

RESEARCH ARTICLE

# Testing for the footprints of stabilization economic policy in forecast errors

Wojciech Charemza<sup>1,2</sup>, Christian Francq<sup>3</sup>, Radu Lupu<sup>4</sup>, Svetlana Makarova<sup>5\*</sup>, Jean-Michel Zakoïan<sup>3</sup>

**1** Vistula University, Poland, **2** University of Leicester, United Kingdom, **3** CREST, Paris, France, **4** Bucharest University of Economic Studies and Institute for Economic Forecasting, Romanian Academy, Romania, **5** University College London, United Kingdom

\* [s.makarova@ucl.ac.uk](mailto:s.makarova@ucl.ac.uk)



## OPEN ACCESS

**Citation:** Charemza W, Francq C, Lupu R, Makarova S, Zakoïan J-M (2025) Testing for the footprints of stabilization economic policy in forecast errors. PLoS One 20(12): e0336495. <https://doi.org/10.1371/journal.pone.0336495>

**Editor:** Nicolò Pecora, Università Cattolica del Sacro Cuore Sede di Piacenza e Cremona Facoltà di Economia; Università Cattolica del Sacro Cuore Facoltà di Economia e Giurisprudenza, ITALY

**Received:** May 18, 2025

**Accepted:** October 27, 2025

**Published:** December 1, 2025

**Copyright:** © 2025 Charemza et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** Data is available in Supplement 3 and in the figshare repository at [https://figshare.com/articles/dataset/Data\\_for\\_Testing\\_Footprints\\_xlsx/30016453?file=57852376](https://figshare.com/articles/dataset/Data_for_Testing_Footprints_xlsx/30016453?file=57852376).

## Abstract

This paper introduces a novel statistical test, the Policy Effects Lagrange Multiplier (PELM) test, to detect stabilization policy effects in the distribution of forecast errors from dynamic financial models. Traditional analyses of policy impact typically rely on explicit policy information or direct intervention data, which are often unavailable or incomplete. In contrast, the proposed PELM test infers policy footprints from the distribution of forecast errors alone. Empirically applied to sovereign bond yield data from 33 countries before the Russian financial crisis of 2014, the test identifies countries showing stabilization policy footprints. Subsequent analysis shows that significant budgetary improvements were observed for years following the crisis in the group of countries where our test statistically confirmed stabilization policies. This confirms the rationale of test foundations and also indicates its predictive properties. Robustness checks further validate these findings across various model specifications and sensitivity scenarios. The proposed PELM test offers policymakers and researchers a powerful tool for evaluating stabilization policies, facilitating better forecasting and assessing policy efficiency in diverse economic contexts without necessitating detailed policy intervention data.

## 1 Introduction

This paper has two primary aims. The first aim is to propose a test for identifying stabilization policy effects (or ‘footprints’) in the distribution of forecast errors generated by a dynamic financial market model with time-varying volatility of a policy-dependent variable. The second aim is to demonstrate how the results of this test can effectively enhance the prediction of fiscal indicator dynamics.

The commonly used approaches to testing policy outcomes mainly focus on analyzing how the policy impacts the conditional mean of the modeled process. These methods typically involve identifying policy effects by using explicit information about

**Funding:** This work was supported by the Economic and Social Research Council and the Agence Nationale de la Recherche (ANR, Project PRAM ANR-10-ORAR-008-01) through the Open Research Area project Probabilistic Approach to Assessing Macroeconomic Uncertainties No. RES-360-25-0003.

**Competing interests:** The authors have declared that no competing interests exist.

the timing or magnitude of interventions and then applying econometric tools such as structural factor analysis, regime-switching techniques, or textual measures of policy communication (see, for example, Basistha and Kurov [1]; Forni and Gambetti [2]; Hanish [3]; Klingelhöfer and Sun [4]; Guidolin and Pedio [5]). Studies like Ehrmann and Fratzscher [6]; Nakamura and Steinsson [7]; and especially Hoesch, Rossi, and Sekhposyan [8], expanded this framework by examining how both information advantage and information channels evolve over time, demonstrating that the effectiveness of policy communication differs across various economic environments and time periods.

Therefore, if one considers the effects of policy decisions on higher moments of market indicators, most notably the variance of ex-ante forecast errors, usually associated with a measure of forecast uncertainty, the problem of data for identifying the factors affecting these becomes more intricate. From a theoretical perspective, the difficulty of identifying the timing and nature of events that reduce forecast uncertainty arises because uncertainty itself, if understood in the Knight [9] sense, is not directly observable. It can, however, be inferred from the distribution of forecast errors (see Rossi, Sekhposyan and Soupre [10]). In the rational expectations framework, information asymmetries imply that policy interventions can affect uncertainty through two channels: by altering the information set available to agents, thereby narrowing the dispersion of beliefs (see Lucas [11]), or by influencing second-order expectations, that is, beliefs about the beliefs of others (see Morris and Shin [12]). In both cases, the effect manifests not in the conditional mean of forecasts but in the higher moments of their distribution, particularly variance. This creates an identification challenge: variance reductions may stem from structural shocks directly targeting uncertainty, such as new regulatory frameworks or systematic changes in central bank communication, or other factors that incidentally dampen volatility. Hence, while the theoretical literature emphasizes the role of credible policy, transparency, and communication in stabilizing expectations, empirical analysis must grapple with the difficulty of mapping such abstract channels onto specific, datable events. In practice, there is often a plethora of such factors that are difficult to quantify or even date, but undoubtedly have some effects on the stabilization of financial markets. This information might either (a) represent direct decisions leading to reduced uncertainty or (b) be the indirect cause of such decisions.

Below are two rather evident examples of (a) above: decisions to stabilize the government bonds market. The first example is the MiFID II (Markets in Financial Instruments Directive II) Implementation Plans by the European Commission. It reduced uncertainty premiums while the regulatory framework remained ambiguous, particularly in corporate bond markets where transparency requirements were changing significantly (see, e.g., Kaya [13]). However, its timing is not clearly identified. It was announced on January 15, 2014 (political agreement reached); final technical standards were published on April 15, 2016. Another example is the evolution of communication between 2001 and 2005 at the European Central Bank (ECB). It occurred through a series of gradual changes to the Bank's communication practices. In late 2001, the ECB began using specific signaling phrases in its statements. Throughout

2002–2003, it standardized its communication patterns, and by 2004–2005, it had established a predictable format for including forward-looking qualitative projections in its statements. The entire process stabilized European bond markets, resulting in a gradual decline in forecast errors for ECB policy decisions and corresponding reductions in yield volatility around ECB announcements (see, e.g., Jansen and De Haan [14]). Many more examples can be given that are related to different countries and markets.

Regarding (b), it is much more challenging to come up with examples. It may be real-time information from experts, politicians, and independent forecasters, which serves as an early warning, creating a stabilizing effect. Sometimes, such information can be identified; during the 2008 financial crisis, Warren Buffett published the op-ed entitled ‘Buy American. I Am’ in the *New York Times* (October 16, 2008) significantly helped to stabilize markets by expressing long-term confidence when panic was widespread. However, this might be very difficult in most cases due to the diversified and sometimes confidential or secret nature of this information.

How do we deal with all this information coming at different times, interacting with each other, and generally not known to the researcher? They surely have some effects on forecast uncertainty, some stabilizing it, and some do not. We propose a sort of reverse engineering approach here. We focus on identifying policy outcomes rather than explicitly tracing the policy interventions themselves. It builds on the findings of Charemza, Díaz, and Makarova [15], who showed that, under reasonable assumptions, the empirical distribution of ex-ante forecast errors from an econometric model of policy-dependent economic indicators can be approximated by the weighted skew-normal (WSN) distribution. The parameters of this distribution can be interpreted as outcomes of actions based on qualitative and quantitative signals received from various sources, such as banks’ statements, communications from independent forecasters, and economic commentators. However, estimating these parameters is challenging due to numerical difficulties and restrictions that must be imposed on the parameters. The test proposed in this paper, which is based on the Lagrange Multiplier principle and is called the PELM test (the Policy Effects Lagrange Multiplier test) throughout this paper, can be used for detecting footprints of economic policy in forecast errors without the need for such estimation.

In other words, we can test for the effects of non-econometric forecasting, e.g., by experts, professional forecasters, influencers, gurus, or economic analysts in cases where such forecasts’ data are unavailable. In fact, except for a few countries, such data are not available or do not exist in an explicit form, as experts’ opinions and advice are often presented not transparently. Alternatively, they might be produced in a fuzzy or indirect form and published in various media.

The quadratic loss function is a natural choice of the objective function in this context because reductions in forecast error variance directly lead to measurable improvements in predictive accuracy. At the same time, we recognize that alternative uncertainty measures have been suggested, including those based on asset-pricing models or non-quadratic loss frameworks (see, for example, David and Veronesi [16]). Our aim is not to dismiss these methods but to provide a complementary, testable mechanism: one that links stabilization policies to changes in the distribution of forecast errors without requiring detailed information about the timing or magnitude of policies.

The significance of the PELM test statistics is, evidently, only the necessary but not sufficient condition for claiming that the test proposed here indeed discovers footprints of a stabilization policy. As is often the case in abductive reasoning, there might be potentially different causes of such a result. This is why we conduct our research in two stages, which correspond to the two aims of this paper. In the first stage, we apply the PELM test to the ex-post forecast errors (forecast uncertainty) of a dynamic volatility model of government bond yields to detect possible policy intervention footprints. Then, in the second stage, we test whether significant budgetary changes confirm the stabilization effect.

Reduction in forecast uncertainty might also positively affect budget deficit through lower borrowing costs, increasing fiscal policy confidence, stimulating economic growth, and allowing more flexibility in central bank policy coordination (for evidence in developed markets, see, e.g., Peppel-Srebrny [17]; and, in particular, Jiang and Li [18], for comprehensive empirical evidence; for a broader context, see Manzli et al [19]).

Our empirical analysis focuses on finding traces of stabilization policy in the time series of sovereign bond yields for 33 countries. First, we apply the PELM test to forecast errors of a dynamic volatility model fitted to daily data on these yields. On the basis of such testing, we divide these countries into two groups: one where the test statistics are significant and the other containing the remaining countries. Concentrating on the aftermath of the Russian financial crisis in December 2014, in the second stage of testing, we apply the paired t-test separately for both groups to the difference in budget deficits in the first year of the crisis and, subsequently, two, three, and four years later, when the stabilization policy effects should be evident. The paired test statistic turns out to be significantly positive for the group of countries where footprints of stabilization policy have been confirmed, but not for the other group. In our view, this confirms the rationale of our testing approach and fulfils the second aim of this paper.

The remainder of the paper is organized as follows: in Section 2, we formalize the research problem and illustrate the rationale of our settings. Section 3 sets out the general assumptions underpinning our testing approach. It also describes two variants of our PELM test; one is computationally easier and has a chi-squared asymptotic distribution, but is based on rather strict assumptions. The other variant is computationally heavy but can be conducted under less restrictive assumptions. This section gives a detailed discussion of the primary asymptotic characteristics of both variants. Section 4 analyzes the practical aspects and properties of these tests in finite samples, particularly the use of simulated critical values and p-values. It also presents some results illustrating power of the tests. Section 5 addresses the second aim of the paper. This section discusses the paired test results for budget deficits separately for the countries with significant and non-significant PELM statistics. Section 6 summarizes the outcomes of various robustness checks. In particular, it demonstrates that our test can be of practical use for analyzing other global financial crises, such as the crisis caused by the unexpected outcome of the UK's Brexit referendum in 2016. This empirical evidence confirms that the significance of the PELM statistics corresponds to creating positive fiscal effects. Section 7 concludes, highlighting broader potential applications and directions for future research. The paper is completed with Supplement 1 with generalizations, proofs of lemmas and propositions; Supplement 2 with additional simulation and empirical results; and Supplement 3 with the data used for empirical example and robustness analysis. The data is also available at data repository *figshare* at [https://figshare.com/articles/dataset/Data\\_for\\_Testing\\_Footprints\\_xlsx/30016453?file=57852376](https://figshare.com/articles/dataset/Data_for_Testing_Footprints_xlsx/30016453?file=57852376).

## 2 Set-up and the testing problem

The problem of detecting the effects of policy interventions on the distribution of forecast errors can be formulated in the following way.

Let  $m_t$  and  $\sigma_t^2$  be the core predictions of the level and variance of an economic or financial variable  $X_t$ . These predictions rely entirely on information available at time  $t-1$ . This information is denoted by  $\mathcal{F}_{t-1}$ , and comprises the historical data of the economic variable, i.e., the values of  $X_u$  for  $u < t$ . Thus, in the absence of economic policy intervention, the variable would be

$$X_t^* = m_t + \sigma_t U_t^*,$$

with an error term  $U_t^*$ , which is assumed to be identically and independently distributed, *iid*,  $\mathcal{N}(0, 1)$ . Next, assume that policymakers have access to a vector  $s_t$  of additional information that includes news, information from policy advisors, or even rumors created by media influencers that arise after the core prediction was made. If such information had been publicly available prior to the time  $t$  it would have been included in  $\mathcal{F}_{t-1}$  and thus in the core prediction. Let  $\mathcal{S}_t$  be the sigma-field generated by  $s_t - E(s_t | \mathcal{F}_{t-1})$ , that is the part of the information conveyed by  $s_t$  that cannot be obtained from the past information. Note that  $\mathcal{S}_t$  is independent of  $\mathcal{F}_{t-1}$ , but not of  $X_t^*$ , and therefore not of  $U_t^*$ .

By contrast to  $\mathcal{F}_{t-1}$ ,  $\mathcal{S}_t$  captures real-time information arriving at time  $t$ , after the core prediction was made at  $t-1$ , and that cannot be extracted from information set  $\mathcal{F}_{t-1}$ . For financial markets, such information can be on order-flow

imbalances, intra-day realized volatility, unexpected news releases, and also on relevant economic, political, weather-related, and other news not incorporated into the core prediction. Importantly, the role of information set  $S_t$  is not to shift the mean forecast but to reduce the dispersion of realized forecast errors by possibly triggering stabilizing actions. Empirical studies of forecast errors in financial markets show low correlation between ex-ante forecasts and subsequent real-time information updates, especially during periods of market stress when stabilizing interventions are most common; see, e. g., Fleming and Remolona [20].

It is relevant that the time distance between  $t$  and  $t-1$  is short. In daily or intra-day settings,  $S_t$  typically consists of genuinely new microstructure signals or unanticipated shocks that are inaccessible at time  $t-1$ , which makes them exogenous to the forecast process. In contrast, when considering lower frequencies, such as monthly, the information environment becomes more dispersed. In such a case, forecast updates are released throughout the month, and policymakers or market participants may directly base their actions on these forecasts. As a result, the chance of feedback between forecasts and stabilization signals increases, making this assumption less credible.

Consequently, and disregarding the expected feedback from the possible policy effect, the prediction augmented by  $S_t$  is  $E(X_t|F_{t-1}, S_t)$ . Let

$$D_t = E(X_t|F_{t-1}, S_t) - m_t \quad \text{and} \quad Z_t = \frac{D_t}{\sigma_t}$$

be the absolute and relative differences between the augmented and core predictions. Assume that  $(U_t^*, Z_t)$  is *iid*, with

$$\begin{pmatrix} U_t^* \\ Z_t \end{pmatrix} \sim N \left\{ 0, \begin{pmatrix} 1 & \rho\sigma_Z \\ \rho\sigma_Z & \sigma_Z^2 \end{pmatrix} \right\}. \quad (1)$$

When  $|D_t|$  is large enough, e.g., it exceeds one standard deviation, the policy makers decide to act in such a way that the actual economic or financial variable becomes

$$X_t = X_t^* + \alpha D_t 1_{Z_t > \nu} + \beta D_t 1_{Z_t < \kappa},$$

where  $\kappa \leq 0 \leq \nu$ ;  $1_{Z_t > \nu}$  is an indicator of event  $\{Z_t > \nu\}$ , that is  $1_{Z_t > \nu} = 1$  if  $Z_t > \nu$  and 0 otherwise; similarly,  $1_{Z_t < \kappa}$  is an indicator of event  $\{Z_t < \kappa\}$ . If the policymaker perceives  $D_t$  as an unexpected shock on the economic or financial variable that needs to be attenuated, the values of  $\alpha$  and  $\beta$  are negative. Note that, by scaling the parameters  $\alpha$ ,  $\beta$ ,  $\nu$  and  $\kappa$ , it is not restrictive to assume  $\sigma_Z = 1$  in (1). The conditional prediction error based on the past is then

$$U_t := \frac{X_t - m_t}{\sigma_t} = U_t^* + \alpha Z_t 1_{Z_t > \nu} + \beta Z_t 1_{Z_t < \kappa} \quad (2)$$

The distribution of  $U_t$  is the *normalized weighted skew-normal* (see Charemza, Díaz and Makarova [15]), where the policy decisions on the distribution of forecast errors are reflected by  $\alpha$  and  $\beta$  in (2), representing the policy's strength. The parameters  $\alpha$  and  $\beta$  show the effects of, respectively, the downward and upward correcting action undertaken when the real-time forecast signal  $Z_t$  exceeds the upper threshold  $\nu$ , or is below the lower threshold  $\kappa$ .

As in the empirical part of the paper, we are adapting the weighted skew-normal distribution to sovereign bond yields, the parameters  $\alpha$  and  $\beta$  have a slightly different interpretation than that in Charemza, Díaz and Makarova [15] model, originally designed for analyzing inflation forecast uncertainty. In our case, the parameter  $\alpha$  captures the adjustment when the observed volatility of bond yields is higher than warranted by economic fundamentals, implying that policy acts to reduce excessive uncertainty. In contrast,  $\beta$  would correspond to a correction when observed volatility is lower than underlying risks might suggest.



However, in the context of bond market stabilization, the objective of policy is typically not to increase uncertainty, even if it appears understated. Therefore, it is reasonable to set  $\beta$  equal to zero or close to zero, reflecting that interventions are designed to reduce excessive uncertainty rather than actively introduce additional uncertainty. Under this specification,  $\alpha$  alone represents the adjustment of forecast uncertainty, and a negative  $\alpha$  would indicate a rational stabilization action. Consequently, testing for  $\alpha = \beta = 0$  is of interest, as rejecting this hypothesis confirms the existence of policy footprints in the distribution of forecast errors.

The parameter  $\varrho$ , the correlation coefficient between  $U_t^*$  and  $Z_t$  captures the proportion of the variation in forecast uncertainty that can be statistically attributed to the external information signal. A higher  $\varrho$  indicates that real-time information influences the distribution of forecast errors stronger and is more relevant for stabilization policy, compared to the case when  $\varrho$  is close to zero.

**Fig 1** demonstrates how economic policies might affect the distribution of ex-post forecast errors. It plots variance  $\text{Var}(U_t)$  and  $E(U_t^2)$ , the second ordinary (non-centered) moment of  $U_t$ , against  $-\alpha$  (for better visibility we use  $-\alpha$  rather than  $\alpha$ ). Unlike variance,  $E(U_t^2)$  also captures the first-moment effects of the policy. The upper panel of **Fig 1** shows the result for  $\varrho = 0.4$ , that is, for the case where the influence of real-time information on the distribution of forecast errors is reasonably weak, and the lower panel for  $\varrho = 0.8$ , where such influence is much stronger. Other parameters of the distributions are  $\nu = -\kappa = 1$ , and either  $\beta = 0$  or  $\beta = \alpha$  for  $0 < -\alpha < 2$ . If the strengths of the policies are identical regardless of their direction ( $\alpha = \beta < 0$ ) then  $E(U_t) = 0$  and  $\text{Var}(U_t) = E(U_t^2)$ . If there are no policy effects, then  $\alpha = \beta = 0$  and distribution of  $U_t$  collapses to the standard normal, which is marked by the dashed horizontal line drawn at 1. The plot shows that the policies, if effective, reduce forecast uncertainty compared to that of the ‘no action’ case ( $\alpha = \beta = 0$ ) by lowering its variance below 1, if  $-\alpha < 2\varrho$ . The maximum stabilization effect is achieved at  $-\alpha = \varrho$ . However, if the policies are too strong, that is, if  $-\alpha > 2\varrho$  they become counter-effective, as the variance exceeds that of the ‘no action’ policy. While comparing the upper and lower panels of **Fig 1**, it can be noticed that with the increase in  $\varrho$ , that is, when the influence of real-time information on the distribution of forecast uncertainty becomes stronger, the uncertainty reduction effect increases and is evident for a wider range of  $\alpha$ . The blue and green lines in **Fig 1** illustrate the extreme asymmetric case where  $\alpha < 0$ ,  $\beta = 0$ , in which the first moment of the series is also affected, as  $\text{Var}(U_t) \neq E(U_t^2)$ . Thus, **Fig 1** explains the relevance of testing for  $\alpha = \beta = 0$ .

In the next section, we consider a general time series model with the WSN innovations and propose the PELM test for the absence of policy interventions; that is, for  $\alpha = \beta = 0$ .

### 3 The model and testing

#### 3.1 Introducing the PELM test

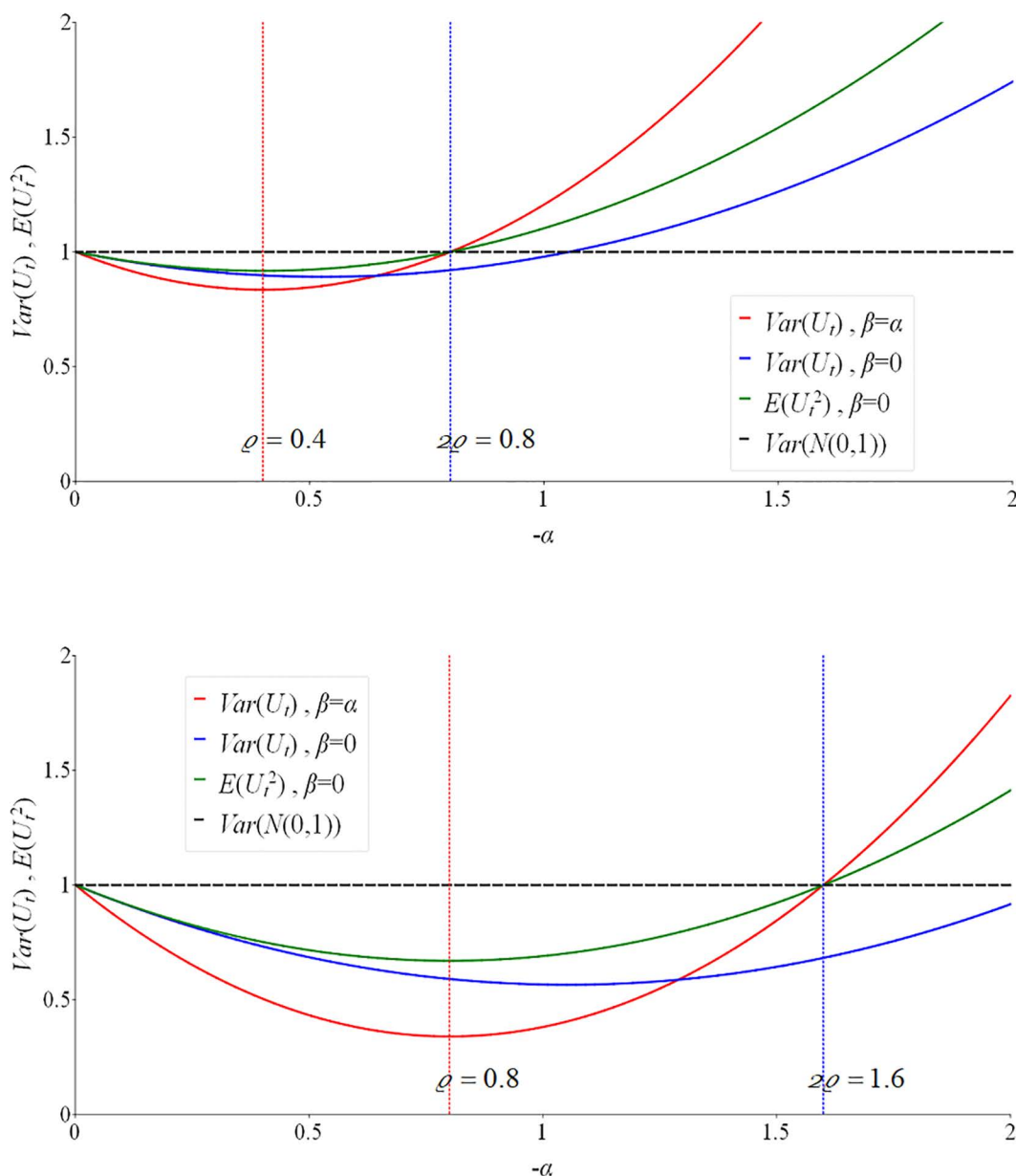
For the purpose of testing, we consider the stationary process  $(X_t)$  satisfying:

$$X_t = m(X_{t-1}, \dots, X_{t-p}; \theta_{0m}) + \sigma(X_{t-1}, \dots, X_{t-q}; \theta_{0\sigma})U_t, \quad (3)$$

where  $m(\cdot)$  and  $\sigma(\cdot)$  are functions valued in  $\mathbb{R}$  and  $\mathbb{R}^+$  respectively, and where  $\theta_{0m}$  and  $\theta_{0\sigma}$  are vectors of unknown parameters belonging to  $\Theta_m \subset \mathbb{R}^{s_m}$  and  $\Theta_\sigma \subset \mathbb{R}^{s_\sigma}$ . We do not include any exogenous variables in  $m(\cdot)$  and  $\sigma(\cdot)$ . However, the model with such variables is discussed in Supplement 1 Sections S1 and S5. It is assumed that  $U_t$  is independent of  $\{X_u, u < t\}$  and that  $(U_t)$  is iid WSN distributed with the parameters:

$$\theta_{0u} = (\alpha_0, \beta_0, \nu_0, \kappa_0, \varrho_0)' \in \Theta_u \subset (-\infty, 0]^2 \times [0, \infty) \times (-\infty, 0] \times (-1, 1). \quad (4)$$

Note that, because the WSN is not centered, the functions  $m(\cdot; \theta_{0m})$  and  $\sigma(\cdot; \theta_{0\sigma})$  do not correspond to the first conditional moments of  $X_t$  (except when  $\alpha_0 = \beta_0 = 0$ ). This is why parameters  $\theta_{0m}$  and  $\theta_{0\sigma}$  cannot be estimated by the quasi-maximum likelihood, but must be estimated by the maximum likelihood (ML) together with  $\theta_{0u}$ . The proof of asymptotic



**Fig 1. Stabilization policy effects in variance of forecast errors.** Legend: the upper panel shows the results for  $\rho = 0.4$  and the lower panel for  $\rho = 0.8$ .

<https://doi.org/10.1371/journal.pone.0336495.g001>

consistency, generalization that allows for adding exogenous variables in (3), and other properties of the ML estimator of  $\vartheta_0 = (\theta'_{0m}, \theta'_{0\sigma}, \theta'_{0u})'$ , are provided in Supplement 1, Section S1. Note that our testing approach, as described further, allows us to circumvent the problem of the ML estimation, which is numerically awkward. Nevertheless, the properties of the ML estimator are needed for the derivation of the asymptotic properties of the test.

As discussed in Section 2, we are interested in testing the hypothesis:

$$H_0 : \alpha_0 = \beta_0 = 0 \text{ against } H_1 : \alpha_0 < 0 \text{ or } \beta_0 < 0, \quad (5)$$

where  $\alpha_0, \beta_0$  are the parameters of the WSN distribution defined in (2) and (4).

Note that, under  $H_0$ , the parameter  $\pi_0 = (\nu_0, \kappa_0, \varrho_0)'$  of the WSN distribution is not identified as the distribution of  $U_t$  collapses to the standard normal; see definition of  $U_t$  in (2), and Lemma 1 on the properties of the ML estimator of  $\vartheta_0 = (\theta'_{0m}, \theta'_{0\sigma}, \theta'_{0u})'$  does not apply (see Supplement 1 Section S1). Denote by  $\Pi$  the parameter space for  $\pi_0$ . Let  $\theta_0 = (\theta'_{0m}, \theta'_{0\sigma})' \in \Theta_m \times \Theta_\sigma$  be the vector of the parameters that remain unknown under the null. Under  $H_0$ , the WSN density  $f_{\theta_{0u}}$  reduces to the density of the standard normal distribution. The maximum likelihood estimator of  $\theta$  is then defined by

$$\hat{\theta} = \arg \max_{\theta \in \Theta_m \times \Theta_\sigma} \tilde{Q}_n^*(\theta), \quad \tilde{Q}_n^*(\theta) = \frac{1}{n} \sum_{t=1}^n \tilde{\ell}_t^*(\theta), \quad (6)$$

where

$$\tilde{\ell}_t^*(\theta) = \log \left\{ \frac{1}{\tilde{\sigma}_t(\theta_\sigma)} \phi \left( \frac{X_t - \tilde{m}_t(\theta_m)}{\tilde{\sigma}_t(\theta_\sigma)} \right) \right\} = -\frac{1}{2} \left\{ \left( \frac{X_t - \tilde{m}_t(\theta_m)}{\tilde{\sigma}_t(\theta_\sigma)} \right)^2 + \log \tilde{\sigma}_t^2(\theta_\sigma) \right\}. \quad (7)$$

Thus  $\hat{\theta} = (\hat{\theta}'_m, \hat{\theta}'_\sigma)'$  is the Gaussian MLE of  $\theta_0$  under  $H_0$ . The strong consistency and asymptotic normality of  $\hat{\theta}$  defined by (6) – (7) are given by **Proposition 1** in Supplement 1 Section S2. To distinguish formulas that appear in the Supplement 1 and are referred to in the main text, we use the prefix 'S' (S1, S2, etc.).

In view of Proposition 1, and in particular (S5), which shows that the Fisher information matrix  $I_\pi$  is invertible when  $\varrho_0 \neq 0$  and  $(\nu_0, \kappa_0) \neq (0, 0)$ , we now restrict the possible values  $\pi$  to a compact set not including zeros, of the form

$$\Pi_0 = [\underline{\nu}, \bar{\nu}] \times [\underline{\kappa}, \bar{\kappa}] \times \{[-\bar{\varrho}, -\underline{\varrho}] \cup [\underline{\varrho}, \bar{\varrho}]\},$$

where  $0 < \underline{\nu} < \bar{\nu}$ ,  $\underline{\kappa} < \bar{\kappa} < 0$  and  $0 < \underline{\varrho} < \bar{\varrho}$ . Now, consider the PELM test statistic

$$\text{PELM}_{n,\pi} = n \frac{\partial}{\partial \underline{\theta}'_u} \tilde{Q}_n \left\{ \hat{\vartheta}(\pi) \right\} \hat{I}^{\underline{\theta}_u} \frac{\partial}{\partial \underline{\theta}'_u} \tilde{Q}_n \left\{ \hat{\vartheta}(\pi) \right\}, \quad (8)$$

where  $\hat{\vartheta}(\pi) = (\hat{\theta}', \theta_u(\pi)')' = (\hat{\theta}', 0, 0, \pi')'$ ,  $\underline{\theta}_u = (\alpha, \beta)'$  and  $\hat{I}^{\underline{\theta}_u}$  is the lower-right block of size  $2 \times 2$  of the matrix  $\hat{I}_\pi^{-1}$ , which exists for  $n$  large enough when  $\pi \in \Pi_0$ , by (S5). The following Proposition 2 provides details for computing the derivatives in (8), see (S10) – (S11), and summarizes the properties of the  $\text{PELM}_{n,\pi}$  test statistic.

**Proposition 2** Under  $H_0$  and the other conditions of Proposition 1, for each  $\pi \in \Pi_0$ , we have

$$\text{PELM}_{n,\pi} \xrightarrow{d} \chi_2^2, \quad \text{as } n \rightarrow \infty,$$

where  $\chi_2^2$  denotes the chi-square distribution with 2 degrees of freedom.

The proof of Proposition 2 is provided in Supplement 1 Section S3. This Proposition facilitates testing substantially, as the p-values and critical values of the  $\chi_2^2$  distribution can be used, if the statistical sample is large enough (see Section 4 for more detailed results). However, the drawback of using the  $\text{PELM}_{n,\pi}$  statistic for testing is that it relies on the somewhat arbitrary choice of  $\pi \in \Pi_0$ . To solve a similar problem Davies [21] and Davies [22] suggested using the supremum test statistic

$$\text{PELM}_n = \sup_{\pi \in \Pi_0} \text{PELM}_{n,\pi}. \quad (9)$$



Supremum Lagrange Multiplier tests are often used for testing nonlinearities in time series; see, e.g., Christou and Fokianos [23]; Francq, Horváth and Zakoïan [24]; Kirch and Kamgaing [25]. Weighted averages of these and other test statistics have also been proposed in the literature (see, e.g., Hansen [26]). Under  $H_0$  and the other conditions of Proposition 1,  $PELM_n$  converges to a limit distribution, which justifies approximating the p-values in finite samples by simulation. The full set of properties of  $PELM_n$  test statistics defined by (9) are formulated and proved in **Proposition 3** in Supplement 1 Section S4.

### 3.2 The PELM test in ARMA-GARCH model

In the empirical part of the paper, in Section 5, we apply the test for one of the most popular dynamic volatility models, the ARMA-GARCH model. In practice, more advanced models with time-varying volatility are often used; see, e.g., Huang, Chiang, and Lin [27], Zou, Xu, and Chen [28], and others, for recent important applications and developments. However, we focus on the ARMA-GARCH model due to its popularity among practitioners and forecasting accuracy, which, in its original form or variations, has frequently proven satisfactory to practitioners (see, for example, Clark and Ravazzolo [29]; Naimy et al. [30]; Wu, Kuang, and Xu et al. [31]).

We also considered other time series models that satisfy (3). In particular, we analyzed the double autoregressive models (see Ling [32], Ling and Li [33]) and, for financial applications, Cai, Montes-Rojas, and Olmo [34]. Although we conducted extensive power and size distortion analysis for these models, we have decided not to use them here, as their interpretation in the context of the sovereign bond market analysis is unclear.

For simplicity, we focus on the first-order ARMA-GARCH model, that is ARMA(1,1)-GARCH(1,1). However, the straightforward generalization holds for higher-order lags in the AR part of the model, which will be utilized in the empirical example in Section 5. Here we discuss the ARMA(1,1)-GARCH(1,1) model given in the form:

$$\begin{cases} X_t = a_0 X_{t-1} + c_0 + \epsilon_t - b_0 \epsilon_{t-1} \\ \epsilon_t = \sigma_t^*(\theta_{0\sigma}) U_t, \quad \sigma_t^{*2}(\theta_{0\sigma}) = \omega_0 + \gamma_0 \epsilon_{t-1}^2 + \delta_0 \sigma_{t-1}^{*2}(\theta_{0\sigma}), \end{cases} \quad (10)$$

where  $\theta_{0m} = (a_0, b_0, c_0)'$  and  $\theta_{0\sigma} = (\omega_0, \gamma_0, \delta_0)'$ . First, note that (10) is not exactly of the form (3) because

$$\sigma_t^{*2}(\theta_{0\sigma}) = \omega_0 + \gamma_0 \left\{ \sum_{i=0}^{\infty} b_0^i (X_{t-i-1} - a_0 X_{t-i-2} - c_0) \right\}^2 + \delta_0 \sigma_{t-1}^{*2}(\theta_{0\sigma}),$$

with  $\theta_0 = (\theta'_{0m}, \theta'_{0\sigma})'$ , does not depend only on  $\theta_\sigma$ , but also on the mean parameter  $\theta_m$  via the approximation of the linear innovation  $\tilde{\epsilon}_t(\theta_m)$ , which is recursively defined by

$$\tilde{\epsilon}_t(\theta_m) = X_t - aX_{t-1} - c + b\tilde{\epsilon}_{t-1}(\theta_m), \quad t = 2, \dots$$

for some fixed initial value  $\tilde{\epsilon}_1(\theta_m) = \tilde{\epsilon}_1$ . However, the results of Propositions 1–3 remain valid, provided that  $\theta_\sigma$  and  $\hat{\theta}_\sigma$  are replaced by corresponding recursive estimators, which account for this dependency, as shown in (S18) – (S20).

Practical implementation of the two PELM tests in the ARMA-GARCH model can be done in three steps described below (additional explanations and formulas are given in Supplement 1):

1. Estimate the ARMA-GARCH model under the null of (5) by the quasi-maximum likelihood estimate (QMLE) as in (6) – (7).
2. For the  $PELM_{n,\pi}$  test, define the parameters  $\pi = (\nu, \kappa, \varrho)'$ . If the residuals are standardized, the natural settings for  $\nu$  and  $\kappa$  are  $\pm$  one standard deviation of the distribution of the ex-post forecast errors. It would correspond to the intuitive presumption that the forecast errors exceeding one standard deviation can be regarded

as 'large', and hence a stabilization policy should be applied. The setting for  $\varrho$  might depend on some external knowledge; the higher  $\varrho$ , the greater is the relevance of the real-time information for explaining forecast uncertainty.

3. For the PELM<sub>n</sub> test, set the upper and lower limits for  $\nu$ ,  $\kappa$  and  $\varrho$  that is, intervals  $[\underline{\nu}, \bar{\nu}]$ ,  $[\underline{\kappa}, \bar{\kappa}]$  and  $[\underline{\varrho}, \bar{\varrho}]$ . The wider intervals cover more combinations of the parameters and hence are safer to apply; on the other hand, narrower intervals increase the power of the test if they are properly set. Also, decide on the number of drawings from these intervals, as the large number of drawings might slow down computations substantially, while a too small number of drawings would distort the size of the test (see Section 4 further on for a discussion of numerical results which help with this decision).
4. Denote the cumulative function and probability density function of the standard normal distribution by  $\Phi(\cdot)$  and  $\phi(\cdot)$  respectively. Let also

$$\tilde{m}_t(\theta_m) = m(X_{t-1}, \dots, X_1, x_0, x_{-1}, \dots; \theta_m) \text{ and } \tilde{\sigma}_t(\theta_\sigma) = \sigma(X_{t-1}, \dots, X_1, x_0, x_{-1}, \dots; \theta_\sigma)$$

are empirical approximations of  $m(\cdot; \theta)$  and  $\sigma(\cdot; \theta)$  in the model specification (3). The derivatives for calculating PELM<sub>n,π</sub> appearing in (8) can be computed (as shown by (S10) – (S11)) as

$$\frac{\partial \tilde{Q}_n\{\hat{\vartheta}(\pi)\}}{\partial \theta_u} = \frac{1}{n} \sum_{t=1}^n g_\pi(\tilde{U}_t), \text{ where}$$

$$\tilde{U}_t = \frac{X_t - \tilde{m}_t(\hat{\theta}_m)}{\tilde{\sigma}_t(\hat{\theta}_\sigma)}, \quad g_\pi(x) = (g_{\alpha,\pi}(x), g_{\beta,\pi}(x))',$$

and

$$\begin{aligned} g_{\alpha,\pi}(x) &= (x^2 - 1)\varrho\Phi\left(\frac{\varrho x - \nu}{\sqrt{1-\varrho^2}}\right) + \phi\left(\frac{\varrho x - \nu}{\sqrt{1-\varrho^2}}\right) \left\{ \frac{(1-\varrho^2)x - \nu\varrho}{\sqrt{1-\varrho^2}} \right\} \\ g_{\beta,\pi}(x) &= (x^2 - 1)\varrho\Phi\left(\frac{-\varrho x + \kappa}{\sqrt{1-\varrho^2}}\right) + \phi\left(\frac{-\varrho x + \kappa}{\sqrt{1-\varrho^2}}\right) \left\{ \frac{\kappa\varrho - (1-\varrho^2)x}{\sqrt{1-\varrho^2}} \right\}, \end{aligned} \quad (11)$$

Matrix  $\hat{l}_{\theta_u}^{\varrho_u}$  in (8) is the lower-right block of size  $2 \times 2$  of the inverse estimated Fisher information matrix,  $\hat{l}_\pi^{-1}$ , with the following components (see (S3), (S6) and (S18) – (S20)):

$$\hat{l}_\pi = \begin{pmatrix} \hat{l}_\theta & \hat{l}_{\theta, \theta_u} \\ \hat{l}_{\theta_u, \theta} & \hat{l}_{\theta_u} \end{pmatrix},$$

where  $\hat{l}_{\theta_u} = E g_\pi(U_t) g'_\pi(U_t)$ , with  $g_\pi(\cdot)$  defined by (11); and

$$\hat{l}_\theta = \frac{1}{n} \sum_{t=1}^n \frac{1}{\tilde{\sigma}_t^2(\hat{\theta})} \frac{\partial \tilde{\epsilon}_t(\hat{\theta}_m)}{\partial \theta} \frac{\partial \tilde{\epsilon}_t(\hat{\theta}_m)}{\partial \theta'} + \frac{1}{2} \frac{1}{n} \sum_{t=1}^n \frac{1}{\tilde{\sigma}_t^4(\hat{\theta})} \frac{\partial \tilde{\sigma}_t^2(\hat{\theta})}{\partial \theta} \frac{\partial \tilde{\sigma}_t^2(\hat{\theta})}{\partial \theta'} \quad (12)$$

and

$$\hat{l}_{\theta, \theta_u} = \frac{-1}{n} \sum_{t=1}^n \frac{1}{\tilde{\sigma}_t(\hat{\theta})} \frac{\partial \tilde{\epsilon}_t(\hat{\theta}_m)}{\partial \theta} E U_1 g'_\pi(U_1) + \frac{1}{2n} \sum_{t=1}^n \frac{1}{\tilde{\sigma}_t^2(\hat{\theta})} \frac{\partial \tilde{\sigma}_t^2(\hat{\theta})}{\partial \theta} E (U_1^2 - 1) g'_\pi(U_1), \quad (13)$$

where the linear innovations  $\tilde{\epsilon}_t(\theta_m)$  being recursively defined by:

$$\tilde{\epsilon}_t(\theta_m) = X_t - aX_{t-1} - c + b\tilde{\epsilon}_{t-1}(\theta_m), \quad t = 2, \dots$$

In (12)-(13) the derivatives are computed recursively (see (S20)) as

$$\frac{\partial \tilde{\epsilon}_t(\theta_m)}{\partial \theta} = \begin{pmatrix} -X_{t-1} \\ \tilde{\epsilon}_{t-1}(\theta_m) \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + b \frac{\partial \tilde{\epsilon}_{t-1}(\theta_m)}{\partial \theta}, \quad \frac{\partial \tilde{\sigma}_t^2(\theta)}{\partial \theta} = \begin{pmatrix} 2\gamma \epsilon_{t-1}(\theta_m) \frac{\partial \tilde{\epsilon}_{t-1}(\theta_m)}{\partial \theta} \\ 1 \\ \epsilon_{t-1}^2(\theta_m) \\ \tilde{\sigma}_{t-1}^2(\theta) \end{pmatrix} + \delta \frac{\partial \tilde{\sigma}_{t-1}^2(\theta)}{\partial \theta}.$$

For  $PELM_n$  test use (9) to obtain a supremum as described in 3 above.

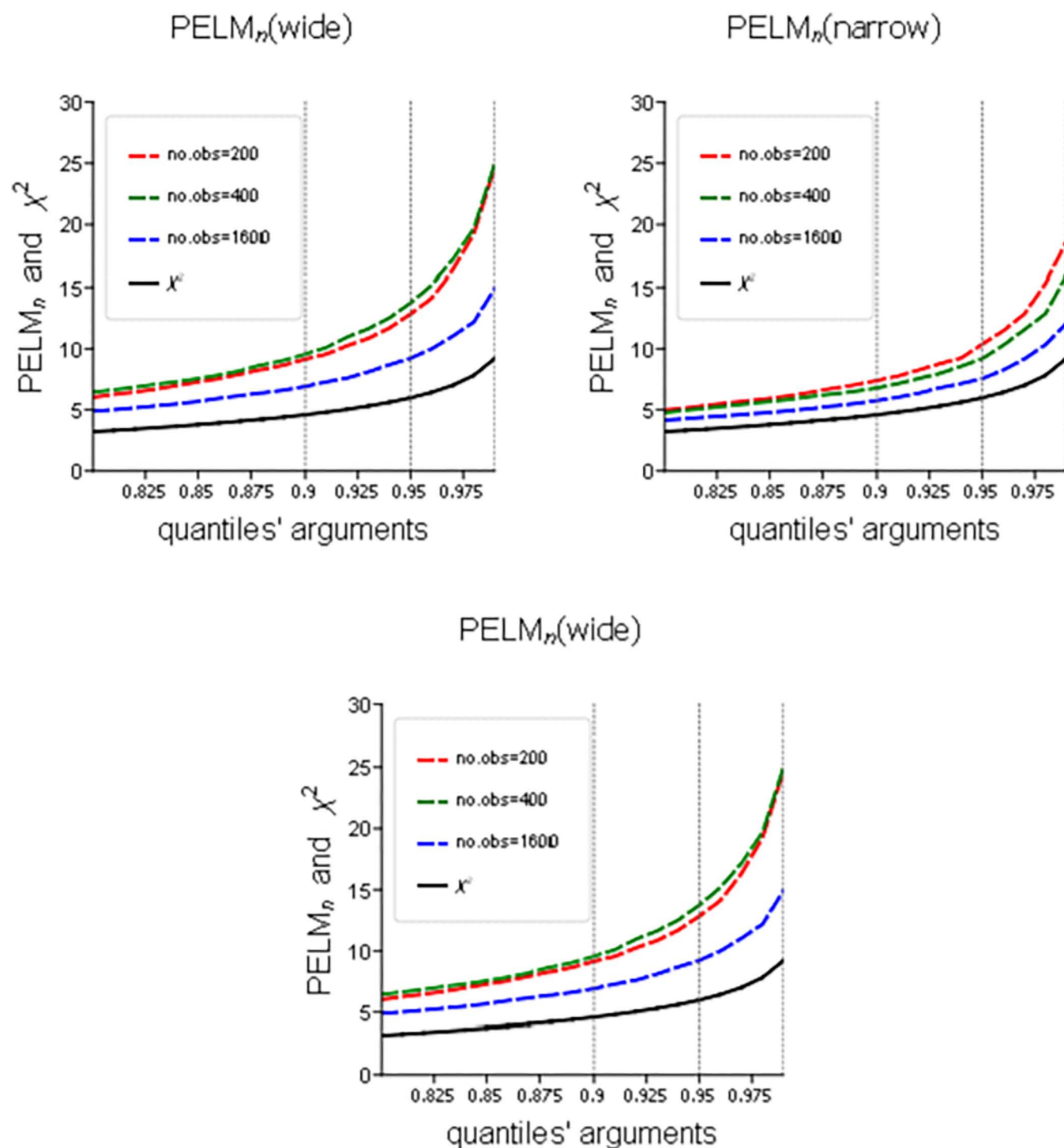
The codes in Aptech Gauss for computing the  $PELM_{n,\pi}$  and  $PELM_n$  are available on GitHub <https://github.com/Svetlana2201/Policy-effects>.

#### 4 Size distortion and power of the tests

This section gives the selected numerical results on the evaluation of size distortion and power of the tests proposed in Section 3. For the analysis of size distortion, we present results obtained for the ARMA-GARCH model (10) with the parameters set as  $(a_0, b_0, c_0, \omega_0, \gamma_0, \delta_0) = (0.5, 0.2, 1, 1, 0.5, 0.3)$  (results for other ARMA-GARCH settings are available upon request; the outcomes of the size distortion analysis are not substantially affected by changes in ARMA-GARCH specification). These results are obtained by simulation. For the  $PELM_{n,\pi}$  test, we set  $\alpha_0 = \beta_0 = 0$  and  $\pi_0 = (\nu_0, \kappa_0, \varrho_0)' = (1, -1, 0.8)'$ . We apply the  $PELM_n$  test in two variants, named respectively  $PELM_n(\text{narrow})$  and  $PELM_n(\text{wide})$ , which differ from each other by the width of the intervals we set for  $\nu_0$  and  $\kappa_0$ . For the  $PELM_n(\text{narrow})$  test, we draw  $\nu_0$  from the interval  $[0.9, 1.1]$ , and  $\kappa_0$  from the interval  $[-1.1, -0.9]$ . For the  $PELM_n(\text{wide})$  test, the drawings are from intervals of  $[0.75, 1.25]$  and  $[-1.25, -0.75]$  respectively. For all tests, we keep  $\varrho_0 = 0.8$ .

As in (1)-(2), for  $PELM_{n,\pi}$ ,  $\nu_0$  and  $\kappa_0$  are equal to one standard deviation of the distribution of the real-time forecast signals. This choice has been motivated by the common presumption that shocks of less than one standard deviation can be regarded as 'small', and, hence, likely to be neglected by policymakers. We choose a reasonably high value of the parameter  $\varrho_0$ , as it corresponds to assuming that the real-time information is highly relevant for explaining forecast uncertainty (see Section 2). The supremum in (9) is computed using 40 independent draws from these intervals. The number of replications is equal to 10,000 for each run. For the  $PELM_n$  test, the practical problem is to decide on the number of drawings made from the intervals  $[\underline{\nu}, \bar{\nu}]$ ,  $[\underline{\kappa}, \bar{\kappa}]$  and  $[\underline{\varrho}, \bar{\varrho}]$ , as, when the number of drawings increases, the computation time increases substantially. Consequently, we conducted some additional simulation experiments and found that increasing this number above 40 does not alter the estimated size of the test in any systematic way. This conclusion stems from the results of the randomness test of the quantiles of the  $PELM_{n,\pi}$  statistic obtained for the case where the number of drawings is gradually increased (see Supplement 2 Part 1). They show that, when the number of drawings exceeds 40, the null hypothesis that the estimated size of the test changes randomly rather than systematically cannot be rejected. Also, the rate of the standard deviations to the means of the computed quantiles of the  $PELM_n$  test statistics obtained for a different number of drawings are in the range of 0.5–3.0%, which is quite small. Hence, increasing the number of drawings above 40 does not add any meaningful information on the true test size. Therefore, to speed up computations, we have decided to use 40 as the number of such drawings.

Fig 2 shows the right tails of the simulated inverted CDFs of the distributions of the  $PELM_{n,\pi}$  and  $PELM_n$  statistics that are computed under the null for different sample sizes, with 10,000 simulations in each case. It also shows the right tail of the  $\chi_2^2$  CDF with the quantiles  $q_{0.8}$ ,  $q_{0.9}$  and  $q_{0.95}$  marked by vertical dotted lines. It demonstrates the downward bias if the  $\chi_2^2$  p-values or critical values are used instead of the true but unknown ones. Such bias is not substantial for the  $PELM_{n,\pi}$  test, particularly for the quantiles smaller than 0.95, which is the conventional level of significance. However, it is slightly bigger for  $PELM_n(\text{narrow})$  and much bigger for  $PELM_n(\text{wide})$ .



**Fig 2.** Size distortions in PELM<sub>n,π</sub>, PELM<sub>n</sub>(narrow), and PELM<sub>n</sub>(wide) test statistics (right tails of the simulated inverted CDFs).

<https://doi.org/10.1371/journal.pone.0336495.g002>

For a quick evaluation of the significance, we constructed [Table 1](#), with the most frequently used critical values that correspond to selected quantiles. For comparison, the corresponding  $\chi^2_2$  critical quantiles are also included. This table is essentially an excerpt from [Fig 2](#), but its traditional critical values form makes it straightforward to use.

For a more accurate approximation of the p-values, we run Monte Carlo simulations for different sample sizes and settings for  $\pi$ , and then interpolate. We based our approximations on the estimation of 35 different data-generating processes with different sample sizes  $n$  ( $n = 200, 400, 600, 1000, 1600, 3500$ ), and intervals for  $\pi$  set as for  $PELM_n(\text{narrow})$  and  $PELM_n(\text{wide})$  above. For the given value of the PELM statistic, we interpolated between adjacent sample sizes and accounted for some (very infrequent and small in magnitude) non-monotonicity by smoothing using the B  zier spline. As a result, we obtained impulse surfaces that produce the p-values with reasonable accuracy.

We illustrate the evolution of power of the tests for different sample sizes. The settings for simulations are the same as for the analysis of size distortion, except for removing the null hypothesis setting of  $\alpha_0 = \beta_0 = 0$ . [Fig 3](#) shows the probabilities of rejecting the null hypothesis for all three tests, that is  $PELM_{n,\pi}$ ,  $PELM_n(\text{narrow})$  and  $PELM_n(\text{wide})$ , under the alternative hypothesis, where  $\alpha_0$  is gradually decreasing from zero to  $-1$  and  $\beta_0 = 0$ . It demonstrates that the power of the tests is monotonously increasing with the increase in the strength of the economic policy shown by  $-\alpha_0$ .

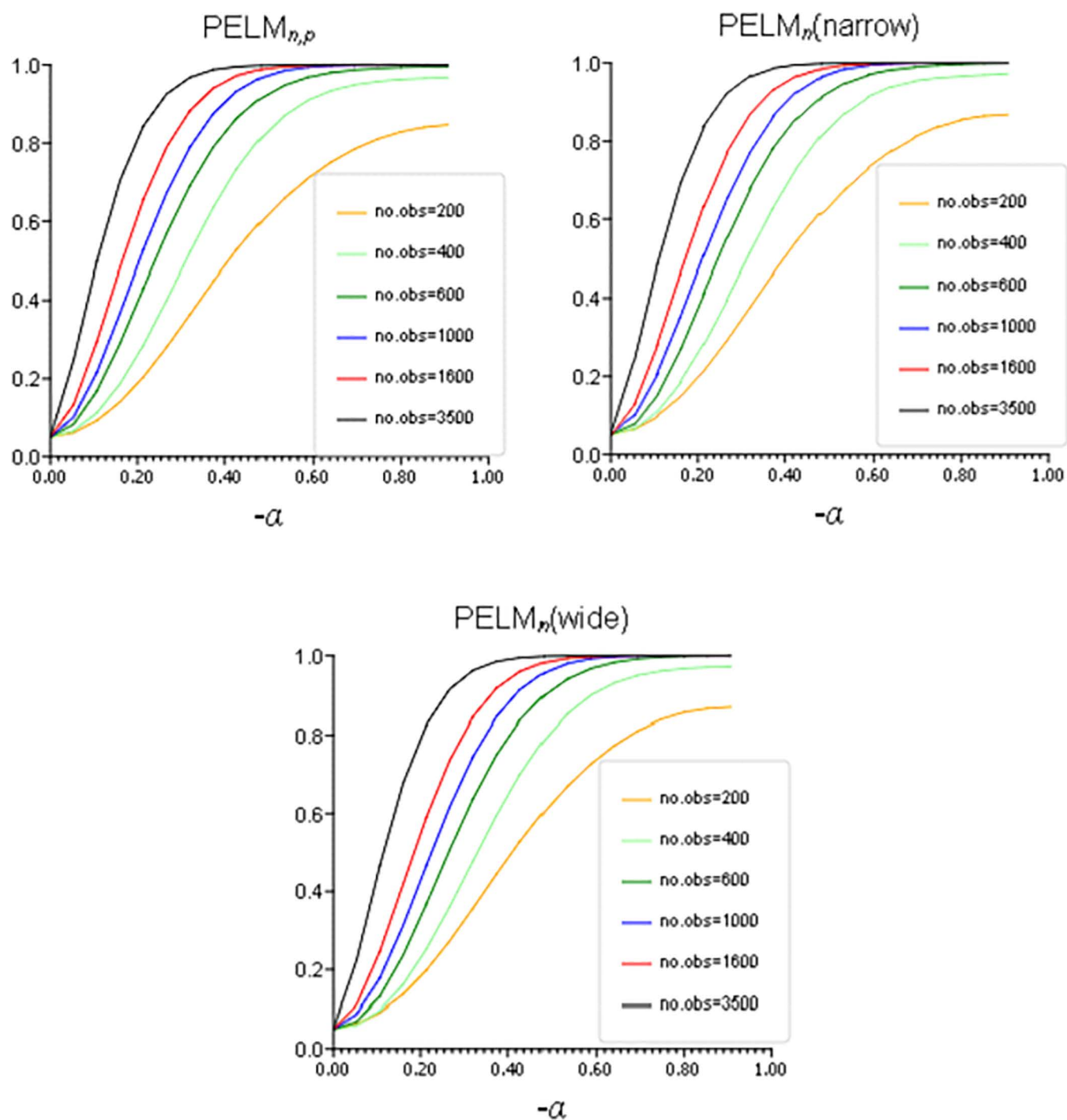
Results of the more detailed simulation analysis (not presented here and available on request) show the power of the  $PELM_{n,\pi}$  and  $PELM_n$  tests do not increase monotonously with the increase in  $-\alpha_0$  and  $-\beta_0$ , and so with the increase in the policy strength for all values of  $\alpha_0, \beta_0 \in (-\infty, 0)$ . The intuition of this is demonstrated by [Fig 1](#) in Section 1 above. As the variance of  $U_t$  initially decreases with the decrease in  $\alpha_0$  and  $\beta_0$ , and then increases, it is prudent to restrict the range of the admissible values of  $\alpha_0$  and  $\beta_0$  such as  $-\alpha_0 \in [0, -\alpha^*]$  and  $-\beta_0 \in [0, -\beta^*]$ , where:

$$(\alpha^*, \beta^*) = \arg \min_{\alpha_0, \beta_0} \text{Var}(U_t).$$

**Table 1. Simulated quantiles for  $PELM_{n,\pi}$ ,  $PELM_n(\text{narrow})$  and  $PELM_n(\text{wide})$  test statistics.**

Probability	$\chi^2_2$ quantiles	Sample size		
		200	400	1600
		PELM $_{n,\pi}$ quantiles		
0.800	3.219	4.280	4.137	3.805
0.850	3.794	5.073	4.944	4.391
0.900	4.605	6.308	5.942	5.247
0.950	5.991	8.550	7.971	6.993
0.990	9.210	15.570	13.150	10.980
		PELM $_n$ (narrow) quantiles		
0.800	3.219	4.980	4.776	4.156
0.850	3.794	5.935	5.684	4.782
0.900	4.605	7.401	6.790	5.756
0.950	5.991	10.360	9.182	7.551
0.990	9.210	18.590	15.960	12.080
		PELM $_n$ (wide) quantiles		
0.800	3.219	6.029	6.436	4.902
0.850	3.794	7.284	7.561	5.689
0.900	4.605	9.146	9.552	6.921
0.950	5.991	12.830	13.740	9.225
0.990	9.210	24.710	25.010	14.920

<https://doi.org/10.1371/journal.pone.0336495.t001>



**Fig 3. Power of  $PELM_{n,p}$ ,  $PELM_n(\text{narrow})$  and  $PELM_n(\text{wide})$  for different sample sizes.**

<https://doi.org/10.1371/journal.pone.0336495.g003>



It can be shown that  $-\alpha^* = -\beta^* = \varrho_0$  for the case where  $\nu_0 = -\kappa_0$ . However, if there is a lack of symmetry between  $\nu_0$  and  $\kappa_0$ , so  $\nu_0 \neq -\kappa_0$ , the relationship can be more complex.

Further results show that the power of the test is reasonably robust to changes of the threshold parameters  $\nu_0$  and  $\kappa_0$ . Analogously, if  $\varrho_0$  rises above 0.8, the power is increased accordingly.

## 5 Empirical application: Can the PELM tests help in predicting fiscal improvement?

If our approach is valid, policy imprints on the bond market should be detectable by the PELM test when applied to time series of government bond yields. If these policies were effective, this should also be reflected in positive movements in relevant economic aggregates, such as reductions in budget deficits.

In this context, we formulate the following working hypothesis:

*If a global shock affects the financial stability of some countries, the negative consequences of such a shock are more likely to diminish in countries where imprints of economic policies have been detected in the bond market than in countries where such imprints are absent.*

To test this hypothesis, we compute the  $PELM_{n,\pi}$  and  $PELM_n$  statistics for time series of data on government bond yields for a broad group of countries. We then divide these countries into two groups: Group A, where the PELM test statistics are statistically significant at a reasonable level of significance (i.e., countries where imprints of policies are confirmed by rejecting the null hypothesis in (5)), and Group B, where the test statistics are not significant. To enable cross-country comparison, we focus on possible consequences of a global shock that might affect countries at approximately the same time, albeit in different ways. As the effects of policy might occur after some time, we define the *aftershock year* as the year in which the results of such a shock are likely to become evident for all selected countries, and the *final year* as the year in which the positive effects of the economic counteractive policy are expected to be evident. On factors affecting the timing of the effects of such policy, see Blanchard and Perotti [35]; Mishkin [36]; Reinhart and Rogoff [37].

We then compare the IMF data on budget deficits as a percentage of GDP for the aftershock and final years. Changes in the budget deficit serve as an indicator of fiscal efficiency. If our hypothesis holds, for countries in Group A where traces of economic policy are found, the differences in budget deficits between the aftershock and final years should be significantly positive; that is, the budget deficit should, on average, decrease. For comparison, we also apply other fiscal indicators, obtaining similar results (see Supplement 2, Part 4).

Our primary example focuses on the effects of policy reactions to the Russian financial crisis that began in mid-2014, culminating on December 16, 2014, and resulting in a drastic devaluation of the Russian ruble. We chose this crisis as a subject of testing rather than the more severe European sovereign debt crisis of 2010–2012 for pragmatic reasons. As we compare budget dynamics across 33 countries, the Russian crisis affected a greater number of countries than the European sovereign debt crisis, making statistical testing more robust. We have not considered the COVID-19 crisis, as its causes were non-financial, its implications were much broader, and it does not fit most definitions of a financial crisis (see, e.g., Babecký et al. [38], for a review of such definitions, and Hu et al. [39], for the analysis of the COVID-19 crisis).

The Russian financial crisis of 2014 developed in mid-2014 and culminated on December 16, 2014, resulting in approximately 50% devaluation of the ruble against the US dollar. This crisis stemmed from two primary factors: international sanctions following Russia's annexation of Crimea and a sharp decline in oil prices from over \$100 to below \$60 per barrel. The Central Bank of Russia's intervention, raising interest rates to 17%, proved insufficient to stabilize the currency, necessitating direct government financial support measures. For a detailed description of the causes and immediate effects of the crisis in Russia, see Dreger et al. [40]; see also Ahmad, Mishra, and Daly [41].

Global sovereign bond markets in 2015–2016 showed varying responses to this crisis. Developed economies experienced yield compression as investors sought low-risk assets (See Bank of International Settlements Annual Report [42]).

Emerging markets reacted differently, depending on their economic fundamentals and commodity exposure. Nations heavily dependent on hydrocarbon exports saw significant yield increases, while economies with diversified export structures maintained relatively stable borrowing costs. Countries that had implemented precautionary stabilization policies by maintaining adequate foreign exchange reserves, establishing fiscal buffers, adopting flexible exchange rate mechanisms, and pursuing economic diversification showed greater resilience in their bond markets during this period. These preventive measures created identifiable policy signals in bond markets prior to the crisis (see Demirer, Ferrer, and Shahzad [43]; Costantini and Sousa [44]).

The Russian ruble crisis can only be partly seen as a reflection of a global financial crisis. On the one hand, it shows several characteristics typical of systemic events: a rapid currency decline, sudden halts in capital inflows, increased volatility in the bond market, and fiscal stress affecting many other countries. Its spillover effects spread worldwide through trade, remittances, and financial connections. From this perspective, the crisis illustrates how a shock in one country, exacerbated by external factors such as falling oil prices, can spread through sovereign debt markets.

On the other hand, its global reach was relatively limited in comparison to the 2008 global financial crisis or the European sovereign debt crisis. The most direct effects were geographically concentrated, with fiscal consequences strongest in Russia and oil-dependent economies, while distant advanced economies experienced only mild second-round volatility. In addition, the crisis was compounded by sanctions and commodity price shocks, which made it less “pure” as a financial crisis. As such, its representativeness is restricted; however, its timing, clarity of onset, and measurable fiscal impacts make it valuable for methodological purposes. In this sense, the Russian crisis can serve as a reasonably instructive, though maybe not fully representative, test case for evaluating the stabilization effects of discovering policy footprints.

Being aware of these limitations, we additionally tested our working hypothesis, focusing on the effect of other financial crises that had worldwide implications, namely the crisis triggered by the unexpected outcome of the Brexit referendum in the UK in June 2016. These results are briefly discussed in Section 6 and documented in Supplement 2, Part 4.

We define the aftershock year as 2015, the year when the fiscal consequences of the crisis became evident, and the final years as 2017, 2018, and 2019, when the results of the eventual fiscal policy actions should affect the budget deficit. Using these years allows us to capture the full policy adjustment cycle and avoid blurring the policy effects by the COVID-19 pandemic.

For testing, we use daily data on the yields on government bonds taken from the Thomson Reuters database (<http://thomsonreuters.com/>). We collected data for 33 countries: Austria, Australia, Belgium, Brazil, Canada, Czech Republic, Chile, Denmark, Finland, France, Germany, Hungary, Iceland, India, Indonesia, Ireland, Italy, Japan, Korea, Mexico, Netherlands, Poland, Portugal, Russia, South Africa, Spain, Slovakia, Slovenia, Sweden, Switzerland, Turkey, United Kingdom, United States. We estimate the ARMA-GARCH models using data from the three-year period 2012–2014, using 783 observations for each country. Estimating the model for this period avoids contaminating the policy efforts by overlapping with the aftermath of the 2008–2009 crisis while capturing relevant pre-2014 crisis stability actions. To avoid, or at least reduce, sample selection bias and enable testing the budgetary effects, we use the broadest possible selection of countries for which data on session-to-session government bond yields are available in the Thomson Reuters database (<http://thomsonreuters.com/>) and that meet the following criteria.

1. Are complete for the estimation period (1 January 2012–31 December 2014).
2. Data for the IMF indicators for 2017–2019 are also available.
3. Data shows sufficient variability; that is, it does not contain duplications for a reasonably long period of time, or it does not have missing observations.

All relevant data is included in Supplement 3; see also data repository *figshare* at [https://figshare.com/articles/dataset/Data\\_for\\_Testing\\_Footprints\\_xlsx/30016453?file=57852376](https://figshare.com/articles/dataset/Data_for_Testing_Footprints_xlsx/30016453?file=57852376)

In Supplement 2 Parts 2 and 3, we show some descriptive statistics for the data and present results of unit root testing, which confirms stationarity of the data used for the ARMA-GARCH estimation.

Such a panel of data is potentially heterogeneous, as some countries might have conducted stabilization policies in the period investigated, while others might not. Hence, we estimate the ARMA( $k, 1$ )-GARCH(1,1) model under the null of (5) using the quasi maximum likelihood method given in (6) – (7). The optimal  $k$  was obtained by applying the Ljung-Box criterion, where  $k$  is equal to the minimum number of lags for which the Ljung-Box test does not reject the null hypothesis of a lack of autocorrelation in the residuals at the 5% level of significance.

As in Section 3, we apply the PELM $_{n,\pi}$  test with  $\nu_0 = -\kappa_0 = 1$  and  $\varrho_0 = 0.8$ . While setting of  $\nu_0$  and  $\kappa_0$  at the levels of one standard deviation seems to be rational, the choice of  $\varrho_0$  might be more difficult. As stated in Section 2, it represents the strength of the influence of real-time information on the distribution of forecast errors. Such strength, in practice, might not be easy to assess, as it would require access to detailed and often confidential data supported by some textual analysis and a thorough behavioural understanding of the market-related decision-making mechanism. Due to these complexities, we leave it for further research and assume a priori the reasonable strength of the influence of such information, which allows us to maintain a relatively high power of the test applied (see Section 4).

We use two versions of PELM $_n$ , namely PELM $_n$  (narrow), where the limits on  $\nu$  and  $\kappa$  needed for computing the supremum in (9) are within the  $\pm 0.1$  intervals around  $\nu_0 = -\kappa_0 = 1$ ; and PELM $_n$  (wide), with limits of  $\pm 0.25$  intervals around  $\nu_0 = -\kappa_0 = 1$ . We are setting  $\rho = 0.8$ , as for the PELM $_{n,\pi}$ .

Next, we divide the countries into two groups. Group A comprises countries where the PELM statistics are significant at the 0.05 nominal level of significance; that is, countries where we found statistical confirmation of the stabilization policy. Other countries with non-significant PELM statistics are left in Group B. The selection for Groups A and B turned out to be identical for the PELM $_{n,\pi}$  and PELM $_n$ (narrow), but it was slightly different if the selection was based on PELM $_n$ (wide). For all tests, countries in Group B are Austria, Brazil, France, Korea, Mexico, South Africa, Spain, Turkiye, and the United Kingdom. Under the PELM $_n$ (wide) classification, Canada is also in Group B. For the remaining countries, the PELM statistics are significant at the 0.05 level, and these countries are in Group A.

While testing simultaneously for the effects of an economic policy across 33 countries, we effectively conducted 33 separate hypothesis tests. Such simultaneous testing increases the probability of Type I errors (rejecting a true null hypothesis) due to the increased likelihood of obtaining significant results by chance alone (the multiple hypothesis testing problem). We alleviate this problem by applying the Truncated Product Method (TPM); see Zaykin et al. [45]. The TPM is a technique designed to combine p-values in multiple hypothesis testing while emphasizing significant results below an a priori given truncation threshold  $\tau$ . It helps mitigate the false discovery rate (FDR), which is the expected proportion of false positives among all discoveries, by focusing on p-values below the threshold and reducing noise from non-significant results. The selection of  $\tau$  influences the ability to detect false positives: lower thresholds mean that only the most significant p-values are combined, increasing sensitivity to true effects but potentially inflating the risk of overlooking genuine but weaker signals.

Table 2 presents the expected proportion of false discoveries of significance across all series by all three PELM tests obtained by TPM for the p-values of the PELM statistics for countries selected for Groups A and B. We also applied more general methods, namely the Modified Truncation Product Method and its bootstrapped counterparts; see Sheng and Yang [46]. The outcomes of these tests are not substantially different, in their essence, from those in Table 2.

As it follows from Table 2, for Group A, the TPM consistently produces an extremely small expected proportion of false discoveries, practically equal to zero, across all tests and truncation thresholds, providing strong evidence against the null hypothesis and supporting the claim that the policy footprints have been found for these countries. In contrast, Group B yields large p-values (mostly 1.000) across all tests and truncation values, suggesting no strong evidence to reject the null hypothesis of the lack of presence of such footprints. This pattern holds even at the strictest truncation threshold  $\tau = 0.050$ . Consequently, we can conclude that we have confirmation of the rationale behind the applied split of countries into Groups A and B.

**Table 2. Expected proportions of false discoveries of the significance.**

Truncation threshold	PELM <sub>n,π</sub>		PELM <sub>n</sub> (narrow)		PELM <sub>n</sub> (wide)	
	Group A	Group B	Group A	Group B	Group A	Group B
$\tau = 0.125$	0.000	1.000	0.000	0.117	0.000	0.131
$\tau = 0.100$	0.000	1.000	0.000	1.000	0.000	0.131
$\tau = 0.075$	0.000	1.000	0.000	1.000	0.000	0.131
$\tau = 0.050$	0.000	1.000	0.000	1.000	0.000	1.000

Legend: truncation parameter  $\tau$  indicates the minimum p-values excluded from joint testing.

<https://doi.org/10.1371/journal.pone.0336495.t002>

The significance of the PELM statistics is only a necessary, but not sufficient, condition for claiming that our test indeed evidences a stabilization policy. The next step is to evaluate the working hypothesis that the economic policies, if effective, should mitigate the effect of the shock and, consequently, cause a reasonably quick fiscal recovery. This can be confirmed by an 'after treatment' statistical test, which, in our case, is the paired difference test. The test checks whether the differences in the budget deficit to GDP ratio (see <http://www.imf.org/external/ns/cs.aspx?id=262>) between the aftershock and final years for countries in each Group are statistically positive; that is, the budget deficit diminishes. If this is the case for Group A, but not for Group B, we may conclude that the 'treatment', if discovered by the PELM tests, indeed evidences the presence of an effective stabilization policy.

Table 3 shows the results of this test: the  $t$ -ratios of the paired difference test statistics between the aftershock and final years for Groups A and B. The bootstrapped p-values are given in brackets below the test statistics. The results in Table 3 confirm our working hypothesis and, hence, the rationale of our testing method. For Group A, the paired  $t$ -test statistics are significant at any sensible level of significance; for Group B, the test statistics are either insignificant or on the verge of 10% significance. Consequently, for countries in Group A, it can be concluded that the economic stabilization policies undertaken prior to 2015 (the aftershock year) are evidenced by the significance of the PELM<sub>n,π</sub> and PELM<sub>n</sub> statistics, resulting in success, that is, in reduction in budget deficits in later years. As this cannot be said for Group B, we may conclude that our working hypothesis is confirmed.

**Table 3. Results of paired difference testing.**

Selection by	Group A	Group B
Final year 2017		
PELM <sub>n,π</sub> & PELM <sub>n</sub> (narrow)	15.201 (0.000)	-1.085 (0.651)
PELM <sub>n</sub> (wide)	15.193 (0.002)	-1.224 (0.664)
Final year 2018		
PELM <sub>n,π</sub> & PELM <sub>n</sub> (narrow)	28.717 (0.000)	0.460 (0.448)
PELM <sub>n</sub> (wide)	27.185 (0.000)	0.984 (0.396)
Final year 2019		
PELM <sub>n,π</sub> & PELM <sub>n</sub> (narrow)	25.714 (0.000)	0.541 (0.431)
PELM <sub>n</sub> (wide)	24.262 (0.000)	1.194 (0.363)

Legend: columns 2 and 3: the  $t$ -ratios of the paired difference test statistics between the aftershock and final year; the bootstrapped p-values are in parentheses.

<https://doi.org/10.1371/journal.pone.0336495.t003>

Overall, the findings of this section are promising. We can say that if footprints of economic policy are detected by the PELM test in a country's sovereign bond yield time series, the chances for a quicker recovery from the crisis increase compared to when such footprints are not confirmed. It is worth noting that the time series used ended before the crisis began. Therefore, the assessment of the recovery chances can be made at the very start of the crisis. In other words, the answer to the question posed in the section's title, whether the PELM tests can help in predicting fiscal improvement, is yes.

So far, these findings are limited to evidence from one financial crisis and single fiscal stability indicator. The next Section 6 presents a wider evidence.

## 6 Robustness of the empirical results

The settings of models that give the results described above have been subjected to various robustness checks. We verified whether a different classification of countries into Groups A and B, achieved by excluding certain countries from the sample or altering the size of Group A, affects the validity of our working hypothesis. We also examined the effects of applying a different bootstrapping technique to compute the p-values on the outcome of testing the working hypothesis.

Below, we present a summary of the robustness analysis, with detailed results available upon request.

### 6.1. Sample size sensitivity

We examined how sample size affects estimation outcomes. While our main results use three years of daily data on sovereign bond yields, we tested periods of different lengths: an extended four-year period (2011–2014, 1,048 observations) and two shorter periods: one with two years of data (2013–2014, 524 observations) and another covering eighteen months (mid-2013–2014, 394 observations). Although extending the sample produced similar testing results, we focused on the three-year period to avoid contamination from the earlier financial crisis. The shorter samples produced slightly less significant results in paired difference testing for Group B.

### 6.2. Flexibility of groups' classification

We tested the sensitivity of our results to group classification by systematically moving countries from Group A to Group B, starting with those showing the least significant PELM statistics (highest p-values). The reclassification of up to five countries from Group A to B maintained the significance of the paired  $t$ -test, supporting our working hypothesis.

### 6.3. Time-varying parameter analysis

We implemented a rudimentary time-varying parameters approach using rolling windows of 250 observations for the ARMA-GARCH model estimation. We calculated PELM statistics for each window and identified Group A using discounted sums of cases where PELM statistics showed p-values of 0.01 or less. Using daily data from early 2011, we applied various time discounting methods (see Laibson [47]; Heilmann [48]), weighting recent observations more heavily than earlier to achieve the time preference effect. The pairwise testing results continued to support our working hypothesis.

### 6.4. Alternative fiscal balance indicators

Supplement 2 Part 4, outlines the results of applying our two-step testing approach and using, instead of the budget deficit, three alternative indicators of fiscal imbalance published by the IMF. These indicators are the difference between budget revenue and expenditures, net lending/borrowing, and primary net lending/borrowing. For all three of these indicators, the paired  $t$ -test statistics are significant, or on the verge of the 10% significance level, for Group A. For Group B, two of these statistics are highly insignificant. Although these results are only fragmentary, they generally confirm the robustness of our approach.



## 6.5. Another example: the effect of the Brexit referendum

As an additional example, we use the effects of the unexpected outcome of the U.K. Brexit referendum, which took place on 23 June 2016. The Brexit referendum shock occurred on June 24, 2016, when the “Leave” victory triggered immediate financial market turmoil. The British pound fell to a 31-year low, equity markets plunged globally, and safe-haven flows intensified toward German bonds and gold. European banks were particularly affected due to their UK exposure, while peripheral eurozone countries faced renewed sovereign stress. For a literature review and comprehensive theoretical analysis, see Drinkwater and Jennings [49].

Both the Russian ruble and Brexit referendum crises shared some common elements. The Russian crisis exhibited classic emerging market vulnerability patterns: currency depreciation, reserve depletion, and contagion to similar economies. The Brexit crisis was characterized by typical uncertainty-driven market responses: a rush to sell riskier assets, currency volatility, and an immediate reassessment of cross-border investment risks. Both crises demonstrated how integrated global financial markets transmit localized shocks internationally through investor sentiment and portfolio rebalancing.

A summary of the testing results for the crisis caused by the Brexit referendum outcome is provided in Supplement 2, Part 4. It shows that the paired  $t$ -test, which uses the budget deficit data, gives insignificant results for Groups A and B, contradicting our working hypothesis. However, testing the changes in the other three financial balance indicators confirms it; the paired  $t$ -statistics are significant for Group A and insignificant for Group B. The discrepancy may stem from different methods used to construct these indicators. The insignificant results when testing with the budget deficit in Group A after the Brexit crisis likely reflect this measure’s inclusion of interest payments and one-off items, which are unrelated to uncertainty shocks. These payments, during the immediate post-referendum period, added measurement noise that hid potential stabilization effects. The significance of the paired  $t$ -test for Group A, while using the other three indicators, suggests that these measures better capture the effects in countries with stronger policy footprints in forecast uncertainty.

## 7 Conclusions

The PELM test we propose in this paper for detecting traces of economic policies in forecast errors of time series models demonstrates, in all its versions, sound asymptotic properties and high power across a wide range of alternatives. Its important advantage is the ability to detect traces of stabilization policy without the need to use additional information on the details of such policy, like textual data or dummy variables. While computationally feasible, its less restrictive version requires combining maximum likelihood estimation with simulation. Nevertheless, the table of critical values given in the text and extensive impulse surface approximations available on request make testing feasible in empirical models. The ARMA-GARCH model we discuss here is reasonably simple, and more advanced alternatives are available.

Our empirical application to modelling 10-year government bond yields produces easily interpretable results. The test proves useful for predicting fiscal balance indicators. Significant PELM test statistics computed for bond yields in a given period serve as a good indicator that the country’s fiscal position might improve in the years ahead. The results are reasonably robust, so our approach can have wider applications in evaluating the effects of stabilization policy.

Although our abductive reasoning approach, understood as forming the most likely explanation from incomplete information, fits a broader tradition in economics that seeks to recover unobserved mechanisms from observable outcomes, it is not within methodological limitations. As is often the case in empirical economics, research based on abductive reasoning is never complete. In our case, other plausible explanations may exist for the significance of the PELM testing and changes in fiscal balance indicators. Hence, further inference on discovering policy footprints in forecast uncertainty is needed, possibly by developing further statistical tests and providing wider empirical evidence.



The test offers several promising applications beyond our current scope. Particularly, it can be applied to discover the stabilization effects of inflation-targeting economic policy and further develop the optimal policy perturbation model. Other potential applications include empirical evaluation of direct and indirect outcomes of forecast-targeting economic policy.

The successful verification of our working hypothesis has further practical consequences, particularly in the context of modern Bayesian forecasting of financial balances (see Martin et al. [50]). Before forecasting financial balances, our PELM test can be applied to time series of government bond yields and other relevant financial indicators. If the test confirms the traces of an economic policy in forecast errors, the priors applied for forecasting can be relatively stronger (with higher weights) than in cases where such effects are not confirmed. Our test can also benefit frequentist or non-econometric forecasting. When the test indicates that economic policy stabilizes the bond market, non-econometric predictions or scenarios can be formulated more confidently.

## Supporting information

### Supplement 1. Main proofs, formulas and generalization.

(PDF)

### Supplement 2. Additional simulation and empirical results.

(PDF)

### Supplement 3. Data used for empirical examples and robustness analysis.

(XLSX)

## Author contributions

**Conceptualization:** Christian Francq, Jean-Michel Zakoïan.

**Data curation:** Radu Lupu.

**Formal analysis:** Christian Francq, Jean-Michel Zakoïan.

**Investigation:** Wojciech Charemza, Radu Lupu, Svetlana Makarova.

**Methodology:** Wojciech Charemza, Christian Francq, Svetlana Makarova, Jean-Michel Zakoïan.

**Software:** Wojciech Charemza, Radu Lupu.

**Writing – original draft:** Wojciech Charemza, Christian Francq, Svetlana Makarova, Jean-Michel Zakoïan.

**Writing – review & editing:** Wojciech Charemza, Svetlana Makarova.

## References

1. Basistha A, Kurov A. Macroeconomic cycles and the stock market's reaction to monetary policy. *Journal of Banking & Finance*. 2008;32(12):2606–16. <https://doi.org/10.1016/j.jbankfin.2008.05.012>
2. Forni M, Gambetti L. The dynamic effects of monetary policy: A structural factor model approach. *Journal of Monetary Economics*. 2010;57(2):203–16. <https://doi.org/10.1016/j.jmoneco.2009.11.009>
3. Hanisch M. The effectiveness of conventional and unconventional monetary policy: Evidence from a structural dynamic factor model for Japan. *Journal of International Money and Finance*. 2017;70:110–34. <https://doi.org/10.1016/j.jimonfin.2016.08.002>
4. Klingelhöfer J, Sun R. China's regime-switching monetary policy. *Economic Modelling*. 2018;68:32–40. <https://doi.org/10.1016/j.econmod.2017.04.017>
5. Guidolin M, Pedio M. Forecasting and trading monetary policy effects on the riskless yield curve with regime switching Nelson–Siegel models. *Journal of Economic Dynamics and Control*. 2019;107:103723. <https://doi.org/10.1016/j.jedc.2019.103723>
6. Ehrmann M, Fratzscher M. Communication by central bank committee members: Different strategies, same effectiveness?. *Journal of Money, Credit Banking*. 2007;39(2–3):509–41. <https://doi.org/10.1111/j.0022-2879.2007.00034.x>

7. Nakamura E, Steinsson J. High-frequency identification of monetary non-neutrality: The information effect. *The Quarterly Journal of Economics*. 2018;133(3):1283–330. <https://doi.org/10.1093/qje/qjy004>
8. Hoesch L, Rossi B, Sekhposyan T. Has the information channel of monetary policy disappeared? Revisiting the empirical evidence. *American Economic Journal: Macroeconomics*. 2023;15(3):355–87. <https://doi.org/10.1257/mac.20200068>
9. Knight FH. Risk, uncertainty and profit. Boston: Houghton Mifflin Company. 1921.
10. Rossi B, Sekhposyan T, Soupre M. Understanding the sources of macroeconomic uncertainty. 11415. CEPR Discussion Papers. 2016. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2816841](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2816841)
11. Lucas RE Jr. Expectations and the neutrality of money. *Journal of Economic Theory*. 1972;4(2):103–24. [https://doi.org/10.1016/0022-0531\(72\)90142-1](https://doi.org/10.1016/0022-0531(72)90142-1)
12. Morris S, Shin HS. Social value of public information. *American Economic Review*. 2002;92(5):1521–34. <https://doi.org/10.1257/000282802762024610>
13. Kaya O. High-frequency trading: reaching the limits. Deutsche Bank Research. 2016. [https://www.smallake.kr/wp-content/uploads/2016/05/20160530\\_042542.pdf](https://www.smallake.kr/wp-content/uploads/2016/05/20160530_042542.pdf)
14. Jansen D-J, de Haan J. Look who's talking: ECB communication during the first years of EMU. *International Journal of Finance and Economics*. 2006;11(3):219–28. <https://doi.org/10.1002/ijfe.294>
15. Charemza W, Díaz C, Makarova S. Quasi ex-ante inflation forecast uncertainty. *International Journal of Forecasting*. 2019;35(3):994–1007. <https://doi.org/10.1016/j.ijforecast.2019.03.002>
16. David A, Veronesi P. A Survey of alternative measures of macroeconomic uncertainty: Which measures forecast real variables and explain fluctuations in asset volatilities better?. *Annual Review Financial Economics*. 2022;14(1):439–63. <https://doi.org/10.1146/annurev-financial-111720-091804>
17. Peppel-Srebrny J. Not all government budget deficits are created equal: Evidence from advanced economies' sovereign bond markets. *Journal of International Money and Finance*. 2021;118:102460. <https://doi.org/10.1016/j.jimonfin.2021.102460>
18. Jiang M, Li J. Government spending shocks and default risk in emerging markets. *PLoS One*. 2023;18(7):e0288802. <https://doi.org/10.1371/journal.pone.0288802> PMID: 37471355
19. Manzli YS, Fakhfekh M, Béjaoui A, Alnafisah H, Jeribi A. On the hedge and safe-haven abilities of bitcoin and gold against blue economy and green finance assets during global crises: Evidence from the DCC, ADCC and GO-GARCH models. *PLoS One*. 2025;20(2):e0317735. <https://doi.org/10.1371/journal.pone.0317735> PMID: 39928701
20. Fleming MJ, Remolona EM. Price formation and liquidity in the U.S. Treasury Market: The Response to public information. *The Journal of Finance*. 1999;54(5):1901–15. <https://doi.org/10.1111/0022-1082.00172>
21. Davies RB. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*. 1977;64(2):247. <https://doi.org/10.2307/2335690>
22. Davies RB. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*. 1987;74(1):33–43. <https://doi.org/10.1093/biomet/74.1.33>
23. Christou V, Fokianos K. Estimation and testing linearity for non-linear mixed poisson autoregressions. *Electronic Journal of Statistics*. 2015;9(1). <https://doi.org/10.1214/15-ejs1044>
24. Francq C, Horvath L, Zakoian J-M. Sup-tests for linearity in a general nonlinear AR(1) model. *Econometric Theory*. 2009;26(4):965–93. <https://doi.org/10.1017/s0266466609990430>
25. Kirch C, Kamgaing JT. Testing for parameter stability in nonlinear autoregressive models. *Journal of Time Series Analysis*. 2012;33(3):365–85. <https://doi.org/10.1111/j.1467-9892.2011.00764.x>
26. Hansen BE. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*. 1996;64(2):413. <https://doi.org/10.2307/2171789>
27. Huang S-F, Chiang H-H, Lin Y-J. A network autoregressive model with GARCH effects and its applications. *PLoS One*. 2021;16(7):e0255422. <https://doi.org/10.1371/journal.pone.0255422> PMID: 34324604
28. Zou Y, Xu J, Chen Y. Volatility, correlation and risk spillover effect between freight rates in BCI and BPI markets: Evidence from static and dynamic GARCH-Copula and dynamic CoVaR models. *PLoS One*. 2025;20(1):e0315167. <https://doi.org/10.1371/journal.pone.0315167> PMID: 39821163
29. Clark TE, Ravazzolo F. Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*. 2014;30(4):551–75. <https://doi.org/10.1002/iae.2379>
30. Naimy V, Haddad O, Fernández-Avilés G, El Khoury R. The predictive capacity of GARCH-type models in measuring the volatility of crypto and world currencies. *PLoS One*. 2021;16(1):e0245904. <https://doi.org/10.1371/journal.pone.0245904> PMID: 33513150
31. Wu Q, Kuang X, Wu B, Xu X. Pricing of Shanghai stock exchange 50 ETF options based on different volatility models. *PLoS One*. 2023;18(12):e0288266. <https://doi.org/10.1371/journal.pone.0288266> PMID: 38060470
32. Ling S. Estimation and testing stationarity for double-autoregressive models. *Journal of the Royal Statistical Society - Series B: Statistical Methodology*. 2003;66(1):63–78. <https://doi.org/10.1111/j.1467-9868.2004.00432.x>
33. Ling S, Li D. Asymptotic inference for a nonstationary double AR(1) model. *Biometrika*. 2008;95(1):257–63. <https://doi.org/10.1093/biomet/asm084>

34. Cai Y, Montes-Rojas G, Olmo J. Quantile double AR time series models for financial returns. *Journal of Forecasting*. 2013;32(6):551–60. <https://doi.org/10.1002/for.2261>
35. Blanchard O, Perotti R. An empirical characterization of the dynamic effects of changes in government spending and taxes on output. *The Quarterly Journal of Economics*. 2002;117(4):1329–68. <https://doi.org/10.1162/003355302320935043>
36. Mishkin FS. Monetary policy strategy: lessons from the crisis. 16755. National Bureau of Economic Research; 2011. <https://www.nber.org/papers/w16755>
37. Reinhart C, Rogoff KS. *This time is different: eight centuries of financial folly*. Princeton University Press; 2009.
38. Babecký J, Havránek T, Matějů J, Rusnák M, Šmídková K, Vašíček B. Banking, debt, and currency crises in developed countries: Stylized facts and early warning indicators. *Journal of Financial Stability*. 2014;15:1–17. <https://doi.org/10.1016/j.jfs.2014.07.001>
39. Hu C, Pan W, Pan W, Dai W-Q, Huang G. The association of COVID-19 nexus on China's economy: A financial crisis or a health crisis?. *PLoS One*. 2022;17(9):e0272024. <https://doi.org/10.1371/journal.pone.0272024> PMID: 36070293
40. Dreger C, Kholodilin KA, Ulbricht D, Fidrmuc J. Between the hammer and the anvil: The impact of economic sanctions and oil prices on Russia's ruble. *Journal of Comparative Economics*. 2016;44(2):295–308. <https://doi.org/10.1016/j.jce.2015.12.010>
41. Ahmad W, Mishra AV, Daly KJ. Financial connectedness of BRICS and global sovereign bond markets. *Emerging Markets Review*. 2018;37:1–16. <https://doi.org/10.1016/j.ememar.2018.02.006>
42. Bank of International Settlements. Annual Report. 2015. <https://www.bis.org/publ/arpdf/ar2015e.pdf>
43. Demirel R, Ferrer R, Shahzad SJH. Oil price shocks, global financial markets and their connectedness. *Energy Economics*. 2020;88:104771. <https://doi.org/10.1016/j.eneco.2020.104771>
44. Costantini M, Sousa RM. What uncertainty does to euro area sovereign bond markets: Flight to safety and flight to quality. *Journal of International Money and Finance*. 2022;122:102574. <https://doi.org/10.1016/j.jimonfin.2021.102574>
45. Zaykin DV, Zhivotovsky LA, Westfall PH, Weir BS. Truncated product method for combining p-values. *Genetic Epidemiology*. 2002;22(2):170–85. <https://doi.org/10.1002/gepi.0042> PMID: 11788962
46. Sheng X, Yang J. Truncated product methods for panel unit root tests. *Oxford Bulletin of Economics and Statistics*. 2013;75(4):624–36. <https://doi.org/10.1111/j.1468-0084.2012.00705.x> PMID: 23869116
47. Laibson D. Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*. 1997;112(2):443–78. <https://doi.org/10.1162/003355397555253>
48. Heilmann C. A representation of time discounting. CPNSS Working Paper, 2008; vol. 4, no. 6. The Centre for Philosophy of Natural and Social Science (CPNSS), London School of Economics. Available from; [http://eprints.lse.ac.uk/23858/1/A\\_Representation\\_of\\_Time\\_Discounting\\_%28LS-ERO%29.pdf](http://eprints.lse.ac.uk/23858/1/A_Representation_of_Time_Discounting_%28LS-ERO%29.pdf)
49. Drinkwater S, Jennings C. The Brexit referendum and three types of regret. *Public Choice*. 2022;193(3–4):275–91. <https://doi.org/10.1007/s11127-022-00997-z>
50. Martin GM, Frazier DT, Maneesoonthorn W, Loaiza-Maya R, Huber F, Koop G, et al. Bayesian forecasting in economics and finance: A modern review. *International Journal of Forecasting*. 2024;40(2):811–39. <https://doi.org/10.1016/j.ijforecast.2023.05.002>