Supplemental material for "*Cross-regional spillover effects of sustainability indices: A heteroscedasticity-robust VAR approach*"

Kaiji Motegi* – Kobe University

Saki Sugano† – Aoyama Gakuin University

This draft: August 31, 2025

In this supplemental material, we provide technical and empirical details omitted from the main paper. To make this document self-contained, we review how the full sample period (01/06/2004 – 07/31/2024, $n = 5040$ days) is partitioned into five subperiods:

Pre-GFC: 01/06/2004 – 08/29/2008 ($n = 1147$).

GFC: 09/01/2008 – 12/30/2008 ($n = 81$).

Bet-GFC&Covid: 01/05/2009 – 02/28/2020 ($n = 2729$).

Covid: 03/02/2020 – 04/30/2020 ($n = 42$).

Post-Covid: 05/01/2020 – 07/31/2024 ($n = 1041$).

Pre-GFC, Bet-GFC&Covid, and Post-Covid are non-crisis periods, and GFC and Covid are crisis periods.

In Section 1, we describe detailed procedures of the heteroscedasticity-robust methods. In Section 2, we perform a variety of unit root tests and cointegration tests for our target variables: the Dow Jones Sustainability Indices of North America, Europe, and Asia Pacific. In Section 3, we check the robustness of empirical results presented in the main paper.

*Corresponding author. Graduate School of Economics, Kobe University. Address: 2-1 Rokkodai-cho, Nada, Kobe, Hyogo 657-8501 Japan. Email: motegi@econ.kobe-u.ac.jp

†College of Economics, Aoyama Gakuin University. Email: sugano@econ.aoyama.ac.jp

1 Heteroscedasticity-robust inference

In this section, we present technical details of the advanced methods that are robust to conditional heteroscedasticity of unknown form. Recall that the vector autoregressive (VAR) model is specified as

$$\boldsymbol{y}_t = \boldsymbol{A}_0 + \sum_{k=1}^{p} \boldsymbol{A}_k \boldsymbol{y}_{t-k} + \boldsymbol{u}_t, \qquad t \in \{1, \dots, n\}, \tag{1}$$

where $\boldsymbol{y}_t$ is a $K \times 1$ vector of target variables; $\boldsymbol{A}_0$ is a $K \times 1$ vector of intercept parameters; $\boldsymbol{A}_k$ is a $K \times K$ matrix of VAR parameters; $\boldsymbol{u}_t$ is a $K \times 1$ vector of error terms; n is the sample size.

In our empirical analysis, the target variables are daily log-returns of the Dow Jones Sustainability Indices (DJSIs) for North America (NA), Europe (EU), and Asia Pacific (AP). The number of target variables is therefore $K = 3$. The BIC-based joint selection algorithm of Cavaliere, De Angelis, Rahbek, and Taylor (2018) suggests that the lag length should be $p = 1$ for the Pre-GFC and Post-Covid periods and $p = 2$ for the Bet-GFC&Covid period. In this section, we focus on the simpler case of $p = 1$ to enhance exposition. Cases with $p \geq 2$ can be handled analogously, although notation would be more involved. We keep K to be an arbitrary natural number, as it does not complicate the exposition.

To facilitate mathematical formulation, we introduce some matrix notation. Model (1) with $p = 1$ can be rewritten as $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{W} + \boldsymbol{U}$, where $\boldsymbol{X} = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_n)$, $\boldsymbol{A} = (\boldsymbol{A}_0, \boldsymbol{A}_1)$, $\boldsymbol{W} = (\boldsymbol{w}_0, \dots, \boldsymbol{w}_{n-1})$, $\boldsymbol{w}_t = (1, \boldsymbol{y}_t^\top)^\top$, and $\boldsymbol{U} = (\boldsymbol{u}_1, \dots, \boldsymbol{u}_n)$. Let $\widehat{\boldsymbol{A}} = (\widehat{\boldsymbol{A}}_0, \widehat{\boldsymbol{A}}_1)$ be the least squares estimator for $\boldsymbol{A}$, then $\widehat{\boldsymbol{A}} = \boldsymbol{X}\boldsymbol{W}^\top(\boldsymbol{W}\boldsymbol{W}^\top)^{-1}$. Let $\widehat{\boldsymbol{U}} = (\widehat{\boldsymbol{u}}_1, \dots, \widehat{\boldsymbol{u}}_n)$ be the residual, then $\widehat{\boldsymbol{U}} = \boldsymbol{X} - \widehat{\boldsymbol{A}}\boldsymbol{W}$. The estimated error covariance matrix is given by $\widehat{\boldsymbol{\Sigma}} = n^{-1}\sum_{t=1}^{n} \widehat{\boldsymbol{u}}_t \widehat{\boldsymbol{u}}_t^\top$. Define $\widehat{\boldsymbol{\Gamma}} = n^{-1}\sum_{t=0}^{n-1} \boldsymbol{w}_t \boldsymbol{w}_t^\top$ and $\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Gamma}}^{-1} \otimes \widehat{\boldsymbol{\Sigma}}$, where $\otimes$ is the Kronecker product.

In Section 1.1, we describe the wild-bootstrap Granger causality tests of Hafner and Herwartz (2009). In Section 1.2, we describe the residual-based moving block bootstrap of Brüggemann, Jentsch, and Trenkler (2016) to compute asymptotically valid confidence intervals of impulse response functions.

1.1 Granger causality tests

Define $\boldsymbol{A}_1 = [a_{ij}]_{i,j}$, then the elements of direct interest for Granger causality are off-diagonals: $(a_{12}, a_{13}, a_{21}, a_{23}, a_{31}, a_{32})$. Granger non-causality from Europe (i.e., the second

variable) to North America (i.e., the first variable) at horizon $h = 1$ is equivalent to $a_{12} = 0$. Similarly, Granger non-causality from North America to Asia Pacific (i.e., the third variable) at $h = 1$ is equivalent to $a_{31} = 0$. Consider $H_0 : a_{12} = 0$ for concreteness; other scenarios can be handled analogously. Construct a selection matrix $\boldsymbol{R}$ which picks a_{12} from $\boldsymbol{a} = \text{vec}(\boldsymbol{A})$, where $\text{vec}(\cdot)$ is the column-wise vectorization operator. Specifically, $\boldsymbol{R}$ is a $1 \times K(K+1)$ vector whose $(2K+1)^{th}$ element is 1 and all others are 0, where the dimension of VAR is $K = 3$ in our present analysis. The Wald test statistic associated with H_0 is given by $\widehat{\mathcal{W}} = n\widehat{\boldsymbol{a}}^\top \boldsymbol{R}^\top (\boldsymbol{R}\widehat{\boldsymbol{\Omega}}\boldsymbol{R}^\top)^{-1}\boldsymbol{R}\widehat{\boldsymbol{a}}$, where $\widehat{\boldsymbol{a}} = \text{vec}(\widehat{\boldsymbol{A}})$.

If the error term $\boldsymbol{u}_t$ is assumed to be serially independent, it can be shown (up to several regularity conditions) that $\widehat{\mathcal{W}} \xrightarrow{d} \chi^2_1$ as $n \to \infty$ under H_0. In this case, the asymptotic p-value is readily available by picking an appropriate percentile of the chi-squared distribution. Once we allow for conditional heteroscedasticity in $\boldsymbol{u}_t$, the asymptotic chi-squared convergence is no longer guaranteed. In this case, the p-value should be computed via the wild bootstrap of Hafner and Herwartz (2009) [HH2009] as follows.

Step 1: Generate $\eta_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ and compute a bootstrap residual $\boldsymbol{u}^*_t = \eta_t \widehat{\boldsymbol{u}}_t$.

Step 2: Generate a bootstrap sample $\{\boldsymbol{y}^*_t\}_{t=1}^n$ according to the rule $\boldsymbol{y}^*_t = \widehat{\boldsymbol{A}}_0 + \widehat{\boldsymbol{A}}_1 \boldsymbol{y}^*_{t-1} + \boldsymbol{u}^*_t$, where $\boldsymbol{y}^*_0 = \boldsymbol{0}$ and the null hypothesis of non-causality is imposed. For example, when $H_0 : a_{12} = 0$ is considered, the $(1,2)$-element of $\widehat{\boldsymbol{A}}_1$ is forced to be 0.

Step 3: Given $\{\boldsymbol{y}^*_t\}_{t=1}^n$, compute a bootstrap Wald test statistic $\widehat{\mathcal{W}}^*$ in the same way as the actual test statistic $\widehat{\mathcal{W}}$ is computed.

Step 4: Repeat Steps 1-3 J times, resulting in $\{\widehat{\mathcal{W}}^*_j\}_{j=1}^J$. Compute the bootstrap p-value $\widehat{p}^*_J = J^{-1}\sum_{j=1}^J \boldsymbol{1}(\widehat{\mathcal{W}}^*_j \geq \widehat{\mathcal{W}})$, using the indicator function. Reject H_0 at the $100\alpha\%$ level if $\widehat{p}^*_J < \alpha$, where $\alpha \in (0,1)$ is a nominal size.

In our empirical analysis, the number of bootstrap iterations is $J = 2000$.

HH2009 establish the asymptotic validity of the wild bootstrap, imposing their Assumptions (A1)-(A6). These assumptions require that $\boldsymbol{u}_t$ should be α-mixing and a martingale difference sequence with finite fourth moment. Various forms of conditional heteroscedasticity are allowed under Assumptions (A1)-(A6), including stationary GARCH. Proposition 2 of HH2009 states that, under H_0, the wild bootstrap is asymptotically valid in the following

sense:

$$\sup_{0<c<\infty} \left| \Pr^*(\widehat{\mathcal{W}}^* \leq c) - \Pr(\widehat{\mathcal{W}} \leq c) \right| \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty,$$

where $\Pr^*$ denotes the bootstrap probability measure and "$\xrightarrow{p}$" signifies *convergence in probability*.

1.2 Impulse response functions

To formulate impulse response functions (IRFs), assume that $\mathrm{E}(\boldsymbol{u}_t) = \boldsymbol{0}$ and $\mathrm{E}(\boldsymbol{u}_t \boldsymbol{u}_t^\top) = \boldsymbol{\Sigma}$, a $K \times K$ positive-definite matrix. Define the Cholesky decomposition $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top$, where $\boldsymbol{L}$ is a $K \times K$ lower triangular matrix. Given the invertibility condition, model (1) with $p = 1$ can be transformed to a vector moving average model of order infinity with an orthogonalized shock:

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \sum_{k=0}^{\infty} \boldsymbol{\Theta}_k \boldsymbol{v}_{t-k},$$

where $\boldsymbol{\mu} = (\boldsymbol{I}_K - \boldsymbol{A}_1)^{-1}\boldsymbol{A}_0$, $\boldsymbol{\Theta}_k = \boldsymbol{A}_1^k \boldsymbol{L}$, and $\boldsymbol{v}_t = \boldsymbol{L}^{-1}\boldsymbol{u}_t$; note that $\mathrm{E}(\boldsymbol{v}_t) = \boldsymbol{0}$ and $\mathrm{E}(\boldsymbol{v}_t \boldsymbol{v}_t^\top) = \boldsymbol{I}_K$ by construction. Write $\boldsymbol{\Theta}_h = [\theta_{ij,h}]_{i,j}$, then $\theta_{ij,h}$ represents the (structural) IRF of variable i to a 1σ shock in variable j at prediction horizon $h \geq 0$. Perform the least squares to get the estimator $(\widehat{\boldsymbol{A}}_0, \widehat{\boldsymbol{A}}_1)$, residual $\widehat{\boldsymbol{u}}_t$, and estimated error covariance matrix $\widehat{\boldsymbol{\Sigma}}$. Implement the Cholesky decomposition to get $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{L}}\widehat{\boldsymbol{L}}^\top$. The point estimator of IRF can be computed as $\widehat{\boldsymbol{\Theta}}_k = \widehat{\boldsymbol{A}}_1^k \widehat{\boldsymbol{L}}$ for $k \geq 0$. To simplify exposition, let $\hat{\theta}$ signify a generic element of $\widehat{\boldsymbol{\Theta}}_k$.

Brüggemann, Jentsch, and Trenkler (2016) [BJT2016] propose the residual-based moving block bootstrap (MBB). They claim that the $100(1-\alpha)\%$ confidence interval for $\hat{\theta}$ should be constructed as follows, where $\alpha \in (0, 1)$ is a nominal size.

Step 1: Choose a block length $\ell < n$ and let $b = \lceil n/\ell \rceil$ be the number of blocks so that $\ell b \geq n$. Define a $K \times \ell$ block $\boldsymbol{B}_{i,\ell} = (\widehat{\boldsymbol{u}}_{i+1}, \ldots, \widehat{\boldsymbol{u}}_{i+\ell})$ for $i \in \{0, \ldots, n-\ell\}$.

Step 2: Draw $(i_0, \ldots, i_{b-1})$ independently and identically from $\{0, 1, \ldots, n-\ell\}$ with uniform probability. Stack blocks $(\boldsymbol{B}_{i_0,\ell}, \ldots, \boldsymbol{B}_{i_{b-1},\ell})$ horizontally, and discard the last $b\ell - n$ columns to get bootstrap residuals $(\widehat{\boldsymbol{u}}_1^*, \ldots, \widehat{\boldsymbol{u}}_n^*)$. Center the bootstrap residuals according to the rule $\boldsymbol{u}_{j\ell+s}^* = \widehat{\boldsymbol{u}}_{j\ell+s}^* - (n-\ell+1)^{-1} \sum_{k=0}^{n-\ell} \widehat{\boldsymbol{u}}_{s+k}$ for $s \in \{1, \ldots, \ell\}$ and $j \in \{0, 1, \ldots, b-1\}$ to ensure $\mathrm{E}^*(\boldsymbol{u}_t^*) = \boldsymbol{0}$ for all $t \in \{1, \ldots, n\}$.

Step 3: Generate a bootstrap sample $(\boldsymbol{y}_1^*, \ldots, \boldsymbol{y}_n^*)$ according to the rule $\boldsymbol{y}_t^* = \widehat{\boldsymbol{A}}_0 + \widehat{\boldsymbol{A}}_1 \boldsymbol{y}_{t-1}^* + \boldsymbol{u}_t^*$, where $\boldsymbol{y}_0^* = \boldsymbol{0}$. Compute IRF $\hat{\theta}^*$ given the bootstrap sample, where the procedure is the same as in the actual sample.

Step 4: Repeat Steps 2-3 J times, resulting in $(\hat{\theta}_1^*, \ldots, \hat{\theta}_J^*)$. Sort to get $\hat{\theta}_{(1)}^* \leq \cdots \leq \hat{\theta}_{(J)}^*$. The $100(1-\alpha)\%$ confidence interval for $\hat{\theta}$ is constructed via Hall's percentile interval: $\widehat{CI}_\alpha = [\hat{\theta} - \hat{\theta}_{\{(1-\alpha/2)J\}}^*,\ \hat{\theta} - \hat{\theta}_{\{(\alpha/2)J\}}^*]$.

In our empirical analysis, the number of bootstrap iterations is $J = 2000$.

BJT2016 establish the asymptotic validity of the MBB-based confidence interval, imposing their Assumptions 2.1 and 4.1 as well as a proper rate of divergence for the block length: $\ell \to \infty$ and $\ell^3/n \to 0$ as $n \to \infty$. A suggested block length of Jentsch and Lunsford (2019), $\ell = \lfloor \kappa n^{1/4} \rfloor$ with $\kappa = 5.03$, satisfies the latter condition. In our empirical analysis, we follow their suggestion and use $\ell = \lfloor 5.03 n^{1/4} \rfloor$. This gives $\ell \in \{29, 36, 28\}$ for the Pre-GFC, Bet-GFC&Covid, and Post-Covid periods, respectively. Assumptions 2.1 and 4.1 of BJT2016 permit a variety of conditionally heteroscedastic errors such as GARCH with a finite eighth moment. Corollary 5.2 of BJT2016 states that, under these regularity conditions, the MBB-based confidence interval is asymptotically valid in the following sense:

$$\sup_{x \in \mathbb{R}} \left| \Pr^* \left\{ \sqrt{n} \left(\hat{\theta}^* - \hat{\theta} \right) \leq x \right\} - \Pr \left\{ \sqrt{n} \left(\hat{\theta} - \theta \right) \leq x \right\} \right| \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty,$$

where $\mathbb{R}$ is the set of real numbers. This equation implies that the 90% confidence interval, for example, contains a true IRF with probability approaching 0.9 as $n \to \infty$.

2 Unit root tests and cointegration tests

Target variables of this paper are the Dow Jones Sustainability North America Index (Bloomberg ticker: `A1SGI`), the Dow Jones Sustainability Europe Index (`DJSEUR`), and the Dow Jones Sustainability Asia Pacific Index (`P1SGI`). The log-value (i.e., $\ln P_t$) and log-difference (i.e., $\Delta \ln P_t$) of these indices are plotted in Figure 1.

In the main paper, we formulate the vector error correction model (VECM) for the log-values of the three sustainability indices, and determine the optimal cointegrating rank r^* and the optimal lag length p^* simultaneously via the Bayesian Information Criterion (BIC). Cavaliere, De Angelis, Rahbek, and Taylor (2018) [CDRT2018] advise this approach, claiming that it satisfies desired properties in both large and small samples when there exists

conditional heteroscedasticity of unknown form. Adopting the BIC approach, we find that the DJSI log-values are non-stationary and non-cointegrated: $r^* = 0$.

In this section, we perform standard unit root tests and cointegration tests to check the robustness of the BIC-based results. The unit root tests are implemented in Section 2.1, and the cointegration tests are implemented in Section 2.2.

2.1 Unit root tests

We consider three unit root tests which are routinely used in the literature of time series analysis. The first test is the Augmented Dickey-Fuller (ADF) test, in which the null hypothesis H_0 is that the target variable is non-stationary. The lag length is chosen by the Schwarz Information Criterion. The second test is the Phillips-Perron (PP) test, in which H_0 is the non-stationarity of the target variable. The Bartlett kernel and the Newey-West automatic bandwidth selection are used. The third test is the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, in which H_0 is the *stationarity* of the target variable. As in the PP test, the Bartlett kernel and the Newey-West automatic bandwidth selection are used for KPSS.[1]

We explore three scenarios for each test. First, the target variable is the log-value $\ln P_t$, and only the constant term is included in the regression model. Second, the target variable is $\ln P_t$, and both the constant term and the linear time trend are included in the model. Third, the target variable is the log-difference $\Delta \ln P_t$, and only the constant term is included in the model. The three tests are applied to each region and to each of the Pre-GFC, Bet-GFC&Covid, and Post-Covid periods in parallel.

In general, it is often taken for granted that log stock indices are $I(1)$ (i.e., non-stationary in level and stationary in the first difference). This consensus corresponds to the weak form efficient market hypothesis, and also means that log stock indices are what Campbell, Lo, and MacKinlay (1997) call Random Walk 3 (i.e., random walk with serially uncorrelated increment). The $I(1)$ property of the log-DJSIs is what we expect in view of Figure 1, and our conjecture is supported by the BIC approach of CDRT2018. We check if the classical ADF, PP, and KPSS tests arrive at the same conclusion.

Results of the unit root tests are summarized in Table 1. Asymptotic p-values of the ADF test are shown in Panel 1, asymptotic p-values of the PP test are shown in Panel 2, and LM test statistics of the KPSS test are shown in Panel 3. For the Post-Covid period,

[1]Detailed procedures of the three unit root tests can be found in Hamilton (1994, Ch. 17) and Hansen (2022, Ch. 16), among others. These are built-in tests in EViews 14, and we use it to execute the tests.

our conjecture is confirmed unanimously. When the target variable is the log-value $\ln P_t$, the ADF tests fail to reject the null hypothesis of unit root at the 10% level for all regions, whether the time trend is included or not. The same goes for the PP tests. The KPSS tests reject the null hypothesis of stationarity at the 10% or stricter level for all regions, whether the time trend is included or not. When the target variable is the log-difference $\Delta \ln P_t$, the ADF and PP tests reject the unit root hypothesis at the 1% level, and the KPSS tests fail to reject the stationarity hypothesis at the 10% level. Thus, for the Post-Covid period, there is overwhelming evidence that the DJSIs are non-stationary in log-levels and stationary in log-differences. We observe almost the same results for the Pre-GFC period.

For the Bet-GFC&Covid period, the ADF and PP tests for the log-values produce unexpected results under some cases. When the time trend is not included, the ADF and PP tests *reject* the unit root hypothesis for Asia Pacific at the 5% level, although they fail to reject H_0 for North America and Europe (Panels 1-2). When the time trend is included, the ADF and PP tests *reject* the unit root hypothesis for all regions at the 10% or even stricter level. These results assert that the DJSI log-values are stationary, contrary to our conjecture. A potential source of these counter-intuitive results is a strong degree of conditional heteroscedasticity. The Bet-GFC&Covid period contains the Greek government-debt crisis around May 2010 and the Brexit turmoil around June 2016. Due to these events, the Bet-GFC&Covid period is the most volatile subperiod of the three (Figure 1). There is a chance that the excessive volatility persistence is amplifying the rate of the type-I error, a false rejection of the true null hypothesis, in the ADF and PP tests.

Keeping the focus on the Bet-GFC&Covid period, the KPSS tests for the log-values lead to expected results. The null hypothesis of stationarity is rejected at the 1% level for all regions, regardless of the presence or absence of time trend (Panel 3). Further, all tests arrive at the expected results when the target variable is the log-difference $\Delta \ln P_t$. The ADF and PP tests reject the unit root hypothesis at the 1% level for all regions (Panels 1-2), and the KPSS test fails to reject the stationarity hypothesis at the 10% level for all regions (Panel 3). Hence, the DJSI log-differences are most likely stationary.

In summary, the consensus that stock indices are non-stationary in log-levels and stationary in log-differences holds in our study too, although there are some exceptions when the ADF and PP tests are applied to log-levels under the heavily volatile period (i.e., Bet-GFC&Covid). The BIC approach of CDRT2018, which is guaranteed to operate well under conditional heteroscedasticity of unknown form, points to the consensus. Thus, we conclude

that the log-DJSIs are $I(1)$ for all regions and subperiods considered.

2.2 Cointegration tests

Assuming the log-values of the DJSI North America, Europe, and Asia Pacific are all $I(1)$, the next crucial step is to determine the number of cointegrating relationships, r. The BIC approach of CDRT2018 finds $\hat{r}^* = 0$ (i.e., non-cointegration). In this section, we perform the Johansen cointegration test for robustness check. The lag length of the VECM is either $p = 1$ or $p = 2$. The test statistic is either the trace statistic or the maximum eigenvalue statistic. We consider three specifications for deterministic components:

Case 1: Cointegrating relationship includes a constant, and short-run dynamics include a constant.

Case 2: Cointegrating relationship includes a constant and trend, and short-run dynamics include a constant.

Case 3: Both the cointegrating relationship and short-run dynamics include a constant and trend.[2]

In Table 2, we report the rank (i.e., the number of cointegrating relationships) selected by the critical values of MacKinnon, Haug, and Michelis (1999), where the nominal size is 5%. For the Pre-GFC and Post-Covid periods, the selected rank is 0 for all cases but one. This is strong evidence of non-cointegration, and it is consistent with the BIC-based result.

For the Bet-GFC&Covid period, the results vary across the specifications of deterministic terms. Fixing $p = 1$, the trace test points to $r = 0$ in Case 1, $r = 1$ in Case 2, and $r = 3$ in Case 3; the max-eigenvalue test points to $r = 0$ in Case 1 and $r = 1$ in Cases 2-3. The same results appear when the lag length is $p = 2$. As discussed in the previous section, the Bet-GFC&Covid period is the most volatile subperiod of the three. The conventional Johansen test is not designed to be robust to conditional heteroscedasticity. Hence, there is a chance that the strong volatility persistence under the Bet-GFC&Covid period is contaminating the inference. Prioritizing the BIC-based results, we conclude that the log-DJSIs are not cointegrated for any subperiods.

[2]Detailed procedures of the Johansen cointegration test can be found in Lütkepohl (2006, Ch. 8) and Enders (2015, Ch. 6), among others. This is a built-in test in EViews 14, and we use it to implement the test.

3 Robustness check of the main empirical results

In this section, we check the robustness of our main empirical results. The main scenario elaborated in the main paper is called Scenario #1 here. In Scenario #1, we use the Dow Jones Sustainability North America Index (U.S. dollars, price return, `A1SGI`), the Dow Jones Sustainability Europe Index (euros, price return, `DJSEUR`), and the Dow Jones Sustainability Asia Pacific Index (U.S. dollars, price return, `P1SGI`). These are defined as variables #1–#3 in Table 3. Taking Scenario #1 as the benchmark, we consider the following alternative scenarios.

Scenario #2: To control for a market portfolio, we add the fourth variable called World (WD) to the VAR model. We use the Dow Jones Global Index (U.S. dollars, price return) for WD. In this scenario, the number of target variables is $K = 4$.[3]

Scenario #3: We change the currency of the Europe Index from euros to U.S. dollars. Consequently, all indices in the model are denominated in dollars, implying that the exchange rate risk is hedged from the U.S. perspective. The number of target variables is $K = 3$.

Scenario #4: We replace the Europe Index with the Dow Jones Sustainability Eurozone Index (euros, price return) to focus on the narrower area of Europe. As of December 31^{st}, 2024, the U.K. accounts for 19.7% and Switzerland accounts for 17.1% in the country breakdown of the Europe Index. The Eurozone Index exclude these and some other countries not using euros as their currency. The number of target variables is $K = 3$.

Scenario #5: The North America Index is replaced with the Dow Jones Sustainability U.S. Index (U.S. dollars, price return). The former consists of the U.S. (94.5%) and Canada (5.5%), while the latter excludes Canada. The number of target variables is $K = 3$.

Scenario #6: We use *net total returns* instead of *price returns* for all regions. Using the net total returns corresponds to a situation where regular cash dividends

[3]Motegi and Iitsuka (2023) also augment their VAR model with a market portfolio, though in a different context of the J-REIT market.

are reinvested at the close on the ex-date after the deduction of applicable withholding taxes. The number of target variables is $K = 3$.

The full details of these variables are listed in Table 3, and the specification of each scenario is summarized in Table 4.

Exploring these additional scenarios serves as an important contribution to the relevant literature. Previous studies do not provide a comprehensive treatment of currency, the definition of regions, and return types. Some studies do not even state how their main scenario is designed in terms of these aspects. Providing these details helps readers get more concrete implications on portfolio strategies and sustainable policies.

Daily log-values (i.e., $Y_t = \ln P_t$) and log-returns (i.e., $y_t = \ln P_t - \ln P_{t-1}$) of all indices are plotted in Figure 1. In Table 5, we report sample statistics of the log-returns during the Post-Covid period. In Table 6, we show contemporaneous correlations of all pairs of the log-returns for the Post-Covid period. Sample statistics and correlations for the Pre-GFC and Bet-GFC&Covid periods are qualitatively similar, hence omitted to save space.

In light of Table 6, the correlation between the Dow Jones Sustainability North America Index and the Dow Jones Global Index is 0.92, indicating the dominant role of the U.S. in NA and WD. The high correlation between the two variables may cause multicollinearity in Scenario #2. The correlation between the Europe Indices in U.S. dollars and euros is 0.91, signaling that Scenario #3 should arrive at similar results to Scenario #1. The same conjecture applies to Scenario #4, since the correlation between the Europe and Eurozone Indices is 0.95. The U.S. Index is almost identical to the North America Index with the correlation being 1.00, hence Scenario #5 is effectively the same as Scenario #1. Finally, the net total return series and the price return series are almost identical to each other with the correlation being 1.00 for each of NA, EU, and AP. Thus, the results of Scenario #1 should be all preserved under Scenario #6.

Not surprisingly, Scenarios #2–#6 lead to almost the same empirical results as Scenario #1. The BIC-based joint selection algorithm of Cavaliere, De Angelis, Rahbek, and Taylor (2018) points to $(\hat{r}^*, \hat{p}^*) = (0, 1)$ under the Pre-GFC and Post-Covid periods, and $(\hat{r}^*, \hat{p}^*) = (0, 2)$ under the Bet-GFC&Covid period. In Table 7, we find significant Granger causality from NA to EU, causality from NA to AP, and causality from EU to AP for most scenarios and subperiods. The statistical significance is weaker for Scenario #2, most likely due to multicollinearity between NA and WD.

In Figures 2-4, we plot impulse response functions of Scenario #1 for the Pre-GFC,

Bet-GFC&Covid, and Post-Covid periods, respectively. These are similar to each other, and impulse responses of Scenarios #2–#6 are also similar. The latter figures are omitted to save space, but available upon request. The forecast error variance decomposition is implemented in Table 8. All scenarios exhibit the "upper triangular" structure, implying that there are large, positive, unidirectional spillover effects from NA to EU and AP as well as small, positive, unidirectional spillover effects from EU to AP.

The robustness of our empirical results is primarily due to the high correlation between the original and alternative variables, but it also stems from our econometric design. First, we are restricting our attention to three regions (i.e., AP, EU, and NA) and the VAR lag length is tightly chosen via BIC. Moreover, we are not fitting a parametric model for the conditional variance-covariance structure, which reduces the number of parameters dramatically. Such a parsimonious specification stabilizes estimation and hypothesis testing. Second, we are excluding the GFC and Covid periods to remove outliers, and each non-crisis subperiod has plenty of observations. Third, AP, EU, and NA have fairly low correlations with each other, which is a desired situation in terms of avoiding multicollinearity (Table 6).

In Scenario #7, we analyze general stock market indices in which ESG scores are not prioritized. We use the Dow Jones U.S. Index (Bloomberg ticker: `DJUS`) in place of the DJSI-NA, the Dow Jones Europe Index (`E1DOW`) in place of the DJSI-EU, and the Dow Jones Asia/Pacific Index (`P1DOW`) in place of the DJSI-AP. These DJIs are all *price return* series denominated in U.S. dollars (Table 3). We formulate a trivariate VAR model for the DJI-NA, DJI-EU, and DJI-AP (Table 4). Time series plots of the log-values and log-returns of each DJI are included in Figure 1. Sample statistics of the DJI log-returns are reported in Table 5. The sample correlation matrix of all variables including the DJIs is presented in Table 6.

While the summary results and interpretation of the DJI subsample analysis are offered in the main paper, we report more detailed results in this supplemental material for completeness. Refer to Table 7 for Granger causality tests, and refer to Table 8 for forecast error variance decomposition. Impulse responses of the DJI system are similar to those of the DJSI system (Figures 2-4), hence they are omitted to save space. Overall, the DJI system (Scenario #7) leads to almost the same results as the DJSI system (Scenario #1). In particular, we find significant, positive, unidirectional spillover effects from DJI-NA to DJI-EU and DJI-AP. The similarity between the DJI and DJSI results suggest that financial factors dominate non-financial factors in the DJSI spillover mechanism. This arguably implies that

ESG investing can contribute to the Sustainable Development Goals while keeping a given level of financial performance.

References

Brüggemann, R., C. Jentsch, and C. Trenkler (2016): "Inference in VARs with conditional heteroskedasticity of unknown form," *Journal of Econometrics*, 191, 69–85.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997): *The Econometrics of Financial Markets*. Princeton University Press.

Cavaliere, G., L. De Angelis, A. Rahbek, and R. A. M. Taylor (2018): "Determining the cointegration rank in heteroskedastic VAR models of unknown order," *Econometric Theory*, 34, 349–382.

Diebold, F. X., and K. Yilmaz (2009): "Measuring financial asset return and volatility spillovers, with application to global equity markets," *The Economic Journal*, 119, 158–171.

Enders, W. (2015): *Applied Econometric Time Series*. John Wiley & Sons, Inc.

Hafner, C. M., and H. Herwartz (2009): "Testing for linear vector autoregressive dynamics under multivariate generalized autoregressive heteroskedasticity," *Statistica Neerlandica*, 63, 294–323.

Hamilton, J. D. (1994): *Time series analysis*. New Jersey: Princeton University Press.

Hansen, B. E. (2022): *Econometrics*. New Jersey: Princeton University Press.

Jentsch, C., and K. G. Lunsford (2019): "The dynamic effects of personal and corporate income tax changes in the United States: Comment," *American Economic Review*, 109, 2655–2678.

Lütkepohl, H. (2006): *New Introduction to Multiple Time Series Analysis*. Springer-Verlag Berlin Heidelberg.

MacKinnon, J. G., A. A. Haug, and L. Michelis (1999): "Numerical distribution functions of likelihood ratio tests for cointegration," *Journal of Applied Econometrics*, 14, 563–577.

Motegi, K., and Y. Iitsuka (2023): "Inter-regional dependence of J-REIT stock prices: A heteroscedasticity-robust time series approach," *North American Journal of Economics and Finance*, 64, #101840.

Table 1: Univariate unit root tests for the main scenario (Scenario #1)

1. Asymptotic p-values of the Augmented Dickey-Fuller tests (H_0: Non-stationarity)

		Pre-GFC			Bet-GFC&Covid			Post-Covid		
Target	Trend	NA	EU	AP	NA	EU	AP	NA	EU	AP
$\ln P_t$	No	0.42	0.57	0.46	0.61	0.27	0.03**	0.58	0.39	0.17
$\ln P_t$	Yes	0.46	1.00	0.82	0.01***	0.03**	0.05*	0.62	0.26	0.44
$\Delta \ln P_t$	No	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***

ADF test: The lag length is chosen by the Schwarz Information Criterion.

2. Asymptotic p-values of the Phillips-Perron tests (H_0: Non-stationarity)

		Pre-GFC			Bet-GFC&Covid			Post-Covid		
Target	Trend	NA	EU	AP	NA	EU	AP	NA	EU	AP
$\ln P_t$	No	0.52	0.58	0.48	0.62	0.29	0.04**	0.62	0.42	0.17
$\ln P_t$	Yes	0.68	1.00	0.87	0.01***	0.06*	0.06*	0.69	0.29	0.44
$\Delta \ln P_t$	No	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***

PP test: The Bartlett kernel and the Newey-West automatic bandwidth selection are used.

3. LM test statistics of the Kwiatkowski-Phillips-Schmidt-Shin tests (H_0: Stationarity)

		Pre-GFC			Bet-GFC&Covid			Post-Covid		
Target	Trend	NA	EU	AP	NA	EU	AP	NA	EU	AP
$\ln P_t$	No	3.38***	2.84***	3.72***	6.22***	5.65***	3.39***	2.28***	2.39***	0.38*
$\ln P_t$	Yes	0.39***	0.77***	0.66***	0.24***	0.44***	0.36***	0.47***	0.38***	0.42***
$\Delta \ln P_t$	No	0.13	0.51**	0.20	0.04	0.07	0.10	0.17	0.06	0.18

KPSS test: The Bartlett kernel and the Newey-West automatic bandwidth selection are used. When the intercept is included, the asymptotic 10%, 5%, 1% critical values are 0.347, 0.463, 0.739, respectively. When the intercept and linear time trend are included, the asymptotic 10%, 5%, 1% critical values are 0.119, 0.146, 0.216, respectively.

The constant term is included for all cases considered. NA: North America (Bloomberg ticker: `A1SGI`). EU: Europe (`DJSEUR`). AP: Asia Pacific (`P1SGI`). Asterisks ***, **, and * indicate a rejection of the null hypothesis at the 1%, 5%, and 10% levels, respectively.

Table 2: Johansen cointegration tests for the main scenario (Scenario #1)

		Pre-GFC			Bet-GFC&Covid			Post-Covid		
p	Stat.	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3	Case 1	Case 2	Case 3
1	Trace	0	0	0	0	1	3	0	0	1
1	Max	0	0	0	0	1	1	0	0	0
2	Trace	0	0	0	0	1	3	0	0	0
2	Max	0	0	0	0	1	1	0	0	0

We perform the Johansen cointegration test for the log-values of the DJSI North America (Bloomberg ticker: `A1SGI`), Europe (`DJSEUR`), and Asia Pacific (`P1SGI`) for each subsample period. The lag length of VECM is either $p = 1$ or $p = 2$. The test statistic is either the trace statistic or the maximum eigenvalue statistic. Case 1: Cointegrating relationship includes a constant, and short-run dynamics include a constant. Case 2: Cointegrating relationship includes a constant and trend, and short-run dynamics include a constant. Case 3: Both the cointegrating relationship and short-run dynamics include a constant and trend. This table reports the rank (i.e., the number of cointegrating relationships) selected by the critical values of MacKinnon, Haug, and Michelis (1999), where the nominal size is 5%.

Table 3: Full details of the stock indices used in this paper

ID	ticker	index name	currency	type
#1	A1SGI	Dow Jones Sustainability North America Index	$	PR
#2	DJSEUR	Dow Jones Sustainability Europe Index	€	PR
#3	P1SGI	Dow Jones Sustainability Asia Pacific Index	$	PR
#4	W1DOW	Dow Jones Global Index	$	PR
#5	DJSEURD	Dow Jones Sustainability Europe Index	$	PR
#6	DJSEUZ	Dow Jones Sustainability Eurozone Index	€	PR
#7	AASGI	Dow Jones Sustainability U.S. Index	$	PR
#8	A1SGITR	Dow Jones Sustainability North America Index	$	NTR
#9	DJSEURT	Dow Jones Sustainability Europe Index	€	NTR
#10	P1SGITR	Dow Jones Sustainability Asia Pacific Index	$	NTR
#11	DJUS	Dow Jones U.S. Index	$	PR
#12	E1DOW	Dow Jones Europe Index	$	PR
#13	P1DOW	Dow Jones Asia/Pacific Index	$	PR

This table lists the ID number, Bloomberg ticker, index name, currency (U.S. dollars or euros), and return type (price return or net total return) of all stock indices used in this paper.

Table 4: Variable selections by alternative scenarios

ID	ticker	#1	#2	#3	#4	#5	#6	#7
#1	A1SGI	1. NA	1. NA	1. NA	1. NA			
#2	DJSEUR	2. EU	2. EU			2. EU		
#3	P1SGI	3. AP	3. AP	3. AP	3. AP	3. AP		
#4	W1DOW		4. WD					
#5	DJSEURD			2. EU				
#6	DJSEUZ				2. EU			
#7	AASGI					1. NA		
#8	A1SGITR						1. NA	
#9	DJSEURT						2. EU	
#10	P1SGITR						3. AP	
#11	DJUS							1. NA
#12	E1DOW							2. EU
#13	P1DOW							3. AP

In our VAR model, the first region is North America (NA), the second region is Europe (EU), and the third region is Asia Pacific (AP). This table summarizes which stock index proxies which region. In Scenario #1, $(\mathrm{NA}, \mathrm{EU}, \mathrm{AP})$ are proxied by stock indices #1–#3, respectively. In Scenario #2, World (WD) proxied by stock index #4 is added to the model as the fourth variable. In Scenario #3, $(\mathrm{NA}, \mathrm{EU}, \mathrm{AP})$ are proxied by stock indices #1, #5, #3, respectively. The remaining scenarios are understood analogously. Details of the stock indices are listed in Table 3.

Table 5: Sample statistics of the daily log-returns of the stock indices (Post-Covid)

		$(\times 10^{-4})$	$(\times 10^{-4})$			$(\times 10^{-2})$		
#	ticker	mean	median	min	max	stdev	skew	kurt
1	A1SGI	6.46	5.40	−0.063	0.053	1.069	−0.341	5.705
2	DJSEUR	3.57	5.72	−0.038	0.047	0.913	−0.046	6.193
3	P1SGI	3.46	5.79	−0.046	0.046	1.049	−0.088	4.112
4	W1DOW	4.79	6.96	−0.047	0.042	0.900	−0.279	5.371
5	DJSEURD	3.45	4.71	−0.054	0.060	1.130	−0.098	6.189
6	DJSEUZ	4.62	5.95	−0.053	0.068	1.152	0.046	7.025
7	AASGI	6.57	5.77	−0.063	0.054	1.091	−0.329	5.673
8	A1SGITR	7.22	5.82	−0.063	0.053	1.069	−0.338	5.705
9	DJSEURT	4.65	7.79	−0.038	0.047	0.911	−0.032	6.146
10	P1SGITR	4.56	6.13	−0.046	0.046	1.049	−0.089	4.116
11	DJUS	6.04	3.75	−0.061	0.056	1.109	−0.389	5.340
12	E1DOW	3.88	5.80	−0.054	0.060	1.179	−0.143	5.943
13	P1DOW	2.34	5.81	−0.040	0.046	0.914	−0.137	4.683

Post-Covid period: 05/01/2020 – 07/31/2024 ($n = 1041$). In this table, we report sample mean, median, minimum, maximum, standard deviation, skewness, and kurtosis of the daily log-return ($y_t = \ln P_t - \ln P_{t-1}$) of each stock index. Details of the stock indices are listed in Table 3.

Table 6: Correlation matrix of the daily log-returns of the stock indices (Post-Covid)

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13
#1	1	0.56	0.26	0.92	0.54	0.58	1.00	1.00	0.56	0.26	0.98	0.56	0.29
#2		1	0.36	0.70	0.91	0.95	0.54	0.56	1.00	0.36	0.54	0.89	0.33
#3			1	0.50	0.47	0.36	0.24	0.26	0.36	1.00	0.25	0.49	0.81
#4				1	0.73	0.72	0.91	0.92	0.70	0.50	0.94	0.76	0.53
#5					1	0.91	0.52	0.54	0.91	0.47	0.53	0.99	0.44
#6						1	0.56	0.58	0.95	0.36	0.56	0.90	0.34
#7							1	1.00	0.55	0.24	0.98	0.53	0.27
#8								1	0.56	0.26	0.98	0.56	0.29
#9									1	0.36	0.54	0.89	0.33
#10										1	0.25	0.49	0.81
#11											1	0.55	0.29
#12												1	0.47
#13													1

Post-Covid period: 05/01/2020 – 07/31/2024 ($n = 1041$). In this table, we report the sample correlation coefficient matrix of the daily log-returns of the stock indices. Details of the stock indices are listed in Table 3.

Table 7: Bootstrap p-values of the Granger causality tests

Scenario #1: Main scenario

Sample period	EU↛NA	AP↛NA	NA↛EU	AP↛EU	NA↛AP	EU↛AP
Pre-GFC	0.753	0.899	0.000***	0.336	0.000***	0.000***
Bet-GFC&Covid	0.200	0.125	0.000***	0.918	0.000***	0.000***
Post-Covid	0.883	0.942	0.000***	0.479	0.000***	0.000***

Scenario #2: World is added as the fourth variable

Sample period	EU↛NA	AP↛NA	NA↛EU	AP↛EU	NA↛AP	EU↛AP
Pre-GFC	0.462	0.342	0.000***	0.504	0.026**	0.000***
Bet-GFC&Covid	0.678	0.920	0.000***	0.641	0.001***	0.000***
Post-Covid	0.820	0.869	0.023**	0.564	0.812	0.004***

Scenario #3: Currency of the Europe Index is changed from euros to dollars

Sample period	EU↛NA	AP↛NA	NA↛EU	AP↛EU	NA↛AP	EU↛AP
Pre-GFC	0.908	0.889	0.000***	0.030**	0.000***	0.000***
Bet-GFC&Covid	0.168	0.270	0.000***	0.704	0.000***	0.000***
Post-Covid	0.702	0.953	0.000***	0.754	0.000***	0.000***

NA: North America. EU: Europe. AP: Asia Pacific. The scenarios are defined in Table 4. We perform Granger causality tests with the wild bootstrap of Hafner and Herwartz (2009). The number of bootstrap iterations is $J = 2000$. This table reports the bootstrap p-values associated with the null hypothesis of Granger non-causality from one region to another. EU↛NA, for example, signifies non-causality from EU to NA. Asterisks ***, **, and * indicate a rejection of the non-causality hypothesis at the 1%, 5%, and 10% levels, respectively.

Table 7: Bootstrap p-values of the Granger causality tests (continued)

Scenario #4: The Europe Index is replaced with the Eurozone Index

Sample period	EU↛NA	AP↛NA	NA↛EU	AP↛EU	NA↛AP	EU↛AP
Pre-GFC	0.879	0.817	0.000***	0.309	0.000***	0.000***
Bet-GFC&Covid	0.173	0.118	0.000***	0.806	0.000***	0.000***
Post-Covid	0.886	0.896	0.000***	0.430	0.000***	0.000***

Scenario #5: The North America Index is replaced with the U.S. Index

Sample period	EU↛NA	AP↛NA	NA↛EU	AP↛EU	NA↛AP	EU↛AP
Pre-GFC	0.754	0.687	0.000***	0.214	0.000***	0.000***
Bet-GFC&Covid	0.173	0.086*	0.000***	0.857	0.000***	0.000***
Post-Covid	0.876	0.922	0.000***	0.544	0.000***	0.000***

Scenario #6: Price returns are replaced with net total returns

Sample period	EU↛NA	AP↛NA	NA↛EU	AP↛EU	NA↛AP	EU↛AP
Pre-GFC	0.709	0.904	0.000***	0.304	0.000***	0.000***
Bet-GFC&Covid	0.212	0.110	0.000***	0.886	0.000***	0.000***
Post-Covid	0.848	0.926	0.000***	0.449	0.000***	0.000***

Scenario #7: The DJSIs are replaced with the DJIs

Sample period	EU↛NA	AP↛NA	NA↛EU	AP↛EU	NA↛AP	EU↛AP
Pre-GFC	0.702	0.645	0.000***	0.065*	0.000***	0.000***
Bet-GFC&Covid	0.119	0.053*	0.000***	0.458	0.000***	0.000***
Post-Covid	0.847	0.862	0.000***	0.815	0.000***	0.000***

NA: North America. EU: Europe. AP: Asia Pacific. The scenarios are defined in Table 4. We perform Granger causality tests with the wild bootstrap of Hafner and Herwartz (2009). The number of bootstrap iterations is $J = 2000$. This table reports the bootstrap p-values associated with the null hypothesis of Granger non-causality from one region to another. EU↛NA, for example, signifies non-causality from EU to NA. Asterisks ***, **, and * indicate a rejection of the non-causality hypothesis at the 1%, 5%, and 10% levels, respectively.

Table 8: Forecast error variance decomposition in the long-run ($h = 4$)

Scenario #1: Main scenario

1. Pre-GFC				2. Bet-GFC&Covid				3. Post-Covid			
(%)	NA	EU	AP	(%)	NA	EU	AP	(%)	NA	EU	AP
NA	100	33.6	25.8	NA	99.5	48.9	34.9	NA	100	35.0	32.7
EU	0.0	66.3	8.2	EU	0.2	51.1	3.2	EU	0.0	64.9	5.5
AP	0.0	0.1	66.0	AP	0.3	0.0	61.9	AP	0.0	0.1	61.8
Total	100	100	100	Total	100	100	100	Total	100	100	100
Spillover Index = 22.6%				Spillover Index = 29.2%				Spillover Index = 24.4%			

Scenario #2: World is added as the fourth variable

1. Pre-GFC				2. Bet-GFC&Covid				3. Post-Covid			
(%)	NA	EU	AP	(%)	NA	EU	AP	(%)	NA	EU	AP
NA	99.8	33.7	25.9	NA	99.3	49.1	34.8	NA	100	35.0	32.7
EU	0.0	66.2	8.2	EU	0.3	50.9	3.3	EU	0.0	64.9	5.5
AP	0.0	0.1	64.8	AP	0.3	0.0	61.5	AP	0.0	0.0	60.2
WD	0.2	0.0	1.2	WD	0.2	0.1	0.4	WD	0.0	0.0	1.6
Total	100	100	100	Total	100	100	100	Total	100	100	100
Spillover Index = 40.1%				Spillover Index = 46.1%				Spillover Index = 42.4%			

NA: North America. EU: Europe. AP: Asia Pacific. WD: World. The scenarios are defined in Table 4. We implement the forecast error variance decomposition at horizon $h = 4$. The forecast error variance of AP for Scenario #1 in the Pre-GFC period, for instance, is explained 25.8% by NA, 8.2% by EU, and 66.0% by AP itself. The spillover index of Diebold and Yilmaz (2009) is the sum of off-diagonal elements divided by the sum of all elements.

Table 8: Forecast error variance decomposition in the long-run ($h = 4$; continued)

Scenario #3: Currency of the Europe Index is changed from euros to dollars

1. Pre-GFC				2. Bet-GFC&Covid				3. Post-Covid			
(%)	NA	EU	AP	(%)	NA	EU	AP	(%)	NA	EU	AP
NA	100	29.6	25.7	NA	99.4	48.0	34.9	NA	100	35.9	32.7
EU	0.0	69.9	11.3	EU	0.4	52.0	6.6	EU	0.0	64.1	10.5
AP	0.0	0.5	63.0	AP	0.2	0.1	58.5	AP	0.0	0.0	56.8
Total	100	100	100	Total	100	100	100	Total	100	100	100
Spillover Index = 22.4%				Spillover Index = 30.0%				Spillover Index = 26.4%			

Scenario #4: The Europe Index is replaced with the Eurozone Index

1. Pre-GFC				2. Bet-GFC&Covid				3. Post-Covid			
(%)	NA	EU	AP	(%)	NA	EU	AP	(%)	NA	EU	AP
NA	100	34.3	25.9	NA	99.5	49.1	34.8	NA	100	36.2	32.8
EU	0.0	65.6	8.1	EU	0.3	50.9	3.2	EU	0.0	63.8	6.1
AP	0.0	0.1	66.0	AP	0.3	0.0	62.0	AP	0.0	0.1	61.2
Total	100	100	100	Total	100	100	100	Total	100	100	100
Spillover Index = 22.8%				Spillover Index = 29.2%				Spillover Index = 25.0%			

NA: North America. EU: Europe. AP: Asia Pacific. The scenarios are defined in Table 4. We implement the forecast error variance decomposition at horizon $h = 4$. The forecast error variance of AP for Scenario #3 in the Pre-GFC period, for instance, is explained 25.7% by NA, 11.3% by EU, and 63.0% by AP itself. The spillover index of Diebold and Yilmaz (2009) is the sum of off-diagonal elements divided by the sum of all elements.

Table 8: Forecast error variance decomposition in the long-run ($h = 4$; continued)

Scenario #5: The North America Index is replaced with the U.S. Index

1. Pre-GFC				2. Bet-GFC&Covid				3. Post-Covid			
(%)	NA	EU	AP	(%)	NA	EU	AP	(%)	NA	EU	AP
NA	100	32.6	24.2	NA	99.4	46.5	33.5	NA	100	33.2	31.5
EU	0.0	67.2	9.3	EU	0.3	53.5	4.2	EU	0.0	66.7	6.3
AP	0.0	0.2	66.5	AP	0.3	0.0	62.3	AP	0.0	0.0	62.2
Total	100	100	100	Total	100	100	100	Total	100	100	100
Spillover Index = 22.1%				Spillover Index = 28.3%				Spillover Index = 23.7%			

Scenario #6: Price returns are replaced with net total returns

1. Pre-GFC				2. Bet-GFC&Covid				3. Post-Covid			
(%)	NA	EU	AP	(%)	NA	EU	AP	(%)	NA	EU	AP
NA	100	33.7	25.8	NA	99.5	48.8	34.8	NA	100	35.1	32.8
EU	0.0	66.2	8.3	EU	0.2	51.2	3.3	EU	0.0	64.8	5.6
AP	0.0	0.1	65.9	AP	0.3	0.0	61.9	AP	0.0	0.1	61.7
Total	100	100	100	Total	100	100	100	Total	100	100	100
Spillover Index = 22.6%				Spillover Index = 29.2%				Spillover Index = 24.5%			

NA: North America. EU: Europe. AP: Asia Pacific. The scenarios are defined in Table 4. We implement the forecast error variance decomposition at horizon $h = 4$. The forecast error variance of AP for Scenario #5 in the Pre-GFC period, for instance, is explained 24.2% by NA, 9.3% by EU, and 66.5% by AP itself. The spillover index of Diebold and Yilmaz (2009) is the sum of off-diagonal elements divided by the sum of all elements.

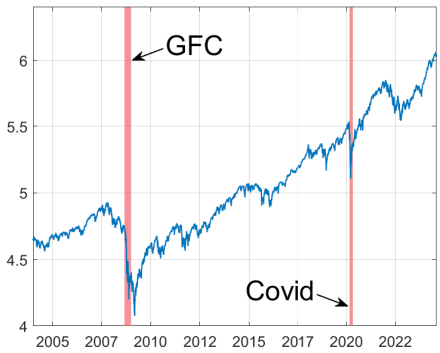
Table 8: Forecast error variance decomposition in the long-run ($h = 4$; continued)

Scenario #7: The DJSIs are replaced with the DJIs

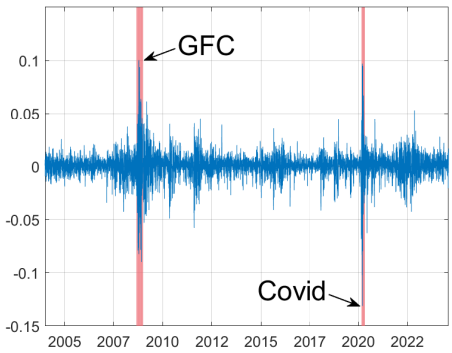
1. Pre-GFC				2. Bet-GFC&Covid				3. Post-Covid			
(%)	NA	EU	AP	(%)	NA	EU	AP	(%)	NA	EU	AP
NA	100	29.5	27.5	NA	99.1	46.7	31.1	NA	100	37.0	31.0
EU	0.0	70.2	11.9	EU	0.6	53.3	7.4	EU	0.0	63.0	9.4
AP	0.0	0.4	60.5	AP	0.4	0.1	61.6	AP	0.0	0.0	59.7
Total	100	100	100	Total	100	100	100	Total	100	100	100
Spillover Index = 23.1%				Spillover Index = 28.7%				Spillover Index = 25.8%			

NA: North America. EU: Europe. AP: Asia Pacific. The scenarios are defined in Table 4. We implement the forecast error variance decomposition at horizon $h = 4$. The forecast error variance of AP for Scenario #5 in the Pre-GFC period, for instance, is explained 24.2% by NA, 9.3% by EU, and 66.5% by AP itself. The spillover index of Diebold and Yilmaz (2009) is the sum of off-diagonal elements divided by the sum of all elements.

Figure 1: Daily log-values and log-returns of the stock indices

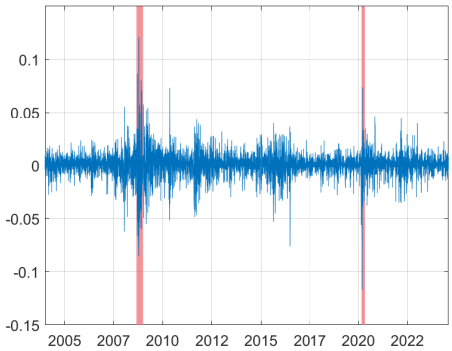


#1. A1SGI (log-value)



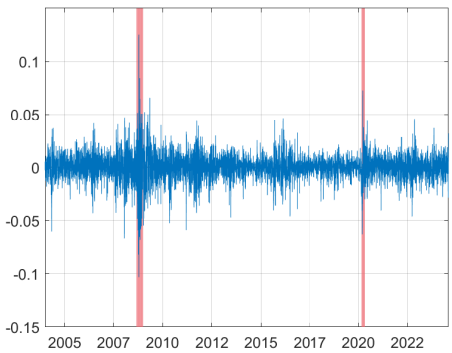
#1. A1SGI (log-return)

#2. DJSEUR (log-value)



#2. DJSEUR (log-return)

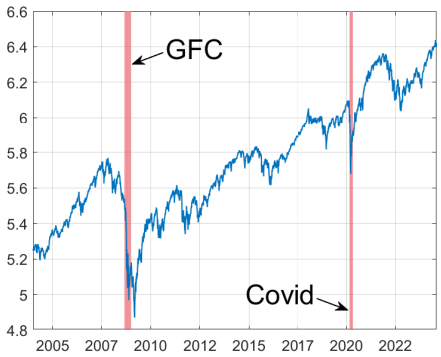
#3. P1SGI (log-value)



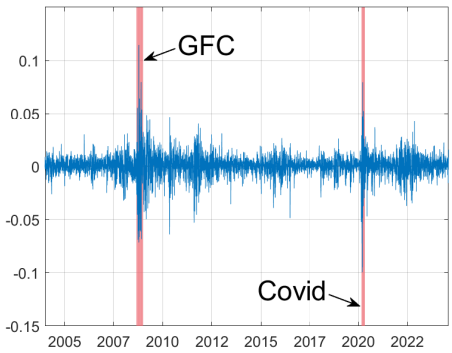
#3. P1SGI (log-return)

This figure plots daily log-values ($Y_t = \ln P_t$) and log-returns ($y_t = \ln P_t - \ln P_{t-1}$) of the stock indices. The first shade represents the Global Financial Crisis (09/01/2008 – 12/30/2008). The second shade represents the COVID-19 crisis (03/02/2020 – 04/30/2020). Details of the stock indices are listed in Table 3.

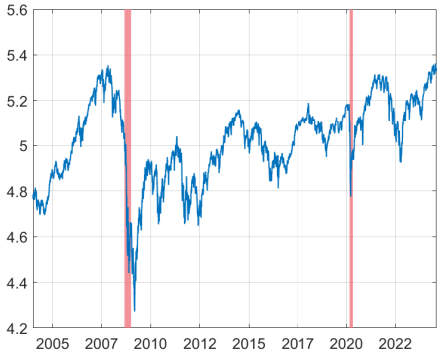
Figure 1: Daily log-values and log-returns of the stock indices (continued)



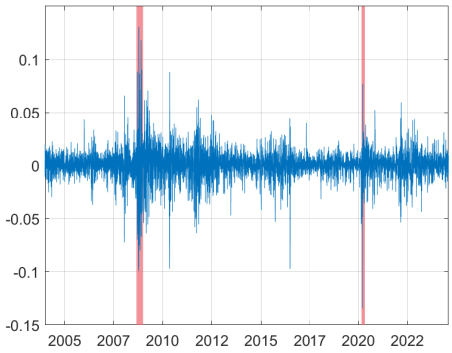
#4. W1DOW (log-value)



#4. W1DOW (log-return)

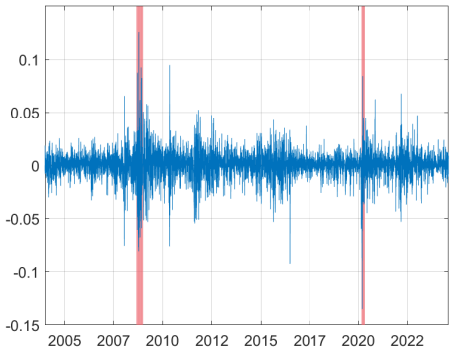


#5. DJSEURD (log-value)



#5. DJSEURD (log-return)

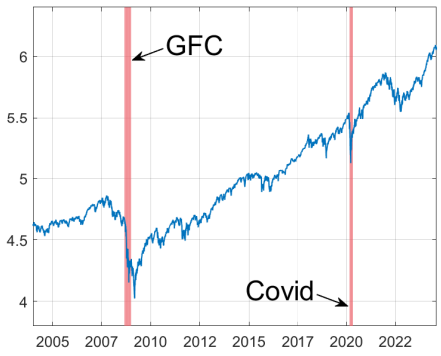
#6. DJSEUZ (log-value)



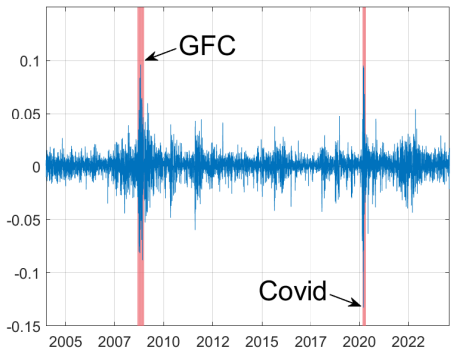
#6. DJSEUZ (log-return)

This figure plots daily log-values ($Y_t = \ln P_t$) and log-returns ($y_t = \ln P_t - \ln P_{t-1}$) of the stock indices. The first shade represents the Global Financial Crisis (09/01/2008 – 12/30/2008). The second shade represents the COVID-19 crisis (03/02/2020 – 04/30/2020). Details of the stock indices are listed in Table 3.

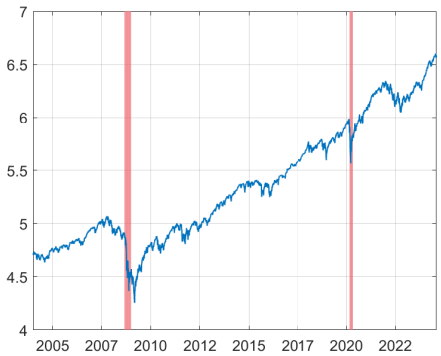
Figure 1: Daily log-values and log-returns of the stock indices (continued)



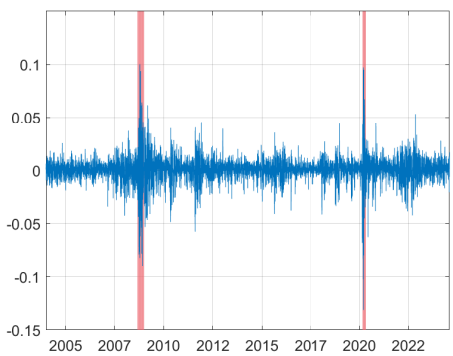
#7. AASGI (log-value)



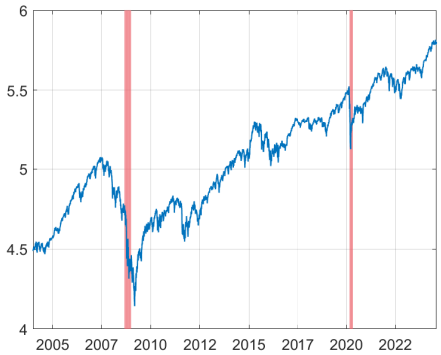
#7. AASGI (log-return)



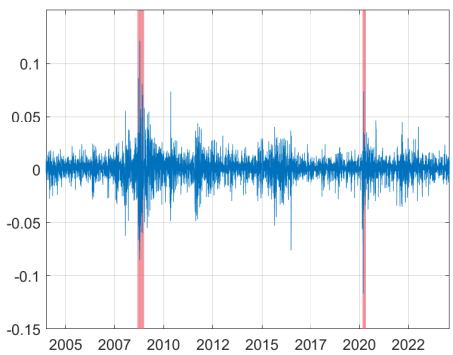
#8. A1SGITR (log-value)



#8. A1SGITR (log-return)



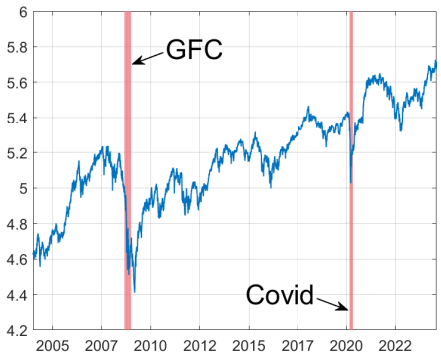
#9. DJSEURT (log-value)



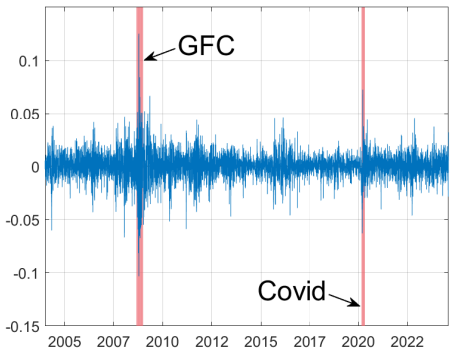
#9. DJSEURT (log-return)

This figure plots daily log-values ($Y_t = \ln P_t$) and log-returns ($y_t = \ln P_t - \ln P_{t-1}$) of the stock indices. The first shade represents the Global Financial Crisis (09/01/2008 – 12/30/2008). The second shade represents the COVID-19 crisis (03/02/2020 – 04/30/2020). Details of the stock indices are listed in Table 3.

Figure 1: Daily log-values and log-returns of the stock indices (continued)

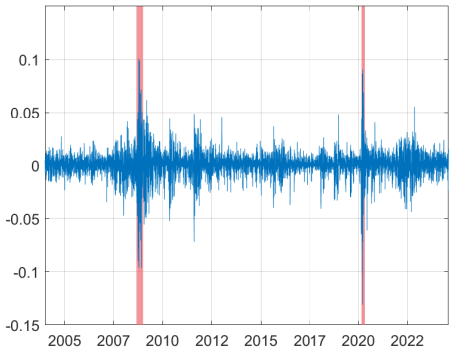


#10. P1SGITR (log-value)



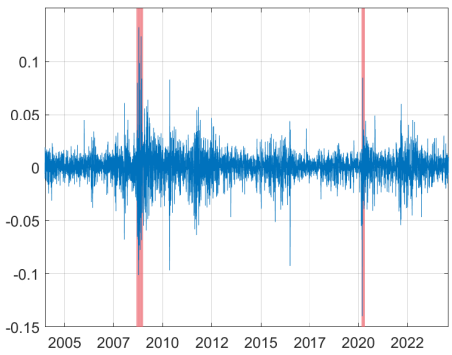
#10. P1SGITR (log-return)

#11. DJUS (log-value)



#11. DJUS (log-return)

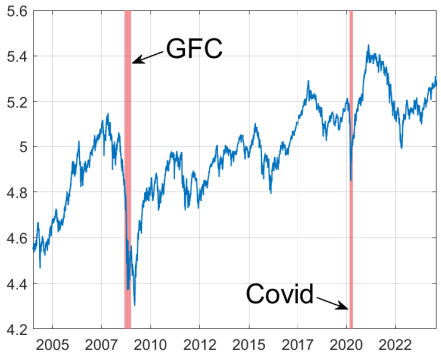
#12. E1DOW (log-value)



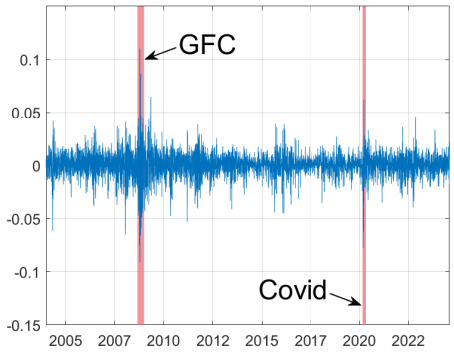
#12. E1DOW (log-return)

This figure plots daily log-values ($Y_t = \ln P_t$) and log-returns ($y_t = \ln P_t - \ln P_{t-1}$) of the stock indices. The first shade represents the Global Financial Crisis (09/01/2008 – 12/30/2008). The second shade represents the COVID-19 crisis (03/02/2020 – 04/30/2020). Details of the stock indices are listed in Table 3.

Figure 1: Daily log-values and log-returns of the stock indices (continued)



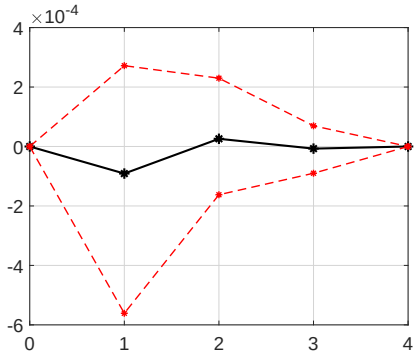
#13. P1DOW (log-value)



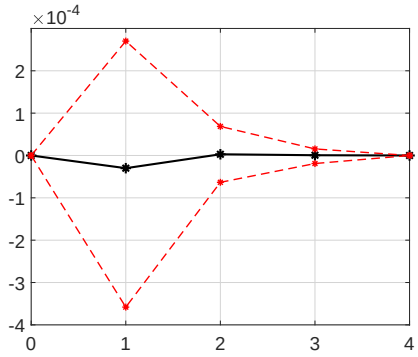
#13. P1DOW (log-return)

This figure plots daily log-values ($Y_t = \ln P_t$) and log-returns ($y_t = \ln P_t - \ln P_{t-1}$) of the stock indices. The first shade represents the Global Financial Crisis (09/01/2008 – 12/30/2008). The second shade represents the COVID-19 crisis (03/02/2020 – 04/30/2020). Details of the stock indices are listed in Table 3.

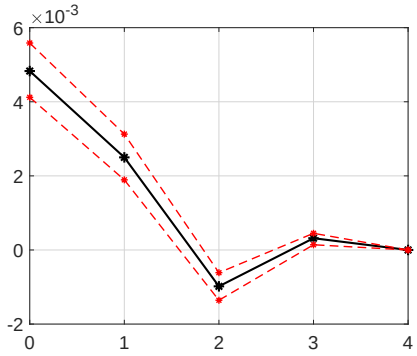
Figure 2: Impulse response functions and confidence intervals (Scenario #1, Pre-GFC)



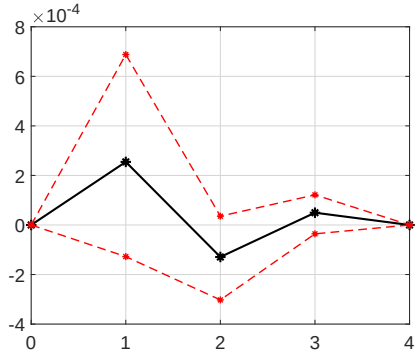
1. IRF from EU to NA ($\hat{\theta}_{12,h}$)



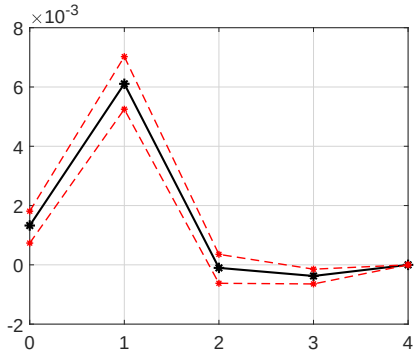
2. IRF from AP to NA ($\hat{\theta}_{13,h}$)



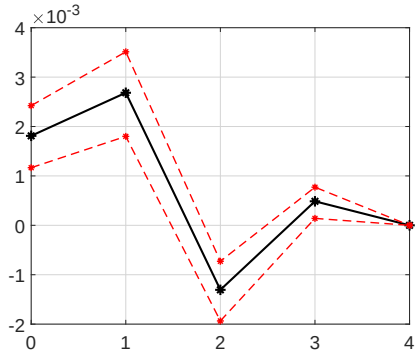
3. IRF from NA to EU ($\hat{\theta}_{21,h}$)



4. IRF from AP to EU ($\hat{\theta}_{23,h}$)



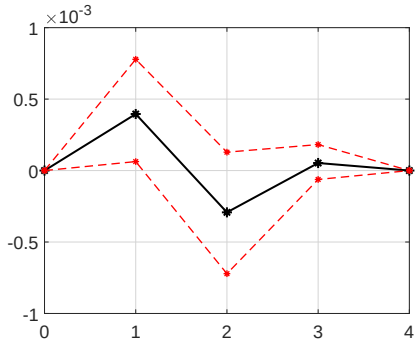
5. IRF from NA to AP ($\hat{\theta}_{31,h}$)



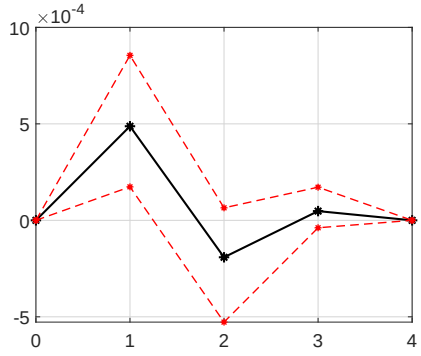
6. IRF from EU to AP ($\hat{\theta}_{32,h}$)

NA: North America. EU: Europe. AP: Asia Pacific. Scenario #1: Main scenario. We plot IRFs from a region to another at horizons $h \in \{0, \ldots, 4\}$ for the Pre-GFC period. Panel 1, for instance, depicts how NA responds to a 1σ shock in EU. The black, solid line depicts the IRF, and the red, dashed lines depict the 90% bootstrap confidence interval. When both the lower and upper bounds of the confidence interval are positive (resp. negative), the corresponding IRF is significantly positive (resp. negative) at the 10% level. The number of bootstrap iterations is $J = 2000$.

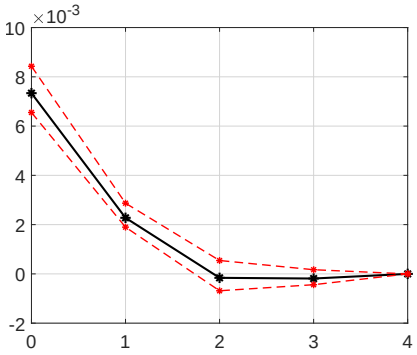
Figure 3: Impulse response functions and confidence intervals (Scenario #1, Bet-GFC&Covid)



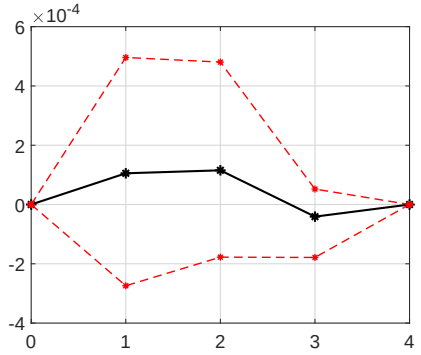
1. IRF from EU to NA ($\hat{\theta}_{12,h}$)



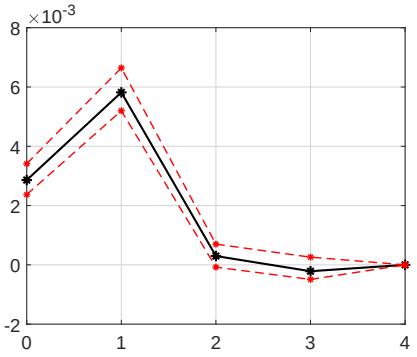
2. IRF from AP to NA ($\hat{\theta}_{13,h}$)



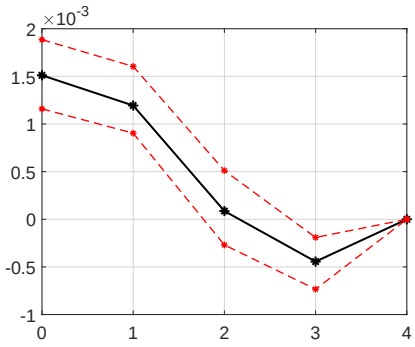
3. IRF from NA to EU ($\hat{\theta}_{21,h}$)



4. IRF from AP to EU ($\hat{\theta}_{23,h}$)



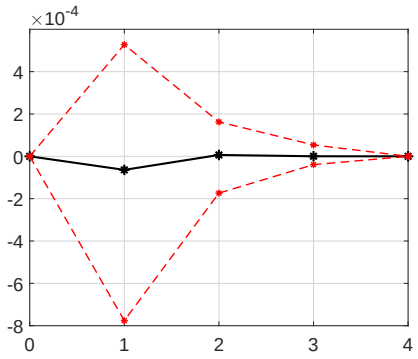
5. IRF from NA to AP ($\hat{\theta}_{31,h}$)



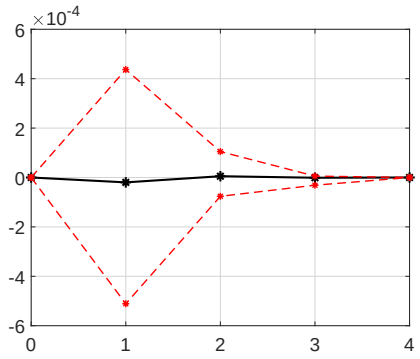
6. IRF from EU to AP ($\hat{\theta}_{32,h}$)

NA: North America. EU: Europe. AP: Asia Pacific. Scenario #1: Main scenario. We plot IRFs from a region to another at horizons $h \in \{0, \ldots, 4\}$ for the Bet-GFC&Covid period. Panel 1, for instance, depicts how NA responds to a 1σ shock in EU. The black, solid line depicts the IRF, and the red, dashed lines depict the 90% bootstrap confidence interval. When both the lower and upper bounds of the confidence interval are positive (resp. negative), the corresponding IRF is significantly positive (resp. negative) at the 10% level. The number of bootstrap iterations is $J = 2000$.

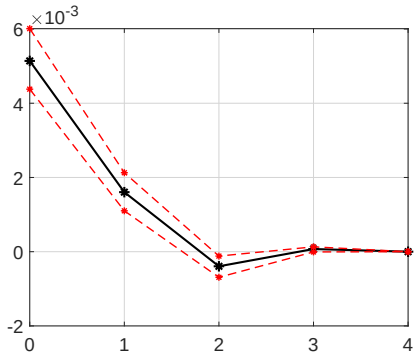
Figure 4: Impulse response functions and confidence intervals (Scenario #1, Post-Covid)



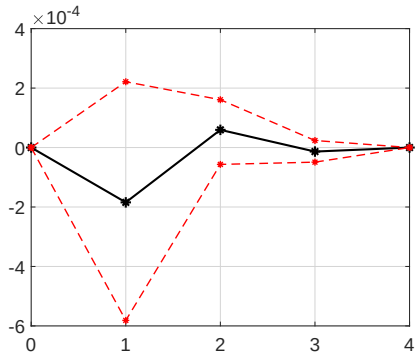
1. IRF from EU to NA ($\hat{\theta}_{12,h}$)



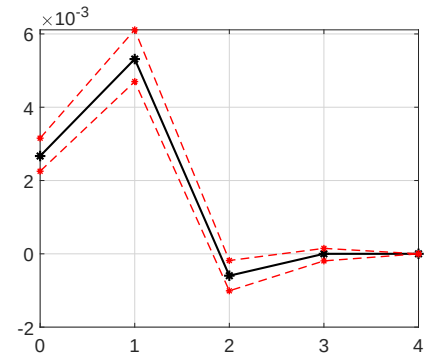
2. IRF from AP to NA ($\hat{\theta}_{13,h}$)



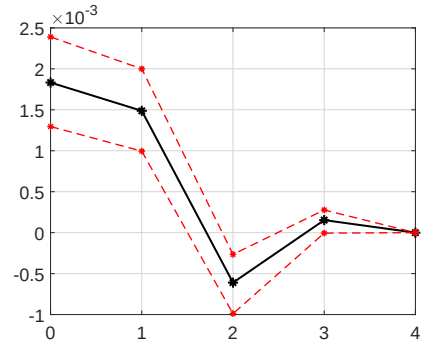
3. IRF from NA to EU ($\hat{\theta}_{21,h}$)



4. IRF from AP to EU ($\hat{\theta}_{23,h}$)



5. IRF from NA to AP ($\hat{\theta}_{31,h}$)



6. IRF from EU to AP ($\hat{\theta}_{32,h}$)

NA: North America. EU: Europe. AP: Asia Pacific. Scenario #1: Main scenario. We plot IRFs from a region to another at horizons $h \in \{0, \ldots, 4\}$ for the Post-Covid period. Panel 1, for instance, depicts how NA responds to a 1σ shock in EU. The black, solid line depicts the IRF, and the red, dashed lines depict the 90% bootstrap confidence interval. When both the lower and upper bounds of the confidence interval are positive (resp. negative), the corresponding IRF is significantly positive (resp. negative) at the 10% level. The number of bootstrap iterations is $J = 2000$.