# Structural Volatility Impulse Response Analysis\*

Matthias Fengler<sup>‡</sup>

Jeannine Polivka<sup>§</sup>

November 4, 2024

#### Abstract

We make three contributions to the volatility impulse response function (VIRF) developed by Hafner and Herwartz (2006). First, we derive its law for multivariate GARCH models of the BEKK type. Second, we present a structural embedding of the VIRF, leveraging recent advancements in the identification of MGARCH models. Third, we show how to endow the VIRF with a causal interpretation. We illustrate the merits of a structural VIRF analysis by investigating the impacts of historical and out-of-sample shock scenarios on the U.S. equity, government bond, and foreign exchange markets.

**Keywords:** causality in volatility, multivariate GARCH models, proxy identification, structural identification, volatility impulse response functions

**JEL Codes:** C32, C58, G17

<sup>\*</sup>This work was supported by the Swiss National Science Foundation [176684], Project "Structural Models of Volatility".

<sup>&</sup>lt;sup>‡</sup>SFI, School of Economics and Political Science, Department of Economics, University of St. Gallen, Bodanstrasse 6, 9000 St. Gallen, Switzerland. Phone: +41 71 224 2457, Email: matthias.fengler@unisg.ch.

<sup>&</sup>lt;sup>§</sup>School of Economics and Political Science, Department of Economics, University of St. Gallen, Bodanstrasse 6, 9000 St. Gallen, Switzerland. Email: jeannine.polivka@unisg.ch.

## 1 Introduction

The impulse response function (IRF) is a primary tool for analyzing how dynamic multivariate systems respond to shocks. It illustrates the effects of an unanticipated disturbance on the modeled variables, revealing how the sign, magnitude, and overall duration of the response evolve across the forecast horizon. In vector autoregressions, IRF analysis focuses on feedback within the mean process (Lütkepohl, 2010; Inoue and Kilian, 2013). However, in volatility models of high-frequency speculative asset returns—where the mean equation often carries little significance—the second-order response becomes the central focus. For example, a mutual fund manager might be interested in how the variance matrix of specific asset classes responds to an unexpected policy shock. In such situations, the volatility impulse response function (VIRF) is the appropriate tool.

The VIRF, initially proposed by Hafner and Herwartz (2006), extends the concept of the generalized IRF (GIRF) due to Koop et al. (1996) to second-order moments. By conditioning on past information and an exogenous shock, it traces that shock's nonlinear impact on volatility dynamics. Among the various VIRF specifications proposed (Gallant et al., 1993; Liu, 2018), the approach of Hafner and Herwartz (2006) is of special significance, as it permits a closed-form solution in one of the most flexible multivariate generalized autoregressive conditional heteroskedasticity (MGARCH) models: the BEKK(p, q) model introduced by Baba, Engle, Kraft, and Kroner (Engle and Kroner, 1995).

In this work, we revisit the VIRF and contribute to the literature in three key ways: First, we derive the VIRF's asymptotic distribution in the BEKK model. Specifically, we show that, like the VIRF itself, its asymptotic variance matrix can be expressed as a function of the forecast horizon in a compact recursive form. This approach enables an efficient numerical computation of confidence intervals, which previously required time-intensive simulation techniques such as the bootstrap. A comparative study of both methods reveals that our newly available asymptotic confidence intervals provide researchers with a valuable tool for assessing the volatility impact of a given shock within an MGARCH model.

As a second contribution, we provide the VIRF with a structural interpretation by leveraging recent advances in the identification of MGARCH models (Hafner et al., 2022; Fengler and Polivka, 2024). This new interpretation materially broadens its range of applications. Traditionally, the VIRF has been used to study the impact of a realized shock on volatility—a use case we term 'historical VIRF analysis.' By incorporating identified and labeled structural shocks derived from a structural MGARCH model, we can construct specific, interpretable shocks for scenario analyses, similar to those in structural vector autoregressive (VAR) models (Amisano and Giannini, 2012; Kilian and Lütkepohl, 2017). This approach allows for defining counterfactual scenarios and assessing the expected volatility impact of well-defined shock scenarios. We introduce the term 'scenario VIRF' to describe this new application framework.

Our third contribution is to equip the VIRF with a causal interpretation. Drawing on advances in causal inference for time series by Rambachan and Shephard (2020, 2021), we examine the conditions under which the microeconometricians' concept of causality can be applied—specifically, when the VIRF can be understood as representing dynamic causal effects of assignments on outcomes. In other words, we provide the conditions under which the VIRF, obtained for a given tail event, measures how changes in shocks (i.e., assignments) dynamically cause movements in second-order moments over a given horizon. These results align with recent efforts in macroeconometrics to provide a causal interpretation of impulse responses in structural VARs and local projections—see, e.g., Cloyne et al. (2023) and Gonçalves et al. (2024).

Studying the impact of shocks on financial volatility is essential in applications across asset pricing, risk management, and portfolio optimization. The development of tools to measure volatility responses has thus been a longstanding focus in financial econometrics, beginning with the work of Gallant et al. (1993) and Lin (1997). However, both

approaches rely on correlated reduced-form errors, which limits the meaningful analysis of individual shock components. Hafner and Herwartz (2006) address this shortcoming by proposing a VIRF, which is based on mutually orthogonal shocks. Nevertheless, their identification assumption rests solely on the principal matrix square root of the variance matrix sequence. This limitation also applies to recent advances, such as the VIRF for Markov-switching MGARCH models in Cavicchioli (2019) and the asymmetric VIRF proposed by Hafner and Herwartz (2023).

Despite this history, structural advancements in volatility impulse response analysis remain in their early stages. Initial identification efforts leveraging time-varying heteroscedasticity were made by Weber (2010) and Liu (2018), both within conditional correlation models. More recently, Hafner et al. (2022) used information from third and fourth-order moments to identify structural shocks based on independence and non-normality assumptions. While these strategies successfully identify orthogonal structural shocks, they lack economic interpretability and thus are limited to using realized model-implied shocks. To address this issue, Fengler and Polivka (2024) propose employing external instrument data for identification. A major advantage of this approach is that it yields economically labeled shocks, at least for those targeted by the instrument data. In this work, we extend that framework by conducting a structural VIRF analysis using labeled structural shocks.

Lastly, few studies address the asymptotic properties of the VIRF for statistical inference, with Liu (2018) being a notable exception, though based on a distinct MGARCH model and a VIRF definition different from that of Hafner and Herwartz (2006). Additionally, we are the first to provide a discussion of causality for VIRFs.

The remainder is organized as follows. Section 2 introduces the VIRF and presents key results on its asymptotic distribution and connections to causal inference. Section 3 discusses the estimation of the structural model and provides both historical and out-of-sample scenario VIRFs for well-defined risk scenarios. Section 4 concludes. Appendix A offers an overview of definitions, relevant matrix algebra results, and the

proofs.

## 2 Volatility impulse response analysis

### 2.1 Modeling framework

We consider a system of *n* speculative (log) returns for  $t \in \mathbb{Z}$ , driven by

$$r_t = \mu_t + \varepsilon_t , \qquad (1)$$

where  $\mu_t = \mathbb{E}[r_t | \mathcal{F}_{t-1}]$  and  $\mathcal{F}_t = \sigma(\{\varepsilon_s : s \leq t\})$  denotes the  $\sigma$ -algebra generated by the n-dimensional innovation process  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  up to and including time t. The innovations  $\varepsilon_t$  have a conditional mean  $\mathbb{E}[\varepsilon_t | \mathcal{F}_{t-1}] = 0$  and a conditional covariance  $\operatorname{Var}[\varepsilon_t | \mathcal{F}_{t-1}] = H_t$ , where  $H_t \in \mathbb{R}^{n \times n}$  is almost surely symmetric and positive definite for all t. They are assumed to follow the structural MGARCH process defined by

$$\varepsilon_t = H_t^{1/2} \tilde{R} \xi_t \,, \tag{2}$$

where  $\tilde{R}$  is an orthogonal matrix (see Definition D.2) and  $\{\xi_t\}_{t\in\mathbb{Z}}$  an *n*-dimensional real-valued strict white noise process of structural shocks with zero mean and identity covariance matrix, i.e.,  $\xi_t \stackrel{\text{iid}}{\sim} SWN(0, I_n)$ . Throughout this work, we denote by  $A^{1/2}$  the principal matrix square root of a symmetric positive definite square matrix A, which is the unique symmetric positive definite matrix such that  $A = A^{1/2}A^{1/2}$  (see Definition D.1).

The structural volatility model presented in (2) was introduced by Hafner et al. (2022), who specify  $\tilde{R}$  as a rotation matrix. As in Fengler and Polivka (2024), we model  $\tilde{R}$  more broadly as an orthogonal matrix. Like a structural VAR model, model (2) establishes a specific shock composition mechanism, which is represented by  $\tilde{R}$  and endows the shocks  $\xi_t$  with a structural interpretation. For example, choosing  $\tilde{R}$  as the identity matrix implies symmetric volatility spillovers, as  $H_t^{1/2}$  preserves the positive definite-

ness and symmetry of  $H_t$ . With other choices of  $\tilde{R}$ , mixtures of the structural shocks emerge, which are then transformed into the reduced-form shocks  $\varepsilon_t$  by  $H_t^{1/2}$ .

Because any choice for  $\tilde{R}$  is observationally equivalent up to second order, additional information is required for identification. Hafner et al. (2022) develop an identification scheme based on the assumption of non-Gaussianity of  $\xi_t$ . Using additional external data sources, known as instrument or proxy variables, Fengler and Polivka (2024) suggest a framework for proxy-identification. Proxy-identification, by not relying exclusively on statistical assumptions, offers the further benefit of allowing for shock labeling, i.e., a meaningful economic interpretation. We sketch the ideas of proxy-identification in Section 3.1.

### 2.2 The volatility impulse response function

In order to assess the impact of a shock  $\xi_t$  on volatility given  $\mathcal{F}_{t-1}$ , Hafner and Herwartz (2006) define the *h*-step ahead VIRF as the difference between the expected *h*-step-ahead covariance conditioned on the shock and past information, and a natural baseline, the expected *h*-step-ahead covariance conditioned on past information only. Note that, although the VIRF bears the term *volatility* in its name, it is, in fact, a vector of impulse responses of the conditional (*co*)*variances* to the shock  $\xi_t = \bar{\xi}^*$ .

**Definition 2.1.** Let  $h \in \mathbb{N}$ . The *h*-step ahead VIRF is given by

$$V_{t+h}(\bar{\xi}^*) \coloneqq E[\operatorname{vech}(H_{t+h})|\mathcal{F}_{t-1}, \xi_t = \bar{\xi}^*] - E[\operatorname{vech}(H_{t+h})|\mathcal{F}_{t-1}]$$
(3)

and its dimension is  $n^* = \frac{n(n+1)}{2}$  (see Definition D.4 for the vech-operator). This definition follows the tradition of the GIRF as developed by Koop et al. (1996).

By definition, the VIRF represents the expectation of possible future scenarios, conditioned on the history and a shock. Depending on the stance a researcher takes regarding the conditioning sets, different use cases of the VIRF can be imagined.<sup>1</sup> In most

<sup>&</sup>lt;sup>1</sup>See Koop et al. (1996, Sections 4.1-4.2) for further discussions of this aspect.

current applications,  $\mathcal{F}_{t-1}$  is treated as the realized history, and the shock occurring at time *t* is selected, i.e.,  $\bar{\xi}^* = \hat{\xi}_t$ , which implies that the conditioning set in the left-hand side expectation equals  $\mathcal{F}_t$ .<sup>2</sup> The VIRF then traces the volatility response of the system to an actual shock, given information accumulated by time t - 1. We want to call this use case the 'historical VIRF.'

Alternatively, one can adopt a scenario perspective, where  $\bar{\xi}^*$  is not equal to  $\hat{\xi}_t$  but is instead specified by the researcher. In this case, the challenge lies in selecting the shock of interest, as impulse responses in multivariate models depend on the entire shock vector composition rather than on a single component. Ideally, one would choose a shock from the sample of estimated shocks. Without an identified model, however, this is impossible because  $\xi_t = \tilde{R}^T H_t^{-1/2} \varepsilon_t$  is unobserved and, up to second order, observationally equivalent for any choice of  $\tilde{R}$ .

Structural MGARCH models address the composition effect problem by offering a strategy to identify and estimate  $\tilde{R}$ , and thus enable the construction of synthetic but economically meaningful structural shocks. Under the conditions provided in Section 2.5, the VIRF can be interpreted causally, allowing these shocks to be treated as counterfactuals. We introduce the term 'scenario VIRF' for this application framework and demonstrate its usefulness in Section 3.3.2.

## 2.3 The BEKK(*p*, *q*) model and its VIRF

To derive a closed-form expression for the VIRF, we must select a model for the dynamics of the conditional variance matrix process. BEKK GARCH models are especially popular (Bauwens et al., 2006) and appear to be the most commonly employed for VIRFs in the literature (Jin et al., 2012; Olson et al., 2014; Hafner and Herwartz, 2023).

<sup>&</sup>lt;sup>2</sup>See, inter alia, Lin (1997), Hafner and Herwartz (2006), Le Pen and Sévi (2010), or Jin et al. (2012).

The *n*-dimensional process  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  in (2) admits a BEKK(*p*, *q*) specification if  $H_t$  satisfies for all  $t \in \mathbb{Z}$ :

$$H_t = CC^{\top} + \sum_{i=1}^p A_i^{\top} \varepsilon_{t-i} \varepsilon_{t-i}^{\top} A_i + \sum_{j=1}^q B_j^{\top} H_{t-j} B_j \quad (p,q \in \mathbb{N}) ,$$
(4)

where *C* is a lower triangular matrix and  $A_i$ , i = 1, ..., p, and  $B_j$ , j = 1, ..., q, are coefficient matrices in  $\mathbb{R}^{n \times n}$  (Engle and Kroner, 1995). The intercept matrix  $CC^{\top}$  is symmetric and positive semi-definite by construction, and strictly positive definite if *C* has full rank, which ensures the positive definiteness of  $\{H_t\}_{t \in \mathbb{Z}}$ .

Denote the vectorized parameter vector of (4) by  $\eta = (\operatorname{vec}(C)^{\top}, \operatorname{vec}(A_i)^{\top}, \operatorname{vec}(B_j)^{\top})^{\top}$ , where  $i = 1, \ldots, p; j = 1, \ldots, q$ .<sup>3</sup> We adopt the following assumptions:

#### Assumption 2.1.

- (1) The population parameter  $\eta_0$  is in the interior of a compact parameter space.
- (2)  $\eta_0$ , and hence  $H_t(\eta_0)$ , are identifiable.
- (3)  $\{\xi_t\}_{t\in\mathbb{Z}}$  is a centered i.i.d. sequence; all  $\xi_t$  are absolutely continuous with respect to the Lebesgue measure and admit a density function such that  $\mathbb{E} \|\xi_t\|^2 \leq \infty$  and  $\operatorname{Var}[\xi_t] = I_n$ .
- (4)  $\{\varepsilon_t\}_{t\in\mathbb{Z}}$  is strictly stationary and ergodic, and  $\mathbb{E} \|\varepsilon_t\|^6 \leq \infty$ .

These assumptions ensure the consistency and asymptotic normality of the quasimaximum likelihood estimator (Hafner and Preminger, 2009). In reference to Assumption 2.1(2), identifying the BEKK(p,q) model necessitates additional sign restrictions on its parameter matrices due to the model's quadratic structure. For example, in the BEKK(1, 1) model, it is usually assumed that the diagonal elements of *C* and the top left matrix entries of  $A_1$  and  $B_1$  be positive—see Engle and Kroner (1995) for a detailed discussion. For Assumption 2.1(4), strict stationarity and ergodicity follow from mild regularity conditions detailed in Hafner and Preminger (2009). Furthermore, as

<sup>&</sup>lt;sup>3</sup>See D.3 for a definition of  $vec(\cdot)$ -operator.

discussed in Boussama et al. (2011, Theorem 2.4), the BEKK GARCH process is both strictly and weakly stationary if the eigenvalues of  $\sum_{i=1}^{p} A_i \otimes A_i + \sum_{j=1}^{q} B_j \otimes B_j$  are less than one in modulus.<sup>4</sup>

The subsequent exposition requires the VMA( $\infty$ ) representation of the 'squared' BEKK process defined by  $X_t := \operatorname{vech}(\varepsilon_t \varepsilon_t^\top)$ . Additionally, define the process  $Y_t := X_t - \operatorname{vech}(H_t)$ , which is a weak white noise under Assumption 2.1. Denote by  $D_n$  the duplication matrix and by  $D_n^+$  its Moore-Penrose inverse (see Definition D.6), and recall that  $n^* = \frac{n(n+1)}{2}$ .

**Proposition 2.1.** The VMA( $\infty$ ) representation of the 'squared' BEKK(p, q) process is given by

$$X_t = \operatorname{vech}(H) + \sum_{i=0}^{\infty} \Psi_i Y_{t-i}$$
(5)

where  $H = \text{Var}[\varepsilon_t]$ . The  $(n^* \times n^*)$  coefficient matrices  $\Psi_i$  are given by  $\Psi_0 = I_{n^*}$  and  $\Psi_i = -\tilde{B}_i + \sum_{j=1}^i (\tilde{A}_j + \tilde{B}_j) \Psi_{i-j}$ , where  $\tilde{A}_j = D_n^+ (A_j \otimes A_j)^\top D_n$  and  $\tilde{B}_j = D_n^+ (B_j \otimes B_j)^\top D_n$ are matrices of size  $n^* \times n^*$ , with the convention that  $\tilde{A}_j = 0$  for j > p and  $\tilde{B}_j = 0$  for j > q.

For the BEKK(1,1) model, these expressions simplify to  $\Psi_0 = I_{n^*}$ ,  $\Psi_1 = \tilde{A}_1$  and  $\Psi_i = (\tilde{A}_1 + \tilde{B}_1) \Psi_{i-1} = (\tilde{A}_1 + \tilde{B}_1)^{i-1} \tilde{A}_1$  where  $i \ge 2$ .

Proof. See Appendix A.2.

A closed-form expression for the *h*-step ahead VIRF exists for the BEKK(p, q) model. For completeness, we provide a rigorous derivation in the appendix, adapted to the case of a structural model as given in (2). This derivation will be of further value in Section 2.4.

**Proposition 2.2.** Assume that  $\varepsilon_t$  in (2) follows a BEKK(p, q) GARCH. Denote by  $\{\Psi_i\}_{i \in \mathbb{N}}$  the coefficients of its VMA( $\infty$ ) representation as provided in Proposition 2.1. Then the h-step

<sup>&</sup>lt;sup>4</sup>For matrices  $A \in \mathbb{R}^{m \times n}$  and  $B \in \mathbb{R}^{p \times q}$ ,  $A \otimes B$  denotes the Kronecker product, which is the  $(mp) \times (nq)$  matrix formed by multiplying each element  $a_{ij}$  of A by the entire matrix B.

ahead VIRF given a structural shock  $\bar{\xi}^*$  is

$$V_{t+h}(\bar{\xi}^*) = \Psi_h D_n^+ \left( H_t^{1/2} \otimes H_t^{1/2} \right) D_n \operatorname{vech} \left( \tilde{R} \bar{\xi}^* \bar{\xi}^{*\top} \tilde{R}^\top - I_n \right) .$$
(6)

*Proof.* See in Appendix A.2.

Notably, the VIRF is a nonlinear but even function of the structural shock. The persistence of a shock to volatility is governed by the moving average matrices  $\Psi_h$ ,  $h \ge 0$ .

*Remark* 1. Owing to the simpler structure of its VMA( $\infty$ ) coefficients (see Proposition 2.1), the VIRF in the BEKK(1, 1) model can be represented by the following recursion (Hafner and Herwartz, 2006):

$$V_{t+1}(\bar{\xi}^*) = \tilde{A}_1 D_n^+ \left( H_t^{1/2} \otimes H_t^{1/2} \right) D_n \operatorname{vech}(\tilde{R}\bar{\xi}^*\bar{\xi}^{*\top}\tilde{R}^{\top} - I_n)$$

$$V_{t+h}(\bar{\xi}^*) = \left( \tilde{A}_1 + \tilde{B}_1 \right) V_{t+h-1}(\bar{\xi}^*) \quad (h \ge 2).$$
(7)

This follows from setting the BEKK(1, 1) model in its vech form—see Equation (30) in the proof of Proposition 2.1—and inserting the VMA( $\infty$ ) coefficients into (6).

*Remark* 2. The VIRF of the BEKK model is invariant to rotations and reflections, and thus independent of the structural model, if, as in the historical VIRF analysis, the realized shock is chosen. If  $\bar{\xi}^* = \hat{\xi}_t$ , we have

$$\mathbf{V}_{t+h}(\hat{\xi}_t) = \Psi_h D_n^+ \left( H_t^{1/2} \otimes H_t^{1/2} \right) D_n \operatorname{vech} \left( \tilde{R} \hat{\xi}_t \hat{\xi}_t^\top \tilde{R}^\top - I_n \right)$$
(8)

$$=\Psi_{h}\operatorname{vech}\left(H_{t}^{1/2}\left(\tilde{R}\hat{\xi}_{t}\hat{\xi}_{t}^{\top}\tilde{R}^{\top}-I_{n}\right)H_{t}^{1/2}\right)$$
(9)

$$=\Psi_h \operatorname{vech}(\hat{\varepsilon}_t \hat{\varepsilon}_t^\top - H_t)$$
(10)

where the second line follows from an application of (18) and (20). Thus, when  $\hat{\varepsilon}_t$  is known or can be estimated, this expression remains unaffected by the structural mechanism encoded by  $\tilde{R}$ . This property no longer holds in the case of the scenario VIRF, where  $\bar{\xi}^* \neq \hat{\xi}_t$ .

### 2.4 Asymptotic results

As central contribution, we complete the notion of the VIRF by providing its asymptotic theory. Estimation of the BEKK model, as with MGARCH models in general, typically proceeds by quasi-maximum likelihood (QML) estimation assuming multivariate normality of  $\xi_t$ . The QML estimator is defined as maximizing the log likelihood given by  $\mathscr{L}_T(\eta) = -\frac{1}{2T} \sum_{i=1}^T \ell_t(\eta)$ , where  $\ell_t(\eta) = \log(\det(H_t)) + \varepsilon_t^\top H_t^{-1} \varepsilon_t$  and  $H_t = H_t(\eta)$ . For the VIRF, we will now use the notation  $V_{t+h}(\bar{\xi}^*;\eta)$  to emphasize the dependence of  $V_{t+h}$  on the parameter vector  $\eta$  of the underlying MGARCH model. We treat  $\tilde{R}$  as known, which, in light of Remark 2, corresponds to the historical VIRF analysis.

### 2.4.1 Consistency of the VIRF

**Proposition 2.3.** Let  $h \in \mathbb{N}$ ,  $\bar{\xi}^* \in \mathbb{R}^n$  arbitrary and fixed, and  $\hat{\eta}$  the QML estimator of  $\eta$ . If  $V_{t+h}(\bar{\xi}^*, \eta)$  continuous in  $\eta$ , it is consistent under Assumptions 2.1, i.e.,

$$\mathbf{V}_{t+h}(\bar{\xi}^*;\hat{\eta}) \xrightarrow{p} \mathbf{V}_{t+h}(\bar{\xi}^*;\eta_0) \quad (h \in \mathbb{N}) .$$
(11)

*Proof.* Follows from the consistency of the QML estimation (Hafner and Preminger, 2009), continuity, and the continuous mapping theorem.

*Remark* 3. Because the VIRF in the BEKK(p,q) model is continuous in  $\eta$ , its VIRF is consistent.

#### 2.4.2 Asymptotic normality of the BEKK VIRF

Given the QML estimator  $\hat{\eta}$  of the parameters of the BEKK model, we can deduce the asymptotic distribution of the VIRF.

**Theorem 1.** Let  $h \in \mathbb{N}$  and  $\overline{\xi}^* \in \mathbb{R}^n$  arbitrary and fixed. Under Assumptions 2.1, we have:

$$\sqrt{T} \left( \mathbf{V}_{t+h}(\bar{\xi}^*; \hat{\eta}) - \mathbf{V}_{t+h}(\bar{\xi}^*; \eta_0) \right) \stackrel{d}{\longrightarrow} N\left( 0, \mathbf{V}_{\eta} \mathcal{H}^{-1} \mathcal{I} \mathcal{H}^{-1} \mathbf{V}_{\eta}^{\top} \right)$$
(12)

where  $V_{\eta}(\bar{\xi}^*;\eta_0) = \frac{\partial V_{t+h}(\bar{\xi}^*;\eta_0)}{\partial \eta^{\top}}$  denotes the  $n^* \times m$  Jacobian matrix of the VIRF with respect to  $\eta \in \mathbb{R}^m$ ;  $\mathcal{H}(\eta_0) = -\mathbb{E}\left[\frac{\partial^2 \ell_t(\eta_0)}{\partial \eta \partial \eta^{\top}}\right]$  is the Hessian matrix of the log-likelihood contribution  $\ell_t$ , and  $\mathcal{I}(\eta_0) = \mathbb{E}\left[\frac{\partial \ell_t(\eta_0)}{\partial \eta}\frac{\partial \ell_t(\eta_0)}{\partial \eta^{\top}}\right]$  the Fisher information matrix.

*The*  $(n^* \times m)$  *Jacobian is given by* 

$$\frac{\partial \mathcal{V}_{t+h}(\bar{\xi}^*;\eta)}{\partial \eta^{\top}} = \left(\mathcal{V}_t(\bar{\xi}^*;\eta)^{\top} \otimes I_{n^*}\right) \frac{\partial \operatorname{vec}(\Psi_h)}{\partial \eta^{\top}} + \Psi_h \frac{\partial \mathcal{V}_t(\bar{\xi}^*;\eta)}{\partial \eta^{\top}} , \qquad (13)$$

where

$$\begin{aligned} \frac{\partial \mathbf{V}_t(\bar{\xi}^*;\eta)}{\partial \eta^{\top}} &= D_n^+ \left\{ \left[ \left( H_t^{1/2} \tilde{R} \bar{\xi}^* \bar{\xi}^{*\top} \tilde{R}^{\top} \otimes I_n \right) + \left( I_n \otimes H_t^{1/2} \tilde{R} \bar{\xi}^* \bar{\xi}^{*\top} \tilde{R}^{\top} \right) \right] \\ & \times \left[ \left( H_t^{1/2} \otimes I_n \right) + \left( I_n \otimes H_t^{1/2} \right) \right]^{-1} - I_{n^2} \right\} \frac{\partial \operatorname{vec}(H_t)}{\partial \eta^{\top}} \end{aligned}$$

and  $\{\Psi_i\}_{i\in\mathbb{N}}$  denote the coefficients of the VMA( $\infty$ ) representation of the BEKK(p,q) model. The expression for  $\frac{\partial \operatorname{vec}(\Psi_h)}{\partial \eta^{\top}}$  is derived in Appendix A.3, and the derivatives of the BEKK(p,q) model with respect to its parameters, i.e.,  $\frac{\partial \operatorname{vec}(H_t)}{\partial \eta^{\top}}$ , are found in Hafner and Herwartz (2008).

*For the BEKK*(1, 1) *model, we have* 

$$\frac{\partial \mathbf{V}_{t+h}(\bar{\boldsymbol{\xi}}^*;\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^{\top}} = \left(\mathbf{V}_{t+h-1}(\bar{\boldsymbol{\xi}}^*;\boldsymbol{\eta})^{\top} \otimes I_{n^*}\right) \frac{\partial \operatorname{vec}(\tilde{A}_1 + \tilde{B}_1 \mathbb{1}_{\{h>1\}})}{\partial \boldsymbol{\eta}^{\top}} + \left(\tilde{A}_1 + \tilde{B}_1 \mathbb{1}_{\{h>1\}}\right) \frac{\partial \mathbf{V}_{t+h-1}(\bar{\boldsymbol{\xi}}^*;\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^{\top}},$$
(14)

where  $\mathbb{1}_{\mathcal{A}}$  denotes the indicator function, which is equal to one if the event  $\mathcal{A}$  is true and zero otherwise.

*Proof.* See Appendix A.3.

The asymptotic distribution of the VIRF estimator allows us to construct pointwise and simultaneous confidence intervals for the volatility impulse responses of asset returns to structural shocks, as we demonstrate in Section 3.

*Remark* 4. Although Theorem 1 is tailored to the BEKK model, it is in fact more general. It applies to any MGARCH model that satisfies the high-level Assumptions 2.1 and has a so-called VEC representation. This is evident from the proof of Theorem 1, where, as a first step, the VEC representation of the BEKK model is derived—see (30).

### 2.5 Causal interpretation of the VIRF

Our volatility impulse response analysis aligns with recent advances in causal inference, where model outcomes are interpreted as resulting from different assignments in treatment variables. By comparing different states of the world, each linked to distinct values of the causing variable, this notion of causality contrasts with Granger-Sims causality, which focuses on predictability.<sup>5</sup>

A causal interpretation requires additional identifying assumptions that allow one to relate the observed data to the distribution of the potential outcome variables. Rambachan and Shephard (2020, 2021) provide conditions under which the GIRF of Koop et al. (1996) can be endowed with a causal interpretation. In contrast to the GIRF, which addresses the conditional mean, the VIRF traces the dynamic effects of a shock on the conditional second-order moments. In this section, we develop the framework that provides a causal interpretation for the VIRF.<sup>6</sup>

#### 2.5.1 Definitions and assumptions

To frame our variables of interest within the context of causal inference, we interpret the structural shocks  $\xi_t$  as treatment variables, with the treatment being continuous. Additionally, we regard the vectorized outer products of the demeaned return vectors  $\varepsilon_t$ , i.e.,  $X_t := \text{vech}(\varepsilon_t \varepsilon_t^\top)$ , as the observable, continuously valued, multidimensional outcomes. The outcomes are linked to the treatment variables through the potential outcome process, which specifies the outcome that would be observed for a given path of the treatment variable. We begin by adapting the definitions and assumptions of Rambachan and Shephard (2020, 2021) to our framework:

**Definition 2.2.** Let  $\xi := {\xi_t}_{t \ge 1}, \xi_t \in \mathcal{W}_{\xi} \subseteq \mathbb{R}^n$ , denote a stochastic treatment path.

<sup>&</sup>lt;sup>5</sup>See Lechner (2010) for an overview of causality in econometrics.

<sup>&</sup>lt;sup>6</sup>In macroeconomics, similar efforts have recently been made for local projections (Cloyne et al., 2023; Gonçalves et al., 2024).

For any deterministic trajectory  $\bar{\xi} \coloneqq {\{\bar{\xi}_t\}_{t \ge 1}, \bar{\xi}_t \in \mathcal{W}_{\xi}}$ , the potential outcome path is given by  $X(\bar{\xi}) \coloneqq {\{X_t(\{\bar{\xi}_s\}_{s \ge 1})\}_{t \ge 1}}$ .

As this notation makes clear, in general, potential outcomes may also be influenced by future treatments. This is ruled out by

Assumption 2.2 (Time series non-interference<sup>7</sup>). For each  $t \ge 1$  and all deterministic  $\{\bar{\xi}_t\}_{t\ge 1}, \{\bar{\xi}'_t\}_{t\ge 1}$  with  $\bar{\xi}_t, \bar{\xi}'_t \in \mathcal{W}_{\xi}$ :

$$X_t(\{\bar{\xi}_s\}_{1\leq s\leq t}, \{\bar{\xi}_s\}_{s\geq t+1}) = X_t(\{\bar{\xi}_s\}_{1\leq s\leq t}, \{\bar{\xi}'_s\}_{s\geq t+1}) \qquad \text{almost surely.}$$

Assumption 2.2 is analogous to the non-interference assumption in classical causal inference (Rubin, 1980). It restricts the potential outcomes to depend only on past and contemporaneous treatments; therefore, we can drop any dependence on future assignments.

The potential outcomes evaluated at the treatments yield the outcome process:

Assumption 2.3 (Outcomes). The outcome at time *t* is defined as  $X_t := X_t(\{\xi_s\}_{1 \le s \le t})$ , and the outcome process is  $X(\xi) := \{X_t(\{\xi_s\}_{0 \le s \le t})\}_{t \ge 0}$ .

Because our treatment variable is continuous, unlike in Rambachan and Shephard (2020, 2021), we formulate the probabilistic assumptions for Borel sets. Let  $\mathcal{B}(\bar{\xi}^*, \epsilon) = \{\xi \in \mathcal{W}_{\xi} : d(\bar{\xi}^*, \xi) < \epsilon\}$  be an  $\epsilon$ -neighborhood of some fixed, structural shock of interest  $\bar{\xi}^* \in \mathcal{W}_{\xi}$ , where d is some metric on  $\mathbb{R}^n$  and  $\epsilon > 0$ , allowing for arbitrarily small sets. The assignment of the treatments is assumed to be sequentially probabilistic, i.e., at time t, any treatment vector  $\xi_t \in \mathcal{B}(\bar{\xi}^*, \epsilon)$  can be realized with positive probability, given the history generated by both  $X(\xi)$  and  $\xi$ :

**Assumption 2.4** (Positivity). For each  $t \in \mathbb{Z}$  and  $h \ge 0$ , the stochastic treatment path satisfies, for any  $\epsilon > 0$  and any  $\bar{\xi}^* \in W_{\xi}$ ,

$$0 < \Pr(\xi_{t+h} \in \mathcal{B}(\bar{\xi}^*, \epsilon) | \mathcal{G}_{t-1}) < 1$$

<sup>&</sup>lt;sup>7</sup>Also termed 'non-anticipating potential outcomes' in Rambachan and Shephard (2020, 2021).

where  $\mathcal{G}_t \coloneqq \sigma(X_s, \xi_s : s \le t - 1)$ .

The next assumption requires that the assignment of the treatment depends solely on past outcomes and past treatments, conditioned on  $\mathcal{G}_{t-1}$ :

**Assumption 2.5** (Time series unconfoundedness). For each  $t \in \mathbb{Z}$ ,  $h \ge 0$ , and  $\bar{\xi}_s \in \mathcal{W}_{\xi}$ ,

$$\xi_t \perp \left( \{\xi_s\}_{t+1 \le s \le t+h}, X_{t+h}(\{\xi_{s'}\}_{1 \le s' \le t-1}, \{\bar{\xi}_s\}_{t \le s \le t+h}) \right) | \mathcal{G}_{t-1} ,$$

where  $\xi_{s'} \in \mathcal{G}_{t-1}$  for  $s' \leq t-1$ .

Assumption 2.5 defines non-anticipating treatment paths conditional on the information available up to time t - 1. From a time series perspective, one could say that future potential outcomes do not Granger-cause the current treatment  $\xi_t$ .

Lastly, we require continuity assumptions to avoid ambiguities in defining conditional distributions, when conditioning on events with measure zero—see Gill and Robins (2001) for an in-depth discussion:

**Assumption 2.6** (Continuity). For any  $t \in \mathbb{Z}$  and  $h \ge 0$ , the law of  $\xi_{t+h} | \mathcal{G}_{t-1}$  is continuous (with respect to weak convergence) in the joint support of the conditioning variables that generate  $\mathcal{G}_t$ . Likewise, the joint law of

$$\left(\{\xi_s\}_{t+1\leq s\leq t+h}, X_{t+h}(\{\xi_{s'}\}_{1\leq s'\leq t-1}, \{\bar{\xi}_s\}_{t\leq s\leq t+h})\right)|\mathcal{G}_{t-1}|$$

can be chosen continuous in the conditioning variables embedded in  $\mathcal{G}_{t-1}$ .

Given this framework, we define our potential outcome process as  $X_t(\bar{\xi}) = \operatorname{vech}(H_t^{1/2}\bar{\xi}_t\bar{\xi}_t^{\top}H_t^{1/2})$ , where  $H_t$  follows (4). It is deterministic, nonlinear in the assignment for every *t*, and satisfies Assumption 2.2, as  $X_t$  depends on past shocks only due to the VMA( $\infty$ ) representation (Proposition 2.1).

We observe the outcome  $(\xi_t, X_t) = (\xi_t, X_t(\xi_t))$ . Assumptions 2.4 and 2.5 hold because, under Assumption 2.1, the structural shocks  $\xi_t$  are independently and identically distributed, serially independent, and the MGARCH process admits a strictly stationary and non-anticipative solution.<sup>8</sup> This guarantees the independence of  $\xi_t$  from future treatments and the associated potential outcomes. Lastly, it holds that  $\mathcal{G}_t \subseteq \sigma(\varepsilon_s : s \leq t) = \mathcal{F}_t$ , implying that the usual filtration for the MGARCH model nests  $\mathcal{G}_t$ . Assumption 2.6 regarding continuity appears natural when modeling return data.

### 2.5.2 Causal effect

A dynamic causal treatment effect compares potential outcomes along an assignment path to those along a counterfactual path. Given the infinite number of possible paths, we define—similarly to Rambachan and Shephard (2020)—a filtered treatment effect that averages over all possible future assignments and contrasts it with the scenario of averaging over all potential interventions, conditional on (t - 1)-information. Thus, the assignment path we consider is given by  $(\{\xi_s\}_{1 \le s \le t-1}, \bar{\xi}^*, \{\xi_s\}_{t+1 \le s \le t+h})$ , while  $X_{t+h}(\bar{\xi}^*) \coloneqq X_{t+h}(\{\xi_s\}_{1 \le s \le t-1}, \bar{\xi}^*, \{\xi_s\}_{t+1 \le s \le t+h})$  is the (t + h)-potential outcome along this path. Observe that  $X_{t+h} = X_{t+h}(\xi_t)$ . This leads to

**Definition 2.3.** (Causal effect) For any  $t \in \mathbb{Z}$ ,  $h \ge 0$ , and any  $\overline{\xi}^* \in \mathcal{W}_{\xi}$ , define the filtered causal effect, conditional on (t - 1)-information, as

$$\mathbb{E}[X_{t+h}(\bar{\xi}^*) - X_{t+h}|\mathcal{F}_{t-1}].$$

The next proposition demonstrates that the VIRF, when applied to a structural shock of interest  $\bar{\xi}^*$ , can be decomposed into the filtered treatment effect and a selection bias term, which vanishes under time series unconfoundedness:

**Proposition 2.4.** Let Assumptions 2.1-2.6 hold, let  $h \ge 0$ , and assume that  $E[X_{t+h}(\bar{\xi}^*) - X_{t+h}|\mathcal{F}_{t-1}] < \infty$ . Then it holds:

$$V_{t+h}(\bar{\xi}^*) = E[X_{t+h}(\bar{\xi}^*) - X_{t+h}|\mathcal{F}_{t-1}] + \Delta_{t+h}(\bar{\xi}^*|\mathcal{F}_{t-1})$$

<sup>&</sup>lt;sup>8</sup>This means that the process  $\{\varepsilon_t\}_{t\in\mathbb{Z}}$  is such that  $\varepsilon_t = H_t^{1/2}\tilde{R}\xi_t$  is a measurable function of  $\xi_{t-s}$  for  $s \ge 0$ , with  $H_t^{1/2} \perp \sigma(\xi_{t+h} : h \ge 0)$  and  $\varepsilon_t \perp \sigma(\xi_{t+h} : h > 0)$ —see Francq and Zakoïan (2010, Thm. 11.5) and Boussama et al. (2011).

where  $\Delta_{t+h}(\bar{\xi}^*|\mathcal{F}_{t-1})$  is a selection bias which vanishes under Assumption 2.5.

*Proof.* The proof is given in Section A.4.

In summary, under the present assumptions, the VIRF is a filtered treatment effect and thus has a causal meaning.

*Remark* 5. Our discussion on providing the VIRF with a causal interpretation is not confined to the BEKK model. For instance, it also extends to the asymmetric VIRF introduced by Hafner and Herwartz (2023), which accounts for the leverage effect commonly observed in financial data. Their assumptions regarding the structural shock system align with those in our framework, and their leverage function is continuous element-wise with respect to the shocks, ensuring that Assumptions 2.2-2.6 are satisfied.

## **3** Empirical Application

We now present a structural VIRF analysis using a system of daily asset returns across three key asset classes: equity, fixed income, and foreign exchange markets. These asset classes are not only central to portfolio optimization but are also vital components in financial stress tests mandated by central banks (Kremer et al., 2012).

## 3.1 Identification and estimation of the structural MGARCH model

To achieve an economically interpretable structural model, Fengler and Polivka (2024) propose identifying and estimating the orthogonal matrix  $\tilde{R}$  using proxy variables (also known as instruments). The approach proceeds as follows: assume the objective is to identify n - 1 structural shocks, denoted by  $\xi_{1t}$ . For this, partition the vector  $\xi_t^{\top} = (\xi_{1t}^{\top}, \xi_{2t})^{\top}$ , where  $\xi_{2t}$  represents the shock that is not of primary interest. Moreover,

assume there exists a centered, (n - 1)-dimensional, stationary instrument process  $Z = (Z_t)_{t \in I}$ , generated as

$$Z_t = \Phi \xi_t + v_t$$
 ,

where  $\Phi = (\Psi, 0_{n-1\times 1})$  is an  $(n-1) \times n$  matrix and  $\Psi$  an  $(n-1) \times (n-1)$  full column rank matrix. The process  $(v_t)_{t\in I}$  is strict white noise,  $v_t \sim (0, \Sigma_v)$ , where  $\Sigma_v$  is a positive-definite covariance matrix. Additionally,  $\xi_t$  is assumed to be independent of  $v_t$ .

It follows that  $E[Z_t \xi_{1t}^{\top}] = \Psi$ , with the rank condition on  $\Psi$  ensuring that the instruments offer non-redundant information about the structural shocks of interest. The fact that  $E[Z_t \xi_{2t}] = \mathbf{0}_{(n-1)\times 1}$  imposes exogeneity restrictions, which formalize the idea that the instrument is uninformative about the non-targeted shock. This framework allows multiple instruments to jointly convey information about a structural shock, so  $\Psi$  is not required to be diagonal. In our empirical application, however, an overidentification test suggests that this assumption is appropriate. This identification approach leverages methods from proxy identification in structural VAR models (see, inter alia, Stock and Watson, 2012, Mertens and Ravn, 2013, Angelini and Fanelli, 2019, and Giacomini et al., 2022).

For estimation, one proceeds in two steps. First, estimate the MGARCH model using QML with  $\tilde{R} = I_n$ , which provides estimates of  $u_t = H_t^{-1/2} \varepsilon_t$ . In the second step, define an augmented model by introducing the expanded system  $\zeta_t = (u_t^{\top}, Z_t^{\top})^{\top}$ , taking values in  $\mathbb{R}^m$ , where m = 2n - 1. The augmented model can be written as

$$\zeta_t = \begin{pmatrix} \tilde{R}_{\bullet,1:n-1} & \tilde{R}_{\bullet,n} & 0_{n \times n-1} \\ \Psi & 0_{n-1 \times 1} & \Sigma_v^{1/2} \end{pmatrix} \begin{pmatrix} \xi_t \\ v_t \end{pmatrix} = G \begin{pmatrix} \xi_t \\ v_t \end{pmatrix}, \quad (15)$$

where  $\tilde{R} = (\tilde{R}_{\bullet,1:n-1}, \tilde{R}_{\bullet,n})$  is partitioned into columns that align with the instrumented and non-instrumented components of  $\xi_t$ , respectively, and  $\Sigma_v^{1/2}$  is the principal matrix square root of  $\Sigma_v$ . Equation (15) represents a VAR(0), where the first *n* elements of the governing shock vector are the uncorrelated, mean-zero, unit-variance structural shocks. The remaining n - 1 elements comprise the shocks driving the instrument process. Notably, the order of variables within either the return system or the instruments does not impact this setup.

Identification requires that m(m-1)/2 constraints be imposed on *G*. This need arises because a total of m(m+1)/2 parameters are governed by the orthogonality of  $\tilde{R}$  and the estimated parameters in  $\Sigma_{\zeta} = \mathbb{E}[\zeta_t \zeta_t^{\top}] = GG^{\top}$ . From (15), (n+1)(n-1) zero constraints are derived from the instrument exogeneity conditions. However, this number may not suffice to meet the required m(m-1)/2, necessitating additional conditions on  $\tilde{R}$ ,  $\Psi$ , or  $\Sigma_v$ . If such conditions can be established and if  $\Psi$  has full column rank, it is possible to identify  $\tilde{R}$  up to column signs.<sup>9</sup> In Section 3.2, we discuss how we derive these additional restrictions within our application.

The quasi log-likelihood of the structural model is given by

$$\mathscr{L}_T^s(\theta) = -\frac{mT}{2}\log(2\pi) - \frac{T}{2}\log\det(GG^{\top}) - \frac{T}{2}\operatorname{tr}\{G^{-1}\widehat{\Sigma}_{\zeta}(G^{-1})^{\top}\}.$$
 (16)

Note that *G* depends on the parameter vector  $\theta = \left(\operatorname{vec}(\tilde{R})^{\top}, \psi^{\top}, \operatorname{vec}(\Sigma_v^{1/2})^{\top}\right)^{\top}$ , where  $\psi$  represents a vector collecting the free (non-zero) parameters of  $\Psi$ . Consequently, the structural parameter matrix  $\tilde{R}$  will be estimated, up to sign, as part of the parameters that maximize the log-likelihood. The challenge of maximizing the log-likelihood arises from the non-convexity of the orthogonality constraints. As high-lighted in Fengler and Polivka (2024), this issue can be efficiently addressed using methods from Riemannian optimization.

### 3.2 Data and estimation

We borrow elements of the empirical analysis from Fengler and Polivka (2024) and examine a system of daily returns computed from the Standard and Poor's 500 Composite Index, the yield of the U.S. constant maturity 10-year Treasury note, and the

<sup>&</sup>lt;sup>9</sup>See Angelini and Fanelli (2019) and Fengler and Polivka (2024) for further discussion of local point identification.



Figure 1: Upper panel: Demeaned daily log returns of the S&P 500 Composite Index, the yield of the U.S. constant maturity 10-year Treasury note, and the USD Index from 1/1/1998 to 12/31/2014. Lower panel: TRMI U.S. stock index sentiment and TRMI U.S. bond sentiment on trading days (demeaned, ARMA filtered, standardized). Data sources: Bloomberg and Thomson Reuters (TRMI).

USD Index.<sup>10</sup> The data cover the period from 1/1/1998 to 12/31/2014—see Figure 1 for the log returns.

We select the proxy variables in accordance with a core tenet of financial econometrics: fundamental news drives stock returns—see Jeon et al. (2021) for evidence on the link between news and significant intraday stock returns. Our goal is to identify two types of shocks—an equity price shock that is informative about the economic fundamentals of equities, and a bond price shock that reflects shifts, e.g., in real interest rates, inflation expectations, or monetary policy. To achieve this, we use two series of news analytics data from the Thomson Reuters MarketPsych Indices (TRMI) as proxies for the underlying structural shocks, specifically the U.S. stock index news sentiment and U.S. bond news sentiment. The TRMIs are constructed by means of

<sup>&</sup>lt;sup>10</sup>The USD Index is a measure of the U.S. Dollar's value relative to a basket of currencies from U.S. trade partners, increasing as the U.S. Dollar strengthens against these currencies.

Descriptive statistics							
	Asset returns				Instrument data		
Statistic	S&P 500	Yield	USDX	Equitie	s Bonds		
Min.	-0.095	-0.185	-0.027	-3.134	4 -4.830		
Max.	0.109	0.089	0.024	3.37	5 5.600		
Median	0.000	0.000	0.000	-0.01	0 -0.047		
Std. Dev.	0.013	0.018	0.005	1.00	1.000		
Skewn.	-0.203	-0.136	-0.034	0.05	6 0.238		
Kurt.	10.929	8.454	4.491	2.53	3 4.231		

Table 1: Descriptive statistics of asset returns and instrument data. The first three columns present the daily log returns of the S&P 500 index, the U.S. 10-year Treasury constant maturity yield, and the USD Index. The final two columns display proxy data based on U.S. stock index sentiment and U.S. bond sentiment (TRMI MarketPsych indices). The sample covers the period from 01/01/1998 to 12/31/2014, with a total of N = 4435 observations. Data sources: Thomson Reuters.

a proprietary supervised natural language processing scheme applied to news items from a broad range of media outlets. Each item is scored for relevance, novelty, and sentiment to construct the indices.<sup>11</sup> The TRMI data are available at a daily frequency, and we apply flexible autoregressive moving average (ARMA) models to extract the unexpected innovations, which we use as proxies.

As discussed above, to achieve identification, we require m(m-1)/2 = 10 constraints, because m = 2n - 1 = 5 with n = 3 returns and n - 1 = 2 instruments. The zero constraints provide (n + 1)(n - 1) = 8 of these—see (15). One additional constraint is obtained by imposing symmetry on  $\Sigma_v^{1/2}$ . Consequently, the system is just-identified by applying a single zero restriction on  $\Psi$ . However, we impose diagonality on  $\Psi$ , which introduces two zero constraints and results in an overidentified system.

For the MGARCH dynamics, we estimate the BEKK(1,1) specification of (4) as re-

<sup>&</sup>lt;sup>11</sup>This includes, but is not limited to, live content delivered via the Thomson Reuters News Feed Direct, LexisNexis, and financial news sites such as The New York Times, The Wall Street Journal, Financial Times, and Seeking Alpha. See https://www.marketpsych.com/ and Peterson (2016, Appendix A) for further details.

ported in Table 2 and Figure 2. Table 3 presents the structural model estimates. The orthogonal matrix diverges from the identity matrix, indicating a departure from the symmetric volatility spillover pattern that would result if  $\tilde{R} = I_n$ . This shift reallocates mass asymmetrically from the unit diagonal entries of the identity matrix to the offdiagonal elements of the orthogonal matrix. Additionally, the significant estimates of the diagonal elements in  $\hat{\Psi}$  confirm that the chosen TRMI series  $Z_1$  and  $Z_2$  are indeed relevant instruments (Table 3). A likelihood ratio test assessing the overidentifying restriction supports the diagonal structure of  $\Psi$ , as the test statistic LR<sub>T</sub> = 0.44 corresponds to a *p*-value under the  $\chi^2(1)$  distribution.

Due to the identification strategy of using stock market news sentiment for  $\xi_1$  and bond price sentiment for  $\xi_2$ , we refer to these shocks as the equity price shock and bond price shock, respectively. The remaining shock is identified too, although it is not directly targeted by our identification approach. As visible in the orthogonal matrix in Table 3, this shock has the largest impact on the USD Index equation, suggesting it may represent a currency shock. However, we refer to it simply as the 'third' shock.

	ر			$\hat{A}_1$			$\hat{B}_1$	
0.0011	0.0000	0.0000	0.2667	0.0265	-0.0121	0.9580	-0.0034	0.0038
-0.0001	0.0010	0.0000	-0.0010	0.2107	0.0030	0.0020	0.9759	-0.0011
-0.0000	0.0001	0.0002	0.0696	-0.0606	0.1442	-0.0125	0.0096	0.9883
(7.270)			(12.119)	(1.198)	(-1.880)	(142.121)	(-0.475)	(2.506)
-0.449)	(5.285)		(-0.140)	(16.758)	(0.748)	(1.161)	(343.911)	(-1.308)
-0.216)	(1.304)	(3.283)	(1.863)	(-1.506)	(14.151)	(-1.399)	(1.335)	(533.093)
Spectral ra	dius is less th	an one						
Akaike crit	erion							
full BEKK:	-88676.66		diagonal B	EKK: -8866	8.45			
Schwarz cr fiil reky.	iterion 		diamonal R	C988771	2 0 Z			

Table 2: Quasi maximum likelihood (QML) parameter estimates of the unrestricted BEKK(1,1) model of the demeaned daily log returns of the S&P 500, the constant maturity yield of U.S. 10-year Treasury notes, and the USD Index in order of appearance over the time period from 01/01/1998 to 12/31/2014. Entries in parentheses are the robust QML *t*-ratios. The R package *BEKKs* of Fülle et al. (2024) is used for estimation.

	proxy MGAR	CH model			
	0.9127	0.3276	0.2443	0	0
$\left( \hat{R} + 2 \hat{R} + \hat{R} + 0 + 2 \right)$	0.3705	-0.9156	-0.1563	0	0
$\hat{G} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 $	-0.1725	-0.2332	0.9570	0	0
$\left( \begin{array}{cc} \Psi & 0_{2\times 1} \downarrow \Sigma_v^{2/2} \end{array} \right)$	0.3347	0	0	0.9405	0.0574
	0	0.1950	0	0.0574	0.9793
				1	
	0.0158	0.0373	0.0478	ı 1	
	0.0339	0.0183	0.0667	1	
Bootstrapped std. errors	0.0414	0.0701	0.0199	 	
	0.0133			0.0095	0.0073
		0.0150		0.0073	0.0136

Table 3: The estimation results of the structural MGARCH model are based on the demeaned daily log returns of the S&P 500, the U.S. 10-year Treasury constant maturity yield, and the U.S. Dollar Index for the period from 01/01/1998 to 12/31/2014, using stock market sentiment ( $Z_1$ ) and bond market sentiment ( $Z_2$ ) TRMIs as proxy variables. The upper panel presents the estimated elements of *G*, including the orthogonal matrix, relevance parameters, and the symmetric decomposition of the instrument error variance. The lower panel provides the bootstrapped standard errors derived from 999 replications.

## 3.3 Structural volatility impulse response analysis

One of the key applications of MGARCH models is analyzing how volatility responds to shocks (Bauwens et al., 2006). By utilizing identified labeled shocks, we can offer a more refined analysis of the VIRF patterns than has been possible to date. We begin with historical VIRFs, followed by scenario-based VIRFs.

### 3.3.1 Historical VIRFs

We consider three historical events from our sample: First, we analyze the NASDAQ crash on April 14, 2000, which marked the bursting of the Dotcom bubble, as an example of a 1% marginal equity price shock. Second, we examine the semiannual mon-

			Event	
Shock and		Dotcom crisis	Greenspan speech	EU debt crisis
ret	urn vector	04/14/2000	03/07/2002	08/04/2011
	Equity	-4.4292	0.6765	-4.4881
$\xi_t$	Bond	-0.0120	-2.5068	0.5808
	Third	-1.6040	-3.3369	0.7663
	S&P 500	-0.0602	-0.0047	-0.0492
ε <sub>t</sub>	Yield	-0.0150	0.0314	-0.0663
•	USD Index	-0.0044	-0.0101	0.0155

Table 4: Structural shock vectors selected for the historical VIRF analysis with corresponding returns.

etary policy report delivered by Federal Reserve Chairman Alan Greenspan to the US Senate on March 7, 2002, illustrating a 1% tail event in both the bond price shock and the third shock. Lastly, we investigate the concerns in the US equity markets over the European debt crisis on August 4, 2011, as another example of an equity price shock. The corresponding shock vectors are documented in Table 4. In Figure 2, the selected case studies are indicated by red vertical lines, placing these days in the context of the filtered variance and covariance evolutions.

Figures 3 to 5 present the resulting VIRFs (solid black lines). Additionally, each plot includes the following 95% confidence intervals: the individual asymptotic pointwise 95% confidence intervals (black dashed line), based on the QML variance matrix; the individual pointwise 95% confidence intervals (light grey area) obtained from the residual bootstrap procedure, as described in Hafner and Herwartz (2023); and, for completeness, the simultaneous (i.e., across all VIRFs) asymptotic confidence intervals (dot-dashed blue line), derived from the  $\chi^2(6)$  approximation of the Wald statistic, also based on the QML variance matrix.<sup>12</sup>

The structural shock of the NASDAQ crash on 04/14/2000 initially leads to a strong

<sup>&</sup>lt;sup>12</sup>The degrees of freedom follow from the dimension of the estimated VIRF vector—see Lütkepohl et al. (2015) for further discussions.

positive response in the predicted variance of the S&P 500, which is statistically significant at the 5% level, regardless of the confidence interval used (see Figure 3). While the significance diminishes after 50 days when judged by the joint confidence interval, the impulse to the level of volatility in the S&P 500 remains significant for 90-100 days based on the individual pointwise intervals. The strong reaction in the covariance of the S&P 500 and the 10-year Treasury yield returns is marginally significant for about 10-15 days according to the simultaneous intervals, and between 50 and 75 days according to the pointwise asymptotic confidence intervals.

The remaining shocks to variance are ambiguous, as the confidence intervals point in contradictory directions. Notably, although the third shock is relatively large at 1.6 standard deviations (see Table 4), its impact on the USD Index and its VIRF is dampened due to the nonlinear nature of shock propagation within the BEKK model.

For our second case study, we focus on 03/07/2002, when the chair of the Federal Reserve Board, Alan Greenspan, made an appearance in Congress that caught markets off guard. In his speech to the Senate, he presented a much more optimistic view of the economic outlook than he had in his testimony to the House of Representatives just seven days earlier. The structural shock associated with this event was moderately positive in the equity coordinate, but very large and negative in the coordinates of the bond price shock and the third shock (see Table 4). This resulted in a minor shock to the S&P 500, a strong increase in yields (as a negative shock to bond sentiment implies higher yields), and a depreciation of the U.S. Dollar relative to the currencies of major trading partners.

This shock vector triggered remarkably complex patterns of shock propagation in the system (see Figure 4). Despite the relatively small equity component of the shock, we observe a slight but statistically significant decrease in the predicted variance of the S&P 500. Based on the pointwise intervals, the effect persisted for about 10-40 days before becoming insignificant and converging to zero. Thus, this serves as a rare example of a shock having a *calming* effect on equity market variance, likely driven by

the outlook for an accelerated recovery in the US economy.

At the same time, the sharp negative shock to the bond market triggered a substantial short-term rise in predicted yield volatility, which remained significant for at least 50 days, likely reflecting market uncertainty regarding future rate hikes. Increases in yield (return) volatility following Federal Open Market Committee (FOMC) announcements and monetary policy reports have, of course, been well-documented in the literature—see, e.g., Jones et al. (1998) and Rosa (2013), along with the references therein.

The predicted covariance impulse responses of the S&P 500 returns with yield returns and with USD Index returns are both significant but relatively short-lived. In contrast, the structural shock had a more lasting impact, with a positive effect on the predicted variance of the USD Index return and a negative effect on the covariance between the USD Index and the yield returns. The latter decline may appear counterintuitive, as standard exchange rate models tend to suggest a positive relationship between yields and the domestic currency—see the lower right panel in Figure 2; however, the VIRF does not provide insights into the levels of covariance, but merely indicates a negative impulse to them.

We now turn to our third case study, the structural shock observed during the European debt crisis (see Figure 5). With the exception of the equity price shock, the other components of the structural shock on 08/04/2011 remain well within one standard deviation around zero (see Table 4). Nevertheless, the shock had strong, statistically significant effects on the predicted (co-)variances of all asset return components.

In Figure 5, we observe a positive impact on the predicted variances of the S&P 500 returns, the Treasury yield returns, and their covariance. This pattern is consistent with heightened market concerns over potential European debt defaults, as well as flight-to-safety investments driven by declining equity markets amid the European debt crisis.

27

The impact on the USD Index return variance is only significant when judged by the pointwise confidence intervals. In contrast, we find a significant and pronounced decrease in the predicted covariance between the S&P 500 and the USD Index returns. This aligns with the literature documenting that volatility spillover effects between equity and foreign exchange markets are typically small during normal market conditions but become stronger in periods leading up to crises (see, e.g., Grobys, 2015 and Cenedese and Mallucci, 2016). The even sharper drop in the predicted covariance between the USD Index and the 10-year Treasury yield returns reflects two factors: a surge in the U.S. Dollar relative to the Euro, as evidenced by the positive USD Index return on the shock date, and a decline in yields, potentially driven by increased demand for safe-haven investments (see Table 4).

In summary, our historical VIRF analysis underscores the importance of considering confidence intervals to fully appreciate volatility impulse responses. The bootstrapped and asymptotic confidence intervals (based on the QML variance matrix) align well, particularly given the highly nonlinear transformation of the model's parameters into the VIRF. The bootstrap intervals, however, tend to be slightly tighter. The largest discrepancies between the two methods arise when, after the initial impact, the bootstrap method samples parameter constellations that result in VIRFs exhibiting always the same sign. In contrast, the asymptotic intervals, being symmetric by construction, tend to eventually cover zero.

Moreover, the discussion underscores the value of having an economic interpretation of the shocks—a feature lacking in the existing VIRF literature. Our structural model provides this by identifying which specific shocks impacted which returns, and thereby offers a broader understanding of portfolio risks.

#### 3.3.2 Scenario VIRFs

Historical VIRFs are valuable for understanding past volatility events, but their utility is limited since events like those highlighted in Section 3.3 are unlikely to recur. However, as discussed in Section 2.5, the structural model enables the analysis of causal effects from meaningful shock scenarios, whether in an out-of-sample context or as counterfactuals.

For illustration, we adopt a risk manager's perspective and investigate the VIRFs under two scenarios: (a) a 1% marginal equity price shock, and (b) a 1% marginal bond price shock, both on the out-of-sample date 01/02/2015. Assuming independence, we estimate the multivariate density of the structural shocks by fitting the marginal distributions with a Gaussian kernel density estimator. We use these estimates to draw independent 10,000 observations from the 1% quantile of the shock component of interest, while sampling from the full distribution for the remaining components. This approach avoids setting other structural shocks to zero. To illustrate the responses, we compute the pointwise median VIRFs, along with the 25% and 75% quantile VIRFs for each forecast horizon *h*. Additionally, we include the analytical asymptotic simultaneous 95% confidence intervals for the median target VIRF, calculated following Fry and Pagan (2011). It is important to note that these confidence intervals reflect the parameter estimation uncertainty of a single VIRF—the median target VIRF—and are therefore unrelated to the interquartile range, which is computed across the VIRFs based on different shock vectors.

Figure 6 displays the VIRFs resulting from shock scenario (a). According to our analysis in Section 2.5, these responses represent the causal effects associated with the 1% quantile of the equity price shock. We observe a pronounced positive median impact on the forecasted conditional (co-)variances of S&P 500 and Treasury yield returns, with a more muted effect on the (co-)variances involving the USD Index. Given the prevailing conditional covariance patterns, such a shock scenario may be expected to increase the variance levels of both equity and fixed income markets, as well as their correlation, over the medium term. Consequently, the risk manager might consider implementing further diversification strategies to reduce exposure to equity price shocks.

In contrast, Figure 7 shows that the bond price shocks sampled from the 1% quantile in scenario (b) have a negligible impact on the predicted conditional covariance between the Treasury yield and the S&P 500 returns, as well as the forecasted conditional variance of the S&P 500 returns. However, there is a strong positive effect on the predicted variance of yield returns and the predicted conditional covariance with the USD Index returns. In this scenario, the shock-induced increase in covariance between the fixed income and FX markets could be a primary concern for risk assessment.

In summary, Figures 6 and 7 demonstrate that different shock scenarios lead to distinct predictions for (co-)variance levels in the asset return system, providing valuable insights for risk managers to take scenario-specific precautions.

## 4 Conclusion

In this paper, we revisited the VIRF introduced by Hafner and Herwartz (2006), a useful tool for analyzing the impact of shocks on conditional variance matrices in MGARCH models. By deriving the asymptotic distribution of the VIRF within the BEKK model, we enhance its potential by offering asymptotic confidence intervals. We demonstrate that the asymptotic variance matrix, similar to the VIRF itself, can be expressed as a function of the forecast horizon in a compact recursive form, enabling efficient numerical evaluation.

Building on recent advances in identifying MGARCH models, we extend the VIRF to take advantage of structural volatility models. With interpretable, labeled shocks and clearly defined structural propagation channels, we broaden the VIRF's application to include counterfactual and out-of-sample scenario analyses. Beyond the structural interpretation, we demonstrate how to provide the VIRF with a causal interpretation. This approach enables the application of the microeconometricians' concept of causality for assessing the impact of well-defined shock scenarios. In an empirical application to an identified system of equity, government bond, and foreign exchange returns, we illustrate two key use cases: the historical VIRF and the scenario VIRF. Our applications highlight the importance of assessing the statistical significance of volatility impulse responses.



Figure 2: Left panel: Filtered variance paths of the BEKK model estimated on the returns of the S&P 500, 10-year Treasury yields, and the USD Index. Right panel: Estimated covariance paths. The red lines mark the selected days for the historical VIRF analysis. From left to right: Dotcom bubble burst (NASDAQ crash on 04/14/2000), Greenspan's speech to the Senate (03/07/2002), and the onset of the European debt crisis (08/04/2011)—see also Table 4.



Figure 3: The predicted 500-step ahead VIRFs (solid black line) are driven by the structural shock on 04/14/2000, marking the NASDAQ crash at the onset of the Dotcom bubble burst (see Table 4). The dotted black line represents the 95% pointwise asymptotic confidence intervals, the grey area denotes the 95% pointwise bootstrapped confidence intervals, and the dot-dashed blue line shows the asymptotic simultaneous confidence intervals across all VIRFs, based on a  $\chi^2(6)$  approximation. The return system includes the S&P 500, the yield of the US constant maturity 10-year Treasury note, and the USD Index.



Figure 4: The predicted 500-step ahead VIRFs (solid black line) are driven by the structural shock on 03/07/2002 in response to the Greenspan testimony to the Senate (see Table 4). The dotted black line represents the 95% pointwise asymptotic confidence intervals, the grey area denotes the 95% pointwise bootstrapped confidence intervals, and the dot-dashed blue line shows the asymptotic simultaneous confidence intervals across all VIRFs, based on a  $\chi^2(6)$  approximation. The return system includes the S&P 500, the yield of the US constant maturity 10-year Treasury note, and the USD Index.



Figure 5: The predicted 500-step ahead VIRFs (solid black line) are driven by the structural shock on 08/04/2011, during the onset of the EU debt crisis (see Table 4). The dotted black line represents the 95% pointwise asymptotic confidence intervals, the grey area denotes the 95% pointwise bootstrapped confidence intervals, and the dot-dashed blue line shows the asymptotic simultaneous confidence intervals across all VIRFs, based on a  $\chi^2(6)$  approximation. The system includes returns of the S&P 500, the yield of the US constant maturity 10-year Treasury note, and the USD Index.



Figure 6: Predicted 500-step ahead VIRFs in response to a scenario family of 1% structural equity price shocks on the out-of-sample date 01/02/2015: median scenario VIRF (solid black line) with pointwise 25% and 75% quantiles (salmon) and corresponding median target VIRF (dashed green line) with asymptotic 95% confidence intervals, simultaneous across all VIRFs (dot-dashed blue). The system includes returns of the S&P 500, the yield of the US constant maturity 10-year Treasury note, and the USD Index.



Figure 7: Predicted 500-step ahead VIRFs in response to a scenario family of 1% structural bond price shocks on the out-of-sample date 01/02/2015: median scenario VIRF (solid black line) with pointwise 25% and 75% quantiles (salmon) and corresponding median target VIRF (dashed green line) with asymptotic 95% confidence intervals, simultaneous across all VIRFs (dot-dashed blue). The system includes returns of the S&P 500, the yield of the US constant maturity 10-year Treasury note, and the USD Index.

# References

- Amisano, G. and Giannini, C. (2012). *Topics in structural VAR econometrics*, Springer Science & Business Media.
- Angelini, G. and Fanelli, L. (2019). Exogenous uncertainty and the identification of structural vector autoregressions with external instruments, *Journal of Applied Econometrics* **34**(6): 951–971.
- Bauwens, L., Laurent, S. and Rombouts, J. V. K. (2006). Multivariate GARCH models: a survey, Journal of Applied Econometrics 21(1): 79–109.
- Boussama, F., Fuchs, F. and Stelzer, R. (2011). Stationarity and geometric ergodicity of BEKK multivariate GARCH models, *Stochastic Processes and their Applications* **121**(10): 2331–2360.
- Cavicchioli, M. (2019). Fourth moment structure of Markov switching multivariate GARCH models, *Journal of Financial Econometrics* **19**(4): 565–582.
- Cenedese, G. and Mallucci, E. (2016). What moves international stock and bond markets?, *Journal of International Money and Finance* **60**: 94–113.
- Cloyne, J., Jordà, Ò. and Taylor, A. M. (2023). State-dependent local projections: Understanding impulse response heterogeneity, *Technical report*, National Bureau of Economic Research.
- Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH, *Econometric Theory* **11**(1): 122–150.
- Fengler, M. R. and Polivka, J. (2024). Proxy-identification of a structural MGARCH model for asset returns, *Research Paper 24-55*, Swiss Finance Institute. Available at SSRN: http://dx.doi.org/10. 2139/ssrn.4990401.
- Francq, C. and Zakoïan, J. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications,* John Wiley & Sons Ltd.
- Fry, R. and Pagan, A. (2011). Sign restrictions in structural vector autoregressions: A critical review, *Journal of Economic Literature* **49**(4): 938–960.
- Fülle, M. J., Lange, A., Hafner, C. M. and Herwartz, H. (2024). BEKKs: Multivariate Conditional Volatility Modelling and Forecasting. R package version 1.4.4. URL: https://CRAN.R-project.org/package=BEKKs
- Gallant, A., Rossi, P. and Tauchen, G. (1993). Nonlinear dynamic structures, Econometrica 61(4): 871–907.
- Giacomini, R., Kitagawa, T. and Read, M. (2022). Robust Bayesian inference in proxy SVARs, *Journal of Econometrics* **228**(1): 107–126.
- Gill, R. D. and Robins, J. M. (2001). Causal inference for complex longitudinal data: the continuous case, *Annals of Statistics* pp. 1785–1811.
- Gonçalves, S., Herrera, A. M., Kilian, L. and Pesavento, E. (2024). State-dependent local projections, *Journal of Econometrics* p. 105702.
- Grobys, K. (2015). Are volatility spillovers between currency and equity market driven by economic states? Evidence from the US economy, *Economics Letters* **127**: 72–75.
- Hafner, C. M. and Herwartz, H. (2006). Volatility impulse responses for multivariate GARCH models: An exchange rate illustration, *Journal of International Money and Finance* **25**(5): 719–740.

- Hafner, C. M. and Herwartz, H. (2008). Analytical quasi maximum likelihood inference in multivariate volatility models, *Metrika: International Journal for Theoretical and Applied Statistics* **67**(2): 219–239.
- Hafner, C. M. and Herwartz, H. (2023). Asymmetric volatility impulse response functions, *Economics Letters* **222**: 110968.
- Hafner, C. M., Herwartz, H. and Maxand, S. (2022). Identification of structural multivariate GARCH models, *Journal of Econometrics* 227(1): 212–227.
- Hafner, C. M. and Preminger, A. (2009). On asymptotic theory for multivariate GARCH models, *Journal* of *Multivariate Analysis* **100**(9): 2044–2054.
- Horn, R. A. and Johnson, C. R. (2012). Matrix Analysis, 2nd Ed, Cambridge University Press.
- Inoue, A. and Kilian, L. (2013). Inference on impulse response functions in structural VAR models, *Journal of Econometrics* **177**(1): 1–13.
- Jeon, Y., McCurdy, T. H. and Zhao, X. (2021). News as sources of jumps in stock returns: Evidence from 21 million news articles for 9000 companies, *Journal of Financial Economics* **145**(2): 1–17.
- Jin, X., Lin, S. X. and Tamvakis, M. (2012). Volatility transmission and volatility impulse response functions in crude oil markets, *Energy Economics* **34**(6): 2125–2134.
- Jones, C. M., Lamont, O. and Lumsdaine, R. L. (1998). Macroeconomic news and bond market volatility, *Journal of Financial Economics* 47(3): 315–337.
- Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*, Cambridge University Press.
- Koop, G., Pesaran, M. H. and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* **74**(1): 119–147.
- Kremer, M., Lo Duca, M. and Holló, D. (2012). CISS—a composite indicator of systemic stress in the financial system, *Working Paper Series* 1426, European Central Bank.
- Le Pen, Y. and Sévi, B. (2010). Volatility transmission and volatility impulse response functions in european electricity forward markets, *Energy Economics* **32**(4): 758–770.
- Lechner, M. (2010). The relation of different concepts of causality used in time series and microeconometrics, *Econometric Reviews* **30**(1): 109–127.
- Lin, W.-L. (1997). Impulse response function for conditional volatility in GARCH models, *Journal of Business & Economic Statistics* **15**(1): 15–25.
- Liu, X. (2018). Structural volatility impulse response function and asymptotic inference, *Journal of Financial Econometrics* **16**(2): 316–339.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis,* Springer Science & Business Media.
- Lütkepohl, H. (2010). Impulse response function, *Macroeconometrics and time series analysis*, Springer, pp. 145–150.
- Lütkepohl, H., Staszewska-Bystrova, A. and Winker, P. (2015). Confidence bands for impulse responses: Bonferroni vs. Wald, *Oxford Bulletin of Economics and Statistics* **77**(6): 800–821.
- Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons Ltd.

- Mertens, K. and Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the United States, *American Economic Review* **103**(4): 1212–1247.
- Olson, E., Vivian, A. J. and Wohar, M. E. (2014). The relationship between energy and equity markets: Evidence from volatility impulse response functions, *Energy Economics* **43**: 297–305.
- Peterson, R. L. (2016). Trading on sentiment: The power of minds over markets, John Wiley & Sons.
- Rambachan, A. and Shephard, N. (2020). Econometric analysis of potential outcomes time series: instruments, shocks, linearity and the causal response function. Available at: https://ideas.repec. org/p/arx/papers/1903.01637.html.
- Rambachan, A. and Shephard, N. (2021). When do common time series estimands have nonparametric causal meaning? Available at: https://scholar.harvard.edu/shephard/publications/ nonparametric-dynamic-causal-model-macroeconometrics.
- Rosa, C. (2013). The financial market effect of FOMC minutes, *Economic Policy Review* (December Issue, Federal Reserve Bank of New York): 67–81.
- Rubin, D. B. (1980). Randomization analysis of experimental data: The Fisher randomization test, *Journal of the American Statistical Association* **75**(371): 591–593.
- Stock, J. H. and Watson, M. (2012). Disentangling the channels of the 2007-09 recession, *Brookings Papers* on *Economic Activity* **43**(1): 81–156.
- Weber, E. (2010). Structural conditional correlation, Journal of Financial Econometrics 8(3): 392–407.

# A Appendix

### A.1 Notation and results from matrix algebra

**Definition D.1.** Principal matrix square root. Any real symmetric  $(n \times n)$  matrix M can be factorized as  $M = \Gamma \Lambda \Gamma^{\top}$  where  $\Gamma$  is an orthogonal  $(n \times n)$  matrix with the normalized eigenvectors of M as columns and  $\Lambda$  the diagonal matrix of the eigenvalues. The principal matrix square root of M is defined as  $\Gamma \Lambda^{1/2} \Gamma^{\top}$  where  $\Lambda^{1/2}$  denotes the diagonal matrix of the square root of the eigenvalues of M. It is the unique matrix square root which has non-negative eigenvalues, see Horn and Johnson (2012, Theorem 7.2.6).

**Definition D.2.** Orthogonal matrix. An orthogonal matrix *R* is a real  $(n \times n)$  matrix satisfying  $R^{\top}R = RR^{\top} = I_n$ .

**Definition D.3.**  $vec(\cdot)$  operator. The operation vec(M) stacks, starting with the first column, the columns of an  $(n \times n)$  matrix M in an  $n^2$ -dimensional vector. It is a linear operator.

**Definition D.4.**  $\operatorname{vech}(\cdot)$  operator. The operation  $\operatorname{vech}(M)$  stacks, starting with the first column, the lower triangular part of a symmetric  $(n \times n)$  matrix M in an  $n^*$ -dimensional vector where  $n^* = \frac{n(n+1)}{2}$ .

**Definition D.5.** Moore-Penrose inverse. The Moore-Penrose inverse of an  $(m \times n)$  matrix M with  $M^{\top}M$  non-singular is defined as

$$M^{+} = (M^{\top}M)^{-1}M^{\top}.$$
(17)

It has size  $(n \times m)$ .

**Definition D.6.** Duplication matrix. For any symmetric  $(n \times n)$  matrix M, the duplication matrix  $D_n$  denotes the unique  $(n^2 \times n^*)$  matrix, where  $n^* = \frac{n(n+1)}{2}$ , such that

$$\operatorname{vec}(M) = D_n \operatorname{vech}(M).$$
 (18)

The Moore-Penrose inverse of the duplication matrix is denoted by  $D_n^+$ .

**Definition D.7.** Commutation matrix. For every  $(m \times n)$  matrix M, the  $(mn \times mn)$  commutation matrix  $K_{mn}$  is defined by

$$K_{mn}\operatorname{vec}(M) = \operatorname{vec}(M^{\top}). \tag{19}$$

For n = m, we use the abbreviation  $K_n$  for  $K_{nn}$ .

**Result R.1.** vec( $\cdot$ ) operations. For appropriately defined matrices *A*, *B*, and *C* and for some ( $n \times n$ ) matrices *M* and *P*:

$$\operatorname{vec}(ABC) = \left(C^{\top} \otimes A\right) \operatorname{vec}(B)$$
 (20)

$$\operatorname{vec}(A^{\top} \otimes A^{\top}) = \operatorname{vec}((A \otimes A)^{\top})$$
 (21)

$$\operatorname{vec}(M \otimes P) = (I_n \otimes K_n \otimes I_n) [\operatorname{vec}(M) \otimes \operatorname{vec}(P)]$$
. (22)

See Magnus and Neudecker (1988, Theorem 3.10) for a proof of (22).

Result R.2. Matrix derivatives results.

1. For  $n \times n$  matrices *X* and *Z*, *Z* symmetric, it holds:

$$\frac{\partial \operatorname{vec}(XZX)}{\partial \operatorname{vec}(X)^{\top}} = (X^{\top}Z \otimes I_n) + (I_n \otimes XZ)$$
(23)

by an application of the product rule in conjunction with two applications of (20).

2. For any symmetric, positive semidefinite  $n \times n$  matrix H with principal square root  $H^{1/2}$ , it holds:

$$\frac{\partial \operatorname{vec}(H^{1/2})}{\partial \operatorname{vec}(H)^{\top}} = \left[ \left( I_n \otimes H^{1/2} \right) + \left( H^{1/2} \otimes I_n \right) \right]^{-1}$$
(24)

which can be derived by solving a Sylvester type equation for the differential.

3. For any  $(n \times n)$  matrices *M* and *P*, it holds:

$$\frac{\partial \operatorname{vec}(M) \otimes \operatorname{vec}(P)}{\partial \operatorname{vec}(M)^{\top}} = I_n \otimes \operatorname{vec}(P).$$
(25)

4. For any  $(n \times n)$  matrix *M*:

$$\frac{\partial \left(\operatorname{vec}(M) \otimes \operatorname{vec}(M)\right)}{\partial \operatorname{vec}(M)^{\top}} = I_n \otimes \operatorname{vec}(M) + \operatorname{vec}(M) \otimes I_n$$
(26)

by (25) and the product rule.

5. For any  $(n \times n)$  matrix *M*:

$$\frac{\partial \operatorname{vec}(M^{\top})}{\partial \operatorname{vec}(M)^{\top}} = K_n.$$
(27)

## A.2 Prerequisites for the VIRF

*Proof of Proposition* 2.1. Apply the vec-operator to (4), which yields

$$\operatorname{vec}(H_t) = \operatorname{vec}(CC^{\top}) + \sum_{i=1}^p \operatorname{vec}(A_i^{\top}\varepsilon_{t-i}\varepsilon_{t-i}^{\top}A_i) + \sum_{j=1}^q \operatorname{vec}(B_j^{\top}H_{t-j}B_j).$$
(28)

Using (20) and (18), we get:

$$D_{n}\operatorname{vech}(H_{t}) = D_{n}\operatorname{vech}(CC^{\top}) + \sum_{i=1}^{p} (A_{i} \otimes A_{i})^{\top} D_{n}\operatorname{vech}(\varepsilon_{t-i}\varepsilon_{t-i}^{\top}) + \sum_{j=1}^{q} (B_{j} \otimes B_{j})^{\top} D_{n}\operatorname{vech}(H_{t-j}).$$
(29)

Multiplication by  $D_n^+$ , the Moore-Penrose inverse of the duplication matrix, yields the VEC representation of the BEKK model:

$$\operatorname{vech}(H_{t}) = \underbrace{\operatorname{vech}(CC^{\top})}_{=:c} + \sum_{i=1}^{p} \underbrace{D_{n}^{+} (A_{i} \otimes A_{i})^{\top} D_{n}}_{=:\tilde{A}_{i}} \operatorname{vech}(\varepsilon_{t-i}\varepsilon_{t-i}^{\top}) + \sum_{j=1}^{q} \underbrace{D_{n}^{+} (B_{j} \otimes B_{j})^{\top} D_{n}}_{=:\tilde{B}_{j}} \operatorname{vech}(H_{t-j})$$
(30)

Using the definitions of  $X_t$  and  $Y_t$ , we can rearrange (30) to yield:

$$X_{t} = c + \sum_{i=1}^{\max(p,q)} \left( \tilde{A}_{i} + \tilde{B}_{i} \right) X_{t-i} - \sum_{j=1}^{q} \tilde{B}_{j} Y_{t-j} + Y_{t}$$
(31)

where  $\tilde{A}_i = 0$  for i > p and  $\tilde{B}_i = 0$  for i > q. By stationarity of (4), this can be rewritten in VMA( $\infty$ ) form using the lag operator *L*:

$$\underbrace{\left(I_{n^{*}}-\sum_{i=1}^{\max(p,q)}\left(\tilde{A}_{i}+\tilde{B}_{i}\right)L^{i}\right)}_{=:\Phi(L)}X_{t}=c+\underbrace{\left(I_{n^{*}}-\sum_{j=1}^{q}\tilde{B}_{j}L^{j}\right)}_{=:\Theta(L)}Y_{t}$$

$$\Leftrightarrow X_{t}=\Phi(1)^{-1}c+\underbrace{\Phi(L)^{-1}\Theta(L)}_{=:\Psi(L)}Y_{t}$$

$$=\operatorname{vech}(H)+\sum_{i=0}^{\infty}\Psi_{i}Y_{t-i},$$
(32)

where *H* satisfying vech(*H*) =  $\Phi(1)^{-1}c$  denotes the long-run covariance matrix. The  $(n^* \times n^*)$  coefficient matrices  $\Psi_i$  are determined recursively by coefficient matching (Lütkepohl, 2005).

*Proof of Proposition* 2.2. Recall that  $X_t = \operatorname{vech}(\varepsilon_t \varepsilon_t^{\top})$  and  $Y_t = X_t - \operatorname{vech}(H_t)$  and  $h \ge 1$ . Then, we have

$$E[\operatorname{vech}(H_{t+h})|\mathcal{F}_{t-1}] = E[E[X_{t+h}|\mathcal{F}_{t+h-1}]|\mathcal{F}_{t-1}] = E[X_{t+h}|\mathcal{F}_{t-1}].$$
(33)

Using the VMA( $\infty$ ) representation of Proposition 2.1, we obtain for the VIRF in (3)

$$\mathbf{V}_{t+h}(\bar{\boldsymbol{\xi}}^*) = \mathbf{E}\left[\sum_{i=0}^{\infty} \Psi_i Y_{t+h-i} \middle| \mathcal{F}_{t-1}, \boldsymbol{\xi}_t = \bar{\boldsymbol{\xi}}^*\right] - \mathbf{E}\left[\sum_{i=0}^{\infty} \Psi_i Y_{t+h-i} \middle| \mathcal{F}_{t-1}\right].$$
 (34)

Under Assumptions 2.1, we have  $Var(Y_t) < \infty$ , implying that the absolute moments of  $Y_t$  are uniformly bounded. Hence, we can interchange the infinite summation and the conditional expectation:

$$V_{t+h}(\bar{\xi}^*) = \sum_{i=0}^{\infty} \Psi_i \left( E\left[ Y_{t+h-i} | \mathcal{F}_{t-1}, \xi_t = \bar{\xi}^* \right] - E[Y_{t+h-i} | \mathcal{F}_{t-1}] \right) = \Psi_h \left( E\left[ Y_t | \mathcal{F}_{t-1}, \xi_t = \bar{\xi}^* \right] - E[Y_t | \mathcal{F}_{t-1}] \right).$$
(35)

This follows from  $E[Y_{t+h-i}|\mathcal{F}_{t-1}, \xi_t = \bar{\xi}^*] - E[Y_{t+h-i}|\mathcal{F}_{t-1}] = 0$  for all  $(t+h-i) \leq t-1$  due to measurability given  $\mathcal{F}_{t-1}$  and  $E[Y_{t+h-i}|\mathcal{F}_{t-1}, \xi_t = \bar{\xi}^*] - E[Y_{t+h-i}|\mathcal{F}_{t-1}] = 0$  for all  $(t+h-i) \geq t+1$  by the tower property. Using predictability<sup>13</sup> and tower

<sup>&</sup>lt;sup>13</sup>Note that  $H_t^{1/2}$  is  $\mathcal{F}_{t-1}$ -measurable. The measurability follows from the  $\mathcal{F}_{t-1}$ -measurability of  $H_t$  and because the principal square root is a (uniformly) continuous operator in the space of positive definite matrices. Matrix multiplication with  $\tilde{\mathcal{R}}$  preserves measurability.

property arguments, and incorporating model (2), we obtain:

$$\begin{aligned} \mathbf{V}_{t+h}(\bar{\xi}^*) &= \Psi_h \left( \mathbf{E} \left[ X_t - \operatorname{vech}(H_t) | \mathcal{F}_{t-1}, \xi_t = \bar{\xi}^* \right] - \mathbf{E} \left[ X_t - \operatorname{vech}(H_t) | \mathcal{F}_{t-1} \right] \right) \\ &= \Psi_h \left( \mathbf{E} \left[ \operatorname{vech}(\varepsilon_t \varepsilon_t^\top) | \mathcal{F}_{t-1}, \xi_t = \bar{\xi}^* \right] - \mathbf{E} \left[ \operatorname{vech}(\varepsilon_t \varepsilon_t^\top) | \mathcal{F}_{t-1} \right] \right) \\ &= \Psi_h \left( \mathbf{E} \left[ \operatorname{vech} \left( H_t^{1/2} \tilde{R} \bar{\xi}^* \bar{\xi}^{*\top} \tilde{R}^\top H_t^{1/2}^\top \right) | \mathcal{F}_{t-1} \right] - \mathbf{E} \left[ \operatorname{vech}(H_t) | \mathcal{F}_{t-1} \right] \right) \\ &= \Psi_h \left( \operatorname{vech} \left( H_t^{1/2} \tilde{R} \bar{\xi}^* \bar{\xi}^{*\top} \tilde{R}^\top H_t^{1/2}^\top \right) - \operatorname{vech} \left( H_t^{1/2} \tilde{R} \tilde{R}^\top H_t^{1/2}^\top \right) \right) \\ &= \Psi_h \operatorname{vech} \left( H_t^{1/2} \left( \tilde{R} \bar{\xi}^* \bar{\xi}^{*\top} \tilde{R}^\top - I_n \right) H_t^{1/2}^\top \right) . \end{aligned}$$

$$(36)$$

By the symmetry of  $(\tilde{R}\xi^*\xi^{*\top}\tilde{R}^{\top} - I_n)$  and  $H_t^{1/2}$  and using (20), we get:

$$V_{t+h}(\bar{\xi}^*) = \Psi_h D_n^+ \left( H_t^{1/2} \otimes H_t^{1/2} \right) D_n \operatorname{vech}(\tilde{R}\bar{\xi}^*\bar{\xi}^{*\top}\bar{R}^{\top} - I_n).$$
(37)

*Proof of Theorem* **1***.* The result follows from applying the Delta method in conjunction with the asymptotic normality of the QML estimator.

To derive the Jacobian, denote the vector of stacked parameters of the BEKK(p,q) model by  $\eta = (\operatorname{vec}(C)^{\top}, \operatorname{vec}(A_1)^{\top}, \dots, \operatorname{vec}(A_p)^{\top}, \operatorname{vec}(B_1)^{\top}, \dots, \operatorname{vec}(B_q)^{\top})^{\top}$ , supposing that  $\eta \in \mathbb{R}^m$ . Then, expressing the VIRF in (9) with the help of the vec-operator yields:

$$\mathbf{V}_{t+h}(\bar{\xi}^*;\eta) = \Psi_h D_n^+ \left( \operatorname{vec}(H_t^{1/2} \tilde{R} \bar{\xi}^* \bar{\xi}^{*\top} \tilde{R}^\top (H_t^{1/2})^\top) - \operatorname{vec}(H_t) \right)$$
(38)

where  $\{\Psi_h\}_{h\in\mathbb{N}}$  are given in Proposition 2.1.

To calculate the derivative of the VIRF with respect to  $\eta$ , we make use of (23) and (24). Then, for the VIRF at time h = 0:

$$\frac{\partial \mathcal{V}_{t}(\bar{\xi}^{*};\eta)}{\partial \eta^{\top}} = D_{n}^{+} \left[ \frac{\partial \operatorname{vec}\left(H_{t}^{1/2}\tilde{R}\bar{\xi}^{*}\bar{\xi}^{*\top}\tilde{R}^{\top}H_{t}^{1/2}\right)}{\partial \operatorname{vec}\left(H_{t}^{1/2}\right)^{\top}} \frac{\partial \operatorname{vec}\left(H_{t}^{1/2}\right)}{\partial \operatorname{vec}\left(H_{t}\right)^{\top}} \frac{\partial \operatorname{vec}(H_{t})}{\partial \eta^{\top}} - \frac{\partial \operatorname{vec}(H_{t})}{\partial \eta^{\top}} \right] \\
= D_{n}^{+} \left\{ \left[ \left(H_{t}^{1/2}\tilde{R}\bar{\xi}^{*}\bar{\xi}^{*\top}\tilde{R}^{\top}\otimes I_{n}\right) + \left(I_{n}\otimes H_{t}^{1/2}\tilde{R}\bar{\xi}^{*}\bar{\xi}^{*\top}\tilde{R}^{\top}\right) \right] \\
\times \left[ \left(H_{t}^{1/2}\otimes I_{n}\right) + \left(I_{n}\otimes H_{t}^{1/2}\right) \right]^{-1} - I_{n^{2}} \right\} \frac{\partial \operatorname{vec}(H_{t})}{\partial \eta^{\top}} . \tag{39}$$

For the analytical expressions of  $\frac{\partial \operatorname{vec}(H_t)}{\partial \eta^{\top}}$  within the BEKK model, see Hafner and Herwartz (2008).

Now let  $h \in \mathbb{N}$ . Utilizing the recursive definition of the VMA coefficients  $\Psi_i$ , i = 1, ..., h, the derivative of the BEKK(p, q) VIRF can be derived using the product rule (Magnus and Neudecker, 1988, Theorem 5.12):

$$\frac{\partial \mathcal{V}_{t+h}(\bar{\xi}^*;\eta)}{\partial \eta^{\top}} = \left(\mathcal{V}_t^{\top} \otimes I_{n^*}\right) \frac{\partial \operatorname{vec}(\Psi_h)}{\partial \eta^{\top}} + \Psi_h \frac{\partial \mathcal{V}_t(\bar{\xi}^*;\eta)}{\partial \eta^{\top}}$$
(40)

Moreover, we can establish the recursion

$$\frac{\partial \operatorname{vec}(\Psi_{h})}{\partial \eta^{\top}} = \frac{\partial \operatorname{vec}(-\tilde{B}_{h})}{\partial \eta^{\top}} + \sum_{j=1}^{h} \frac{\partial \operatorname{vec}\left(\left(\tilde{A}_{j}+\tilde{B}_{j}\right)\Psi_{h-j}\right)}{\partial \eta^{\top}} \\
= \frac{\partial \operatorname{vec}(-\tilde{B}_{h})}{\partial \eta^{\top}} + \sum_{j=1}^{h} \left\{ \left(\Psi_{h-j}^{\top} \otimes I_{n^{*}}\right) \frac{\partial \operatorname{vec}\left(\tilde{A}_{j}+\tilde{B}_{j}\right)}{\partial \eta^{\top}} \\
+ \left[I_{n^{*}} \otimes \left(\tilde{A}_{j}+\tilde{B}_{j}\right)\right] \frac{\partial \operatorname{vec}\left(\Psi_{h-j}\right)}{\partial \eta^{\top}} \right\}.$$
(41)

For evaluation, we derive  $\frac{\partial \operatorname{vec}(\tilde{A}_j)}{\partial \eta^{\top}}$  and  $\frac{\partial \operatorname{vec}(\tilde{B}_j)}{\partial \eta^{\top}}$ ,  $j = 1, \ldots, h$ . To achieve the first, insert the definition of  $\tilde{A}_j = D_n^+ (A_j \otimes A_j)^{\top} D_n$  and utilize (20), (21), and (22) to perform the following transformations:

$$\frac{\partial \operatorname{vec}\left(\tilde{A}_{j}\right)}{\eta^{\top}} = \frac{\partial \operatorname{vec}\left(D_{n}^{+}\left(A_{j}\otimes A_{j}\right)^{\top}D_{n}\right)}{\partial\eta^{\top}} \\
= \frac{\left(D_{n}^{\top}\otimes D_{n}^{+}\right)\partial \operatorname{vec}\left(A_{j}^{\top}\otimes A_{j}^{\top}\right)}{\partial\eta^{\top}} \\
= \left(D_{n}^{\top}\otimes D_{n}^{+}\right)\left(I_{n}\otimes K_{n}\otimes I_{n}\right)\frac{\partial\left(\operatorname{vec}\left(A_{j}^{\top}\right)\otimes \operatorname{vec}\left(A_{j}^{\top}\right)\right)}{\partial\eta^{\top}}, \quad (42)$$

where  $K_n$  denotes the commutation matrix defined in Definition D.7. Finally, applying (26) and (27) yields:

$$\frac{\partial \left(\operatorname{vec}(A_j^{\top}) \otimes \operatorname{vec}(A_j^{\top})\right)}{\partial \eta^{\top}} = K_n \otimes \operatorname{vec}(A_j^{\top}) + \operatorname{vec}(A_j^{\top}) \otimes K_n .$$
(43)

The derivations for  $\tilde{B}_j = D_n^+ (B_j \otimes B_j)^\top D_n$  follow a similar manner. Therefore,

$$\frac{\partial \operatorname{vec}\left(\tilde{B}_{j}\right)}{\eta^{\top}} = \left(D_{n}^{\top} \otimes D_{n}^{+}\right)\left(I_{n} \otimes K_{n} \otimes I_{n}\right)\left[K_{n} \otimes \operatorname{vec}(B_{j}^{\top}) + \operatorname{vec}(B_{j}^{\top}) \otimes K_{n}\right] \,. \tag{44}$$

In summary, based on (39), the recursion for the derivative of the BEKK(p,q) VIRF is given by (40). It can be implemented utilizing (41), (42), (43), and (44).

For the BEKK(1, 1) model, there is a more compact recursion for  $\frac{\partial V_{t+h}(\bar{\xi}^*;\eta)}{\partial \eta^{\top}}$  based on (7). Let  $h \in \mathbb{N}$ . By an application of the product rule, it holds:

$$\begin{aligned} \frac{\partial \mathcal{V}_{t+h}(\bar{\xi}^*;\eta)}{\partial \eta^{\top}} &= \frac{\partial \operatorname{vec}\left(\left(\tilde{A}_1 + \tilde{B}_1 \mathbb{1}_{\{h>1\}}\right) \mathcal{V}_{t+h-1}(\bar{\xi}^*;\eta)\right)}{\partial \eta^{\top}} \\ &= \left(\mathcal{V}_{t+h-1}(\bar{\xi}^*;\eta)^{\top} \otimes I_{n^*}\right) \frac{\partial \operatorname{vec}\left(\tilde{A}_1 + \tilde{B}_1 \mathbb{1}_{\{h>1\}}\right)}{\eta^{\top}} \\ &+ \left(\tilde{A}_1 + \tilde{B}_1 \mathbb{1}_{\{h>1\}}\right) \frac{\partial \mathcal{V}_{t+h-1}(\bar{\xi}^*;\eta)}{\partial \eta^{\top}}.\end{aligned}$$

In this case, we need to calculate  $\frac{\partial \operatorname{vec}(\tilde{A}_1 + \tilde{B}_1 \mathbb{1}_{\{h>1\}})}{\eta^{\top}}$  only once to establish the recursion for  $\frac{\partial V_{t+h}(\bar{\xi}^*;\eta)}{\partial \eta^{\top}}$ . This completes the proof.

### A.4 Causality of the VIRF

*Proof of Proposition 2.4.* Denote by  $E_{\xi_{t+1},...,\xi_{t+h}|(\xi_t=\bar{\xi}^*)}[X_{t+h}|\mathcal{F}_{t-1}]$  an expectation operator that integrates over the arguments of  $\xi_{t+1},...,\xi_{t+h}$ , using the  $\mathcal{F}_{t-1}$ -conditional joint density of all  $\xi_t,...,\xi_{t+h}$ , however, with the argument referring to  $\xi_t$  set to  $\bar{\xi}^*$ . Furthermore, denote by  $p_{\xi_t}(\bar{\xi}^*|\mathcal{F}_{t-1})$  the marginal density of  $\xi_t$ , conditioned on  $\mathcal{F}_{t-1}$  and evaluated at  $\bar{\xi}^*$ .

We have

$$\begin{split} \mathbf{V}_{t+h}(\bar{\xi}^*) &= \mathbf{E}[\mathbf{vech}(H_{t+h})|\mathcal{F}_{t-1}, \xi_t = \bar{\xi}^*] - \mathbf{E}[\mathbf{vech}(H_{t+h})|\mathcal{F}_{t-1}] \\ &= \mathbf{E}[X_{t+h}|\mathcal{F}_{t-1}, \xi_t = \bar{\xi}^*] - \mathbf{E}[X_{t+h}|\mathcal{F}_{t-1}] \\ &= \mathbf{E}[X_{t+h}(\xi_t, \{\xi_s\}_{t+1 \le s \le t+h})|\mathcal{F}_{t-1}, \xi_t = \bar{\xi}^*] - \mathbf{E}[X_{t+h}|\mathcal{F}_{t-1}] \\ &= \frac{\mathbf{E}_{\xi_{t+1}, \dots, \xi_{t+h}|(\xi_t = \bar{\xi}^*)}[X_{t+h}(\bar{\xi}^*)|\mathcal{F}_{t-1}]}{p_{\xi_t}(\bar{\xi}^*|\mathcal{F}_{t-1})} - \mathbf{E}[X_{t+h}|\mathcal{F}_{t-1}] \\ &= \frac{\mathbf{E}_{\xi_{t+1}, \dots, \xi_{t+h}|(\xi_t = \bar{\xi}^*)}[X_{t+h}(\bar{\xi}^*)|\mathcal{F}_{t-1}]}{p_{\xi_t}(\bar{\xi}^*|\mathcal{F}_{t-1})} \\ &+ \frac{p_{\xi_t}(\bar{\xi}^*|\mathcal{F}_{t-1}) \mathbf{E}[X_{t+h}(\bar{\xi}^*)|\mathcal{F}_{t-1}] - p_{\xi_t}(\bar{\xi}^*|\mathcal{F}_{t-1}) \mathbf{E}[X_{t+h}(\bar{\xi}^*)|\mathcal{F}_{t-1}]}{p_{\xi_t}(\bar{\xi}^*|\mathcal{F}_{t-1})} \\ &= \mathbf{E}[X_{t+h}(\bar{\xi}^*) - X_{t+h}|\mathcal{F}_{t-1}] + \Delta_{t+h}(\bar{\xi}^*|\mathcal{F}_{t-1}) \end{split}$$

where  $\Delta_{t+h}(\bar{\xi}^*|\mathcal{F}_{t-1}) \coloneqq \frac{\mathbb{E}_{\xi_{t+1},\dots,\xi_{t+h}|(\xi_t=\bar{\xi}^*)}[X_{t+h}(\bar{\xi}^*)|\mathcal{F}_{t-1}]}{p_{\xi_t}(\bar{\xi}^*|\mathcal{F}_{t-1})} - \mathbb{E}[X_{t+h}(\bar{\xi}^*)|\mathcal{F}_{t-1}]$ . The first two lines apply the definition and properties of the VIRF. Line three invokes Assumption 2.3, while line four utilizes the definition of conditional densities. The final lines involve rearranging the terms.

Assumption 2.5 asserts that the contemporaneous treatment  $\xi_t$  is jointly independent of all future treatments and potential outcomes. This ensures that for all h > 0, the joint density factors such that

$$\mathbf{E}_{\xi_{t+1},\dots,\xi_{t+h}|(\xi_t=\bar{\xi}^*)}[X_{t+h}(\bar{\xi}^*)|\mathcal{F}_{t-1}] = p_{\xi_t}(\bar{\xi}^*|\mathcal{F}_{t-1})\,\mathbf{E}[X_{t+h}(\bar{\xi}^*)|\mathcal{F}_{t-1}]\,,$$

implying  $\Delta_{t+h}(\bar{\xi}^*|\mathcal{F}_{t-1}) = 0$ . Lastly, it is important to note that, under the maintained continuity conditions in Assumption 2.6, the conditional densities are uniquely defined—see Gill and Robins (2001). Hence, the VIRF identifies a meaningful causal treatment effect.