid mapping between DSGE models and the raw data, arguing that standard filtering and detrending choices can bias structural estimates and the policy conclusions drawn from them. The COVID-19 pandemic is an extreme version of this general problem: a small number of quarters generated deviations an order of magnitude larger than anything in the pre-pandemic sample, making the choice of how to treat them a first-order concern for estimation, not a robustness footnote.

COVID-19 in the DSGE literature A first strand of the response to this problem, developed initially outside DSGE models, treats extreme observations as identified shocks with a distinct volatility process. Lenza and Primiceri (2022) propose estimating a Bayesian VAR in which pandemic-quarter residuals are assigned a period-specific scale factor on their variance rather than being dropped or treated as ordinary innovations; they show that discarding the observations may be acceptable for parameter estimation, but is inappropriate for forecasting, since it understates uncertainty going forward. Carriero et al. (2024) refine this idea by combining transitory (outlier) and persistent stochastic volatility components in a BVAR, so that a model can

distinguish a one-off spike from a lasting shift in the volatility regime. These VAR-based results carry over directly to DSGE estimation, where Faria-e-Castro (2024, June) describes how the St. Louis Fed’s medium-scale New Keynesian model accounts for COVID-19 through dedicated pandemic shocks, and Ferroni et al. (2022, September) and Ferroni et al. (2024) propose an explicit *Covid shock* that loads onto the wedges underlying a model’s usual shocks and evolves according to real-time professional forecasts, allowing standard business-cycle models to be estimated across the pre-, during-, and post-pandemic sample without re-specifying their core structure. In a complementary literature on high-frequency nowcasting, Antolín-Díaz et al. (2024) show that allowing for fat-tailed idiosyncratic shocks in a dynamic factor model is essential to track the size of the pandemic collapse without letting it distort the estimated trend and business-cycle components, a concern that maps closely onto our Treatments 3 and 4 below.

Missing observations and the Kalman filter A second, less commonly used strand of the DSGE literature exploits the fact that the Kalman filter does not require a complete observation vector at every date: when an observable is missing, the filter simply skips the corresponding update step and lets the state evolve according to the model’s own transition dynamics, using whatever other observables remain available at that date to inform the update (Durbin and Koopman, 2012; Harvey, 1989). This machinery was designed for genuine data unavailability, for example in mixed-frequency nowcasting applications where higher-frequency indicators are used to fill in unobserved values of lower-frequency aggregates (Antolín-Díaz et al., 2024). Our Treatment 2 repurposes it for a different end: rather than a variable being unavailable, we deliberately treat real-side observables (GDP, consumption, investment, trade flows) as missing during a high-volatility window, and let the model’s cross-equation restrictions, disciplined by the observables that remain (policy rates, inflation), generate a model-consistent counterfactual path for the real economy. This is closer in spirit to conditional forecasting and counterfactual analysis than to a standard missing-data problem, and to our knowledge it has not been used systematically as a volatility-treatment device and compared against the alternatives considered here.

Time-varying volatility / heteroskedastic shocks A third strand modifies the DSGE model’s shock processes directly, allowing their variances to change over time rather than treating volatility as constant. Justiniano and Primiceri (2008) provide early evidence, in the context of the Great Moderation, that the volatility of the structural shocks driving U.S. business cycles is itself time-varying, and that estimating this variation is informative about the sources of macroeconomic stability. Fernández-Villaverde et al. (2011) apply the same logic to a small open economy setting closely related to ours, showing that stochastic volatility in the interest rate faced by emerging economies has quantitatively important effects on output, consumption, investment, and hours, which is directly relevant given CUMBIA’s small-open-economy structure and its debt-elastic interest rate premium. Cúrdia et al. (2014) take an alternative route to the same problem, replacing the Gaussian shock distribution with a Student’s t to allow rare, large shocks without adding explicit time variation in volatility; we do not pursue this route here, but it represents a natural robustness check for future work. Dynare operationalizes the time-varying-volatility approach through its `heteroskedastic_shocks` block (Adjemian et al., 2024), which lets the researcher rescale the standard deviation of selected shocks in selected periods without re-specifying the model’s equations, and which is exactly the tool used by Ferroni et al. (2022, September) and Ferroni et al. (2024) and by CUMBIA’s own baseline Covid treatment. The remaining question this literature does not fully resolve is what persistence to assume for the added volatility itself: our

Treatment 3 follows the standard practice of scaling the variance of an otherwise AR(1) shock, while our Treatment 4 draws on the precedent of Smets and Wouters (2007), who append a transitory MA(1) term to one of their shock processes precisely to absorb short-lived disturbances without distorting the estimated persistence of the underlying AR(1) component, and asks whether this extra flexibility improves on a purely AR(1) heteroskedastic specification.

Positioning of this paper Each of these strands, structural-break-robust filtering, event-study Covid shocks, missing-data counterfactuals, and stochastic-volatility shock processes, has been proposed and validated on its own, typically on a different model, a different country, and against a different set of benchmarks. What is missing is a head-to-head comparison of these treatments, implemented on a single estimated model and a single dataset, evaluated consistently across the dimensions that matter for a policy institution: impulse responses, posterior parameter estimates, historical shock decompositions, and out-of-sample forecast accuracy. This paper fills that gap using CUMBIA, Banco de la República’s small-open-economy DSGE model for Colombia, comparing no treatment, missing real data, heteroskedastic AR(1) shocks, and heteroskedastic ARMA(1,1) shocks against one another, with the goal of providing concrete, actionable guidance for modeling teams that will inevitably face the next high-volatility episode.

3 The CUMBIA Model

CUMBIA is a small-open-economy New Keynesian DSGE model estimated on Colombian data, developed at Banco de la República as an update to the institution’s previous policy model, PATACON (González et al., 2011). Full derivations, the analytical steady state, and the parameter list are reported in the companion technical documentation (Ávila-Montealegre et al., 2026); this section summarizes only the equations and structure needed to understand how each volatility treatment in Section 4 interacts with the model, and reports the remaining equilibrium conditions in Appendix A.

3.1 Overview

A representative household chooses consumption, labor supply, capital investment, and domestic and foreign bond holdings, with internal habit formation in consumption and a debt-elastic risk premium that closes the small open economy (Garcia-Cicco et al., 2010; Schmitt-Grohé and Uribe, 2003). Labor is intermediated through three layers, an employer firm, a competitive labor packer, and a continuum of monopolistically competitive labor unions that set wages subject to Rotemberg adjustment costs, producing a New Keynesian wage Phillips curve. On the production side, a continuum of monopolistically competitive firms produce differentiated domestic intermediate varieties under Rotemberg price rigidities and are aggregated by a competitive firm into a homogeneous domestic good. Consumption, investment, and raw materials are each produced by combining this domestic good with imported inputs, subject to the input adjustment costs of Ávila-Montealegre and Mix (2024) and Drozd et al. (2021), which generate a dynamic (time-varying) trade elasticity and a gradual pass-through of foreign prices. Imported goods mirror the domestic block: heterogeneous import varieties are produced under monopolistic competition and price rigidities, then aggregated into a homogeneous imported good. Exports respond to relative export prices and foreign demand with partial adjustment, and the central bank follows a

Taylor-type rule targeting core inflation and an output gap. Nearly every exogenous driving force in the model, productivity, preferences, investment efficiency, trade efficiency, foreign demand, foreign prices, and the foreign interest rate, is modeled as a univariate autoregressive process, described next.

3.2 Exogenous shock processes

Every exogenous process in CUMBIA is specified, in logarithms, as either a multiplicative AR(1) process or, where the code allows for a transitory component, a multiplicative ARMA(1,1) process. Writing $\zeta_t \equiv \log(Z_t)$ for a generic exogenous variable Z_t with steady state $\bar{\zeta}$ and persistence ρ , the two baseline specifications are

$$\zeta_t = \rho \zeta_{t-1} + (1 - \rho) \bar{\zeta} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2), \quad (3.1)$$

$$\zeta_t = \rho \zeta_{t-1} + (1 - \rho) \bar{\zeta} + \epsilon_t + \epsilon_t^{MA} - \rho \epsilon_{t-1}^{MA}, \quad \epsilon_t^{MA} \sim \mathcal{N}(0, \sigma^2), \quad (3.2)$$

where the transitory (moving-average) innovation ϵ_t^{MA} shares the same standard deviation as the structural innovation ϵ_t but is fully unwound one period later, so that it contributes no persistence beyond the model's own general-equilibrium propagation. Roughly thirty exogenous processes in CUMBIA, covering productivity, demand, investment and trade efficiency, the foreign block, and the inflation and policy-rate targets, take one of these two forms. This is precisely the object each of the four treatments in Section 4 acts on: Treatment 1 leaves (3.1)/(3.2) untouched with a constant σ ; Treatment 2 instead removes real-side observables from the Kalman filter's information set during the volatility episode; and Treatments 3 and 4 add an episode-specific, otherwise-zero heteroskedastic term to σ in (3.1) and (3.2), respectively, as detailed in Sections 4.3 and 4.4.

3.3 Observables and measurement equations

CUMBIA is estimated on 27 quarterly observables, grouped in Table 3.1 by whether they are real-side national-accounts series, nominal/price series, or purely exogenous foreign series. This grouping matters directly for Treatment 2: the real-side block is the set of series we treat as missing during a high-volatility episode, while the price, rate, and foreign blocks remain observed throughout and are what the Kalman filter uses to reconstruct a model-consistent counterfactual for the real economy.

The mapping from model variables to observables is direct. For example, GDP and foreign demand growth are observed as

$$\Delta y_t^{obs} = \frac{y_t}{y_{t-1}} (1 + g_t^{nz}) - 1 + \epsilon_t^{disc_y}, \quad (3.3)$$

$$\Delta y_t^{\star, obs} = \frac{y_t^{\star}}{y_{t-1}^{\star}} (1 + g_t^{\star}) - 1, \quad (3.4)$$

and the nominal policy rate and headline inflation map one-for-one onto their model counterparts, $i_t^{nom, obs} = i_t^{nom}$ and $\pi_t^{obs} = \pi_t$. The remaining measurement equations, including the national-accounts discrepancy terms $B\Delta x_t^{obs}$, follow the same logic and are reported in full in the companion documentation (Ávila-Montealegre et al., 2026).

Block	Observables
Real activity (<i>missing under Treatment 2</i>)	GDP, consumption, investment, exports, and imports growth rates; foreign GDP growth; output trend growth
Prices and rates (<i>always observed</i>)	Headline, core, food, and regulated inflation; nominal policy rate; risk premium
Foreign block (<i>always observed, exogenous</i>)	Fed funds rate, foreign inflation, relative import/export prices, remittances growth
Discrepancies	Consumption, investment, imports, exports, and interest-rate discrepancies

Table 3.1: CUMBIA’s observables, grouped by role in Treatment 2

4 Four Treatments of High-Volatility Periods

Let τ denote the set of quarters comprising a high-volatility episode (e.g., $\tau = \{2020Q1, \dots, 2021Q4\}$ for Covid-19; Section 5 also considers the 2008–09 global financial crisis and the 2014–15 oil-price collapse). All four treatments below are estimated on the same sample, the same observables, and the same priors for the model’s structural parameters; they differ only in how the shock processes of Section 3.2 or the observation equation of Section 3.3 are modified during τ . This design isolates the effect of the volatility treatment itself from any other difference in specification.

4.1 Treatment 1 — No treatment (baseline)

The baseline treatment estimates CUMBIA exactly as specified in Section 3, over the full 2000Q1–2025Q4 sample, with every exogenous process following its plain AR(1) or ARMA(1,1) law of motion, (3.1) or (3.2), and a *constant* innovation variance σ throughout, including the quarters in τ . No observable is treated differently during the episode.

This treatment is not a straw man: it is the estimation strategy implied by simply running an existing, undocumented DSGE model through the full sample without adjustment, and it remains common practice when a volatility episode is not (yet) recognized as requiring special treatment. Its implicit assumption is strong, however: it treats a Covid-quarter realization of, say, the preferences or investment-efficiency shock as a draw from the *same* distribution that generates an ordinary quarter’s fluctuations. Because the likelihood must rationalize the 2020Q2 collapse and 2020Q3 rebound using only the structural parameters and shock processes calibrated for tranquil times, this can bias estimated persistences and variances toward whatever shocks are best positioned to absorb the extreme observations (Canova, 2014; Lenza and Primiceri, 2022), and, since the Kalman smoother uses the whole sample, distort the smoothed shock decomposition for quarters well before and after τ itself. Treatment 1 serves as the control group against which Treatments 2–4 are benchmarked in Section 6.

4.2 Treatment 2 — Missing data on real variables

Treatment 2 leaves every shock process unchanged from Treatment 1, homoskedastic AR(1) or ARMA(1,1) throughout, and instead modifies the observation equation: for $t \in \tau$, the real-activity

observables in Table 3.1, the five domestic national-accounts growth rates (GDP, consumption, investment, exports, imports) together with foreign GDP growth and output trend growth, are declared missing, while the price, rate, and remaining foreign-block observables continue to be observed as usual. Foreign demand growth and the trend-growth observable are included alongside the domestic block because both are themselves real-activity measures: the former is foreign real GDP growth, subject to the same pandemic-era dislocation as its domestic counterpart, and the latter is constructed from a one-sided filter of domestic real GDP, so leaving either observed would smuggle real-side pandemic information back into the filter through the back door. Mechanically, the Kalman filter simply omits the update step for the missing entries of the observation vector at each $t \in \tau$, so the state, and with it the model’s implied path for the unobserved real variables, evolves purely from the transition equations and from whatever the remaining observables (policy rate, inflation, the Fed funds rate, foreign inflation, and so on) reveal about the underlying state (Durbin and Koopman, 2012; Harvey, 1989).

The output of this exercise is a model-consistent *counterfactual*: what CUMBIA would have inferred about output, consumption, investment, and trade during the episode had it observed only prices, rates, and the non-GDP foreign block, without ever seeing the realized collapse in domestic or foreign real activity. This is informative in two ways. First, comparing this counterfactual to the actual data quantifies how much of the swing in real activity is already implied by co-movement with prices and rates under the model’s cross-equation restrictions, versus how much is genuinely new information contained in the real variables. Second, because the shock processes are never asked to rationalize the realized real-side collapse, Treatment 2 sidesteps the parameter contamination risk of Treatment 1 entirely, at the cost of having, by construction, no in-sample fit to evaluate against the actual real data during τ ; Treatment 2 is therefore judged mainly through its shock decomposition and out-of-sample forecast performance in Section 6, not through in-sample fit over the missing window.

4.3 Treatment 3 — Heteroskedastic shocks, AR(1) processes

Treatment 3 keeps every process in its plain AR(1) form, (3.1), but rescales the standard deviation of the process’s own persistent innovation during τ , using Dynare’s `heteroskedastic_shocks` block (Adjemian et al., 2024):

$$\zeta_t = \rho \zeta_{t-1} + (1 - \rho) \bar{\zeta} + \epsilon_t, \quad \epsilon_t \sim \begin{cases} \mathcal{N}(0, \sigma^2) & t \notin \tau \\ \mathcal{N}(0, (s_\tau \sigma)^2) & t \in \tau, \end{cases} \quad (4.1)$$

where σ is the shock’s own baseline standard deviation, estimated jointly with every other parameter exactly as in Treatment 1, and s_τ is a *calibrated scale factor* ($s_{GFC} = s_{Oil} = 2$, $s_{Covid} = 3$, the latter matching the more severe dislocation documented in Section 5), following the logic of Ferroni et al. (2022, September) and Ferroni et al. (2024) and Faria-e-Castro (2024, June) for DSGE models and Carriero et al. (2024) and Lenza and Primiceri (2022) for the analogous BVAR treatment. s_τ is calibrated rather than estimated for a purely computational reason documented in Appendix B: Dynare’s `heteroskedastic_shocks` scales option is not re-evaluated at each trial parameter vector during Bayesian mode-finding, so a version of this specification with s_τ nominally free (given a prior) produced a likelihood that was, by construction, exactly flat in s_τ . s_τ is *shared* across every shock Table 5.1 assigns to episode τ , both to keep the specification parsimonious and to reflect the reading that a crisis raises the whole economy’s volatility by a common factor rather than each shock idiosyncratically. Because the rescaled innovation still enters through the same

AR(1) recursion as any ordinary quarter's shock, an unusually large realization of ϵ_t for $t \in \tau$ propagates forward at the process's normal persistence ρ : the extra pandemic-quarter variance leaves an echo, decaying geometrically at rate ρ , in the quarters immediately following the episode. This is a deliberate feature of Treatment 3, not an oversight, it is the natural DSGE analogue of the VAR-based outlier treatments in Lenza and Primiceri (2022), and whether that residual persistence is desirable or a nuisance is precisely one of the questions Treatment 4 is designed to probe.

4.4 Treatment 4 — Heteroskedastic shocks, ARMA(1,1) processes

Treatment 4 rescales the variance of a *transitory* innovation added on top of the AR(1) process, rather than the persistent innovation itself, following the ARMA(1,1) form (3.2) and matching CUMBIA's existing Covid-quarter specification:

$$\zeta_t = \rho \zeta_{t-1} + (1 - \rho) \bar{\zeta} + \epsilon_t + \epsilon_{t \in \tau}^{Het} - \rho \epsilon_{(t-1) \in \tau}^{Het}, \quad \epsilon_t^{Het} \sim \mathcal{N}(0, (\sigma^{Het})^2), \quad (4.2)$$

where ϵ_t^{Het} is set to zero outside τ and σ^{Het} is estimated only for quarters within the episode. The coefficient $-\rho$ on the lagged term is chosen so that it exactly cancels the contribution of ϵ_{t-1}^{Het} to the process one period later: a burst of extra variance during τ is, by construction, fully unwound the following period and leaves no persistent mark on the state beyond the model's own general-equilibrium propagation across variables. This is the same logic Smets and Wouters (2007) use to attach a transitory MA(1) term to their price-markup shock, and it is the specification the Kalman smoother uses to isolate CUMBIA's Covid-period shocks as an additional, self-contained source of volatility (see the companion documentation, (Ávila-Montealegre et al., 2026)). Relative to Treatment 3, Treatment 4 trades away the (arguably realistic) persistent echo of an extreme episode for a cleaner, surgical separation between “pandemic volatility” and the model's ordinary business-cycle dynamics; Section 6 evaluates whether that trade is worth making.

4.5 Summary comparison table

Table 4.1 summarizes the four treatments along the dimensions most relevant for a modeling team choosing among them: what is modified, how many additional parameters are estimated, whether the episode's extra volatility persists beyond τ , and what each treatment assumes about the nature of an extreme episode.

Treatment	What changes	Added parameters	Echo beyond τ ?	Implicit assumption
1. No treatment	Nothing	None	—	Episode is an ordinary draw from the model's usual shock distribution
2. Missing real data	Observation equation (real block dropped for $t \in \tau$)	None	N/A (no shock added)	Prices and rates alone are informative enough to pin down a counterfactual real path
3. Heteroskedastic AR(1)	$\sigma \rightarrow s_\tau \sigma$ in (3.1) for $t \in \tau$	One calibrated scale factor s_τ per episode (3 total, shared across shocks; see Appendix B)	Yes, decays at rate ρ	Episode is a large draw of the model's usual shocks, propagating normally
4. Heteroskedastic ARMA(1,1)	Transitory term added per (3.2) for $t \in \tau$	One σ^{Het} per shock/episode	No, self-cancels after one period	Episode is a distinct, self-contained disturbance that should not contaminate ordinary dynamics

Table 4.1: Comparison of the four volatility treatments

5 Data and Empirical Strategy

5.1 Data

All four treatments are estimated on the identical dataset: the 27 quarterly observables of Table 3.1, drawn from DANE, Banco de la República, Bloomberg, and FRED as detailed in Section 3 and the companion documentation (Ávila-Montealegre et al., 2026), over the sample 2000Q1–2025Q4. Data vintages, transformations, and calibration targets are unchanged from the companion CUMBIA specification; only the treatment of the shock processes (Section 3.2) or the observation equation (Treatment 2) varies across specifications, so that any difference in results is attributable to the volatility treatment itself rather than to a difference in the underlying data.

5.2 Identification of high-volatility windows

We consider three high-volatility episodes over the sample, each identified from its own macroeconomic footprint in Colombia rather than a common statistical rule, since the three episodes disturbed different parts of the model. The 2008–09 global financial crisis was transmitted to Colombia mainly through external channels, with real GDP growth slowing from 2.5% in 2008 to near zero in 2009 without the deep domestic contraction seen in many advanced economies (International Monetary Fund, 2010); we therefore date this episode 2008Q3–2009Q2 and restrict its heteroskedastic treatment to the foreign block. The 2014–16 oil-price collapse operated principally through the exchange rate and imported inflation: the peso depreciated sharply from July 2014, with the nominal exchange rate up 81% by its February 2016 peak, while headline inflation rose from around 3% in mid-2014 to a peak of roughly 9% in July 2016 (Banco de la República, 2017); we date this episode 2014Q3–2016Q4 to span the full collapse-depreciation-inflation sequence, and restrict its treatment to terms-of-trade and inflation-related shocks. The Covid-19 episode, 2020Q1–2021Q4, is the broadest of the three, disrupting demand, supply, and the foreign block

simultaneously, and follows the shock list already used in CUMBIA's baseline Covid treatment (Ávila-Montealegre et al., 2026).

Episode	Window	Shocks treated (Treatments 3–4)	Treatment 2
GFC	2008Q3– 2009Q2	Foreign interest rate, risk premium, foreign demand, foreign inflation	Not applied
Oil-price collapse	2014Q3– 2016Q4	Relative export and import prices (foreign currency), risk premium, food inflation, regulated inflation	Not applied
Covid-19	2020Q1– 2021Q4	Risk premium, foreign bonds demand, foreign demand, foreign inflation, output trend growth, domestic producer efficiency, investment-specific technology, imported aggregation efficiency, export demand efficiency, labor efficiency, preferences, marginal efficiency of investment, remittances, food inflation, regulated inflation	Applied

Table 5.1: High-volatility episodes: windows and treated shocks

Each shock's heteroskedastic term is switched on only during the episode(s) to which it is assigned in Table 5.1 and is zero otherwise, so Treatments 3 and 4 are each estimated once, over the full sample, with all three episodes' shock-specific variances active simultaneously in their respective windows, rather than as three separate re-estimations. Treatment 2 is applied only to the Covid-19 window: unlike the GFC and the oil-price collapse, which left Colombia's real national accounts comparatively undisturbed while operating mainly through foreign and relative-price shocks, Covid-19 produced a genuine dislocation in real activity, which is precisely the case Treatment 2's missing-data counterfactual is designed to address.

5.3 Common estimation setup

All four treatments share the same priors, the same sample, and the same computational settings, so that differences in results are attributable to the treatment alone. Priors for the structural and shock parameters are unchanged from the companion CUMBIA specification (Appendix A and (Ávila-Montealegre et al., 2026)); the additional σ^{Het} parameters introduced by Treatments 3 and 4 are estimated with the same prior family used for their corresponding baseline shock standard deviations.

To keep re-estimating four treatments (and, within Treatments 3–4, three episodes at once) computationally tractable, we estimate the posterior *mode* rather than sampling the full posterior: in Dynare terms, `mode_compute=5` (a quasi-Newton hill-climbing algorithm) with `mh_replic=0`, so no Metropolis-Hastings chain is run. This mode-only approach, standard in the Bayesian DSGE literature when the object of interest is the maximum of the posterior kernel rather than the full distribution (An and Schorfheide, 2007), means that Section 6 reports point estimates at the posterior mode together with asymptotic standard errors from the inverse Hessian at that mode, rather than posterior means and credible intervals from an MCMC sample; we flag this explicitly since

it is a simplification relative to full Bayesian estimation, and revisit its implications in Section 7. The sample spans 2000Q1–2025Q4, with the first 12 quarters (2000Q1–2002Q4) used only to initialize the Kalman filter’s recursions and excluded from the likelihood, following standard Dynare presampling practice (`presample=12`).

6 Results

6.1 Treatment 2: A missing-data counterfactual for real activity

Figure 6.1 plots, for each of the seven real-activity observables Treatment 2 declares missing over τ_{Covid} (Table 3.1), the Kalman-smoothed path CUMBIA infers using only prices, rates, and the non-GDP foreign block against the realized data. The smoothed and observed series coincide almost exactly outside τ_{Covid} , since the filter is fed the actual data there; the gap that opens up during the shaded window is therefore entirely attributable to withholding the real-activity observables, and is the counterfactual described in Section 4.

We summarize the size of that gap, variable by variable, with the root-mean-squared and mean-absolute deviation between the counterfactual and the realized data over exactly the eight withheld quarters:

$$\text{RMSE}_v = \sqrt{\frac{1}{|\tau_{Covid}|} \sum_{t \in \tau_{Covid}} \left(y_{v,t}^{obs} - y_{v,t}^{smoothed} \right)^2}, \quad \text{MAE}_v = \frac{1}{|\tau_{Covid}|} \sum_{t \in \tau_{Covid}} \left| y_{v,t}^{obs} - y_{v,t}^{smoothed} \right|, \quad (6.1)$$

where v indexes the seven variables, $y_{v,t}^{obs}$ is the realized (quarter-on-quarter) growth rate, $y_{v,t}^{smoothed}$ is Treatment 2’s smoothed counterfactual, and $\tau_{Covid} = \{2020Q1, \dots, 2021Q4\}$, so $|\tau_{Covid}| = 8$. This is not a forecast-error metric in the usual sense, since the counterfactual is a full-sample *smoothed* estimate rather than a real-time projection; it measures how much of the pandemic-quarter swing in each variable is left unexplained once the model is denied the real-side data and must infer real activity purely from co-movement with prices, rates, and the foreign block, the same RMSE construction is reused for genuine out-of-sample comparisons across all four treatments in Section 6.5.

The gap is far from uniform across variables. It is smallest for foreign GDP growth (RMSE = 4.06 pp) and output trend growth (RMSE = 1.70 pp), consistent with both being either exogenous to the domestic block or constructed from a smoothed filter of GDP itself, and largest for investment growth (RMSE = 16.13 pp) and imports growth (RMSE = 13.43 pp), the two national-accounts components with the most volatile pandemic-era realizations. GDP and consumption growth fall in between (RMSE = 7.73 and 8.12 pp, respectively).

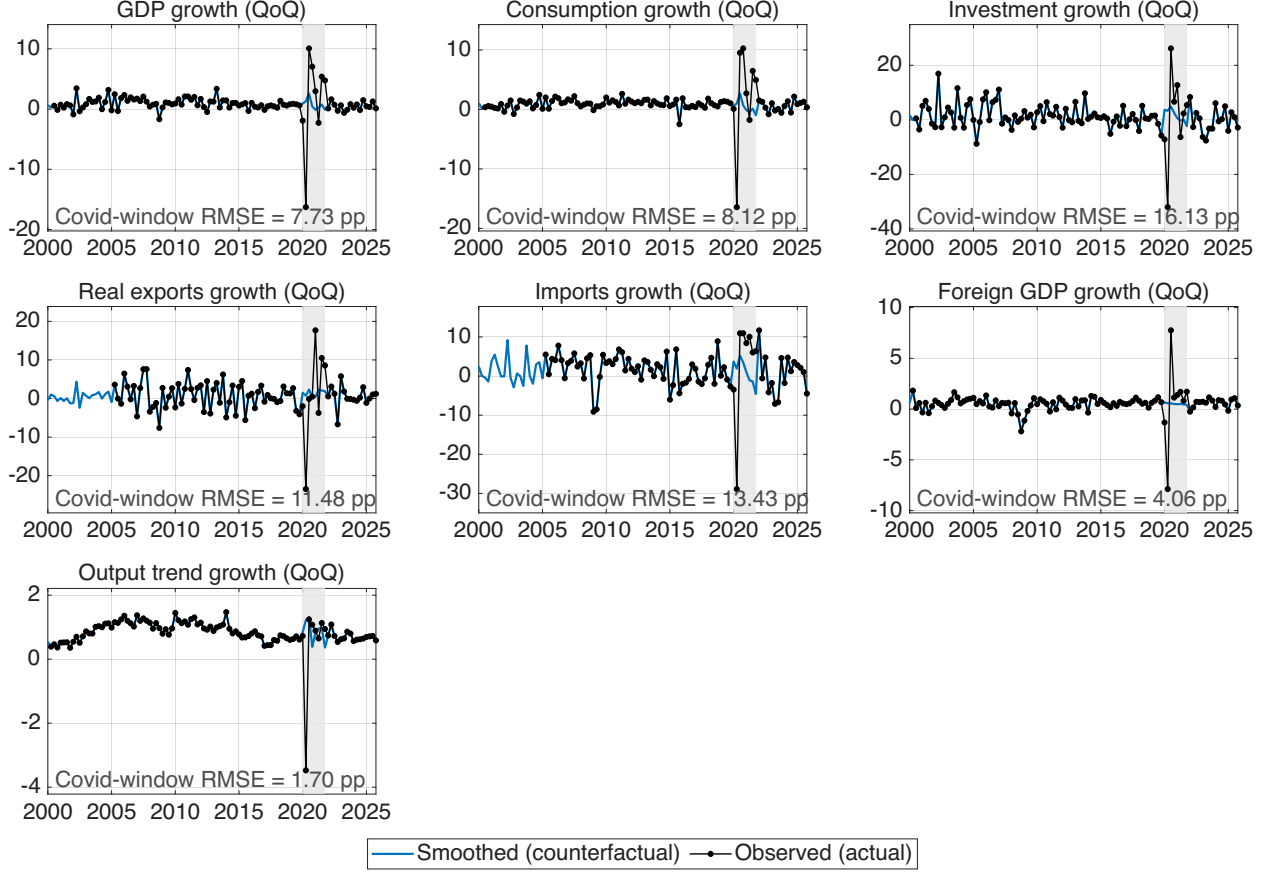


Figure 6.1: Treatment 2: Kalman-smoothed counterfactual vs. observed data for the seven real-activity observables declared missing over 2020Q1–2021Q4 (shaded). RMSE reported per panel is computed over the shaded window only, per (6.1).

6.2 Bayesian Estimation

Section 4 motivated Treatments 2–4 partly as a defense against parameter contamination: forcing Treatment 1’s constant-variance shocks to rationalize the 2020Q2 collapse and rebound, the argument went, should bias whichever shocks are best positioned to absorb it. This section checks that claim directly, comparing the Bayesian posterior mode (Section 5, `mode_compute=5`) and its Hessian-based standard error for every structural parameter and shock standard deviation the four treatments estimate in common, 38 structural parameters and 19 shock standard deviations, 57 in total.¹ For parameter θ and treatment $k \in \{2, 3, 4\}$, we standardize the shift in its posterior mode relative to Treatment 1 by Treatment 1’s own posterior standard error,

$$z_{\theta,k} = \frac{\hat{\theta}_k - \hat{\theta}_1}{\text{se}(\hat{\theta}_1)}, \quad (6.2)$$

so that $z_{\theta,k} = 0$ means treatment k reproduces Treatment 1’s estimate exactly, and $|z_{\theta,k}|$ greater than roughly two means the re-estimated value falls outside Treatment 1’s own 95% interval around its

¹Treatment 3’s calibrated scale factors s_τ and Treatment 4’s additional σ^{Het} shock standard deviations have no Treatment-1 counterpart and are not part of this comparison; they are discussed on their own terms in Section 4.3/4.4 and Appendix B.1.

mode.

Figure 6.2 plots the 15 parameters with the largest $|z_{\theta,k}|$ across any treatment. Three shock standard deviations dominate the ranking by a wide margin, and move in the same direction under all three alternative treatments: the preferences shock ($z = -6.95, -5.71, -6.57$ under Treatments 2–4), foreign demand ($z = -6.85, -5.46, -6.78$), and wage efficiency ($z = -4.90, -3.55, -5.51$), each falling four to seven Treatment-1 standard errors below Treatment 1’s own estimate, with the corresponding persistences (wage efficiency, preferences) moving a further two to three standard errors. This is exactly the signature Section 4 anticipated: preferences, foreign demand, and wages are three of the most flexible channels available to a broad-based, economy-wide collapse and rebound, and under Treatment 1 they are forced to inflate their baseline variance well beyond what tranquil-period data alone would support, in order to rationalize 2020Q2–2021Q4 using nothing but the ordinary constant-variance shock distribution. Every one of Treatments 2, 3, and 4, whatever their other differences, corrects this in the same direction and by a comparable magnitude; contaminated variance estimates for the shocks most exposed to Covid-19 are not a Treatment-1 quirk that any one remedy happens to fix, they are a shared, load-bearing distortion that any credible volatility treatment must remove.

A second, more nuanced pattern separates Treatment 3 from Treatments 2 and 4. The capital adjustment cost (κ), the debt-elastic risk-premium elasticity, and the export price elasticity shift by a comparable one-and-a-half to two Treatment-1 standard errors under *all three* treatments, a treatment-invariant correction. But the standard deviations of the regulated-inflation, remittances, and food-inflation shocks, three of the shocks Table 5.1 assigns to the 2014–16 oil-price episode, shift substantially only under Treatment 3 ($z = -1.89, -1.69, -1.51$) while staying within half a standard error of Treatment 1 under Treatments 2 and 4. This is the estimation-side counterpart of the mechanism distinction already drawn in Section 6.4.2: because Treatment 3 rescales a shock’s *own* AR(1) innovation rather than routing crisis-period variance through a separate channel, inflating that innovation’s variance during τ_{Oil} measurably drags along the same shock’s estimated *baseline* standard deviation for the whole sample; Treatment 4’s dedicated ϵ^{Het} channel, by construction, leaves the baseline estimate for these shocks close to untouched. The policy-rate-smoothing coefficient ϕ_{inm} shows a third, distinct footprint: it moves under Treatment 2 ($z = -1.59$) but barely at all under Treatments 3 or 4 ($z = -0.35, -0.71$), consistent with it being sensitive mainly to whether real-activity data is withheld from the filter rather than to how shock variances are treated.

The full parameter-by-parameter comparison, split by persistences, other structural parameters, discrepancy constants, and shock standard deviations, together with posterior density overlays (Laplace approximation) for the nine parameters most directly tied to the paper’s mechanism, is reported in Appendix C.3 for reference, along with a relative-deviation robustness check on the same 57 parameters.

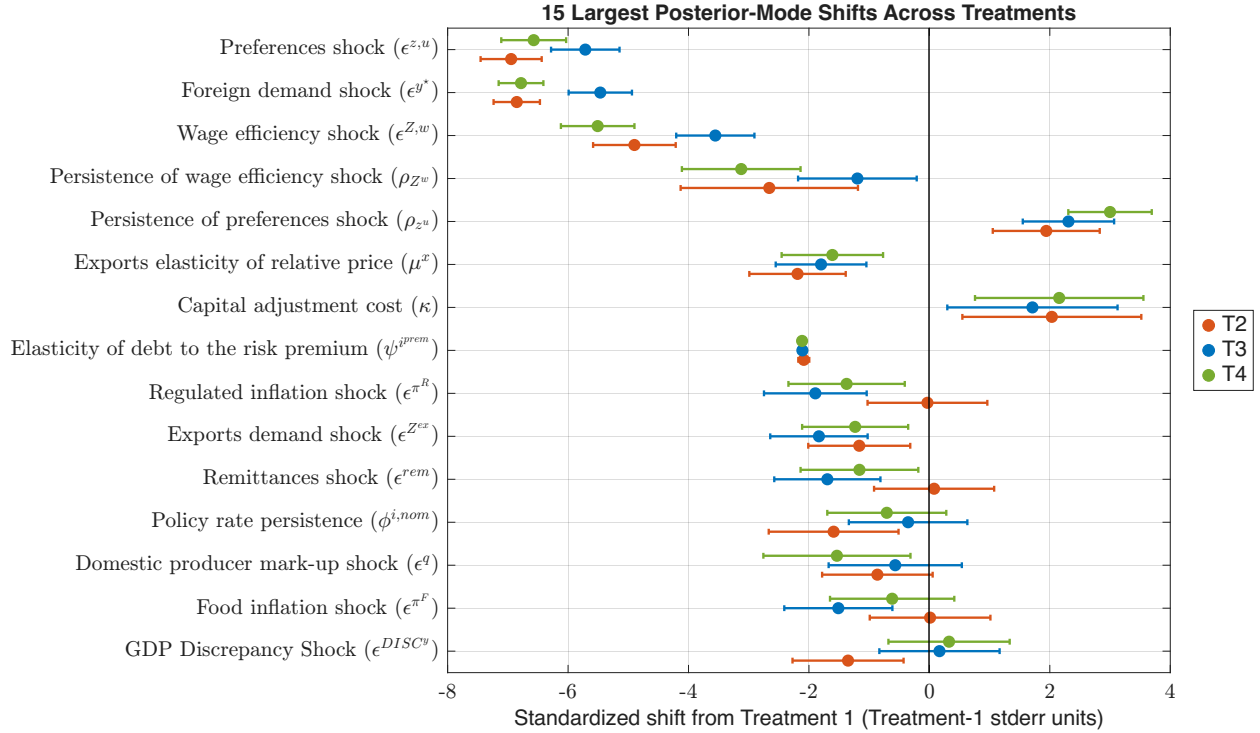


Figure 6.2: The 15 parameters with the largest standardized posterior-mode shift, $z_{\theta,k} = (\hat{\theta}_k - \hat{\theta}_1)/se(\hat{\theta}_1)$, under any of Treatments 2–4 relative to Treatment 1. Dots are posterior modes; whiskers are ± 1 standard error, both in Treatment-1-standard-error units. A whisker not crossing zero means that treatment’s estimate lies outside Treatment 1’s own approximate 68% interval around its mode.

6.3 Impulse Response Functions

A natural concern raised by Section 6.2’s parameter shifts is whether they also change the model’s qualitative transmission mechanism, not just the size of a few shocks. We compute impulse responses under each treatment’s own posterior mode (`stoch_simul(order=1)`) for five shocks spanning both Covid-treated and untreated processes, preferences (ζ^u), domestic producer efficiency (Z^q), the domestic mark-up shock (ϵ^q), the monetary policy shock, and the risk premium ($\zeta^{i,prem}$), and rescale every treatment’s response to the size of Treatment 1’s own estimated standard deviation before comparing, so that differences in the plotted lines reflect propagation dynamics rather than the shock-size differences Section 6.2 already documents. Rather than checking this against a handful of headline aggregates, we trace each shock’s response through fifteen variables spanning CUMBIA’s real, nominal, and financial blocks: the GDP components (output, consumption, investment, exports, imports), six disaggregated inflation measures, real wages, the real exchange rate, and both policy rates, so that a treatment-specific distortion confined to one corner of the model, a particular GDP component or inflation sub-index, cannot hide behind an aggregate that happens to look robust. The finding is reassuring rather than dramatic: for every one of the five shocks, all four treatments produce impulse responses of the same sign, shape, and approximate timing across this entire variable set, movements consistent with standard New Keynesian intuition throughout, peaking and decaying over a broadly similar number of quarters regardless of treatment. Differences are confined to amplitude, and track exactly the shocks Section 6.2 flags as treatment-sensitive: the monetary policy shock’s response is nearly indistinguishable across all

four treatments (it is not rescaled under any treatment during any episode), while the preferences and risk-premium shocks, both actively rescaled during Covid-19, show visibly different amplitudes and, for the risk premium, different decay speeds, with Treatment 1 the outlier in both cases, a pattern that extends past GDP itself to investment and real wages for the risk-premium shock specifically. The single largest amplitude discrepancy we find across all five shocks and fifteen variables, in fact, is not a GDP-related channel at all: the preferences shock’s effect on real depreciation is roughly 2.4 times larger on impact under Treatment 4 (−0.32 pp) than under Treatment 1 (−0.14 pp), with Treatments 2 and 3 in between, all four sharing the same sign and shape, an impact depreciation followed by a reversal that fully unwinds within ten to fifteen quarters. [Figure C.8](#) in [Appendix C.2](#) isolates this comparison directly. We also compare, for Treatment 4 only, each rescaled shock’s ordinary AR(1) impulse response against its own heteroskedastic counterpart’s response where one exists (preferences and domestic efficiency have one; the mark-up, monetary-policy, and risk-premium shocks in this set do not, since none is assigned a heteroskedastic channel in any episode per [Table 5.1](#)): the heteroskedastic shock’s impulse is larger on impact but collapses within one to two quarters, visibly confirming the self-cancelling construction of [Section 4.4](#) as an actual impulse response rather than only an algebraic claim. The full set of figures is reported in [Appendix C.2](#).

6.4 Shock Decomposition

We group CUMBIA’s structural and heteroskedastic shocks into the eleven categories of [Table 4.1](#)’s underlying taxonomy (demand, supply, monetary policy, foreign, financial conditions, and measurement error, each split into a structural and a heteroskedastic component) and compute, for every treatment and every quarter, each group’s contribution to the smoothed path of a given observable. This section asks two questions of that decomposition. First, restricting attention to the annual (year-on-year) transformations of GDP growth, headline inflation, and the policy rate that a policymaker actually reads off a monitor, how much does the choice of volatility treatment matter relative to not treating volatility at all? Second, and more specifically, what does each treatment implicitly assume about whether Covid-19 left a persistent mark on real activity, and why does that assumption matter most for Treatment 4, the specification we adopt going forward? The figures below zoom to the 2018–2025 window most relevant to both questions and compare three treatments at a time for legibility; [Appendix C.4](#) reports the same decomposition for all four treatments together over the full 2005–2025 sample, including all three high-volatility episodes, for reference.

6.4.1 How much do the treatments matter, relative to not treating them at all?

[Table 6.1](#) reports, for annual GDP growth ($\Delta_4 y_t$), annual headline inflation (π_t^4), the annualized policy rate ($i_t^{nom,A}$), and detrended GDP in efficiency units (y_t , CUMBIA’s stationary measure of the level of activity relative to trend), the mean absolute difference between each treatment’s smoothed path and Treatment 1’s, separately over the Covid-19 window (τ_{Covid}) and the four years following it (2022Q1–2025Q4).

Two patterns stand out. First, relevance is concentrated almost entirely in the real block. Annual headline inflation and the annualized policy rate are, to two decimal places, identical across all four treatments in every period: both series are directly, tightly observed, so whatever a treatment does to the variance or persistence of the shocks that feed them, the Kalman smoother is still anchored

Variable	Treatment 2		Treatment 3		Treatment 4	
	Covid	Post-Covid	Covid	Post-Covid	Covid	Post-Covid
GDP growth, YoY ($\Delta_4 y_t$)	10.13 pp	1.36 pp	<0.01 pp	<0.01 pp	<0.01 pp	<0.01 pp
GDP, efficiency units (y_t)	5.46 pp	3.04 pp	0.75 pp	0.82 pp	0.64 pp	0.43 pp
Headline inflation, YoY (π_t^4)	<0.01 pp	<0.01 pp	<0.01 pp	<0.01 pp	<0.01 pp	<0.01 pp
Policy rate, annualized ($i_t^{nom,A}$)	<0.01 pp	<0.01 pp	<0.01 pp	<0.01 pp	<0.01 pp	<0.01 pp

Table 6.1: Mean absolute difference from Treatment 1’s smoothed path, by treatment, variable, and period. “Covid” = 2020Q1–2021Q4; “Post-Covid” = 2022Q1–2025Q4.

to the same data point and returns essentially the same smoothed value. Colombia’s 2021–23 inflation surge, in CUMBIA’s reading, is a story told by the exchange-rate and foreign-price shocks common to all four treatments, not by the Covid-specific shocks any of them rescales; the choice of volatility treatment is, in this sense, irrelevant for the inflation and policy-rate narrative.

Second, relevance for the real block splits into two very different kinds of “mattering.” For annual GDP growth itself, only Treatment 2 moves the smoothed path at all, a mean gap of 10.13 pp during τ_{Covid} and a lingering 1.36 pp afterward, because it is the only treatment that ever withholds the GDP data from the filter; Treatments 3 and 4 leave the smoothed value of $\Delta_4 y_t$ statistically indistinguishable from Treatment 1’s, since neither ever asks the filter to disbelieve an observed data point, it only changes which shock is credited with explaining it. But for the latent, unobserved measure of activity, y_t , all three treatments diverge from Treatment 1, and by economically meaningful amounts even outside τ_{Covid} (0.75–0.82 pp for Treatment 3, 0.64–0.43 pp for Treatment 4): re-estimating the model under a different volatility treatment shifts the full posterior mode, and with it CUMBIA’s entire read of the underlying state, not just its account of the pandemic quarters themselves. Treatment 1’s implicit assumption, that a Covid-quarter shock realization is drawn from the same distribution as any other quarter’s, is therefore not a free simplification: it measurably contaminates the model’s structural narrative for the whole sample, exactly as Section 4 anticipated, even though it leaves the fitted values of the observed series themselves nearly untouched.

Figure 6.3 and Figure 6.4 make the “same value, different story” half of this finding visible directly, plotting Treatment 2’s, Treatment 3’s, and Treatment 4’s own shock decompositions of annual headline inflation and the annualized policy rate side by side, zoomed to 2018–2025 so the four years following τ_{Covid} are actually legible; the 2008–09 and 2014–16 episodes are not this comparison’s story and are left out of frame. The black smoothed-value line is, as Table 6.1 already implied, essentially identical across all three panels in both figures; what differs is the coloring underneath it, and the three treatments tell three distinct stories rather than a single reallocation. Treatment 4’s heteroskedastic shock groups switch on during τ_{Covid} , a mean absolute contribution of 0.97 pp to inflation and 0.92 pp to the policy rate; Treatments 2 and 3 have no such channel, by construction, so those categories stay at zero in both their panels. But “no heteroskedastic channel” means two very different things for Treatments 2 and 3. Treatment 3’s structural groups swell past *both* Treatment 1’s and Treatment 4’s: for inflation, the demand-structural contribution during τ_{Covid} rises from 1.02 pp under Treatment 1 to 1.48 pp under Treatment 3 (falling to 0.35 pp under Treatment 4), and the supply-structural contribution goes from 1.15 pp to 1.64 pp (versus 1.00 pp); the policy rate shows the same ordering (demand-structural 1.21 → 1.88 pp versus 0.49 pp; supply-structural 0.61 → 1.29 pp versus 0.49 pp). Treatment 2’s structural groups move the *opposite* way, shrinking well below Treatment 1’s own baseline (inflation demand-structural: 1.02 pp under Treatment 1 versus 0.42 pp under Treatment 2; policy-rate demand-structural: 1.21 pp versus 0.58 pp): with

the real-activity data that would reveal the collapse withheld from the likelihood, the filter has less reason to invoke an unusual shock realization at all, so its account of τ_{Covid} looks, if anything, *calmer* than an untreated estimation's. Treatment 3 amplifies the channel that explains the crisis; Treatment 4 relies on that channel less and adds a dedicated one instead; Treatment 2 does not get to see there was a crisis to explain in the first place. Re-estimating under any of the three leaves what CUMBIA believes inflation or the policy rate *did* during τ_{Covid} essentially unchanged (the black line barely moves), only which shocks it credits for it (or fails to credit at all, for Treatment 2) – a reallocation invisible to any metric that only looks at fitted values, which is precisely why the shock decomposition, and not just the smoothed-path comparison, is necessary to see the treatments' full effect.

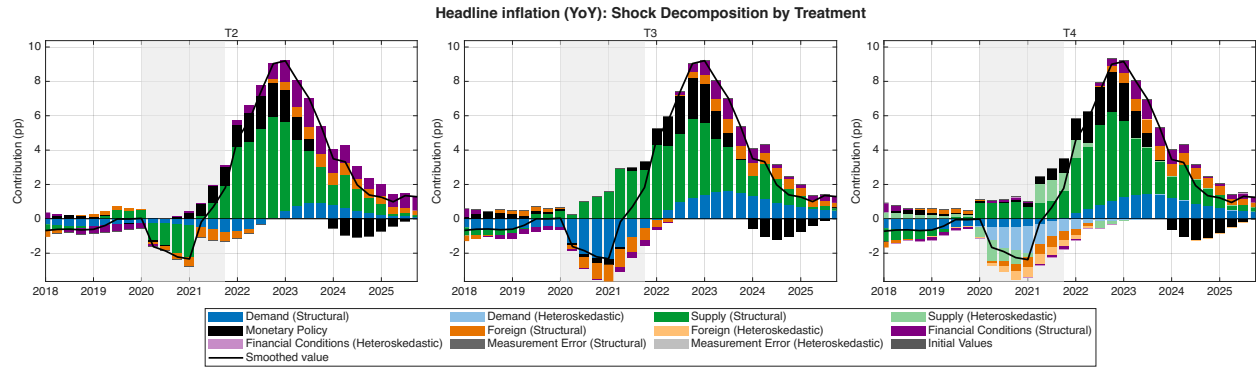


Figure 6.3: Shock decomposition of annual headline inflation (π_t^4), 2018–2025, under Treatments 2, 3, and 4, individually. The smoothed-value line is nearly unchanged across treatments; the composition of shock groups underneath it is not. Treatment 2's structural groups shrink below Treatment 1's own baseline (not shown) during τ_{Covid} , since the filter never observes the collapse; Treatment 3's structural groups grow past it instead; Treatment 4 routes the equivalent variance into a dedicated heteroskedastic channel (lighter shade).

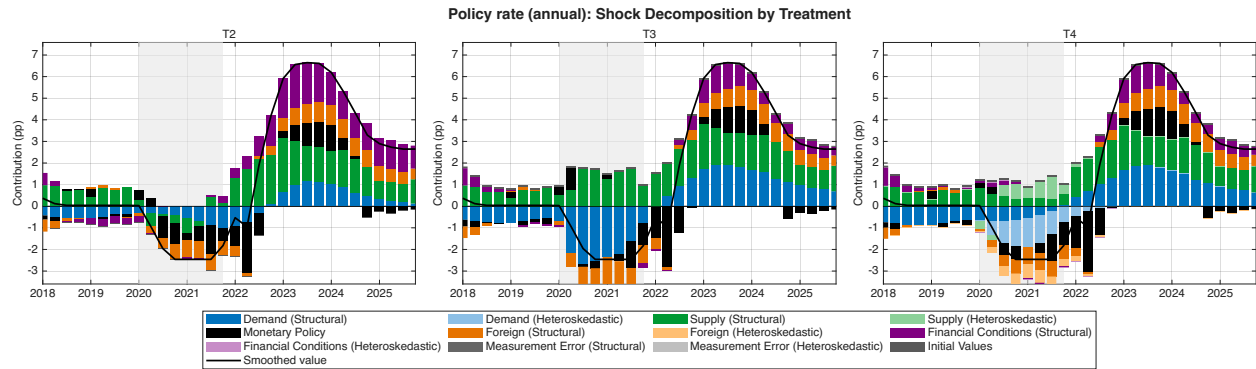


Figure 6.4: Shock decomposition of the annualized policy rate ($i_t^{nom,A}$), 2018–2025, under Treatments 2, 3, and 4, individually. The three smoothed-value lines are nearly indistinguishable; the three treatments nonetheless disagree sharply on which shocks, if any, explain τ_{Covid} .

6.4.2 The persistence assumption: Treatments 2, 3, and 4

Treatments 2 and 4 are, on the surface, unrelated: one drops data, the other adds a variance term to the shock process. Underneath, they share an assumption that Treatment 3 does not. Treatment 2 never lets a single Covid-quarter observation of real activity enter the likelihood, so the AR(1) persistence parameters governing consumption, investment, trade, and GDP growth are identified entirely from non-Covid quarters; nothing in the estimation could tell CUMBIA that the pandemic ought to leave a longer or shorter echo than an ordinary shock would, because the pandemic itself is invisible to that part of the model. Treatment 4 arrives at the same place by a different route: the $-\rho \epsilon_{(t-1) \in \tau}^{Het}$ term in (4.2) is not estimated to cancel, it is *constructed* to cancel, so that an arbitrarily large heteroskedastic shock realization during τ is, by the algebra of the ARMA(1,1) process itself, fully unwound one period later. Both treatments therefore assume, one through an information gap and one through a modeling restriction, that Covid-19 left no lasting mark on the estimated law of motion for real activity. Treatment 3 is the only one of the four that permits the opposite: because its heteroskedastic term rescales the process's own persistent innovation, an extreme realization during τ propagates forward at the shock's ordinary persistence ρ , anywhere from a single quarter (export efficiency, $\rho \approx 0$) to nearly two years (the Fed funds rate, $\rho \approx 0.92$; the demand shock, $\rho \approx 0.88$) before it fully decays.

This is not a minor technical footnote given the paper's overall recommendation. Treatment 4 is preferred for good reasons, chiefly the estimation pathologies documented in Appendix B.1 that a naive design of Treatment 3 ran into, and the cleaner narrative separation it offers between pandemic and ordinary-time dynamics (Section 7). But adopting it also means adopting Treatment 2's no-persistence assumption for real activity, dressed up in more sophisticated machinery. A researcher whose priors favor genuine Covid-era scarring, hysteresis in investment or labor markets, for instance, will not find that hypothesis represented in either Treatment 2's or Treatment 4's shock decomposition; only Treatment 3 can surface it, at the cost of the fragility documented in Appendix B.1.

Figure 6.5 and Figure 6.6 make this concrete by plotting all three treatments' own shock decompositions, not their difference, for annual GDP growth and for GDP in efficiency units, again zoomed to 2018–2025 so the post- τ_{Covid} years are readable. During τ_{Covid} itself the three panels already diverge exactly as Section 6.4.1's inflation and policy-rate comparison would predict, and now with a third data point. For annual GDP growth, Treatment 4's heteroskedastic shock groups contribute a mean absolute 7.16 pp over the eight-quarter window (peaking at –17.2 pp in 2020Q2), while its own structural (demand + supply) groups contribute only 2.16 pp combined; Treatment 3, with no heteroskedastic channel to lean on, routes the equivalent crisis variance entirely through those same structural groups, whose combined contribution, 7.49 pp, is larger than Treatment 1's own baseline, 6.73 pp; Treatment 2's structural groups, by contrast, contribute only 1.15 pp combined, well *below* Treatment 1, because the filter is never shown the collapse it would otherwise need an unusual shock to explain. The GDP-level panel (y_t) shows the same three-way ordering (structural contribution – Treatment 1: 4.32 pp, Treatment 2: 2.68 pp, Treatment 3: 5.23 pp, Treatment 4: 1.26 pp – with Treatment 4 adding a further 5.14 pp through its heteroskedastic groups). Three treatments, three answers to “how much did the pandemic disturb real activity”: Treatment 2 answers “barely any, because I never saw it”; Treatment 3 answers “more than an untreated estimation would say, and through the same channel that always explains real-activity swings”; Treatment 4 answers “about as much as Treatment 1, plus a large but separately-tracked, temporary add-on.” Where the three treatments separate further is what happens *after* τ_{Covid} : Treatment 4's heteroskedastic contribution to annual GDP growth collapses to under 1.5 pp by 2022Q3 and under a tenth of a

percentage point by 2023, the “already gone” pattern its self-cancelling construction guarantees, and Treatment 2’s gap from Treatment 1 (Table 6.1) closes on contact with the first post-pandemic data point in 2022Q1, since the filter has no mechanism to carry forward anything Covid-specific once real data resumes. Treatment 3 offers no equivalently isolable “pandemic bucket” to track in the same figure, its crisis-period variance is absorbed into structural groups active in every ordinary quarter too, so the case that its crisis-window shocks decay more slowly rests on the estimated persistences themselves (up to $\rho \approx 0.92$ for the policy-rate shock, cited above) rather than on a directly readable post-sample decomposition residual. Both Treatment 2 and Treatment 4, in short, converge back to business as usual almost immediately after τ_{Covid} ends, each for its own reason; Treatment 3 is the only one of the three whose crisis-window shocks are still visibly capable of decaying, rather than already gone, a year or more later, at the cost of the opacity just described and the fragility documented in Appendix B.1.

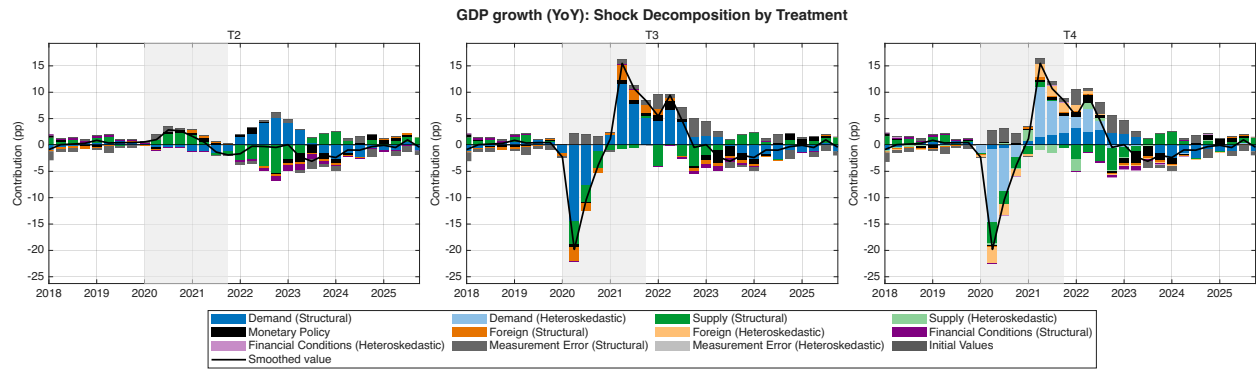


Figure 6.5: Shock decomposition of annual GDP growth ($\Delta_4 y_t$), 2018–2025, under Treatments 2, 3, and 4, individually (not differenced). Shaded region marks τ_{Covid} . Treatment 2’s structural groups shrink below Treatment 1’s own baseline (not shown) since the filter never observes the collapse; Treatment 3’s structural groups grow past it; Treatment 4 keeps its structural groups close to Treatment 1’s baseline and adds a dedicated, self-cancelling heteroskedastic group (lighter shade) instead.

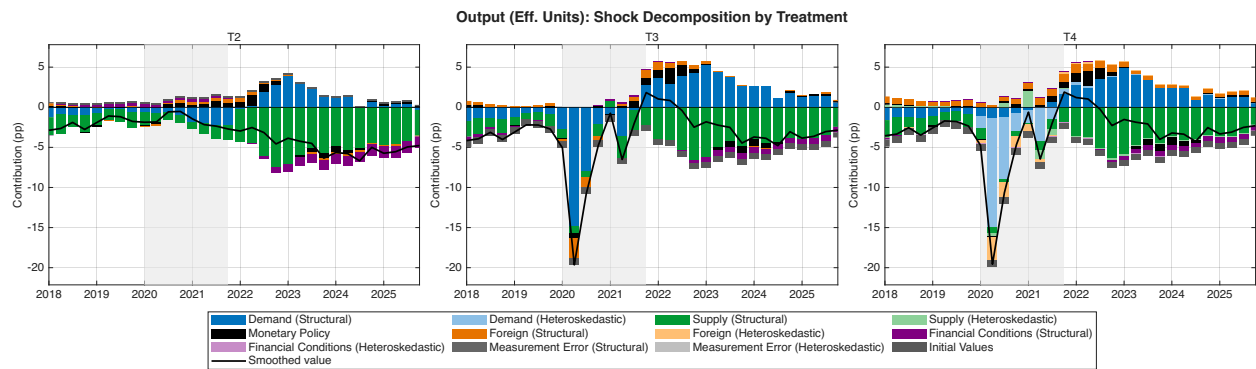


Figure 6.6: Shock decomposition of detrended GDP in efficiency units (y_t), 2018–2025, under Treatments 2, 3, and 4, individually. Treatment 4’s heteroskedastic contribution is large during τ_{Covid} but transitory by construction; Treatment 3 shows the same crisis-period variance absorbed into its structural shock groups, with no separate channel to isolate and track its subsequent decay; Treatment 2 shows the muted, uninformed account that results from never observing the collapse at all.

6.5 Forecast Evaluation

The comparisons so far, smoothed paths (Section 6.1), posterior modes (Section 6.2), and shock decompositions (Section 6.4), all describe what each treatment believes happened. This section asks whether that belief is useful going forward: does correcting for Covid-19 also produce better real-time forecasts once the pandemic is over, or is it a purely narrative improvement that leaves predictive accuracy unchanged? We construct a genuine, fixed-parameter pseudo out-of-sample exercise. For each treatment k and each origin o in $\mathcal{O} = \{2014Q4, \dots, 2023Q4\}$ (37 origins), we hold k 's posterior mode fixed at its Section 6.2 estimate (mode_compute=0, no re-estimation) and run the Kalman filter/forecast recursion forward from o using only data available through o , producing a forecast $\hat{y}_{v,o+h}^k$ for each of the three variables a policymaker actually monitors, annual GDP growth, headline inflation, and the policy rate, at every horizon $h = 1, \dots, 8$ quarters.² "Ground truth" for computing forecast errors is Treatment 1's full-sample Kalman-smoothed path for each variable, held fixed across every treatment and benchmark being evaluated, since a genuinely model-free source of quarterly truth does not exist for constructed variables like the annualized policy rate; this makes the comparison internally consistent (every method is scored against the same yardstick) at the cost of that yardstick itself being one treatment's own construction, a caveat we flag rather than paper over. Reusing the RMSE/MAE construction of (6.1), but now averaging over origins rather than quarters within a single episode,

$$\text{RMSE}_{v,k}(h; \mathcal{W}) = \sqrt{\frac{1}{|\mathcal{O}_{\mathcal{W}}|} \sum_{o: o+h \in \mathcal{W}} \left(y_{v,o+h} - \hat{y}_{v,o+h}^k \right)^2}, \quad (6.3)$$

where $\mathcal{W} \in \{\text{Pre-Covid}, \text{Covid}, \text{Post-Covid}\}$ partitions the *target* quarter $o + h$ exactly as in Table 6.1 (Covid = $\tau_{\text{Covid}} = 2020Q1\text{--}2021Q4$; Post-Covid = 2022Q1–2025Q4), and Pre-Covid covers the remaining, earliest targets the origin range can reach (2015Q1–2019Q4). Alongside the four treatments, we compute two atheoretical benchmarks at every origin using only data available at that origin: a no-change (random-walk) forecast, and a univariate ARMA(1,1) fit by exact Kalman-filter maximum likelihood.

Variable	Pre-Covid			Covid			Post-Covid		
	T1	T3	T4	T1	T3	T4	T1	T3	T4
GDP growth, YoY ($\Delta_4 y_t$)	0.81 pp	0.94 pp	0.96 pp	9.28 pp	9.28 pp	9.27 pp	3.14 pp	2.73 pp	2.94 pp
Headline inflation, YoY (π_t^4)	1.13 pp	1.17 pp	1.19 pp	1.92 pp	2.07 pp	2.15 pp	4.32 pp	3.82 pp	3.99 pp
Policy rate, annualized ($i_t^{\text{nom},A}$)	1.12 pp	1.01 pp	1.02 pp	1.64 pp	1.94 pp	2.01 pp	3.15 pp	2.82 pp	2.98 pp
Average Δ (T3)		−3.2%			−8.7%			+11.7%	
Average Δ (T4)		−5.1%			−11.2%			+6.4%	

Table 6.2: Root-mean-squared forecast error, Treatments 1, 3, and 4, by variable and target window, averaged across horizons $h = 1, \dots, 8$ and all origins reaching that window, per (6.3). Δ is the average percentage RMSE reduction of Treatment 3 (resp. 4) relative to Treatment 1 across the three variables (positive = more accurate than Treatment 1).

Table 6.2 is the section's central result, and it complicates the paper's preferred specification rather than simply confirming it. Neither Treatment 3 nor Treatment 4 is uniformly more accurate than

²Fixing parameters at each treatment's already-estimated mode isolates the effect of the volatility treatment's *filtering/forecasting* mechanism on accuracy, holding the Section 6.2 parameter differences constant; it is not a fully real-time exercise, since a genuine real-time forecaster would also re-estimate at every origin. We revisit this scope limitation in Section 7.

Treatment 1, both are on average 3–11% *less* accurate pre-Covid and during Covid itself, but both are more accurate, for all three variables without exception, once the pandemic is over: average RMSE reductions of 11.7% (Treatment 3) and 6.4% (Treatment 4) across GDP growth, inflation, and the policy rate over 2022Q1–2025Q4, the exact period the paper’s motivating question (Section 1) is about. This is precisely the pattern the estimation-side evidence in Section 6.2 predicts rather than a coincidence: Treatment 1 leaves its constant-variance shocks free to over-explain the pandemic’s realized fluctuations at essentially no cost during the crisis itself (hence its edge over both alternatives within τ_{Covid}), but the contaminated persistence and variance estimates that come with doing so then degrade every forecast made afterward; both alternative treatments’ heteroskedastic channels absorb the pandemic’s variance where it belongs, at some cost to in-crisis fit, and are rewarded for it exactly once the crisis is over. Pooling all origins together regardless of target window, the ranking still favors both treatments over Treatment 1, but only modestly (RMSE improvements of +1.4%/+7.8%/+6.9% for Treatment 3 and +0.5%/+4.3%/+2.3% for Treatment 4, GDP growth/inflation/policy rate respectively): the post-Covid gain is large enough to survive averaging over the pre-Covid and Covid-era quarters where both treatments underperform, but it does not show up as a dramatic unconditional improvement, only as a targeted one, precisely where the paper’s motivation says it should matter most.

The uncomfortable finding is that Treatment 3, not Treatment 4, is the more accurate forecaster on every single row of Table 6.2, pre-Covid, during Covid, and post-Covid alike, and by a wide margin on the metric the paper’s motivating question cares most about (11.7% vs. 6.4% post-Covid). This is not a contradiction of Section 6.4.2’s argument for preferring Treatment 4, it is the accuracy side of exactly the trade-off that section describes. Treatment 3’s persistent AR(1) rescaling lets an extreme pandemic-quarter realization propagate forward at each shock’s own ρ , anywhere from a single quarter to nearly two years (Section 6.4.2); if the underlying pandemic disturbance genuinely had some persistent echo, rather than being a purely transitory event with no lasting mark on real activity, a specification that allows for that echo will fit better going forward than one that (by construction, in Treatment 4’s case) rules it out. Treatment 3’s forecast advantage is therefore evidence *for* some genuine post-Covid scarring in the Colombian data, which is exactly what Treatment 4’s ARMA(1,1) cancellation term assumes away. The trade-off is that Treatment 3, as implemented here, achieves this accuracy gain riding on the estimation fragilities documented in Appendix B.1, a shared scale factor calibrated rather than estimated (for a purely computational reason, not a substantive one) and shock-channel design choices that took three attempts to get past Dynare’s own singularity checks. Section 7 weighs this accuracy-vs-robustness trade-off directly rather than resolving it here.

The same ranking holds under mean absolute error, with similar average post-Covid improvements (12.8% for Treatment 3, 7.9% for Treatment 4, across the three variables) and similar pre-Covid/Covid shortfalls; full MAE figures are reported in Appendix C.5. Both atheoretical benchmarks perform markedly worse than any CUMBIA treatment throughout the Covid and post-Covid windows, for GDP growth during τ_{Covid} , for instance, the random-walk and ARMA(1,1) RMSEs (13.8 pp and 12.3 pp) run 33–47% above any of the three treatments’ (9.3–9.4 pp each), confirming that the structural model’s advantage over Treatment 1 is a second-order refinement on top of a first-order advantage the DSGE specification already holds over simple time-series alternatives. Figure 6.7–6.9 show the full horizon profile behind Table 6.2: Treatment 3’s post-Covid advantage over Treatment 1 is positive at essentially every horizon for all three variables (the sole exception is a –1.2% blip for the policy rate at $h = 1$), and is both larger and more stable across the horizon than Treatment 4’s, whose GDP-growth advantage fades and mildly reverses by $h = 7$ –8; Treatment 3’s own advantage, by contrast, is still 5–15% at $h = 8$ for every variable, consistent with a specification

whose persistent shock echo keeps paying off further into the forecast horizon rather than being a purely short-lived correction.

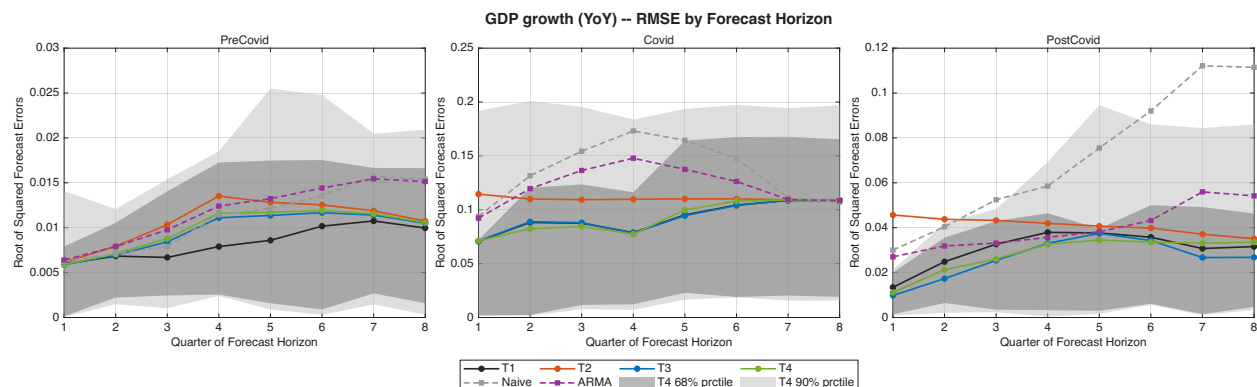


Figure 6.7: Root-mean-squared forecast error vs. horizon for annual GDP growth ($\Delta_4 y_t$), by treatment and benchmark, tiled by target window. Shaded bands are the 68%/90% percentile range of Treatment 4's absolute forecast errors.

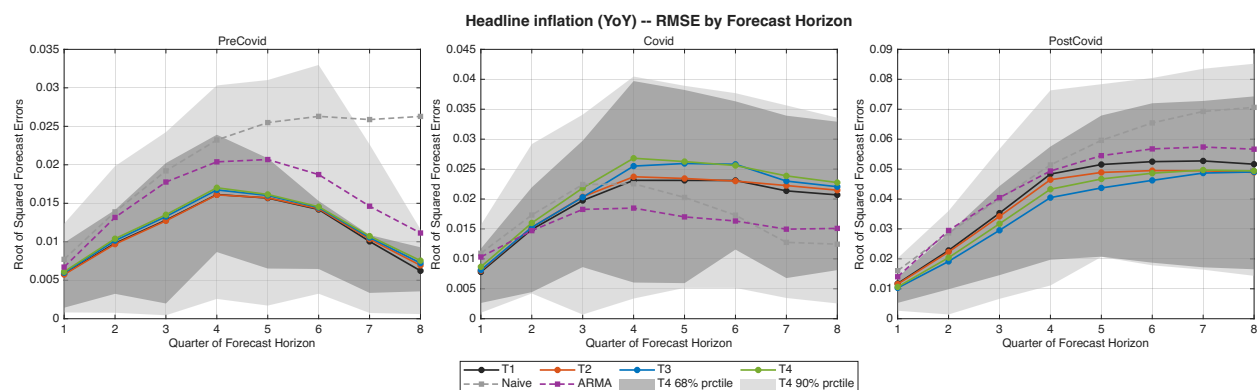


Figure 6.8: Root-mean-squared forecast error vs. horizon for annual headline inflation (π_t^4), by treatment and benchmark, tiled by target window.

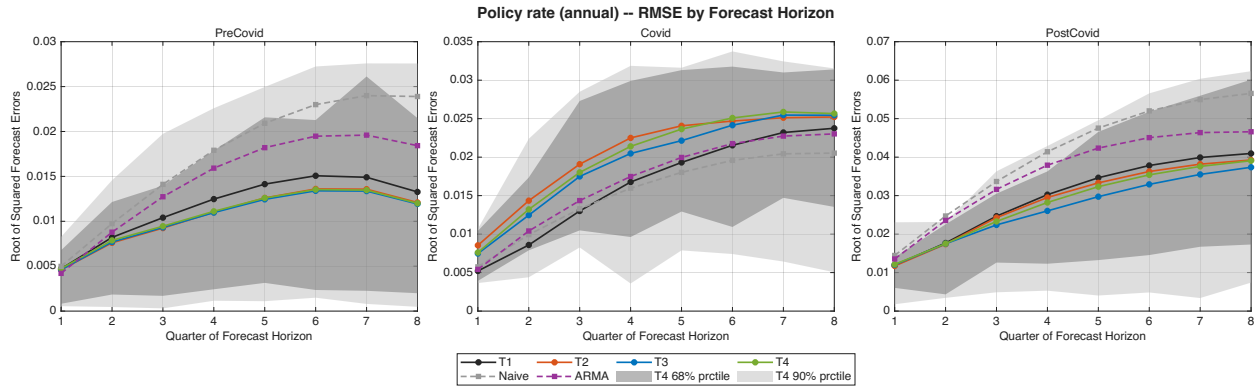


Figure 6.9: Root-mean-squared forecast error vs. horizon for the annualized policy rate ($i_t^{nom,A}$), by treatment and benchmark, tiled by target window.

7 Discussion

Section 4 introduced four treatments as different answers to the same question, how should a DSGE model treat a small number of extreme observations, without committing in advance to which answer is best. The Results section has now given each of them a track record on four dimensions that matter for a policy institution choosing a protocol for the next extreme-volatility episode: post-Covid forecast accuracy (Section 6.5), the robustness of the implementation itself, the computational cost of running it, and how cleanly it separates crisis dynamics from ordinary business-cycle dynamics in the model’s own narrative. Table 7.1 summarizes where each treatment lands.

Treatment	Post-Covid accuracy	Implementation robustness	Comp. cost	Theoretical coherence
1: No treatment	Worst on every variable (baseline)	Trivial – no extra machinery	Lowest	Contaminates parameter estimates (Section 6.2); no separation of crisis vs. ordinary dynamics
2: Missing data	Worse than T1 on GDP growth (–28% RMSE post-Covid); roughly flat on inflation/policy rate	Simple (native Kalman filter missing-data handling)	Low	Clean for the withheld variables’ own dynamics, but degrades real-activity forecasts precisely when real data resumes
3: Heteroskedastic AR(1)	Best of the four, +11.7% average RMSE reduction post-Covid	Fragile: three implementation pitfalls (Appendix B.1); scale factor calibrated, not estimated	Highest (episode-specific re-tuning, sensitivity checks)	Allows genuine persistent scarring, at the cost of an unresolved, non-identified scale parameter
4: Heteroskedastic ARMA(1,1)	Second-best, +6.4% average RMSE reduction post-Covid	Robust: one self-contained shock channel, cleared Dynare’s checks on first correct design	Moderate	Cleanest narrative separation (crisis is genuinely transitory by construction), at the cost of ruling out scarring <i>a priori</i>

Table 7.1: Treatment trade-offs revisited in light of Section 6. “Comp. cost” is relative to Treatment 1’s single, unmodified estimation.

The central tension the table exposes, Treatment 3’s forecast advantage against Treatment 4’s implementation robustness, is not just a difference in degree; the two treatments encode genuinely different economic claims. To check that Treatment 3’s advantage is not merely an artifact of the particular calibrated scale $s_\tau = (2, 2, 3)$ used throughout Section 6, we re-estimated Treatment 3 twice more, halving every episode’s scale (“conservative,” $s_\tau = (1, 1, 1.5)$) and doubling it (“aggressive,” $s_\tau = (4, 4, 6)$), holding everything else fixed. The three shocks most tied to the paper’s mechanism are reassuringly stable: investment efficiency’s persistence and standard deviation move by less than 0.001 across all three scales ($\rho^{Z_i} = 0.295\text{--}0.295$), the risk premium’s persistence stays within a 0.03 band (0.686–0.715), and even the most scale-sensitive of the three, the preferences shock’s persistence, stays within a single, moderate regime ($\rho^{\zeta_u} = 0.606\text{--}0.729$) rather than swinging toward a near-unit-root or near-zero estimate under either alternative. The scale factor itself remains formally unidentified by the likelihood for the reason documented in Appendix B.1, but comparing the three specifications’ posterior density at the mode is still informative: the baseline calibration achieves the highest posterior density of the three (log posterior kernel 6096.7, vs. 6051.7 for the aggressive and 6028.2 for the conservative scale), so although the data cannot pin down s_τ by direct estimation, the calibrated choice used throughout this paper is not simply an arbitrary point on a flat surface, it is the best-fitting of the three natural candidates we could check. Figure C.1 and Figure C.2 in Appendix C.1 show both halves of this claim directly: the posterior densities of the three mechanism shocks barely shift across a fourfold range in s_τ , and the baseline’s posterior-kernel bar is visibly the tallest of the three. This is evidence, not proof, that Treatment 3’s forecast edge reflects a genuine feature of the post-Covid data rather than a fragile artifact of one specific scale choice, though a full forecast re-evaluation of the two alternative scales, which we leave to future work, would be the definitive check.

Section 6.3’s finding that the four treatments share essentially the same impulse-response shapes lowers the stakes of this choice considerably: whichever treatment a modeling team adopts, CUMBIA’s qualitative transmission mechanism, the sign, timing, and general-equilibrium propagation of every shock we checked, is not on the table. What is at stake is a narrower, more specific question: how much of an extreme episode’s variance to attribute to a channel that persists at the shock’s ordinary ρ (Treatment 3) versus one constructed to vanish within a quarter or two (Treatment 4), and how much forecast accuracy a modeling team is willing to trade for the cleaner, more robust implementation and narrative the latter provides.

Our practical recommendation is conditional rather than a single winner. For a central bank’s production forecasting process, where a model must be re-estimated routinely, handed off between staff, and defended to a policy committee under time pressure, Treatment 4 is the safer default: it is easier to implement correctly (Appendix B.1 exists precisely because Treatment 3 is not), it does not carry a non-identified scale parameter, and its narrative, “the pandemic was a large but ultimately transitory disturbance”, is easier to communicate and defend than Treatment 3’s implicit claim of a persistent structural echo. But a team specifically concerned with post-crisis forecast accuracy, or one that suspects genuine scarring in the variable it cares most about, should not dismiss Treatment 3 on robustness grounds alone: our sensitivity check suggests its forecast advantage is not a fragile artifact of one calibration choice, and Table 6.2 shows it is the more accurate specification on every single row. A reasonable compromise, and the one we would recommend in practice, is to adopt Treatment 4 as the default specification while reporting Treatment 3 forecasts as a supplementary, scarring-robust check, particularly in the first several years after any future extreme-volatility episode, exactly the period where Table 6.2 shows the two treatments’ recommendations diverge most.

Four limitations qualify these conclusions. First, this is single-country, single-model evidence:

CUMBIA is one small-open-economy specification estimated on Colombian data, and neither the size of the post-Covid forecast gains nor the direction of the Treatment 3-vs-4 trade-off need carry over to a different model or a different economy’s pandemic experience. Second, every episode’s window (τ_{GFC} , τ_{Oil} , τ_{Covid} , Table 5.1) was dated using external, judgment-based evidence on when each crisis’s macroeconomic footprint began and ended, rather than a statistical break-detection procedure; shifting any window by a quarter or two in either direction could plausibly move the estimates in Table 6.1 and Table 6.2 by a similar order of magnitude to the differences we report between treatments, though we would not expect it to reverse the direction of any of this paper’s main comparisons. Third, Treatment 4’s ARMA(1,1) specification, one lag of self-cancellation, was adopted following Smets and Wouters (2007)’s precedent rather than selected by comparing higher-order alternatives; a two-period cancellation window, for instance, would let an extreme shock’s echo persist an additional quarter, and we have not checked whether this changes either the forecast-accuracy or the shock-decomposition results materially. Fourth, our forecast evaluation (Section 6.5) holds each treatment’s parameters fixed at its full-sample posterior mode rather than re-estimating at every origin, and scores every treatment against Treatment 1’s own smoothed path rather than a model-free ground truth; both were deliberate scope choices, isolating the filtering mechanism from parameter re-estimation and holding every method to a common yardstick, respectively, but a fully real-time exercise with an independent measure of the truth remains a natural, more demanding extension.

8 Conclusion

COVID-19 exposed a gap in standard DSGE practice: the constant-variance, mean-reverting shock processes these models rely on have no principled way to absorb a handful of quarters an order of magnitude more volatile than anything in the estimation sample, and practitioners facing the next such episode, whether a pandemic, a commodity-price collapse, or a financial crisis, need a general, reusable protocol rather than an ad-hoc patch. This paper compared four candidate protocols, no treatment, missing real-side data, heteroskedastic AR(1) shocks, and heteroskedastic ARMA(1,1) shocks, on CUMBIA, Banco de la República’s estimated small-open-economy model for Colombia, across every dimension a modeling team would need to evaluate before adopting one: impulse responses, posterior parameter estimates, historical shock decompositions, and out-of-sample forecast accuracy.

The results converge on a consistent picture. CUMBIA’s qualitative transmission mechanism is robust to the choice of treatment: every shock we checked, across fifteen real, nominal, and financial variables, produces impulse responses of the same sign, shape, and approximate timing under all four treatments, so this choice is not a referendum on the model itself. What it does change is the estimated size of the shocks most exposed to the pandemic and, more consequentially, the story CUMBIA tells about what happened during it: Treatment 1 forces ordinary shocks to rationalize an extraordinary collapse, measurably contaminating their estimated variances; Treatment 2 avoids contamination but only by refusing to look at the real-side data altogether, implicitly assuming the pandemic left no trace once observation resumes; Treatment 3 and Treatment 4 both correct the contamination while keeping the real-activity data in view, but disagree on where the resulting variance should live, inside the ordinary structural shocks or inside a dedicated, self-cancelling channel, and on how long a pandemic-quarter disturbance is allowed to echo forward. That disagreement is not merely conceptual: Treatment 3 is the more accurate post-Covid forecaster on every variable and horizon we evaluated, a robustness check across three alternative episode-scale

calibrations suggests this is not an artifact of one calibration choice, yet Treatment 4 remains easier to implement correctly and easier to communicate to a policy committee under time pressure.

Our practical recommendation follows directly: adopt Treatment 4 as the production default, and report Treatment 3 alongside it as a supplementary, scarring-robust forecast check, particularly in the years immediately following any future extreme-volatility episode. Extending this comparison to multivariate heteroskedasticity, to regime-switching alternatives, or to a second country's pandemic experience would be natural next steps for a literature that, as of this writing, still lacks a settled answer to a question every DSGE-based policy institution has now had to face at least once.

References

- Adjemian, S., Juillard, M., Karamé, F., Mutschler, W., Pfeifer, J., Ratto, M., Rion, N., & Villemot, S. (2024). *Dynare: Reference manual, version 6* (Dynare Working Papers No. 80). CEPREMAP. <https://www.dynare.org>
- An, S., & Schorfheide, F. (2007). Bayesian Analysis of DSGE Models. *Econometric Reviews*, 26(2-4), 113–172. <https://doi.org/10.1080/07474930701220071>
- Antolín-Díaz, J., Drechsel, T., & Petrella, I. (2024). Advances in nowcasting economic activity: The role of heterogeneous dynamics and fat tails. *Journal of Econometrics*, 238(2), 105634. <https://doi.org/10.1016/j.jeconom.2023.105634>
- Avila-Montealegre, O., & Mix, C. (2024). Common trade exposure and business cycle comovement. *Journal of International Economics*, 152, 103998.
- Ávila-Montealegre, Ó. I., Castañeda-Valdés, F. A., Jaulín-Méndez, O., Pulido, J., & Rincón-Galvis, J. A. (2026). *CUMBIA: Colombia updated model for business cycles and inflation analysis — a structural forecast model for colombia* [Companion technical documentation, Banco de la República, mimeo].
- Banco de la República. (2017). Box 1: Exchange-rate and inflationary disturbances in Colombia [Inflation Report box]. <https://www.banrep.gov.co/en/publications/inflation-report/box-1-exchange-rate-and-inflationary-disturbances-in-colombia>
- Canova, F. (2014). Bridging dsge models and the raw data. *Journal of Monetary Economics*, 67, 1–15. <https://doi.org/https://doi.org/10.1016/j.jmoneco.2014.06.003>
- Carriero, A., Clark, T. E., Marcellino, M., & Mertens, E. (2024). Addressing COVID-19 outliers in BVARs with stochastic volatility. *The Review of Economics and Statistics*, 106(5), 1403–1417. https://doi.org/10.1162/rest_a_01213
- Cúrdia, V., Del Negro, M., & Greenwald, D. L. (2014). Rare shocks, great recessions. *Journal of Applied Econometrics*, 29(7), 1031–1052. <https://doi.org/10.1002/jae.2395>
- Drozd, L. A., Kolbin, S., & Nosal, J. B. (2021). The trade-comovement puzzle. *American Economic Journal: Macroeconomics*, 13(2), 78–120.
- Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (2nd). Oxford University Press.
- Faria-e-Castro, M. (2024, June). *The St. Louis Fed DSGE Model* (Working Papers No. 2024-014). Federal Reserve Bank of St. Louis. <https://doi.org/10.20955/wp.2024.014>
- Fernández-Villaverde, J., Guerrón-Quintana, P., Rubio-Ramírez, J. F., & Uribe, M. (2011). Risk matters: The real effects of volatility shocks. *American Economic Review*, 101(6), 2530–2561. <https://doi.org/10.1257/aer.101.6.2530>
- Ferroni, F., Fisher, J. D. M., & Melosi, L. (2022, September). *Usual Shocks in our Usual Models* (Working Paper Series No. WP 2022-39). Federal Reserve Bank of Chicago. <https://doi.org/10.21033/wp-2022-39>
- Ferroni, F., Fisher, J. D. M., & Melosi, L. (2024). Unusual shocks in our usual models. *Journal of Monetary Economics*, 147, 103598. <https://doi.org/https://doi.org/10.1016/j.jmoneco.2024.103598>
- Garcia-Cicco, J., Pancrazi, R., & Uribe, M. (2010). Real business cycles in emerging countries? *American Economic Review*, 100(5), 2510–2531.
- González, A., Mahadeva, L., Prada, J. D., & Rodríguez, D. (2011). Policy analysis tool applied to colombian needs: Patacon model description. *Ensayos sobre Política Económica*, 29(66), 222–245.
- Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press.

- International Monetary Fund. (2010). *Colombia: 2010 article IV consultation—staff report* (IMF Country Report No. 10/105). International Monetary Fund. <https://www.imf.org/external/pubs/ft/scr/2010/cr10105.pdf>
- Justiniano, A., & Primiceri, G. E. (2008). The time-varying volatility of macroeconomic fluctuations. *American Economic Review*, 98(3), 604–641. <https://doi.org/10.1257/aer.98.3.604>
- Lenza, M., & Primiceri, G. E. (2022). How to estimate a vector autoregression after march 2020. *Journal of Applied Econometrics*, 37(4), 688–699. <https://doi.org/10.1002/jae.2895>
- Schmitt-Grohé, S., & Uribe, M. (2003). Closing small open economy models. *Journal of International Economics*, 61(1), 163–185. [https://doi.org/https://doi.org/10.1016/S0022-1996\(02\)00056-9](https://doi.org/10.1016/S0022-1996(02)00056-9)
- Smets, F., & Wouters, R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, 97(3), 586–606. <https://doi.org/10.1257/aer.97.3.586>

Appendix

A CUMBIA Model Equations

This appendix reports, without derivation, the remaining equilibrium conditions of CUMBIA referenced in Section 3. Full derivations, the analytical steady state, and the complete parameter table are available in the companion technical documentation (Ávila-Montealegre et al., 2026); all variables are expressed in efficiency units of labor unless noted otherwise.

A.1 Households and the foreign block

Household preferences over consumption (with internal habit $\bar{h}$) and hours are

$$u(\widehat{C}_t, \widehat{C}_{t-1}, h_t) = \frac{1}{1-\sigma} \left(\widehat{C}_t - \bar{h} \widehat{C}_{t-1} \right)^{1-\sigma} - \mathcal{Z}_t^{1-\sigma} \frac{\psi_h}{1+\eta} h_t^{1+\eta}, \quad (\text{A.1})$$

and the capital stock accumulates with endogenous depreciation and adjustment costs,

$$\delta_t = \bar{\delta}_0 + \delta_1(u_t^k - 1) + \frac{\delta_2}{2}(u_t^k - 1)^2, \quad (\text{A.2})$$

$$k_t = \left(1 - \delta_t - \frac{\kappa}{2} \left(\frac{i_t(1 + g_t^{nz})}{k_{t-1}} - (\delta + g^{nz}) \right)^2 \right) \frac{k_{t-1}}{1 + g_t^{nz}} + i_t. \quad (\text{A.3})$$

The economy is closed with a debt-elastic risk premium on the foreign interest rate,

$$(1 + i_t^{nom, \star}) = (1 + i_t^{nom, FED})(1 + i_t^{prem}), \quad (\text{A.4})$$

$$i_t^{prem} = \zeta_t^{iprem} + \psi^{iprem} \left(z_t \frac{b_t^{\star}}{y_t^{Ann}} - z \frac{b^{\star}}{y^{Ann}} \right), \quad (\text{A.5})$$

where $i_t^{nom, FED}$ and ζ_t^{iprem} follow (3.1), and the real exchange rate and foreign price level are $z_t = s_t P_t^{\star} / P_t$ and $P_t^{\star} = P_{t-1}^{\star} (1 + \pi_t^{\star})$, with $\pi_t^{\star}$ following (3.1).

A.2 Labor market

Absent wage adjustment costs, labor unions would set $\bar{\omega}_t = \frac{\varepsilon_{w,t}}{\varepsilon_{w,t}-1} \frac{\tilde{w}_t}{Z_t^w}$, a markup over the marginal cost of labor. With Rotemberg wage-adjustment costs and indexation to lagged core inflation, this markup condition becomes a forward-looking New Keynesian wage Phillips curve (reported in full in Appendix A of the companion documentation), and union profits are $\Pi_t^w = [\bar{\omega}_t - \tilde{w}_t] h_t - \Upsilon_t^w$.

A.3 Consumption, investment, and materials aggregation

Consumption combines domestic and imported inputs subject to trade adjustment costs (Ávila-Montealegre and Mix, 2024; Drozd et al., 2021):

$$c_t^g + \frac{\theta_D}{2} \left(\frac{c_t^D}{c_{t-1}^D} - 1 \right)^2 + \frac{\theta_{IM}}{2} \left(\frac{c_t^{IM}}{c_{t-1}^{IM}} - 1 \right)^2 = Z_t^c \left[(\gamma^c)^{\frac{1}{\omega_c}} (c_t^D)^{\frac{\omega_c-1}{\omega_c}} + (1 - \gamma^c)^{\frac{1}{\omega_c}} (c_t^{IM})^{\frac{\omega_c-1}{\omega_c}} \right]^{\frac{\omega_c}{\omega_c-1}}, \quad (\text{A.6})$$

with Z_t^c following (3.1). Investment and raw-materials aggregation follow the identical structure with their own home-bias and elasticity parameters (γ^i, ω_i) and (γ^m, ω_m) and efficiency shocks Z_t^i, Z_t^m .

A.4 Domestic goods

Heterogeneous domestic varieties are produced with capital, labor, and materials,

$$Q_t(j) = \frac{Z_t^q \psi_q}{Z^{\pi^{nfr}}} (\mathcal{K}_{t-1}^s(j))^\alpha (\mathcal{M}_t(j))^\mu (\mathcal{Z}_t \mathcal{N}_t h_t(j))^{1-\alpha-\mu}, \quad (\text{A.7})$$

with real marginal cost

$$mc_t^q = \frac{Z^{\pi^{nfr}}}{Z_t^q \psi_q} \left(\frac{r_t^k}{\alpha} \right)^\alpha \left(\frac{p_t^m}{\mu} \right)^\mu \left(\frac{\omega_t}{1-\mu-\alpha} \right)^{1-\mu-\alpha}, \quad (\text{A.8})$$

and, under Rotemberg pricing with indexation, the New Keynesian domestic Phillips curve

$$\begin{aligned} \frac{\varepsilon_{q,t} mc_t^q}{p_t^q} &= (\varepsilon_{q,t} - 1) + \phi_q \left(\frac{p_t^q (1 + \pi_t)}{p_{t-1}^q (1 + \pi_{ind,t-1})} - 1 \right) \left(\frac{p_t^q}{p_{t-1}^q} \frac{1 + \pi_t}{1 + \pi_{ind,t-1}} \right) \\ &\quad - \mathbb{E}_t SDF_{t,t+1} \frac{\zeta_{t+1}^u}{\zeta_t^u} \phi_q \left(\frac{p_{t+1}^q (1 + \pi_{t+1})}{p_t^q (1 + \pi_{ind,t})} - 1 \right) \left(\frac{p_{t+1}^q}{p_t^q} \right)^2 \frac{1 + \pi_{t+1}}{1 + \pi_{ind,t}} \frac{q_{t+1}}{q_t}, \quad (\text{A.9}) \end{aligned}$$

with profits $\Pi_t^q = (p_t^q - mc_t^q)q_t - \Upsilon_t^q$. Imported goods are produced by an analogous two-stage structure (heterogeneous, monopolistically competitive import varieties aggregated by a competitive importer), yielding an import price Phillips curve of the same form; it is omitted here and reported in full in the companion documentation.

A.5 Exports and monetary policy

Real exports adjust gradually to relative export prices and foreign demand,

$$ex_t^{real} = Z_t^{ex} \left(ex_{t-1}^{real} \right)^{\rho_{ex}} \left[\left(z_t \frac{p_t^{ex,\star}}{p_t^q} \right)^{\mu_x} (y_t^\star)^{\mu_y} \right]^{1-\rho_{ex}}, \quad ex_t = z_t p_t^{ex,\star} ex_t^{real}, \quad (\text{A.10})$$

and the central bank sets the policy rate according to

$$\frac{1 + i_t^{nom,A}}{1 + i_t^{nom,n,A}} = \left(\frac{1 + i_{t-1}^{nom,A}}{1 + i_{t-1}^{nom,n,A}} \right)^{\phi_i} \left[(\tilde{\pi}_t)^{\phi_\pi} (\tilde{y}_t)^{\phi_y} \right]^{1-\phi_i} \exp(\epsilon_t^{i^{nom}}), \quad (\text{A.11})$$

$$\tilde{\pi}_t = \frac{1 + \mathbb{E}_t \pi_{t+4}^{nfr,A}}{1 + \tilde{\pi}_t^A}, \quad \tilde{y}_t = \frac{1 + \Delta y_t^A}{1 + g_t^{nz,A}}, \quad (\text{A.12})$$

where the time-varying inflation target $\tilde{\pi}_t$ follows (3.1).

A.6 Aggregates

The trade balance, GDP resource constraint, real exchange rate variation, current account, and total profits are

$$tb_t = ex_t - \frac{z_t p_t^{IM,\star} im_t}{Z_t^{im}}, \quad q_t = c_t^D + i_t^D + m_t^D + ex_t^{real} + \frac{\Upsilon_t}{p_t^q}, \quad (\text{A.13})$$

$$y_t = c_t + p_t^i i_t + tb_t, \quad \frac{z_t}{z_{t-1}} = 1 + \Delta z_t, \quad (\text{A.14})$$

$$ca_t = -z_t \zeta_t^{B^\star} b_t^\star + z_t \frac{b_{t-1}^\star}{(1 + \pi_t^\star)(1 + g_t^{nz})}, \quad \Pi_t = \Pi_t^q + \Pi_t^{im} + \Pi_t^w + \Pi_t^{ex} + p_t^r c_t^r + p_t^f c_t^f, \quad (\text{A.15})$$

where $\Upsilon_t = \Upsilon_t^\omega + \Upsilon_t^q + \Upsilon_t^{IM}$ collects the Rotemberg adjustment costs on wages, domestic prices, and import prices.

B Dynare Implementation

This appendix documents Dynare implementation details for the volatility treatments of Section 4 that are not obvious from the equations alone and are worth recording for reproducibility. Treatment 2's missing-observation setup and Treatment 4's ARMA(1,1) `heteroskedastic_shocks` block use standard, well-documented Dynare features and are not discussed further here; Treatment 3's AR(1) rescaling ran into three implementation pitfalls along the way, documented below, that a modeling team attempting a similar specification would otherwise have to rediscover from scratch.

B.1 Treatment 3: three implementation pitfalls

Implementing (4.1) in Dynare required care on three points that are easy to get wrong and worth documenting for reproducibility. The first two are Kalman filter stochastic-singularity symptoms; the third is a silent non-identification issue that produces no error at all. None of the three surface until the estimation stage, after the model otherwise compiles and solves without complaint.

First, `heteroskedastic_shocks` must target the model's bare, persistent shock (e.g. `eps_zeta_u`) using its `scales` syntax and declare *only* the episode window itself; every other period is left undeclared so that Dynare retains the shock's normal, freely estimated baseline σ throughout. An earlier design instead pinned the standard deviation to zero for every quarter *outside* the window (the mirror image of the intended design, and the natural first attempt given how Treatment 4's transitory shock is switched off outside τ). For a shock that is its process's *sole* innovation, however, this makes the corresponding state variable literally deterministic over roughly four-fifths of the sample, since it has no other source of variance to draw on; with that many state variables reduced to zero variance for most of the sample, the rank of the Kalman filter's forecast-error covariance collapses, and Dynare's `initial_estimation_checks` refuses to proceed with "too many shocks have been calibrated with a zero variance."

Second, s_τ must rescale the shock's *existing* AR(1) innovation rather than an additional, separately-declared auxiliary shock. An earlier design instead reused CUMBIA's existing `eps_X_het` auxiliary shocks (already wired into every process for Treatment 4) and simply dropped the $-\rho \epsilon_{t-1}^{Het}$ cancellation term from (3.2) whenever Treatment 3 was active, so that the auxiliary shock would echo

forward like an ordinary AR(1) innovation instead of self-cancelling. Once the cancellation term is dropped, however, `eps_X_het` enters its process with exactly the same persistent, non-cancelling loading as the process’s own baseline innovation `eps_X_obs`: the two shocks become observationally equivalent, and the Kalman filter cannot distinguish which one generated any given quarter’s variance. This produced a different, later-stage error (“the forecast error variance in the multivariate Kalman filter became singular”), since the model solves and even simulates without issue but the likelihood evaluation itself is degenerate under this exact collinearity. The design that eventually cleared both checks – scales on the bare shock, declaring only the episode window – avoids both problems: there is only ever one shock channel per process, and its baseline variance is never zero, so nothing collapses in rank and nothing duplicates an existing channel.

Third, and least obvious, s_τ cannot itself be estimated through the standard `estimated_params` mechanism. An earlier design declared `scale_gfc`, `scale_oil`, and `scale_covid` as ordinary parameters with a weakly informative Gamma prior and referenced them inside `heteroskedastic_shocks` as `scales` (`scale_gfc`), expecting Dynare to re-evaluate the expression at each trial parameter vector during mode-finding, the same way it would for any other estimated parameter appearing in the model’s equations. Estimation completed without any error, and posterior mode estimates were reported for all three, but inspecting the Hessian at the mode revealed the problem: the rows corresponding to `scale_gfc`, `scale_oil`, and `scale_covid` were, up to numerical noise, pure prior curvature (off-diagonal entries $\sim 10^{-6}$; diagonal entries matching the analytic second derivative of the Gamma prior’s log-density at its own mode to three decimal places), and each parameter’s posterior mode coincided with its prior mode exactly. That is the signature of a likelihood that is completely flat in a parameter: Dynare’s `heteroskedastic_shocks` `scales` expression is evaluated once, against whichever value is in the calibration block at parse time, and is not part of the computational graph the optimizer differentiates through, so varying `scale_gfc` during mode-finding never touched the likelihood at all. The three “estimated” values reported were therefore artifacts of the prior alone. s_τ is calibrated in the final specification (Section 4.3) precisely because of this: reporting a posterior mode that the data had no opportunity to inform would misrepresent the estimation as more data-driven than it is. This is worth flagging for other Dynare users attempting a similar heteroskedastic-AR(1) specification, since nothing in the `heteroskedastic_shocks` documentation states that `scales/values` expressions are evaluated only once rather than per iteration.

C Additional Tables and Figures

This appendix collects the full-detail companion figures behind results the main text summarizes or shows only in part: the Treatment 3 episode-scale robustness check (Section 7), the complete impulse-response comparison (Section 6.3), the full 57-parameter Bayesian estimation comparison (Section 6.2), the shock decomposition for all four treatments at once rather than the main text’s zoomed, three-treatment view (Section 6.4), the mean-absolute-error counterpart to the main text’s RMSE-based forecast evaluation (Section 6.5), and the quantile-quantile companion to the cross-country motivating figure in the Introduction (Section 1).

C.1 Treatment 3 episode-scale sensitivity

Section 7 summarizes this robustness check; the supporting figures are reported here. Figure C.1 overlays the posterior of the three mechanism shocks (preferences, investment efficiency, risk

premium) and their standard deviations under the conservative ($s_\tau = (1, 1, 1.5)$), baseline ($s_\tau = (2, 2, 3)$), and aggressive ($s_\tau = (4, 4, 6)$) episode-scale calibrations. **Figure C.2** plots the log posterior kernel value at the mode for the same three re-estimations.

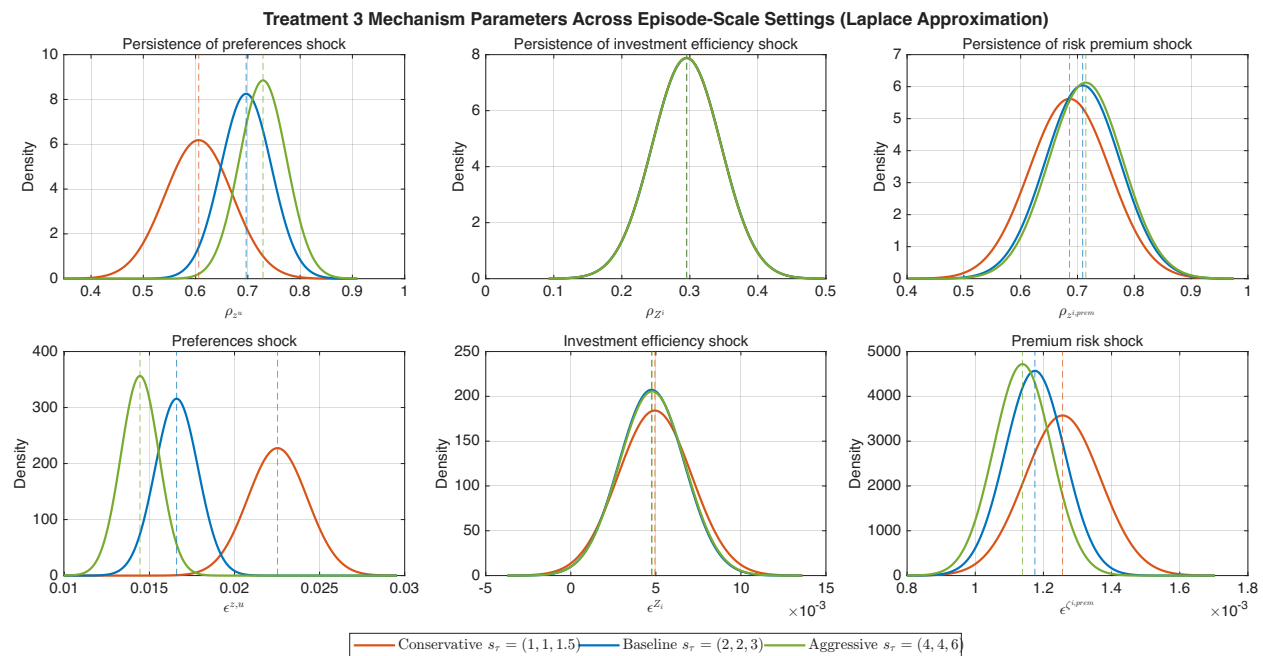


Figure C.1: Posterior mode $\pm$ stderr (Laplace approximation) of Treatment 3's three mechanism shocks and their standard deviations, under the conservative, baseline, and aggressive episode-scale calibrations. All three shocks' persistences and standard deviations stay within a narrow band across a fourfold range in s_τ .

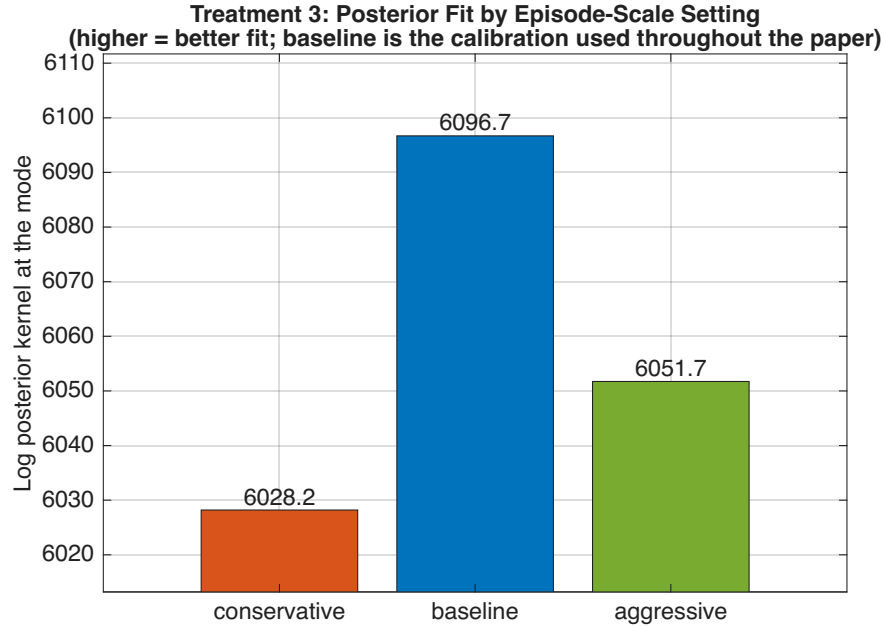


Figure C.2: Log posterior kernel value at the mode for Treatment 3 under each of the three episode-scale calibrations. The baseline calibration used throughout the paper is the best-fitting of the three, though the scale factor itself remains formally non-identified by the likelihood (Appendix B.1).

C.2 Impulse response comparison

Section 6.3 summarizes these results; the figures themselves are reported here. Figure C.3–C.7 plot each of the five shocks’ impulse response under all four treatments, rescaled to a common shock size (Treatment 1’s own estimated standard deviation) so that differences in the plotted lines reflect propagation dynamics rather than the estimated shock-size differences already documented in Section 6.2. Each figure tiles the full response of fifteen variables rather than a handful of headline series, so the robust-transmission-mechanism finding can be checked against CUMBIA’s complete real, nominal, and financial block rather than taken on faith from a summary measure: the five GDP components (output, consumption, investment, exports, and imports, all in efficiency units), the six disaggregated inflation measures CUMBIA tracks quarter-on-quarter (headline, core, food, regulated, domestic, and imported-goods inflation), real wages, the real exchange rate, and both the nominal and real policy rate. The grid is sized to stay close to square as this variable list grows rather than fixed at a particular layout.

A systematic scan of all five shocks’ responses across all fifteen variables turned up one cross-treatment amplitude discrepancy larger than any other in the set, including the risk-premium/GDP case just discussed: the preferences shock’s effect on real depreciation. Figure C.8 isolates this single shock-variable pair on its own axis, since it is easy to miss as one small tile inside Figure C.3’s fifteen-variable grid.

Figure C.9 shows, for Treatment 4 only, each shock’s ordinary AR(1) impulse response against its own heteroskedastic (ε^{Het}) counterpart’s response, for GDP in efficiency units, at each shock’s own native (unrescaled) scale; the mark-up and monetary policy shocks have no heteroskedastic counterpart in any episode and are shown with the structural response only.

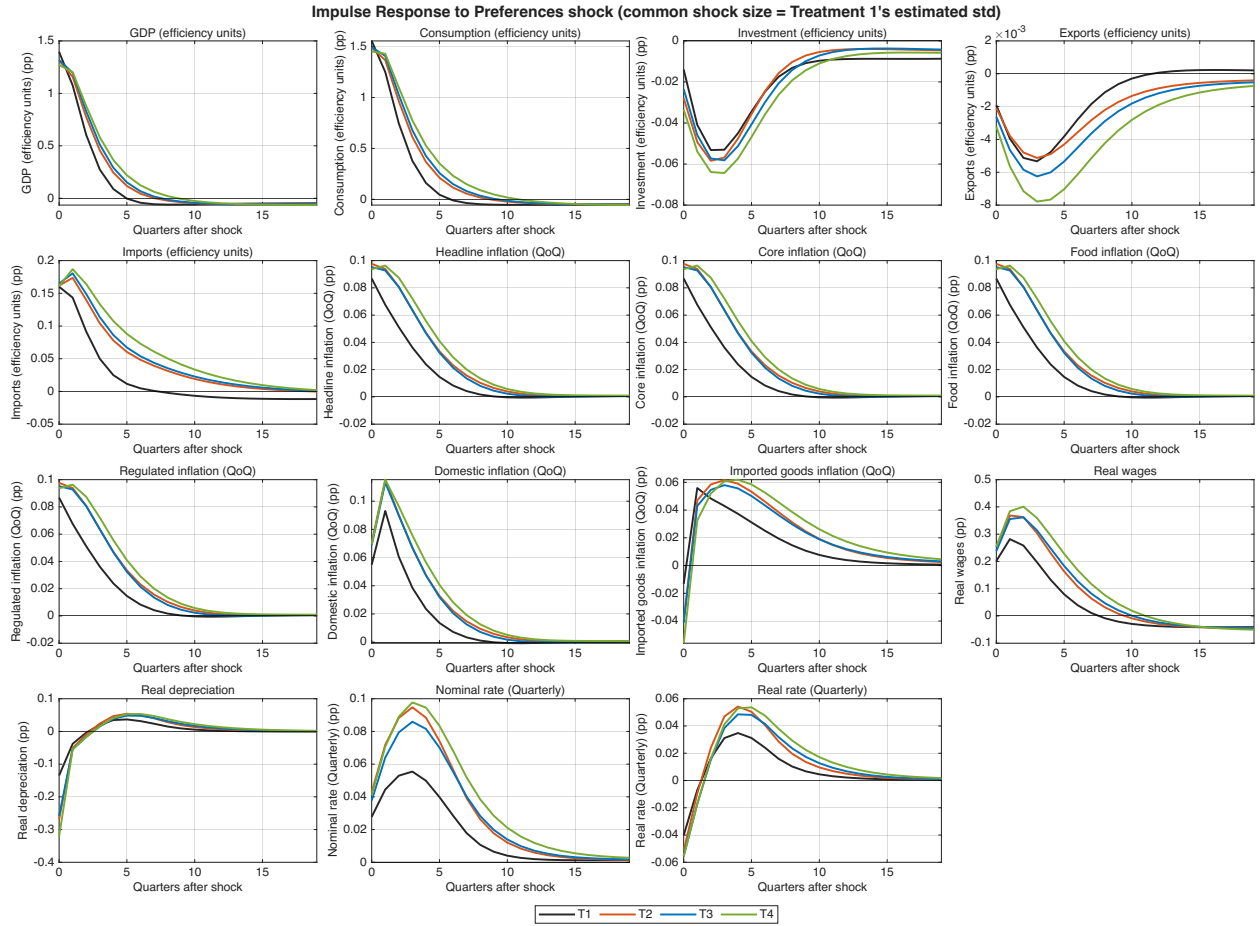


Figure C.3: Impulse response to the preferences shock (ζ^u), all four treatments, common shock size normalized to Treatment 1's estimated standard deviation.

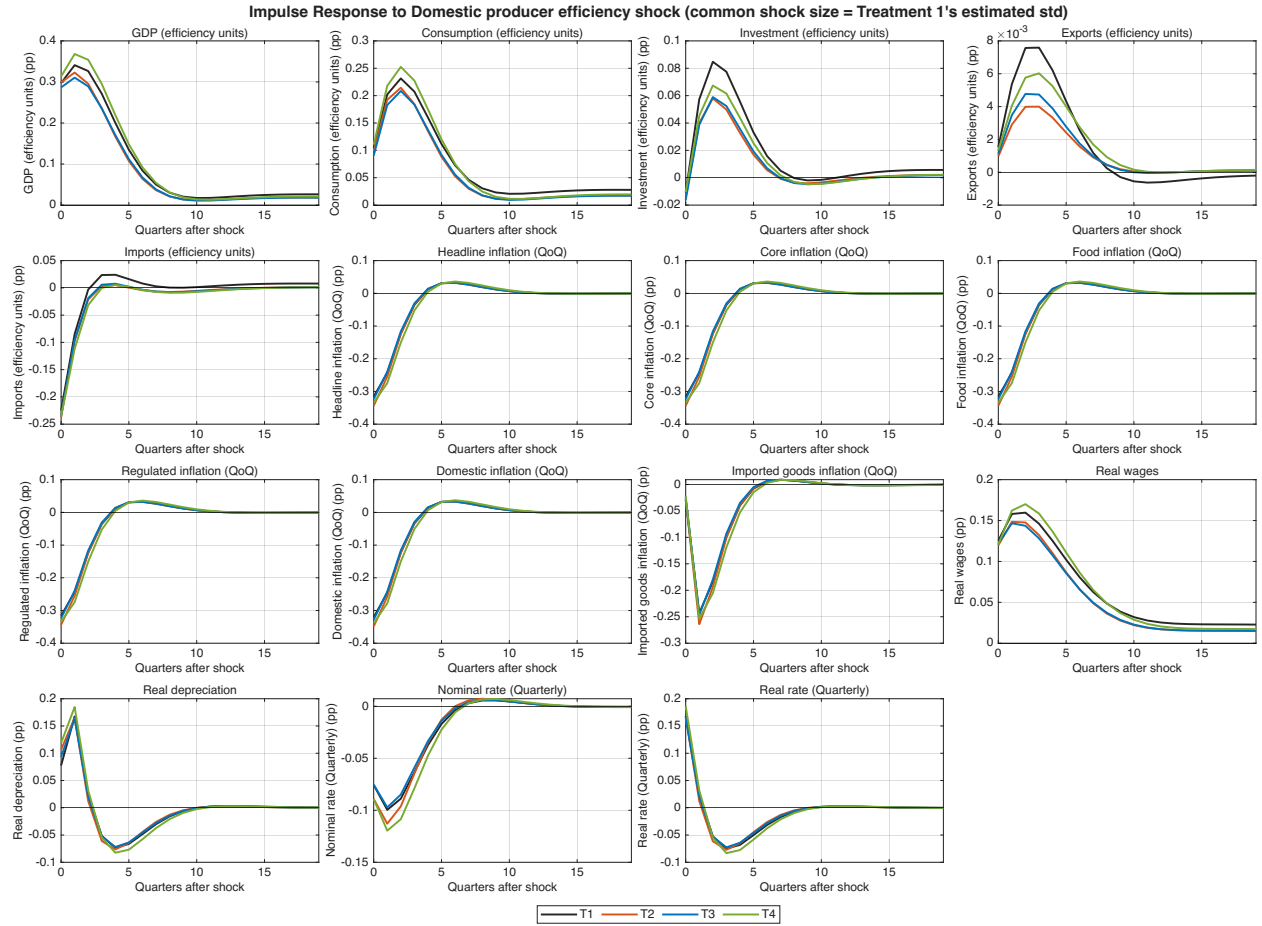


Figure C.4: Impulse response to the domestic producer efficiency shock (Z^q), all four treatments, common shock size normalized to Treatment 1's estimated standard deviation.

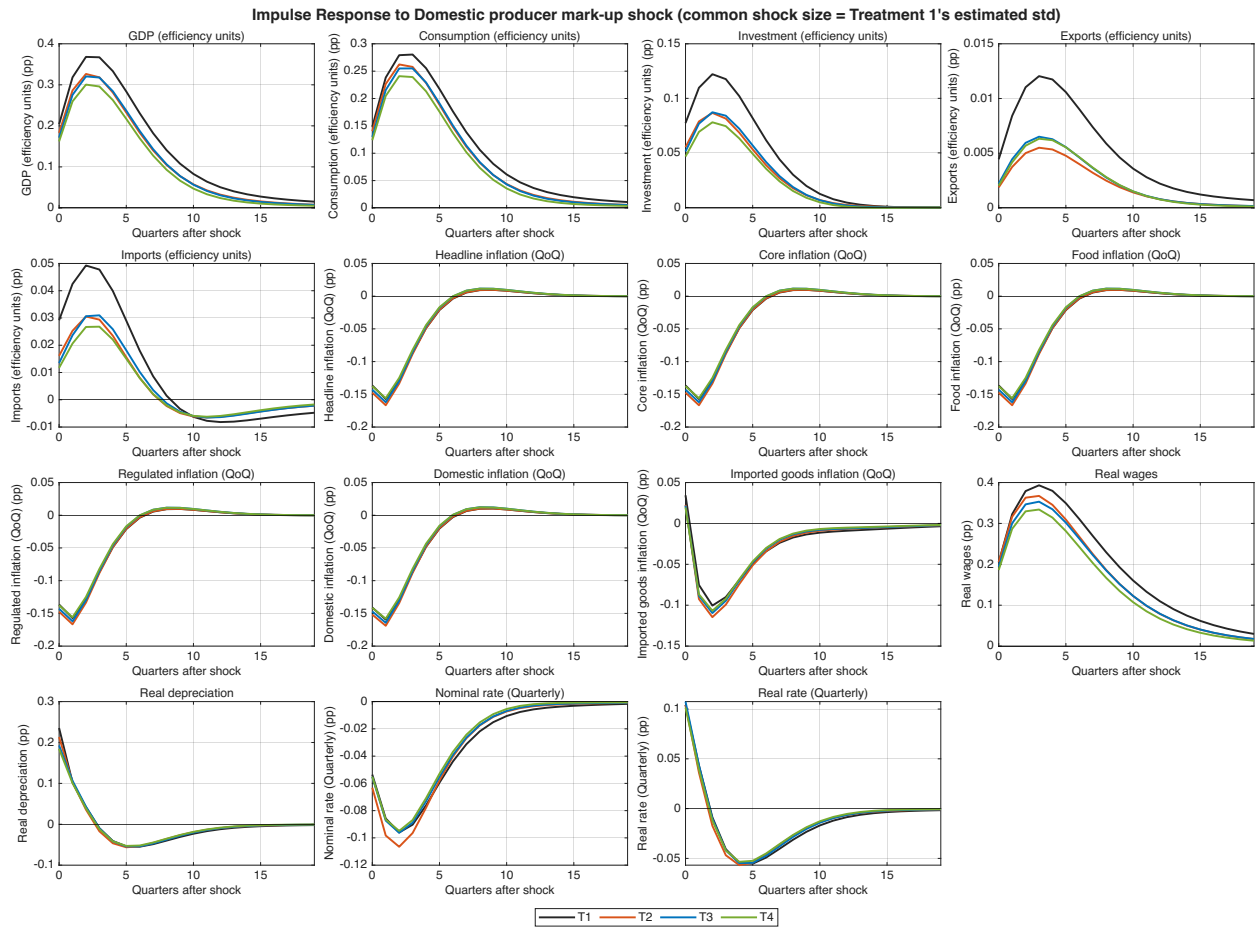


Figure C.5: Impulse response to the domestic producer mark-up shock (ε^q), all four treatments, common shock size normalized to Treatment 1's estimated standard deviation. This shock is not rescaled under any treatment in any episode (Table 5.1), so it serves as a placebo of sorts alongside the monetary policy shock.

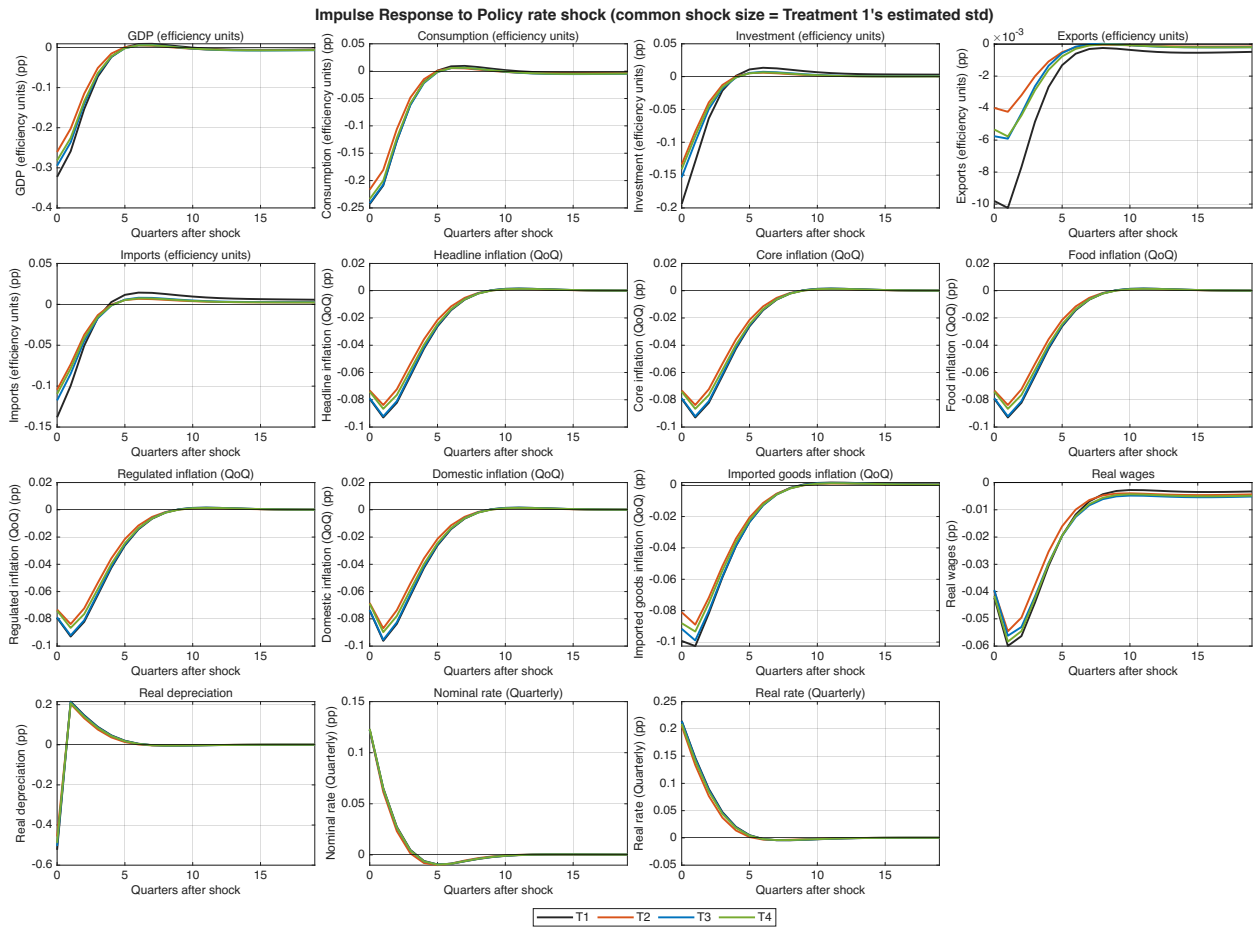


Figure C.6: Impulse response to the monetary policy shock, all four treatments, common shock size normalized to Treatment 1's estimated standard deviation. Like the mark-up shock, this is never rescaled under any treatment, and the four treatments' responses are correspondingly close to indistinguishable.

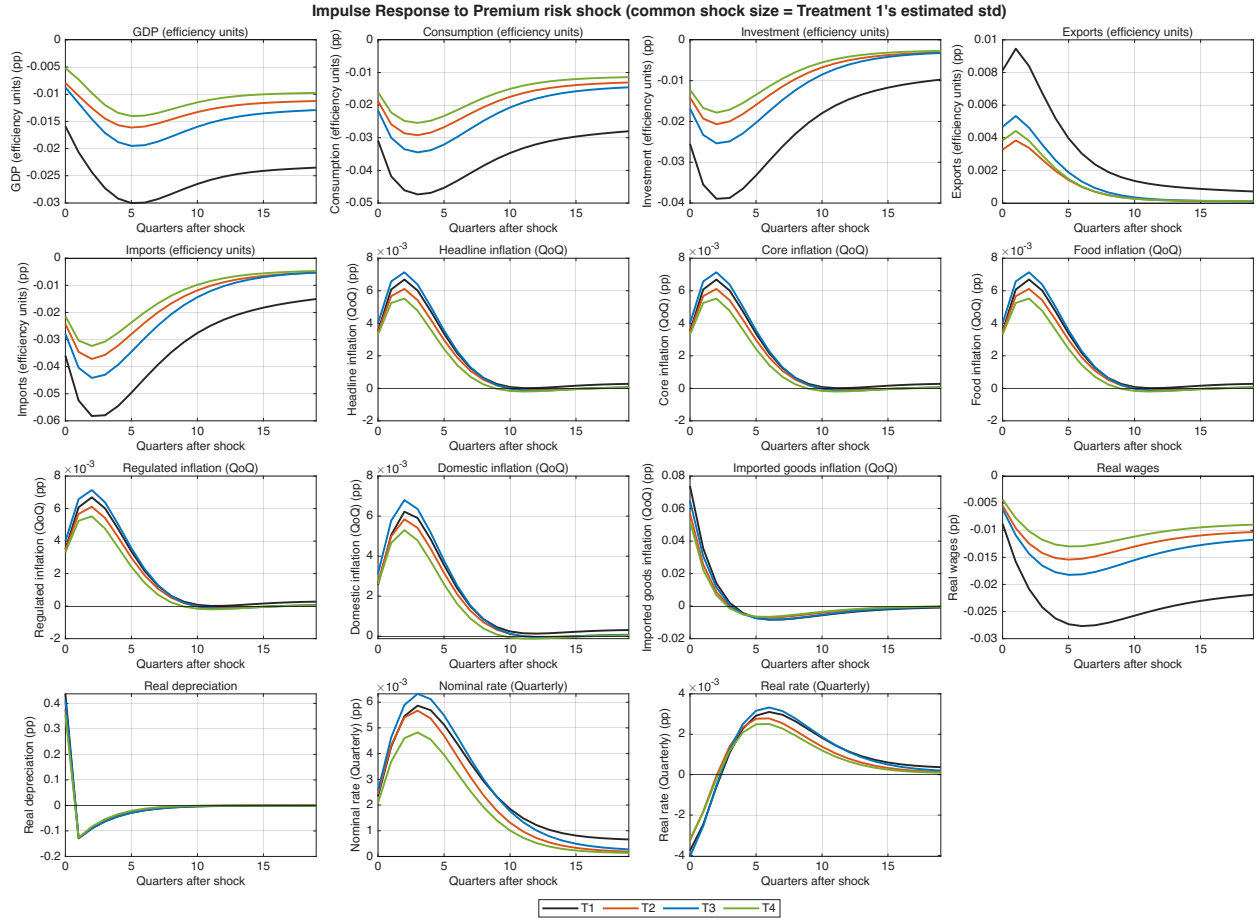


Figure C.7: Impulse response to the risk-premium shock ($\zeta^{i,prem}$), all four treatments, common shock size normalized to Treatment 1's estimated standard deviation.

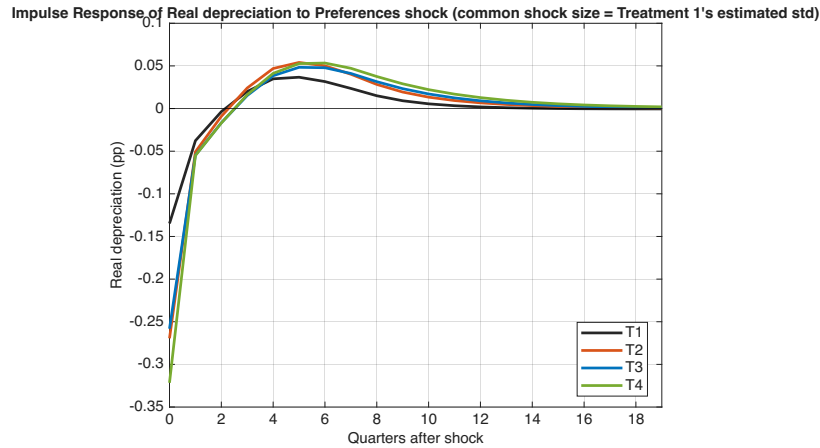


Figure C.8: Impulse response of real depreciation (Δz) to the preferences shock (ζ^u), all four treatments, common shock size normalized to Treatment 1's estimated standard deviation. The impact response is roughly 2.4 times larger under Treatment 4 than under Treatment 1, with Treatments 2 and 3 in between; all four share the same sign and shape, an impact depreciation followed by a reversal that fully unwinds within ten to fifteen quarters.

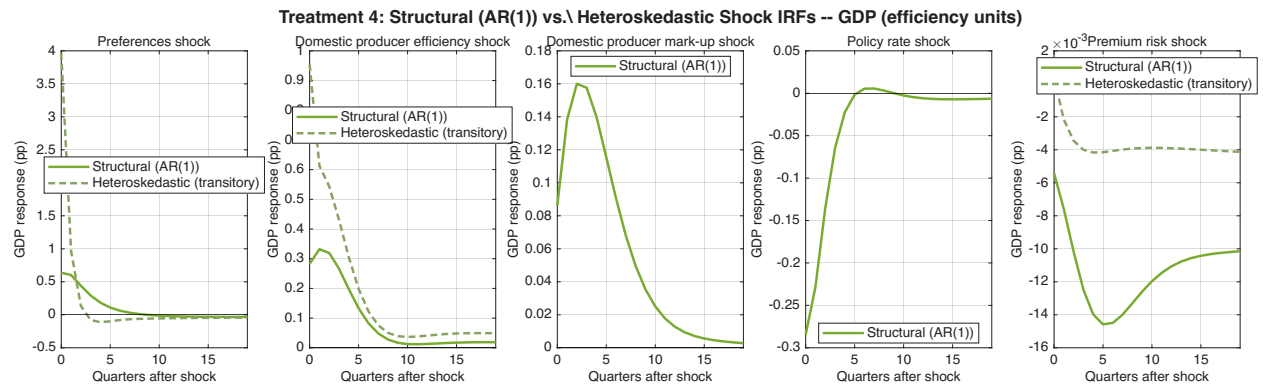


Figure C.9: Treatment 4: structural (AR(1)) vs. heteroskedastic shock impulse responses, GDP in efficiency units, each shock at its own native scale.

C.3 Full parameter comparison

Section 6.2 reports only the 15 largest standardized shifts across all four treatments. Figure C.10–C.13 report the same standardized comparison, $z_{\theta,k} = (\hat{\theta}_k - \hat{\theta}_1)/\text{se}(\hat{\theta}_1)$, for every one of the 57 parameters and shock standard deviations the four treatments estimate in common, split into the four natural blocks: shock persistences, other structural parameters, national-accounts discrepancy constants, and shock standard deviations. Rows within each block are sorted by the largest absolute standardized shift across Treatments 2–4.

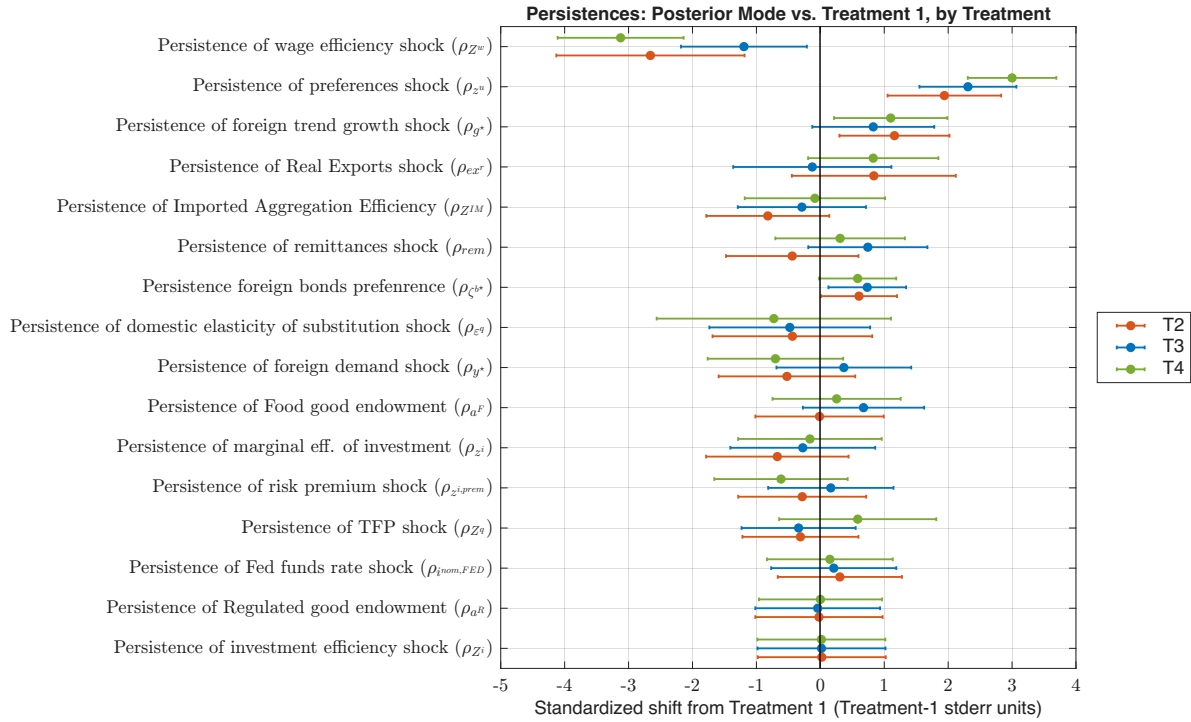


Figure C.10: Standardized posterior-mode shift, all 16 shock persistence parameters (ρ) common to all four treatments.

Figure C.14 shows the same comparison a second way for a curated set of nine parameters most directly tied to the paper’s mechanism, the persistence and standard deviation of the three shocks Treatments 3–4 actually rescale during Covid-19 (preferences, investment efficiency, risk premium) plus the three Taylor rule coefficients, overlaying each treatment’s posterior as a Laplace-approximation normal density $\mathcal{N}(\hat{\theta}_k, \text{se}(\hat{\theta}_k)^2)$ rather than reducing it to a single standardized point estimate.

As a robustness check on the standardized measure above, Figure C.15–C.19 report the same 57 parameters’ shift a second way: as a relative deviation from Treatment 1’s own posterior mode, $r_{\theta,k} = (\hat{\theta}_k - \hat{\theta}_1)/|\hat{\theta}_1|$, expressed as a percent, rather than standardized by Treatment 1’s stderr. The two measures answer different questions and do not always agree on which parameters moved most. The standardized measure asks whether a shift is large relative to how *precisely* Treatment 1 pinned that parameter down, which requires the reader to translate stderr units into a familiar probability statement (a shift of two Treatment-1 stderrs corresponds to roughly the edge of a 95% interval around Treatment 1’s own mode) before the magnitude means anything intuitively. The relative measure sidesteps that translation entirely: a parameter reported at –92% moved to

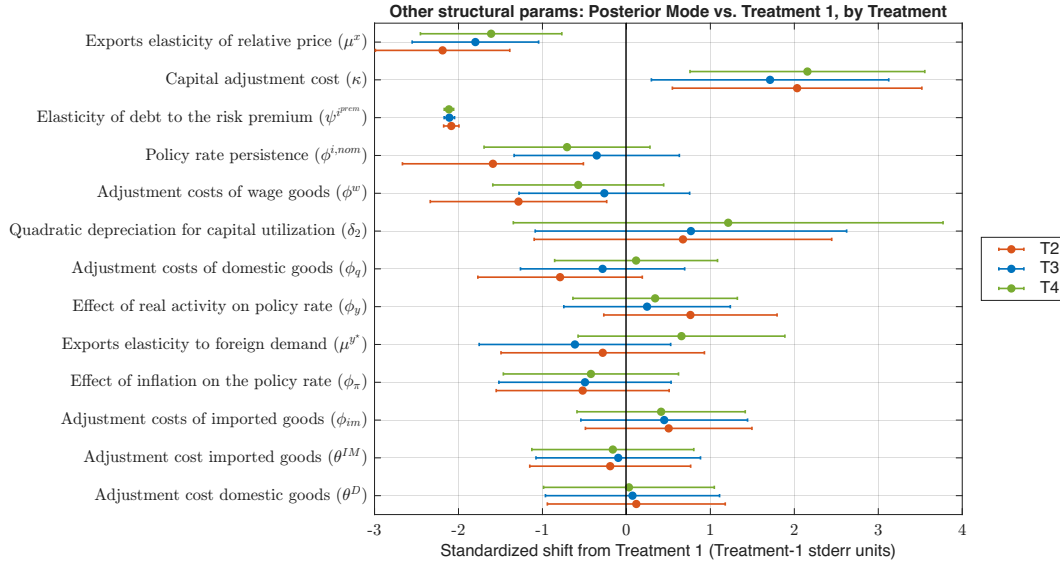


Figure C.11: Standardized posterior-mode shift, the 13 remaining structural parameters (Taylor rule, adjustment costs, trade elasticities) common to all four treatments.

roughly a tenth of its Treatment-1 level, full stop, independent of how tightly either treatment estimated it. For the same reason, [Figure C.15–C.19](#) plot the posterior mode shift only, with no accompanying uncertainty whisker: the natural whisker for this measure, a treatment’s own posterior stderr divided by Treatment 1’s mode, mixes an absolute measure of precision with a level-based denominator, so its length reflects how small the baseline happens to be rather than how far the estimate actually moved, and can dwarf the point estimate for any parameter with a small Treatment-1 level. [Figure C.10–6.2](#) remain the appendix’s source for posterior uncertainty; these figures answer a narrower question, the size of the mode shift itself, deliberately without one. One parameter, a national-accounts discrepancy constant whose Treatment-1 mode is too close to zero to divide by safely, is excluded from this measure only (retained in the standardized figures above). The debt-elastic risk-premium elasticity is the standout case where the two measures diverge in emphasis: the standardized figures above report it as a moderate 1.5–2 stderr shift, but on a relative basis it is the most uniformly large mover of all 57 parameters, falling by roughly 92–93% of its Treatment-1 level under all three alternative treatments alike, a reminder that “statistically distinguishable” and “practically large” are not the same claim.

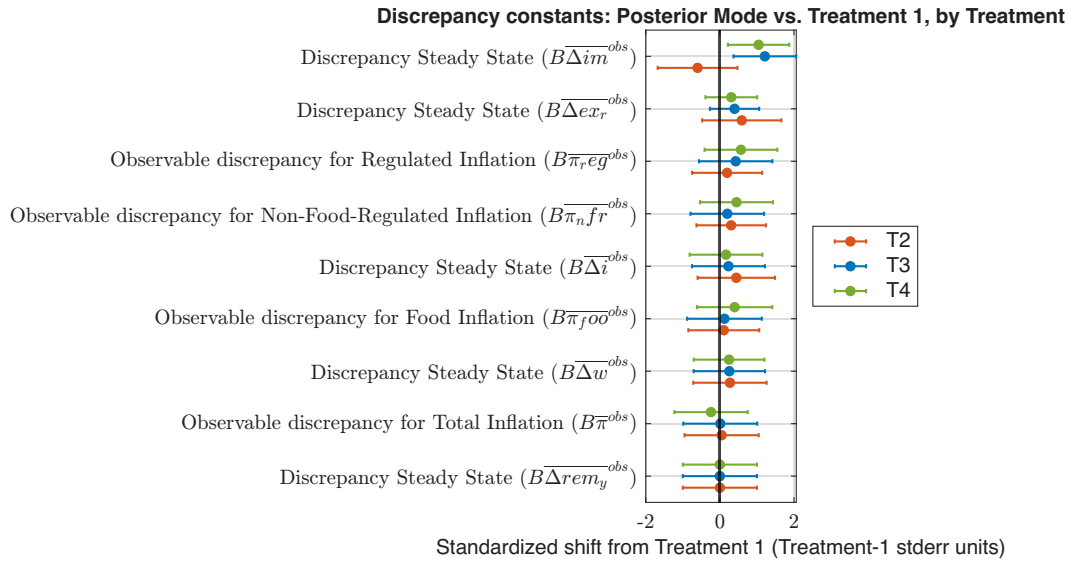


Figure C.12: Standardized posterior-mode shift, the 9 national-accounts discrepancy constants common to all four treatments.

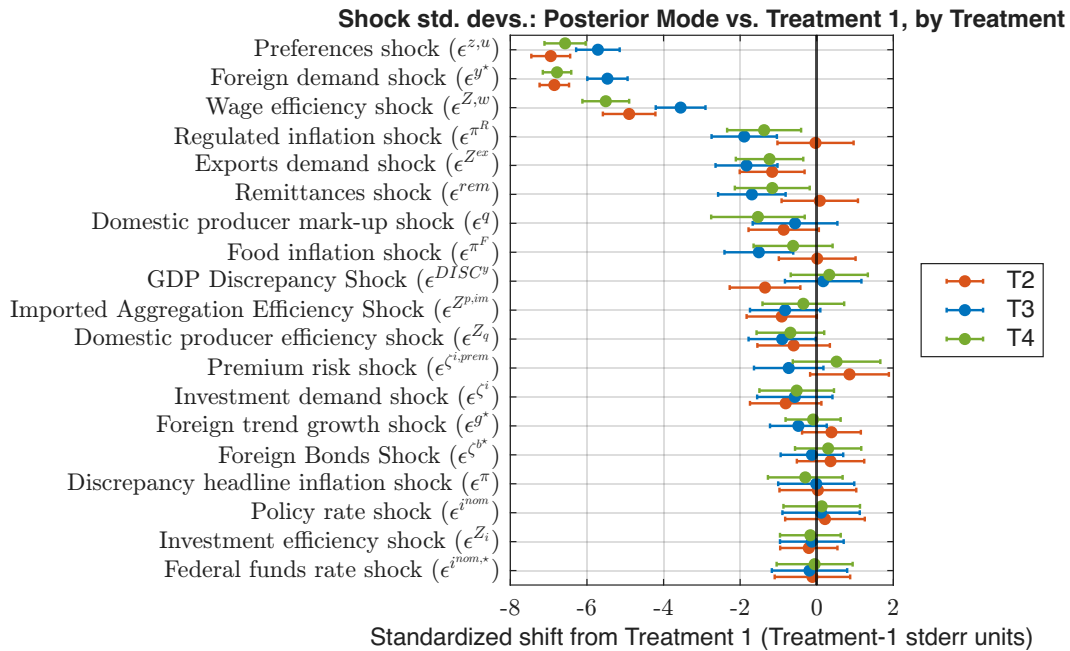


Figure C.13: Standardized posterior-mode shift, all 19 shock standard deviations common to all four treatments.

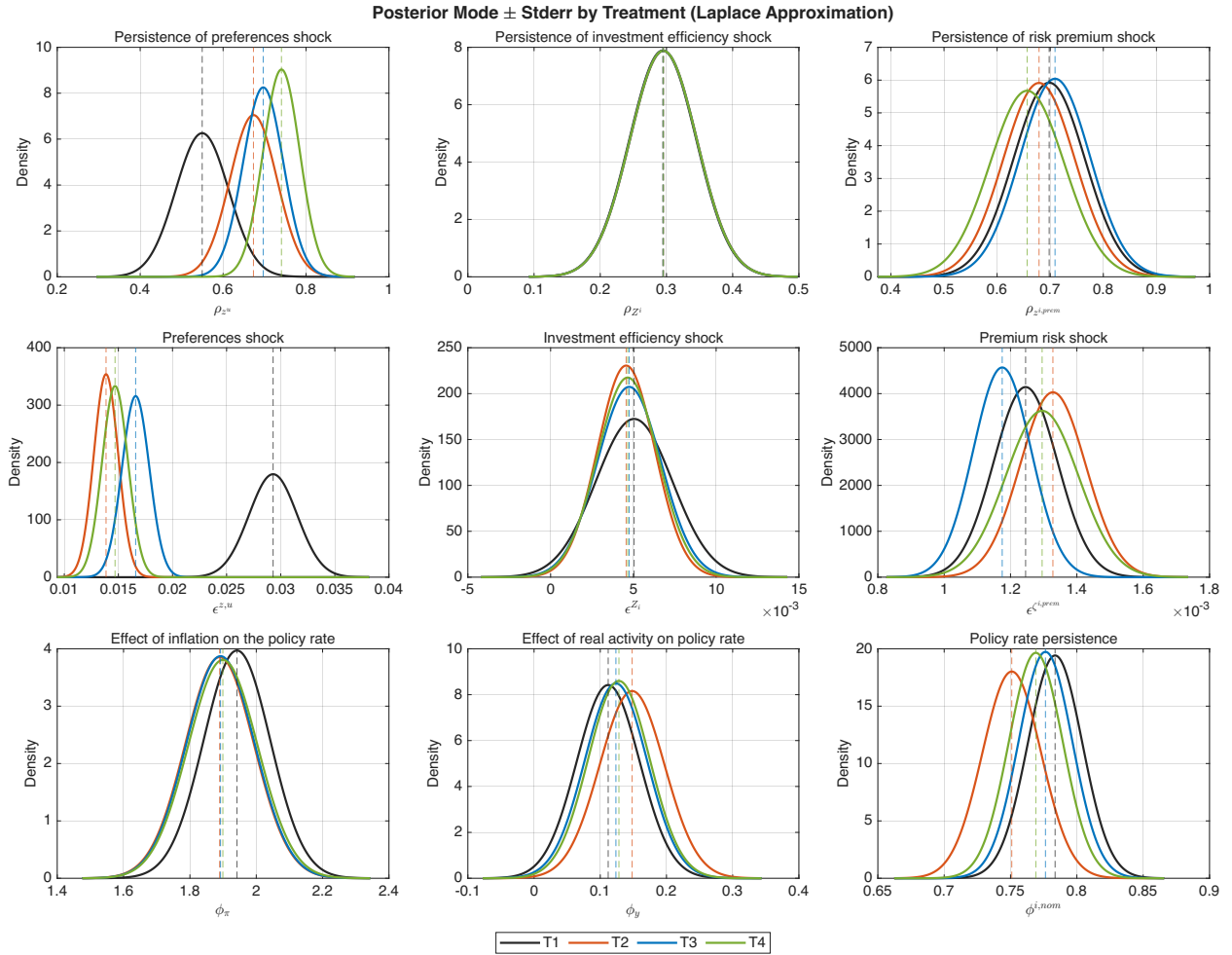


Figure C.14: Posterior mode and standard error (Laplace-approximation normal density) by treatment, for the three Covid-rescaled shocks' persistence and standard deviation, and the Taylor rule.

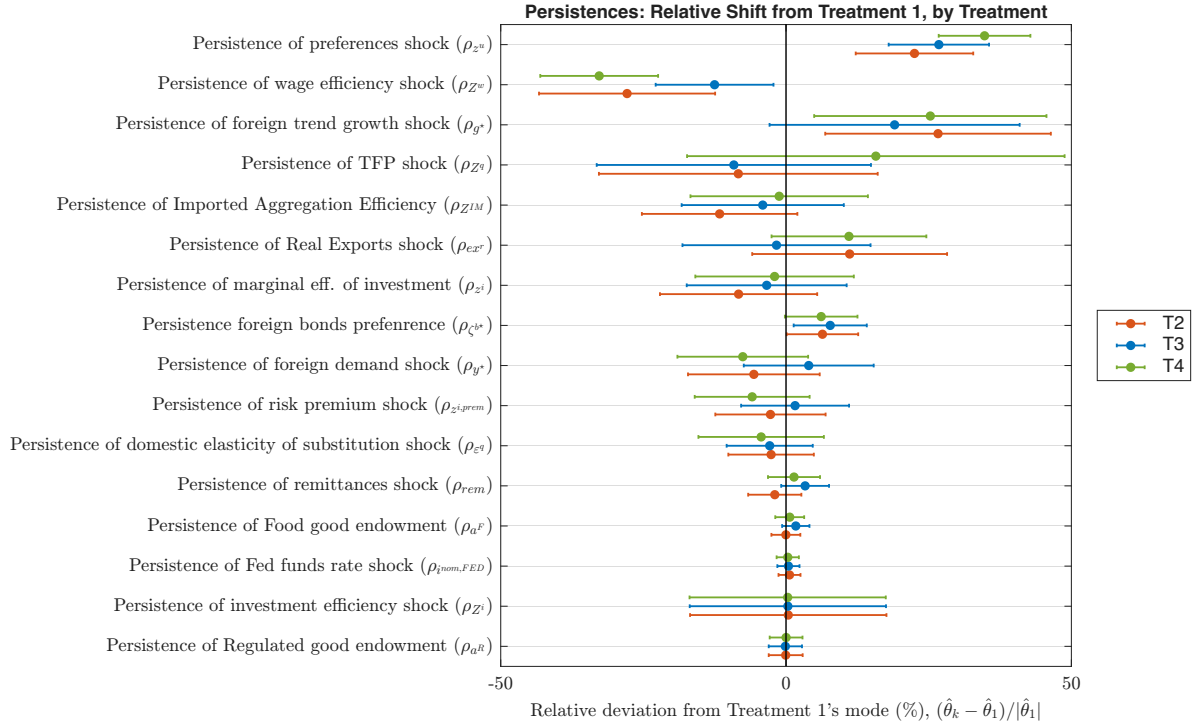


Figure C.15: Relative posterior-mode shift, $(\hat{\theta}_k - \hat{\theta}_1)/|\hat{\theta}_1|$, all 16 shock persistence parameters (ρ) common to all four treatments.

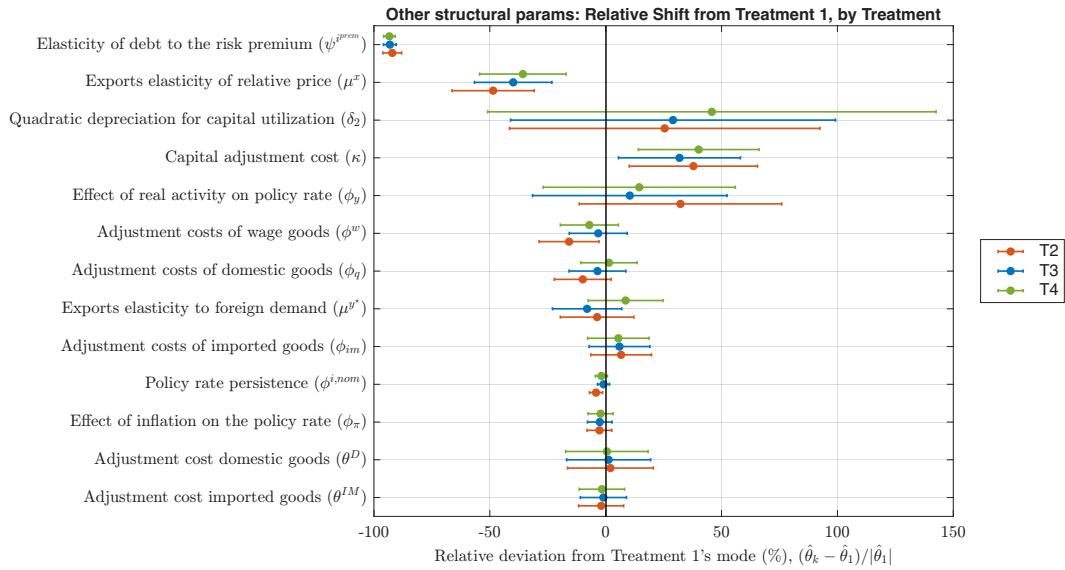


Figure C.16: Relative posterior-mode shift, the 13 remaining structural parameters common to all four treatments.

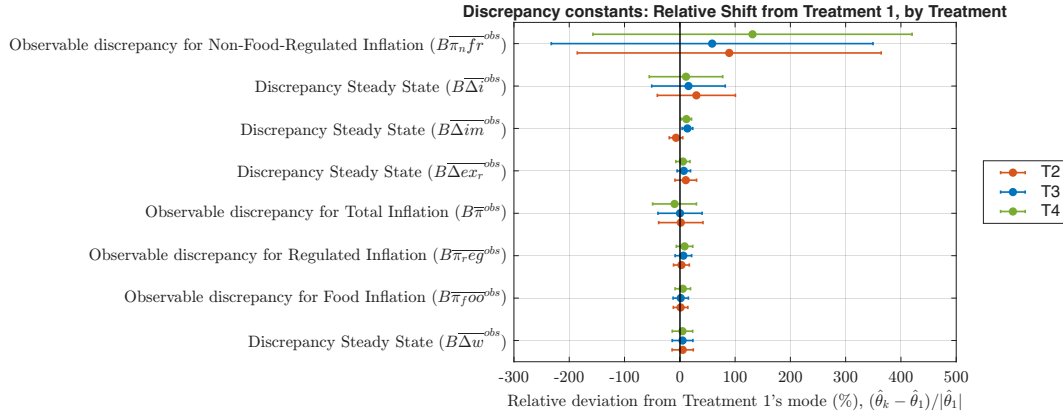


Figure C.17: Relative posterior-mode shift, 8 of the 9 national-accounts discrepancy constants common to all four treatments (one excluded: Treatment 1's own mode is too close to zero to divide by safely).

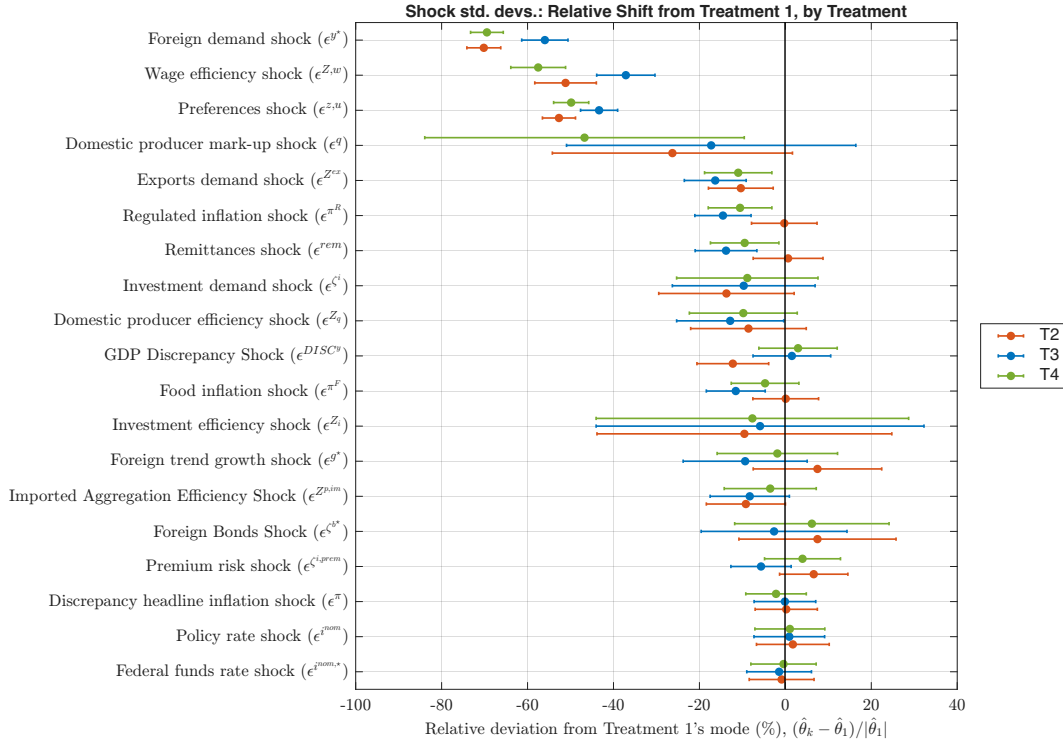


Figure C.18: Relative posterior-mode shift, all 19 shock standard deviations common to all four treatments.

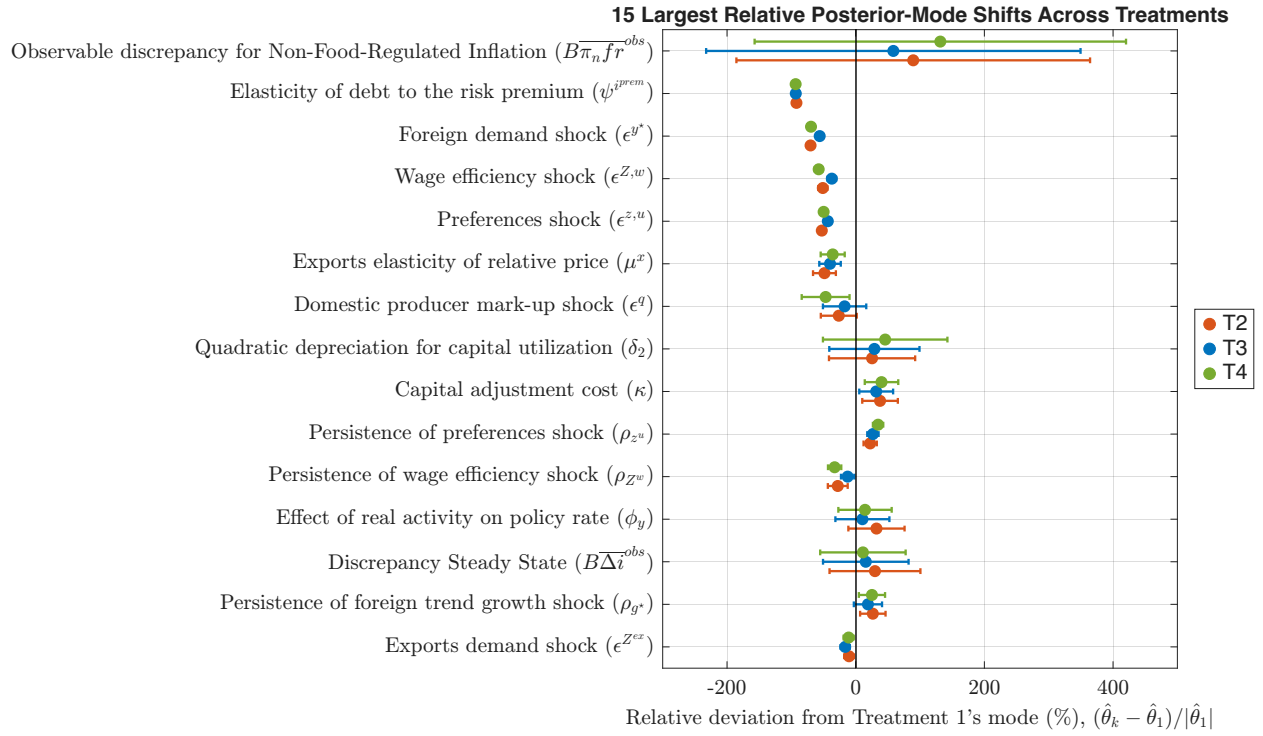


Figure C.19: The 15 parameters with the largest relative posterior-mode shift under any of Treatments 2–4 relative to Treatment 1; compare against Figure 6.2's standardized ranking of the same 57 parameters.

C.4 Full shock decomposition, all four treatments

Section 6.4's main-text figures zoom to a 2018–2025 window and compare three treatments at a time (Treatments 2–4 for the persistence-assumption story), the narrower view needed to keep those figures legible alongside body text. For completeness, Figure C.20–C.23 report the same grouped shock decomposition for all FOUR treatments, including the untreated Treatment 1 baseline omitted from the main text, over the FULL 2005–2025 sample, so all three high-volatility episodes (the 2008–09 financial crisis, the 2014–16 oil-price collapse, and Covid-19) are visible together for every variable. Treatments are stacked vertically here rather than side by side, since a four-treatment row would otherwise be too narrow to read; this layout is used only here; the main text's side-by-side figures remain the more legible choice when discussing a specific window in surrounding prose.

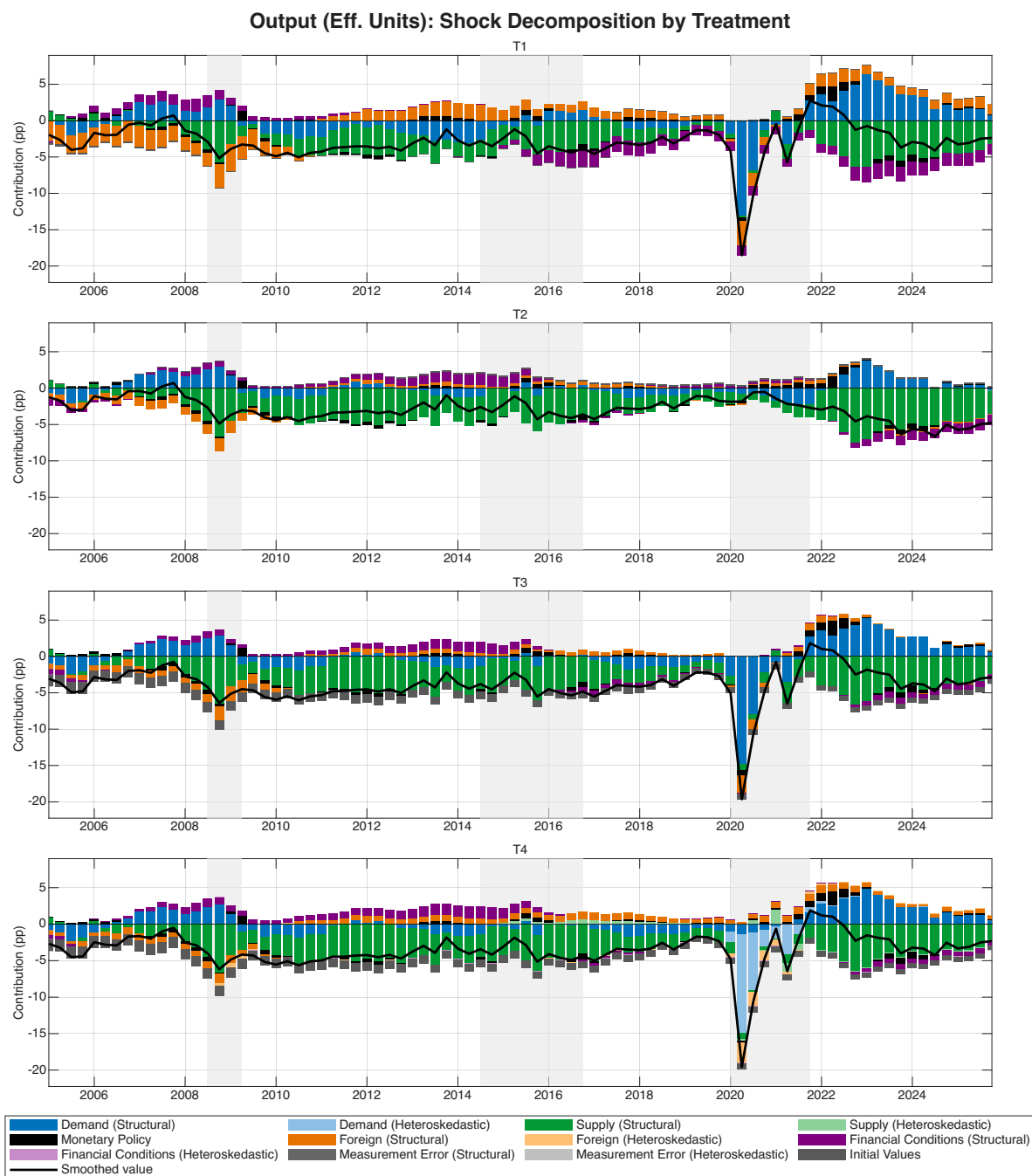


Figure C.20: Shock decomposition of detrended GDP in efficiency units (y_t), full 2005–2025 sample, all four treatments. Shaded regions mark τ_{GFC} , τ_{Oil} , and τ_{Covid} .

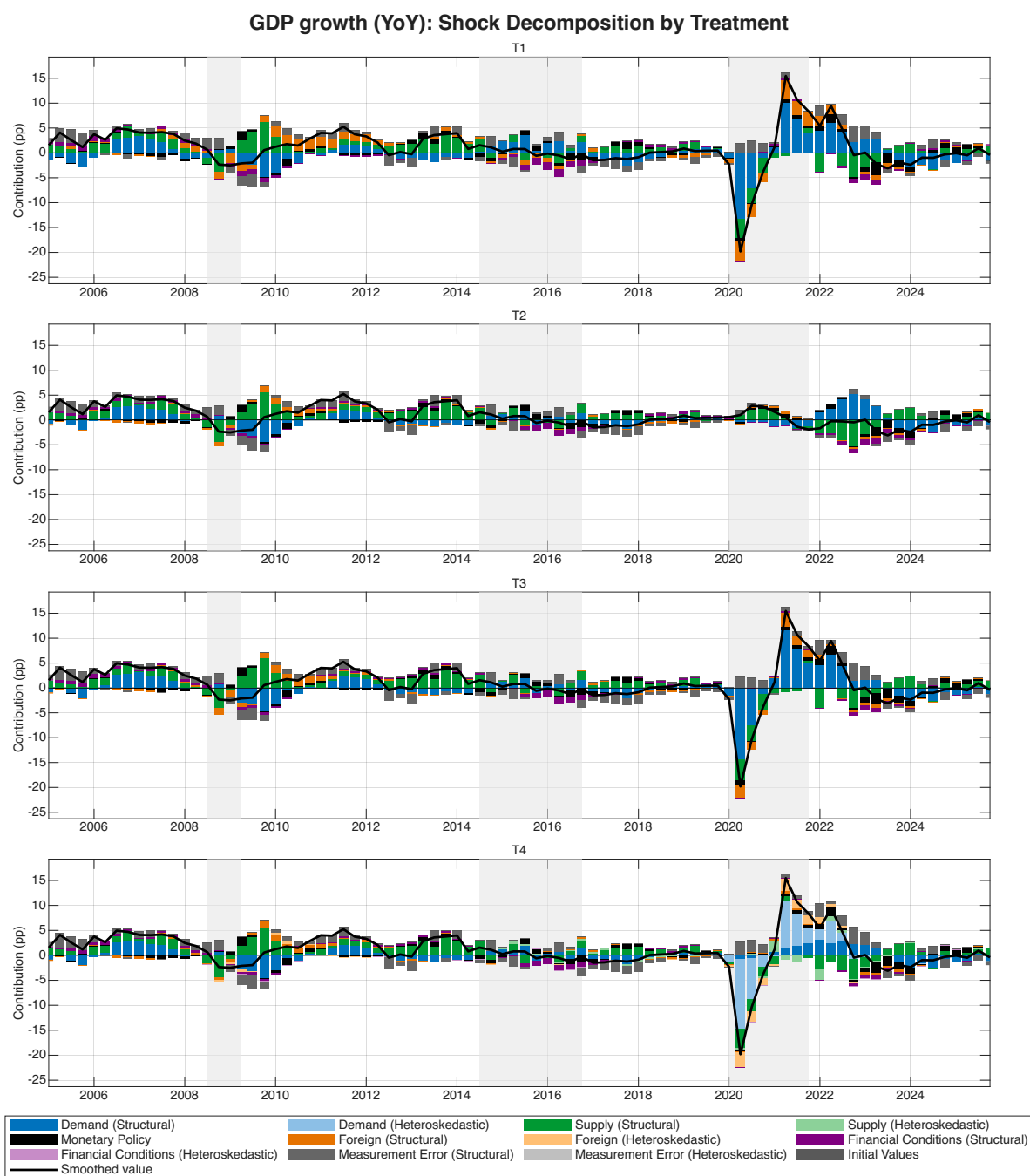


Figure C.21: Shock decomposition of annual GDP growth ($\Delta_4 y_t$), full 2005–2025 sample, all four treatments.

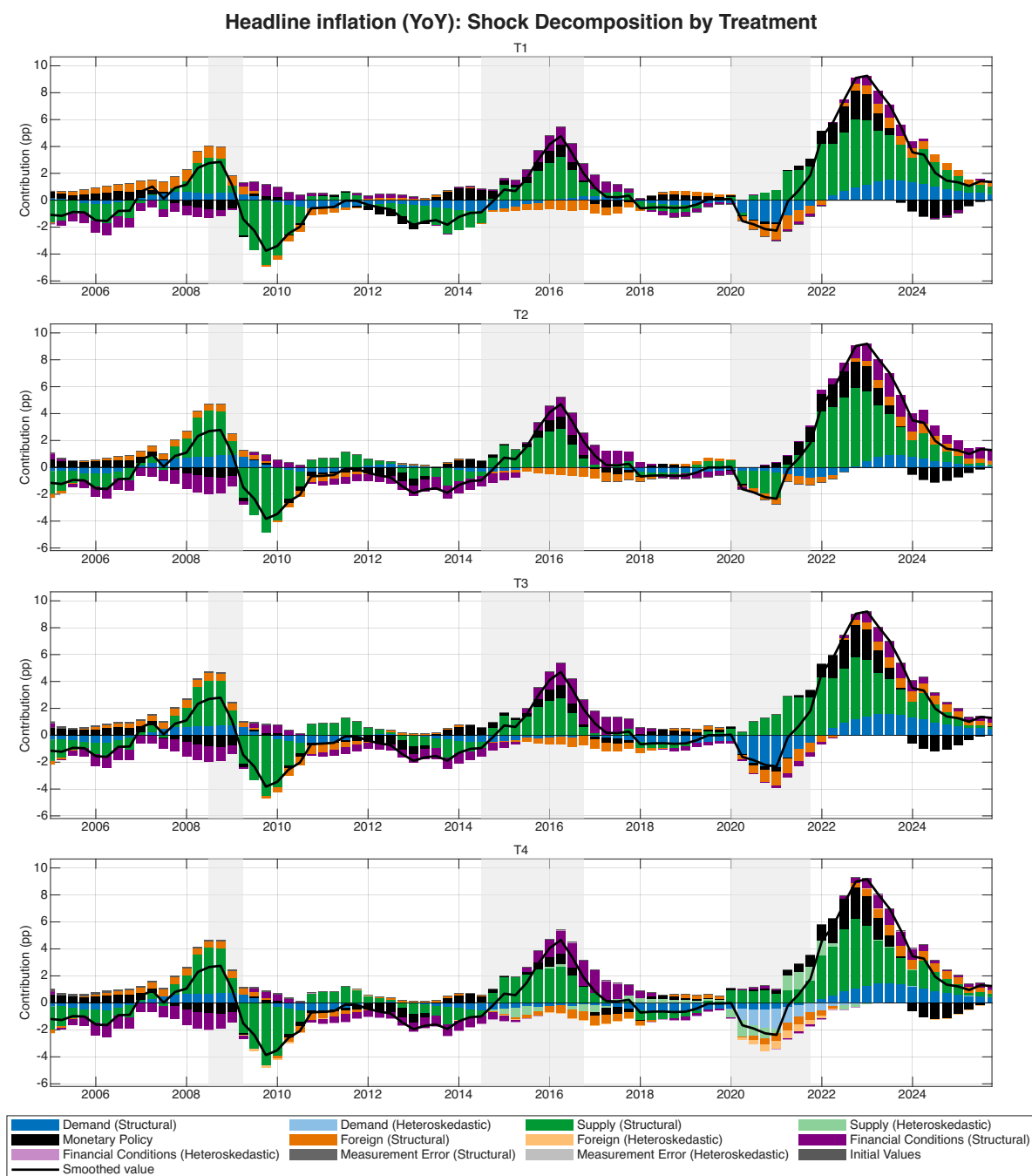


Figure C.22: Shock decomposition of annual headline inflation (π_t^4), full 2005–2025 sample, all four treatments.

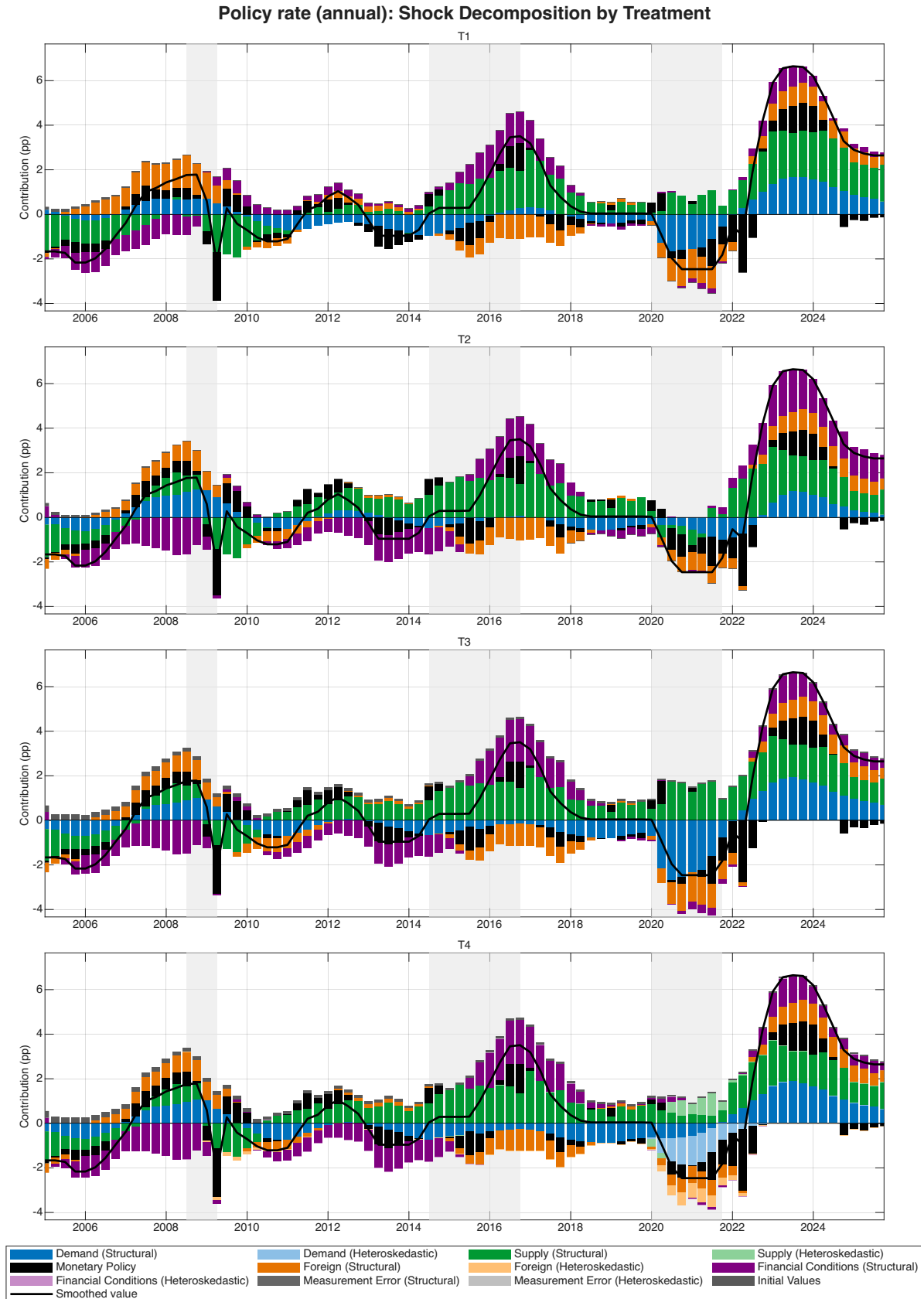


Figure C.23: Shock decomposition of the annualized policy rate ($i_t^{nom,A}$), full 2005–2025 sample, all four treatments.

C.5 Full forecast evaluation results (MAE)

Section 6.5 reports the root-mean-squared forecast error (Table 6.2, Figure 6.7–6.9) as the headline metric. Figure C.24–C.26 report the same horizon-by-horizon, window-by-window comparison under mean absolute error, for all four treatments and both atheoretical benchmarks; the ranking and the post-Covid-specific pattern are unchanged from the RMSE results, MAE is simply less sensitive to the largest individual forecast misses.

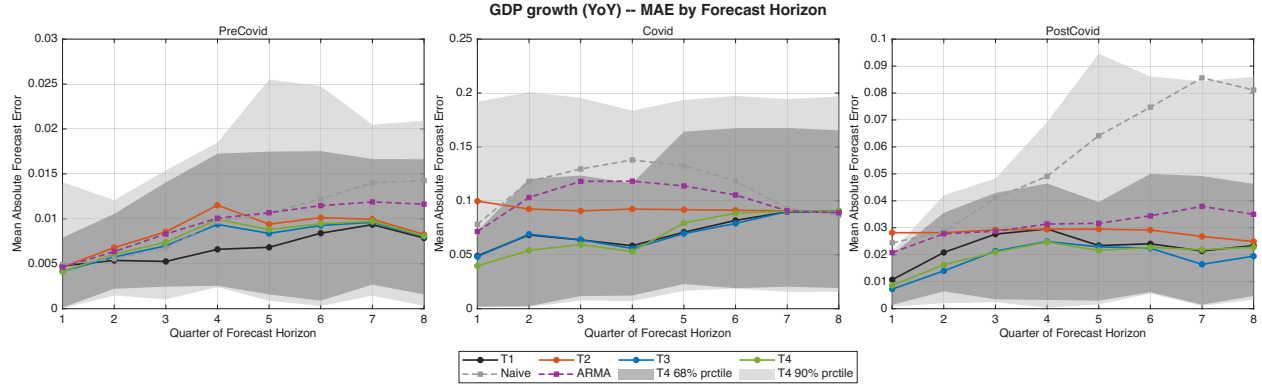


Figure C.24: Mean absolute forecast error vs. horizon for annual GDP growth ($\Delta_4 y_t$), by treatment and benchmark, tiled by target window.

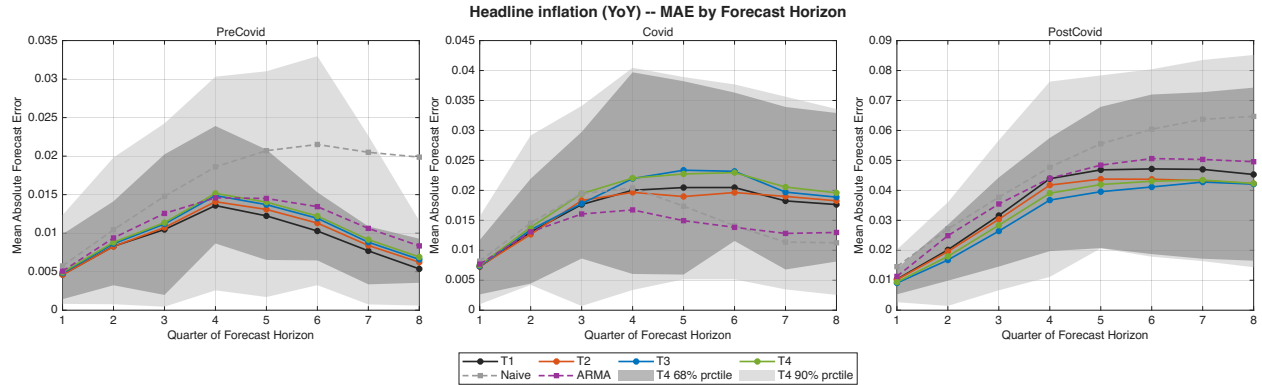


Figure C.25: Mean absolute forecast error vs. horizon for annual headline inflation (π_t^4), by treatment and benchmark, tiled by target window.

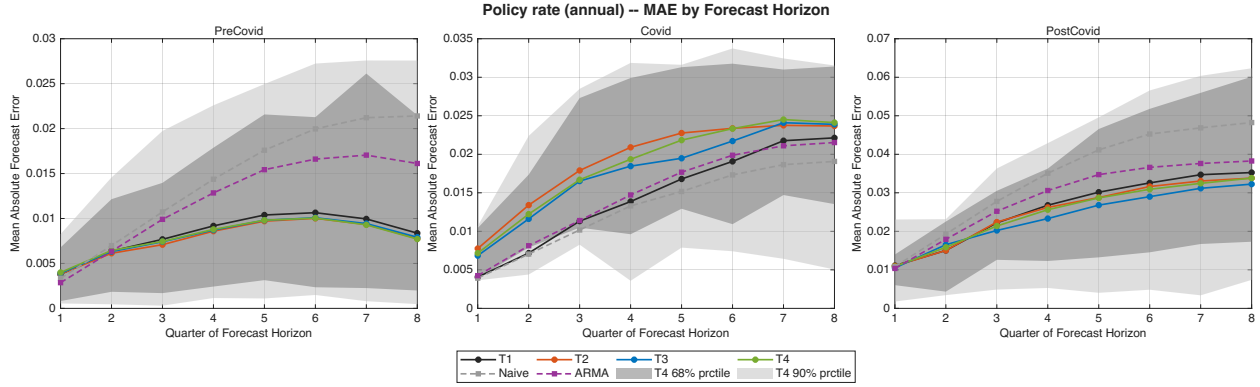


Figure C.26: Mean absolute forecast error vs. horizon for the annualized policy rate ($i_t^{nom,A}$), by treatment and benchmark, tiled by target window.

C.6 Cross-country evidence on pandemic volatility

The scatter comparison in [Figure 1.1](#) is built from quarterly GDP growth, CPI inflation, and short-term interest rate data pulled from the OECD Data Explorer for the same panel of economies, using the τ_{GFC} , τ_{Oil} , and τ_{Covid} windows already defined in [Table 5.1](#) to separate normal quarters from all three high-volatility episodes uniformly across countries, a simplification defensible on the grounds that the GFC and the 2014–16 oil-price collapse were themselves globally synchronized episodes rather than Colombia-specific ones. The two figures below develop the same comparison further as quantile-quantile (QQ) plots, which show the shape, not just the scale, of the pandemic's effect on each series' distribution: each country's observations are first standardized by that country's own normal-quarter mean and standard deviation before being pooled across countries, so that a small, low-volatility economy does not get swamped by a larger or more volatile one.

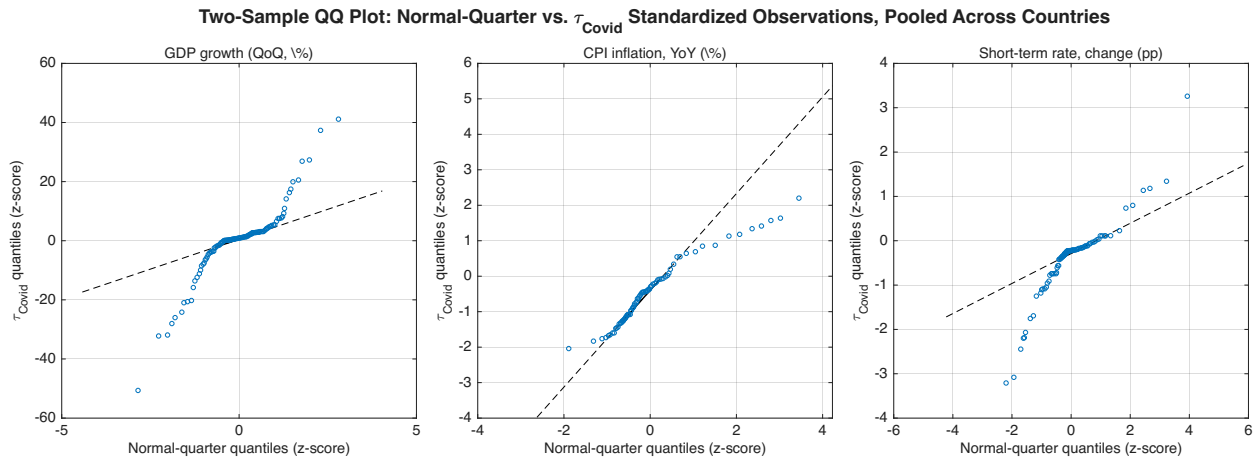


Figure C.27: Two-sample QQ plot comparing pooled, standardized normal-quarter observations against pooled, standardized τ_{Covid} observations, for each of the three series in Figure 1.1. A straight line through the origin would indicate the pandemic quarters were simply a rescaled version of the same distribution; curvature away from the 45-degree reference, most pronounced for GDP growth, indicates the pandemic's added dispersion is concentrated in a small number of extreme quarters rather than spread uniformly across the whole distribution. For the interest rate, most observations are in fact compressed toward the origin below the reference line, consistent with policy rates being held near the effective lower bound through most of the window, but a handful of extreme outliers, the initial emergency cuts of March 2020, still pull the upper tail away from the line.

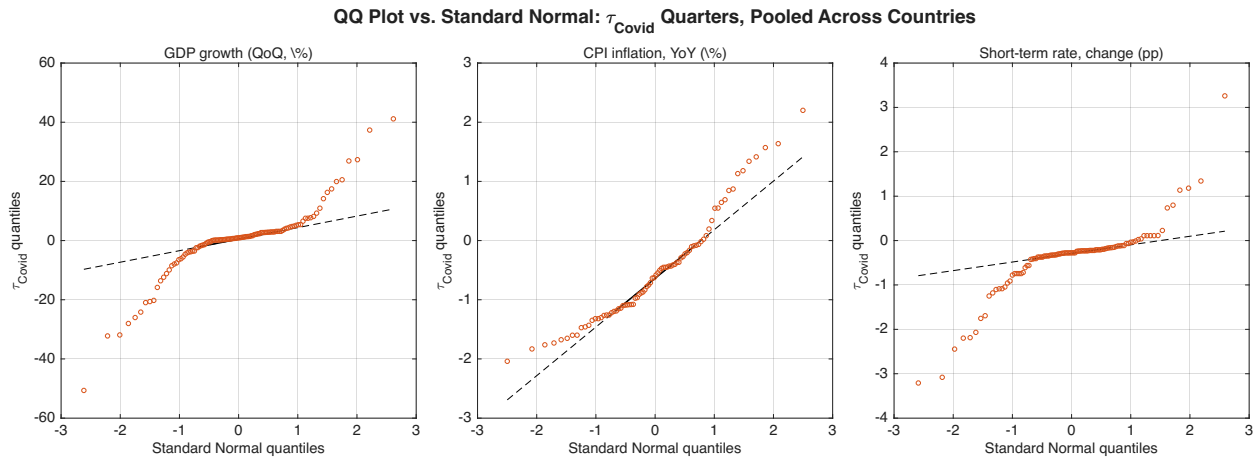


Figure C.28: QQ plot of pooled, standardized τ_{Covid} observations against a standard Normal reference, for the same three series. Points hugging the 45-degree line indicate the pandemic-quarter distribution is close to Gaussian; the pronounced S-shaped bow at both tails for GDP growth, and the milder version of the same pattern for CPI inflation and the interest rate, is a direct picture of the constant-variance-shock assumption underlying Treatment 1 breaking down during τ_{Covid} : a small number of quarters are far more extreme than a Normal distribution calibrated to the bulk of the sample would predict.