Forecast Relative Error Decompositions with an Application to Cyber Risk*

GOURIÉROUX, C.,†and Q., LEE‡

Revised, January 6, 2026.

Abstract

We introduce a nonlinear extension to the Forecast Error Variance Decomposition (FEVD) with a new class of risk decomposition measures called Forecast Relative Error Decompositions (FRED). Intuitively, it is a general decomposition formula that can be applied to positive transforms of dynamic processes and incorporate moment information beyond the second order. We focus our discussion on two FRED decompositions: [1] The Forecast Error Kullback Decomposition (FEKD), a FRED based on transition densities, which allows for the analysis of risk at different points of the predictive distribution. [2] The Forecast Error Laplace Decomposition (FELD), a FRED based on conditional log Laplace transforms, which allows for the analysis of risk at different levels of risk aversion. We demonstrate the relevance of our new measures in the context of cyber risk by modeling cyber attack frequency count data in the United States.

Keywords: Nonlinear Forecast, Predictive Distribution, Learning, FEVD, FRED, Kullback Measure, Laplace Transform, Count Data, Stochastic Volatility, Cyber Risk.

JEL Codes: C01, C32, C53

*We thank Burda, M., Chausse, P., Daouia, A., Hamilton, J., Jasiak, J., Snudden, S., Wan, Y., the participants of the 2024 NBER-NSF Time Series Conference, the 2024 Conference on Real-Time Data Analysis, Methods and Applications, the 39th Canadian Econometrics Study Group Meeting, the Econometrics Seminar at the University of Toronto, the CIREQ Econometrics seminar in Montreal, and the 18th International CFE-CM Statistics Joint Conference for their helpful comments.

†University of Toronto, Toulouse School of Economics and CREST, email: Christian.Gourieroux@ensae.fr.

‡University of Toronto, Email: qt.lee@mail.utoronto.ca.

1 Introduction

Since its introduction in the time series literature [see e.g. Doob (1953), Whittle (1963)], the Forecast Error Variance Decomposition (FEVD) has been used extensively in both macroeconomics and finance to quantify the importance of economic shocks in a dynamic model. In the linear framework, it can be easily constructed and applied to models such as the Structural Vector Autoregression (SVAR) with i.i.d. Gaussian innovations. However, a recent strand of the literature has appealed to nonlinearities introduced in the underlying data generating process of a dynamic model. Indeed, the assumption of linearity can be restrictive and provide an inaccurate depiction of reality[1]. Unfortunately, the presence of nonlinearities will complicate the construction of the FEVD, and its extension into the nonlinear dynamic framework models is still in its infancy.

Different notions of decomposition in the literature have to be distinguished [see Lee (2025) for a review]. The first is a "counterfactual" type analysis corresponding to a hypothetical environment of shocks on specific structural innovations components [see e.g. Lanne and Nyberg (2016), Isakin and Ngo (2020)]. Because these individual shocks themselves are inherently unobservable in practice, this type of analysis relies on a set of additional restrictions for identification. The second is the analysis on "realized" data whereby the interest is not on individual shocks, but rather all shocks at a particular time or maturity horizon. A decomposition along this dimension is always identifiable and is typically done in a finance or insurance setting where the objective is to understand the term structure of risk. The interest in this paper is on this latter notion.

We begin by arguing that that there are several drawbacks with the traditional FEVD approach in a nonlinear setting. First, the notion of variance on the absolute forecast errors is based on a Taylor approximation at order 2 and is valid only for small risks. By

[1]For instance, the linear model exhibits history invariance, which implies that the FEVD remains the same regardless of the current state in the economy.

construction, it fails to account for nonlinear features such as asymmetric or extreme risks. Secondly, the FEVD only exists if a process is square integrable, which rules out data with (conditional) fat tails. Lastly, the FEVD is only applied on pointwise predictions of a process, but not for any of its nonlinear transforms [e.g. the FEVD for prices is not the same for the log of prices]. As such, a nonlinear extension of the FEVD needs to account for higher-order moment information to mitigate some of these concerns.

We focus on the extension of the term structure of risks and propose a new class of measures called Forecast Relative Error Decompositions (FRED). For small risks, these are variance decompositions written on the relative forecast errors of positive, univariate stochastic processes. We show that the FRED provides a general decomposition formula which can be applied to multivariate processes by considering their univariate positive transformations. Our focus will be on two special cases of the FRED: [1] The Forecast Error Kullback Decomposition (FEKD), which is a FRED written on the conditional predictive density of a process. [2] The Forecast Error Laplace Decomposition (FELD), which is a FRED written on the conditional Laplace transform of a process. We showcase our new measures with an application to the modeling of cyber attack frequency breach counts in the United States through the lens of a Negative Binomial Autoregressive framework. We show how the term structures of risks of cyber attacks depend on the levels of risk aversions (respectively on our interests in extreme values) and on the environment.

There are clear advantages of our decomposition measures in a nonlinear setting in comparison to the FEVD. The most important benefit is that our measures provide a decomposition at different points of the predictive density [in the case of the FEKD] and different levels of risk aversion [in the case of the FELD]. Clearly, the decomposition of risk may not be the same at the tails versus the centre of a distribution, or at varying risk attitudes. The FEVD on the other hand, provides only one decomposition in this respect. A second key advantage is that our decomposition measures incorporate higher-order moment information by working with the updating density forecasts and Laplace transforms. This is in contrast to the variance on updating pointwise predictions for the FEVD. A third benefit of our approach is that the FEKD and FELD may exist in situations where the FEVD does not. This situation arises in a number of scenarios, particularly in finance where data tends to exhibit fatter tails.

The organization of the paper is as follows: Section 2 contains a definition of the FEVD and its interpretation for small risks. In Section 3, we introduce the FRED formula and its application to conditional predictive densities and conditional Laplace transforms to yield the FEKD and FELD, respectively. We provide examples in which the decompositions have closed form expressions in Section 4.. The application to cyber attack counts is discussed in Section 5, and we conclude in Section 6. Technical derivations are provided in the (online) Appendix.

2 Variance Analysis as a Local Approximation

2.1 Forecast Error Variance Decomposition

Let us consider a multivariate stochastic process (Y_t) and its increasing sequence of information sets (or filtration) (I_t), where $I_t = (\underline{Y}_t)$ is the σ-algebra generated by the present and lagged values of the process. The forecast errors at horizon h can be written as a sum of multivariate forecast updates:

$$Y_{t+h} - \mathbb{E}(Y_{t+h}|I_t) = \sum_{k=0}^{h-1} [\mathbb{E}(Y_{t+h}|I_{t+k+1}) - \mathbb{E}(Y_{t+h}|I_{t+k})]. \tag{2.1}$$

By the optimality of the conditional expectations, the forecast updatings are uncorrelated conditional on I_t. Then, we get the FEVD [see e.g. Gallant, Rossi and Tauchen (1993), p. 878], also known as the conditional variance profile.

Definition 1 (Forecast Error Variance Decomposition (FEVD)). *Suppose* (Y_t) *is a stochastic process which is square integrable with information set* (I_t). *The FEVD is given by:*

$$\mathbb{V}[Y_{t+h}|I_t] = \sum_{k=0}^{h-1} \mathbb{V}\{[\mathbb{E}(Y_{t+h}|I_{t+k+1}) - \mathbb{E}(Y_{t+h}|I_{t+k})]|I_t\} = \sum_{k=0}^{h-1} \mathbb{E}\left\{\mathbb{V}\left[\mathbb{E}\left(Y_{t+h}|I_{t+k+1}\right)|I_{t+k}\right]|I_t\right\}. \quad (2.2)$$

This matricial decomposition is conditional on I_t, takes into account the conditional heteroscedastic features and is valid for both stationary and non-stationary processes. For a strictly stationary linear process (Y_t), the FEVD can be deduced from its infinite strong moving average representation:

$$Y_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}, A_0 = Id, \quad (2.3)$$

where ε_t is a strong white noise, that is a sequence of i.i.d. random vectors, with zero mean and variance-covariance Σ, and the moving average coefficients satisfy the square integrability condition $\sum_{j=0}^{\infty} ||A_j||^2 < \infty$. Then, the FEVD becomes:

$$\mathbb{V}\left[Y_{t+h}|\underline{Y_t}\right] = \mathbb{V}[Y_{t+h}|\underline{\varepsilon_t}] = \sum_{k=0}^{h-1} A_k \Sigma A_k'. \quad (2.4)$$

In this special moving average case with i.i.d. noise, the decomposition is history invariant, that is, independent on the values of process (Y_t) before time t. [2].

2.2 Local Analysis of Risk

The comparison of multivariate quantitative risks (i.e. the losses) X and Y is based on the notion of stochastic dominance at order 2. Y is riskier than X if and only if $\mathbb{E}[\upsilon(Y)] \geq \mathbb{E}[\upsilon(X)]$, for any increasing convex function υ (such that the expectations exist) [see e.g.

[2]This decomposition has been initially analyzed in a systematic way by Doob (1953). See also Whittle (1963), where the best predictor $\mathbb{E}\left[Y_t|\underline{Y_{t-1}}\right]$ is replaced by linear projection of Y_t on the space of linear functions of $Y_{t-1}, Y_{t-2}, \ldots$. In the strict linear stationary case, the two approaches coincide. They differ in the general framework where the conditional mean can be a nonlinear function of the past.

Rothschild and Stiglitz (1970), Fishburn and Vickson (1978)]. In the univariate case, it is well known that the quadratic function $y \to y^2$ that underlies the definition of variance is convex, but not increasing (except if X and Y are nonnegative). Then the variance is not directly appropriate for measuring risk. However, we have the following local expansion [see Appendix A.1.1. for proof]:

Lemma 1. *Let us assume that v is increasing and convex such that $v(0) = 0$, and the variable Y is close to 0, with 0 mean. Then $\mathbb{E}[v(Y)] \geq 0$ and $\mathbb{E}[v(Y)] \approx \frac{1}{2} \operatorname{Tr} \left[\frac{\partial^2 v(0)}{\partial y \partial y'} \mathbb{V}(Y) \right]$, where* Tr *denotes the trace operator.*

Since Y is assumed close to zero, the variance-covariance matrix is seen as a local risk measure for small risks and by construction does not account for asymmetric risks (which would require an expansion up to order 3), or to extreme risks (which would require an expansion up to order 4). This second–order expansion has been the basis for defining the local version of absolute risk aversion[3] [Arrow (1965)], or for justifying the mean-variance portfolio management in finance [Markowitz (1952), Markowitz and Todd (2000)][4]. Since the risk on a price evolution $y = p_{t+h} - p_t$ increases generally with the term, the mean-variance management is appropriate in the short run.

3 Forecast Relative Error Decomposition (FRED)

As mentioned in the preceding section, the FEVD is only a local measure of small risks and does not account for nonlinear features such as conditional asymmetries or extremes. We now propose a general decomposition formula that can be applied to any univariate

[3]In the multidimensional framework, the risk aversion is a matrix directly linked to the Hessian at 0 of function v, that is, $\frac{\partial^2 v(0)}{\partial y \partial y'}$ [Karni (1979)]

[4]In the standard financial application with Y as a vector of asset returns, the interest is in the risk of the portfolio returns $\alpha' Y$, where α is the vector of portfolio allocations in values. Then the risk becomes scalar, that is we have $v(Y) = u(\alpha' Y)$, where u is defined on $\mathbb{R}$. We deduce that the matrix of risk aversion $\frac{\partial^2 v(0)}{\partial y \partial y'} = \frac{d^2 u(0)}{dw^2} \alpha \alpha'$ is of rank 1. It involves the effect of portfolio allocation and a scalar risk aversion.

positive transformation of multivariate processes. When applied to transition densities and conditional Laplace transforms, we get the Forecast Error Kullback Decomposition (FEKD) and the Forecast Error Laplace Decomposition (FELD), respectively.

3.1 Relative Forecast Updating

Consider a univariate positive stochastic process (Z_t) with a sequence of increasing information sets (I_t) such that Z_t is measurable with respect to I_t. We can construct an analogue of the FEVD by considering the relative forecast update:

$$\frac{\mathbb{E}(Z_{t+h}|I_{t+k})}{\mathbb{E}(Z_{t+h}|I_{t+k+1})},\ \ k = 0, ..., h-1.$$

Then we have:

$$\frac{\mathbb{E}(Z_{t+h}|I_t)}{Z_{t+h}} = \prod_{k=0}^{h-1} \left[\frac{\mathbb{E}(Z_{t+h}|I_{t+k})}{\mathbb{E}(Z_{t+h}|I_{t+k+1})} \right].$$

By taking the log of both sides and then the conditional expectation given I_t, we get the Forecast Relative Error Decomposition [see Appendix A.1.2 for proof]:

Definition 2 (Forecast Relative Error Decomposition (FRED))**.**

$$\mathbb{E}\left\{ \log \left[\frac{\mathbb{E}(Z_{t+h}|I_t)}{Z_{t+h}} \right] \middle| I_t \right\} = \sum_{k=0}^{h-1} \mathbb{E}\left\{ \log \left[\frac{\mathbb{E}(Z_{t+h}|I_{t+k})}{\mathbb{E}(Z_{t+h}|I_{t+k+1})} \right] \middle| I_t \right\}, \tag{3.1}$$

where each term in the decomposition is nonnegative.

For small risks, the FRED is a variance decomposition written on the relative error forecasts, since:

$$\mathbb{E}\left\{ -\log \left[\frac{Z_{t+h}}{\mathbb{E}(Z_{t+h}|I_t)} \right] \right\} \approx \frac{1}{2} \mathbb{V} \left[\frac{Z_{t+h} - \mathbb{E}(Z_{t+h}|I_t)}{\mathbb{E}(Z_{t+h}|I_t)} \middle| I_t \right].$$

Therefore, for small risks, the main difference between the FRED and the FEVD is that the variance expansion is written on the relative forecast errors instead of the absolute forecast errors.

In general, the FRED (3.1) can be written as:

$$\gamma(h|I_t) = \sum_{k=0}^{h-1} \gamma(k, h|I_t),$$

where the left hand side of equation (3.1) $\gamma(h|I_t)$ measures the risk on the prediction errors at horizon h, and the generic term of the right hand side $\gamma(k, h|I_t)$ has a forward interpretation. More precisely, we have:

$$\gamma(k, h|I_t) = \mathbb{E}\left\{ \log\left[\frac{\mathbb{E}(Z_{t+h}|I_{t+k})}{\mathbb{E}(Z_{t+h}|I_{t+k+1})} \right] \middle| I_t \right\} = \mathbb{E}\left[\mathbb{E}\left\{ \log\left[\frac{\mathbb{E}(Z_{t+h}|I_{t+k})}{\mathbb{E}(Z_{t+h}|I_{t+k+1})} \right] \middle| I_{t+k} \right\} \middle| I_t \right].$$

Therefore, $\gamma(k, h|I_t)$ is the expectation at date t of the risk on the short run prediction error of Z_{t+h} at date $t+k$. This forward interpretation involves three dates: t, $t+k$ and $t+h$.

A decomposition of risk in the multivariate framework can be deduced from the FRED by considering univariate positive transformations of the process. Suppose $(Y_t) \in \mathbb{R}^n$ and $Z_t = b(Y_t) > 0$, where $b : \mathbb{R}^n \to \mathbb{R}$. Then the FRED formula can be applied on Z_t to produce a valid decomposition. For example, if (Y_t) is discrete, then the transformation $Z_t = \mathbf{1}(Y_{t+h} = y)$ will lead to a FRED on conditional forecast densities since:

$$\mathbb{E}(Z_{t+h}|I_t) = \mathbb{E}\left[\mathbf{1}(Y_{t+h} = y|I_t)\right] = f(y, h|I_t).$$

Alternatively, the transformation $Z_t = \exp(-u'Y_{t+h})$ will lead to a FRED on the conditional Laplace transform:

$$\mathbb{E}(Z_{t+h}|I_t) = \mathbb{E}\left[\exp(-u'Y_{t+h})|I_t\right] = \Psi(u, h|I_t).$$

Other positive nonlinear transformations of the conditional distribution can be more appealing for financial applications when (Y_t) is a vector of financial returns. A typical example is the Required Capital (RC) associated with a given portfolio with allocation a and a risk level α. For date t and horizon h, it is given by:

$$RC_t(a, \alpha, h) = -VaR_t(a, \alpha, h), \tag{3.2}$$

where the Value-at-Risk (VaR) is the opposite of the conditional quantile of $a'Y_{t+h}$ given $I_t = (\underline{Y_t})$ at level α. This required capital is usually positive if the risk level α is small. Then this decomposition can be used to assign the total required capital by terms in the balance sheet of banks and financial institutions. Then, we have as many decompositions as allocation a and risk level α. In the next two subsections, we formally introduce the FRED applied to these transformations.

3.2 Forecast Error Kullback Decomposition (FEKD)

Let us consider the application of the FRED to predictive densities. Suppose we have a strictly stationary Markov process (Y_t) with transition density at horizon h and time t given by $f(y, h|Y_t)$. When Y_t is a (multivariate) continuous variable, this density is with respect to the Lebesgue measure. When Y is discrete, the density is with respect to the counting measure and coincides with the probability mass function. Clearly, these conditional densities are univariate positive transforms of Y_t. They are well defined for horizons $h \geq 1$, but not at horizon $h = 0$, where the value Y_t is perfectly known and the predictive distribution degenerates into a point mass at Y_t. We get the Forecast Error Kullback Decomposition, which is a measure of risk on the updating of predictive densities.

Definition 3 (Forecast Error Kullback Decomposition (FEKD))**.** *For any value y, we* *have:*

$$\mathbb{E}\left\{\log\left[\frac{f(y,h|I_t)}{f(y,1|I_{t+h-1})}\right]\Bigg|I_t\right\} = \sum_{k=0}^{h-2}\mathbb{E}\left\{\log\left[\frac{f(y,h-k|I_{t+k})}{f(y,h-k-1|I_{t+k+1})}\right]\Bigg|I_t\right\}. \tag{3.3}$$

The generic term on the right hand side of the FEKD can be represented as:

$$\gamma(k,h|y,I_t) = \mathbb{E}\left\{\mathbb{E}\left[\log\frac{f(y,h-k|I_{t+k})}{f(y,h-k-1|I_{t+k+1})}\Bigg|I_{t+k}\right]\Bigg|I_t\right\}.$$

It is easily checked by applying the Bayes formula that the term within the brackets $\{\cdot\}$ is the conditional Kullback proximity measure (or contrast, or divergence) between the two conditional densities. This is the intuition behind the name of the decomposition.

The formula (3.3) takes on an argument of $y \in D$, where D is the domain of the transition density. Thus, the FEKD provides a decomposition of risk at any point of the predictive distribution. The decomposition may not be the same at the tails of the distribution compared to other points, such as at the centre. Conversely, we only have a single FEVD, which may not be a suitable representation for all points in the predictive density. For a univariate (Y_t) and T observations $Y_1, ..., Y_T$ of the process, we can derive the sample distributions, then the sample deciles, and compare the FEKD evaluated at different deciles.

When (Y_t) is continuous, it can also be shown that the FEKD is invariant by one-to-one differentiable transformations. This can be seen in the change-of-variables formula, $f_Y(y) = f_X(b^{-1}(y))\left|\frac{d}{dy}(b^{-1}(y))\right|$, where the term $\left|\frac{d}{dy}(b^{-1}(y))\right|$ corresponds to the Jacobian effect. This effect is independent of the information set and will disappear since we are taking the ratio of two predictive densities at each horizon. Moreover,the FEKD can exist in cases where the FEVD has no meaning. Indeed, the FEVD requires that the observed Y_t are square integrable, so it cannot be applied to data with fat tails.

3.3 Forecast Error Laplace Decomposition (FELD)

Now we consider the application of the FRED to conditional Laplace transforms. Suppose (Y_t) is a stochastic process of dimension n with conditional Laplace transform at horizon h defined by:

$$\Psi(u, h|I_t) = \mathbb{E}\left[\exp(-u'Y_{t+h})|I_t\right], \tag{3.4}$$

where $u \in D \subset \mathbb{R}^n$ such that the expectation exists on domain D.

Definition 4 (Forecast Relative Error Decomposition (FELD))**.**

$$\mathbb{E}\left\{\log\left[\frac{\Psi(u,h|I_t)}{\exp(-u'Y_{t+h})}\right]\middle| I_t\right\} = \sum_{k=0}^{h-1}\mathbb{E}\left\{\log\left[\frac{\Psi(u,h-k|I_{t+k})}{\Psi(u,h-k-1|I_{t+k+1})}\right]\middle| I_t\right\}, \tag{3.5}$$

for any $u \in D$, where D is the domain of arguments that ensures the existence of the conditional Laplace transforms.

When (Y_t) satisfies some "positivity" restrictions, the Laplace transform with positive argument u characterizes the distribution. It is known that the knowledge of the Laplace transform is equivalent to the knowledge of the conditional distribution for the Gaussian case, or if the process satisfies some positivity restrictions [see Feller (1991) and the examples in Section 4]. Moreover, we have $\mathbb{E}\left[\exp(-u'Y_{t+h})|I_t\right] \leq 1$, and the FELD always exists, even if the distribution of Y_t has very fat tails.

Similar to the FEKD, the FELD in (3.5) also defines several decompositions which depend on the argument u. In some applications, u can have economic interpretations, such as the payoff to claim derivatives (after the exponential form of the utility function is applied to the future claim indicator) or the price of a risky asset (after a transformation leading to the definition of the certainty equivalent, as seen below).

3.3.1 Decomposition of Risk Premium

The FELD can be interpreted as a decomposition of risk premiums for spot prices. Consider a decision maker with exponential (CARA) utility function $u(y) = -\exp(-uy)$ and a one-dimensional claim y. The certainty equivalent $\pi(u)$ is the value is:

$$\pi(u) = -\frac{1}{u} \log \mathbb{E}\left[\exp(-uY)\right] \equiv -\frac{1}{u} \log \Psi(u).$$

This is a function of parameter u, interpretable as the Arrow-Pratt scalar measure of absolute risk aversion. Equivalently, for a risky asset $Y_t > 0$, its value is:

$$\pi(u, h|I_t) = -\frac{1}{u} \log \Psi(u, h|I_t), \tag{3.6}$$

which is the spot value of an asset at time t and horizon h. This corresponds to a contract written at time t, which captures the value of delivering the asset at some future horizon h. For an investor with risk aversion parameter u, the FELD in (3.5) is:

$$\underbrace{\pi(u, h|I_t) - \mathbb{E}(Y_{t+h}|I_t)}_{\text{Risk premium of the asset}} = \sum_{k=0}^{h-1} \underbrace{\mathbb{E}\left[\pi(u, h-k-1|I_{t+k+1}) - \pi(u, h-k|I_{t+k}) \middle| I_t\right]}_{\text{Difference in value of forward contracts}}. \tag{3.7}$$

The term $\pi(u, h|I_t) - \mathbb{E}(Y_{t+h}|I_t)$ is the difference between the value (price) of Y_{t+h} at date t and its historical conditional expectation. This difference is positive (by Jensen's inequality) and usually interpreted as a risk premium. Therefore, (3.7) provides a decomposition of this risk premium at varying levels of risk aversion u. More precisely, the term $\pi(u, k, h|I_t) = \mathbb{E}\left[\pi(u, h-k|I_{t+k}) \middle| I_t\right]$ is the value of a forward contract of the asset, written at time t, for a payment at time $t+k$ and delivery at time $t+h$. Hence, the generic term on the RHS of (3.7) captures the difference in values π_f of the forward contracts for payment at time $t+k$ and $t+k+1$ for the delivery of the asset at time $t+h$. As h varies, we get a decomposition of the term structure for the risk premium, or equivalently of the spot value (price) as:

$$\pi(u,h|I_t) = \mathbb{E}(Y_{t+h}|I_t) + \sum_{k=0}^{h-1} \mathbb{E}\left[\pi_f(u,k+1,h|I_{t+k+1}) - \pi_f(u,k,h|I_{t+k})\Big| I_t\right], \tag{3.8}$$

which is compatible with the no dynamic arbitrage condition between spot and forward contracts. There is a debate on the valuation approach to be chosen for contingent assets [see e.g. Embrechts (2000)]. For financial assets traded on very liquid markets, this is usually done by introducing a stochastic discount factor to satisfy the no dynamic arbitrage opportunity condition. The situation is different for individual insurance contracts or for operational risks [see the discussion of cyber risk in Section 7]. The certainty equivalent principle is a more appropriate valuation approach in such frameworks and we have shown ex-post that it is compatible with the no dynamic arbitrage opportunity assumption.

Remark 1: The interpretations above are also valid when Y_t is multivariate with components being the future values of assets or of contracts. Then, the certainty equivalent for a portfolio is:

$$\pi(A,a) = -\frac{1}{A}\log\mathbb{E}\left[\exp(-Aa'Y)\right],$$

where A is the absolute risk aversion and a the porfolio allocation.

4 Examples

4.1 Examples of FEKD

4.1.1 The Gaussian VAR(1)

Let us assume that the n-dimensional stationary process (Y_t) satisfies:

$$Y_t = \Phi Y_{t-1} + \varepsilon_t, \tag{4.1}$$

where the eigenvalues of Φ have a modulus strictly smaller than 1 and the ε_t's are i.i.d. Gaussian $\varepsilon_t \sim N(0, \Sigma)$. Then, the conditional distribution of Y_{t+h} given Y_t is Gaussian with mean $\Phi^h Y_t$ and variance-covariance matrix $\Sigma_h = \Sigma + \Phi\Sigma\Phi' + ... + \Phi^{h-1}\Sigma(\Phi')^{h-1}$. The following proposition provides the closed form FEKD for the Gaussian VAR(1).

Proposition 1. *In the Gaussian VAR(1) model, the FEKD is of the form:*

$$a(h|Y_t) + b(h|Y_t)y + y'c(h|Y_t)y = \sum_{k=0}^{h-2} \left[a(h,k|Y_t) + b(h,k|Y_t)y + y'c(k,h|Y_t)y\right],$$

where:

$$\begin{aligned} a(h,k|Y_t) =& \frac{1}{2}\log\left[\frac{\det \Sigma_{h-k-1}}{\det \Sigma_{h-k}}\right] - \frac{1}{2}Y_t'\left(\Phi^h\right)'\left(\Sigma_{h-k}^{-1} - \Sigma_{h-k-1}^{-1}\right)\Phi^h Y_t \\ &+ \frac{1}{2}\operatorname{Tr}\left[\Sigma_{h-k-1}^{-1}\Phi^{h-k-1}\Sigma_{k+1}\left(\Phi^{h-k-1}\right)' - \Sigma_{h-k}^{-1}\Phi^{h-k}\Sigma_k\left(\Phi^{h-k}\right)'\right] \\ b(h,k|Y_t) =& Y_t'\left(\Phi^h\right)'\left(\Sigma_{h-k}^{-1} - \Sigma_{h-k-1}^{-1}\right),\ c(h,k|Y_t) = -\frac{1}{2}\left(\Sigma_{h-k}^{-1} - \Sigma_{h-k-1}^{-1}\right). \end{aligned}$$

Proof: See Appendix A.2.1.

Given a horizon h and conditioning value Y_t, the left hand side of the FEKD for the VAR(1) model is a functional quadratic form of the argument y, which can be decomposed into a sum of smaller quadratic forms of y at each horizon $k = 0, ..., h-2$. The linear component $b(h,k|Y_t)$ is a function of Y_t, which corresponds to the forward interpretation of the elements in the decomposition. The quadratic component $c(h,k|Y_t)$ is independent of Y_t, since the Gaussian VAR(1) model is conditionally homoscedastic. The intercept $a(h,k|Y_t)$ is just to balance the changes in the tails, since all predictive densities have unit mass.

In Figure 1 we plot the FELD for a bivariate VAR(1) with autoregressive parameter

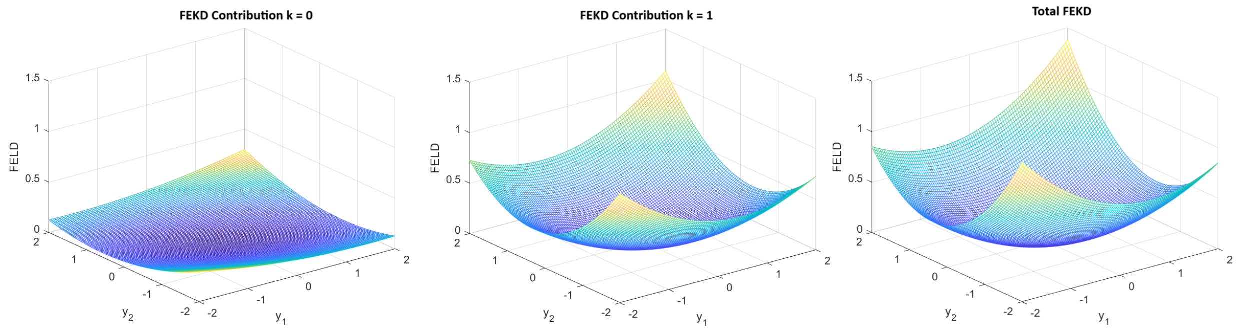


Figure 1: A visual representation of the FELD for the VAR(1) model. Since it is a quadratic functional of argument y, it provides a decomposition of the larger, right most paraboloid (the Total FELD) into two smaller paraboloids (the contribution of $k = 0$ and $k = 1$).

$\Phi = \begin{bmatrix} 0.5 & 0.1 \\ 0.2 & 0.6 \end{bmatrix}$ and variance covariance matrix $\Sigma = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}$. The total FELD depicted in the right most graph of Figure 1 takes a parabolic form as shown in Proposition 3. It can be decomposed into two smaller paraboloids which capture the risk associated with updating between $t+1$ and $t+2$ ($k = 1$ in the middle graph) and the risk associated with updating between t and $t+1$ ($k = 0$ in the left graph). The total risk (i.e. the left hand side the FEKD) and its corresponding decomposition can depend on selected values of y.

We can also compare the FEKD to the FEVD for the Gaussian VAR(1). For the FEVD:

$$\Sigma_h - \Sigma = \sum_{k=0}^{h-2} \left[\Sigma_{h-k} - \Sigma_{h-k-1}.\right]. \tag{4.2}$$

it is written on the prediction variances of the process, whereas the FEKD is written on their inverses. This difference is the analogue of the two equivalent filters, introduced by Kalman for the linear state space model. The standard covariance filter is based on the direct updating of Σ_h, whereas the information form of the filter coincides with the direct updating of the inverse Σ_h^{-1} (called information).

4.1.2 Markov Chains

Suppose Y_t is a stationary Markov chain with n possible states. Let $X_t = (X_{1,t}, ..., X_{n,t})'$, with $X_{i,t} = 1$, if Y_t is in state i, and zero otherwise, for $i = 1, ..., n$. The knowledge of Y_t is equivalent to the knowledge of the vector of binary indicators X_t, which can take on the values: $(1, 0, ...0)'$, $(0, 1, 0, ..., 0), ..., (0, ..., 0, 1)'$. The h-step transition probabilities are the elements of the matrix P^h, defined as $p_{ij}^{(h)} = P(Y_{t+h} = i|Y_t = j)$ for all $t \geq 0$.

Proposition 2. *For a Markov chain with n states and transitiona matrix P, the generic term on the right hand side of the FEKD is:*

$$\gamma(h, k|I_{t+k}) = \left[\widetilde{\log}(P^{h-k})_y P^{k+1} - \widetilde{\log}(P^{h-k-1})_y P^k\right] X_t,$$

where $\widetilde{\log}(P^h)$ is a matrix whose elements are the logged elements of P^h and A_y denotes the y-th row of matrix A.

Proof: See Appendix A.2.2.

4.1.3 The Cauchy AR(1)

Let us consider the stationary univariate process (Y_t) defined by:

$$Y_t = \varphi Y_{t-1} + \sigma\varepsilon_t, \quad |\varphi| < 1,$$

where (ε_t) is a Cauchy distributed strong white noise. Its conditional distribution of Y_{t+h} given Y_t is also Cauchy, with drift $\varphi^h Y_t$ and scale $\sigma\frac{1-|\varphi|^h}{1-|\varphi|}$. Hence, we deduce that:

$$\frac{f(y, h-k|I_{t+k})}{f(y, h-k-1|I_{t+k+1})} = \frac{1-|\varphi|^{h-k-1}}{1-|\varphi|^{h-k}} \frac{1 + \left[(y - \varphi^{h-k-1}Y_{t+k+1})/\left(\sigma\frac{1-|\varphi|^{h-k-1}}{1-|\varphi|}\right)\right]^2}{1 + \left[(y - \varphi^{h-k}Y_{t+k})/\left(\sigma\frac{1-|\varphi|^{h-k}}{1-|\varphi|}\right)\right]^2}.$$

When this expression is logged and expectations are taken, it is integrable with respect to the Cauchy distribution. Hence, the FEKD exists in this context. However, the conditional variance does not exist, and therefore the FEVD does not either.

4.2 Examples of FELD

The FELD has a simple form and is easily applied to the dynamic affine model, which is called the Compound Autoregressive (CaR) model in discrete time [see Duffie et al. (2003), Darolles et al. (2006)]. We review these models briefly and demonstrate how it can be applied to the VAR(1) model.

4.2.1 FELD for Dynamic Affine Models

The process is assumed Markov of order 1 with a conditional log-Laplace transform which is affine in the conditioning value[5] Y_t. Thus, the Laplace transform at horizon 1 can be written as:

$$\Psi(u, 1|I_t) = \exp\left\{-a(u)'Y_t + c(u) - c[a(u)]\right\}, \tag{4.3}$$

where $c(u) = \log \mathbb{E}[\exp(-u'Y_t)]$ is the unconditional log-Laplace transform of Y_t and the function $a(\cdot)$ captures all nonlinear serial dependence features. The affine property remains satisfied at any forecast horizon. More precisely, we have:

$$\Psi(u, h|Y_t) = \exp\left\{-a^{\circ h}(u)'Y_t + c(u) - c\left[a^{\circ h}(u)\right]\right\}, \tag{4.4}$$

where $a^{\circ h}(\cdot)$ is function $a(\cdot)$ compounded h times with itself. Then the FELD is given by the following proposition [see Appendix A.3.1 for proof]:

[5] The extension to a Markov process of order p is straightforward.

Proposition 3. *For dynamic affine models, the FELD takes the form:*

$$
\begin{aligned}
&\left\{u'\left[\frac{da'}{du}(0)\right]^h - a^{\circ h}(u)'\right\} Y_t - u'\frac{dc}{du}(0) + u'\left[\frac{da'}{du}(0)\right]^h\left[\frac{dc}{du}(0)\right] + c(u) - c\left[a^{\circ h}(u)\right] \\
&= \sum_{k=0}^{h-1}\left\{a^{\circ(h-k-1)}(u)'\left[\frac{da'}{du}(0)\right]^{k+1} - a^{\circ(h-k)}(u)'\left[\frac{da'}{du}(0)\right]^{k}\right\} Y_t \\
&\quad + \left(a^{\circ(h-k)}(u)' - a^{\circ(h-k-1)}(u)'\right)\left[\frac{dc}{du}(0)\right] \\
&\quad + \left(a^{\circ(h-k)}(u)'\left[\frac{da'}{du}(0)\right]^{k} - a^{\circ(h-k-1)}(u)'\left[\frac{da'}{du}(0)\right]^{k+1}\right)\left[\frac{dc}{du}(0)\right] \\
&\quad + c\left[a^{\circ(h-k-1)}(u)\right] - c\left[a^{\circ(h-k)}(u)\right], \ \forall u.
\end{aligned}
$$

Proof: See Appendix A.3.1.

Therefore, we get a decomposition of the type:

$$
\alpha(h,u)'Y_t + \beta(h,u) = \sum_{k=0}^{h-1}\left[\alpha(h,k,u)'Y_t + \beta(h,k,u)\right], \tag{4.5}
$$

or equivalently decompositions of functions $\alpha(h,u)$ and $\beta(h,u)$ into $\sum_{k=0}^{h-1}\alpha(h,k,u)$ and $\sum_{k=0}^{h-1}\beta(h,k,u)$, respectively. These decompositions depend on both the term h and the argument u. The decomposition of function $\alpha(h,u)$ is especially appealing due to its interpretation in terms of nonlinear Impulse Response Functions (IRF). Indeed, let us consider a given shock of magnitude δ on the level Y_t. Note that in our nonlinear framework, the magnitude of the shock δ is constrained by the domain of Y_t. It can be multivariate if Y_t is multivariate, constrained to be either 0, or -1 (resp. 0, or 1) if Y_t is binary with value 1 (resp. value 0), and so on. We do not discuss the sources of the shock and if they are identifiable or controllable. The effect on the predictive distribution at horizon h is $\alpha(h,u)'\delta$. Compared to the standard linear approach of the IRF, we see that the IRF depends on the argument u. In other words, this measure changes with the preference (risk aversion) of the analyst.

4.2.2 Strong Linear VAR(1) Model

Consider the strong VAR(1) model given by: $Y_t = \Phi Y_{t-1} + \varepsilon_t$, where ε_t is an iid strong white noise. Then we have [see Appendix A.3.2]:

Corollary 1. *For a strong linear VAR(1) model, the FELD is:*

$$\mathbb{E}\left\{\log\left[\frac{\mathbb{E}\left[\exp\left(-u'Y_{t+h}\right)|I_t\right]}{\exp(-u'Y_{t+h})}\right]\middle| I_t\right\} = \sum_{k=0}^{h-1} b\left[\left(\Phi'\right)^k u\right]. \tag{4.6}$$

Due to the linear dynamic, the FELD does not depend on the value Y_t of the conditioning variable. However, there are still effects on the decomposition of the cross-sectional heterogeneity, that is the non-Gaussian distribution of the errors (ε_t). If the error is Gaussian $\varepsilon_t \sim IIN(0, \Sigma)$, we get $b(u) = \frac{u'\Sigma u}{2}$. Then, the right hand side of the decomposition becomes:

$$\sum_{k=0}^{h-1} b\left[\left(\Phi'\right)^k u\right] = \frac{1}{2}\sum_{k=0}^{h-1}\left(u'\Phi^k\Sigma\left(\Phi'\right)^k u\right) = \frac{1}{2}u'\sum_{k=0}^{h-1}\Phi^k\Sigma\left(\Phi'\right)^k u.$$

From equation (2.4), we know that the FEVD for the Gaussian VAR(1) is $\sum_{k=0}^{h-1}\Phi^k\Sigma\left(\Phi'\right)^k$. Hence, the FELD is equivalent to the FEVD on a portfolio $u'Y_{t+h}$, where the weights on each asset are assigned based on their risk aversion level, for the special case of the VAR(1) with Gaussian errors.

5 Application to Cyberrisk

The prevalance of the internet in our daily lives means that individuals are now able to communicate, transfer and store large amounts of information with just a mobile device. This has significantly improved the way in which businesses operate on a daily basis and has transformed the structure of our modern economy. For example, hospitals have adopted digital patient records which can be accessed by any institution in their network, and alleviated the need to maintain or deliver physical patient files. However, this offers an opportunity for bad faith actors to intercept or steal information, leading to potentially disastrous outcomes for the victims involved. As such, there is demand for businesses and government agencies to model and quantify the risk of cyber attacks in order to insure against these prospects. In this section, we demonstrate how the FELD and the FEKD can be used in the context of a multivariate Negative Binomial Autoregressive framework to study the term structure of risks on cyber attack frequency counts.

5.1 The Challenges and the Available Data

Data on cyber attacks are difficult to obtain. Firstly, there is no consensus on what a cyber attack is. Indeed, there are many definitions, and in a sense it is an umbrella term which includes a variety of illicit behaviours or acts to obtain digital information. Secondly, the collection of data on cyber attacks is rather limited. Although the internet is available almost everywhere today, its accessibility was much less so even just 20 years ago. Hence, tracking cyber attacks has eased only in recent years. Moreover, recording reliable and confirmed cyber attacks is a challenging task, since firms or organizations that are subject to these attacks have an incentive to hide their occurrences. Nonetheless, recent research has appealed to the Privacy Rights Clearinghouse (PRC) dataset, which includes information on publicly reported data breaches across the United States between 2005 and 2022[6]. The PRC was founded in 1992 by the University of San Diego School of Law. The data are gathered from different sources including the Attorney General offices, government agencies, nonprofit websites and media. The reports do not follow a consistent procedure, which may lead to a lack of accuracy and representativeness of the data. Nevertheless, this

[6]The data that support the findings of this study are available here: https://privacyrights.org/data-breaches. Other cyber databases are Advisen and SAS Oprisk [see Eling et al. (2023) for a comparison]

database is usually employed to analyze cyberrisk [Eling and Loperfido (2017), Eling and Jung (2018), Barati and Yankson (2022), Lu et al. (2024)].

We model cyber breach[7] frequency counts and adopt a partial sample of the one used in Lu et al. (2024). In particular, we focus on two types of breaches defined by the PRC: [1] HACK - Hacked by an outside party or infected by malware. [2] INSD - Breach due to an insider, such as an employee, contractor, or customer. The plots of these two time series and their histograms are shown in Figure 2 below.

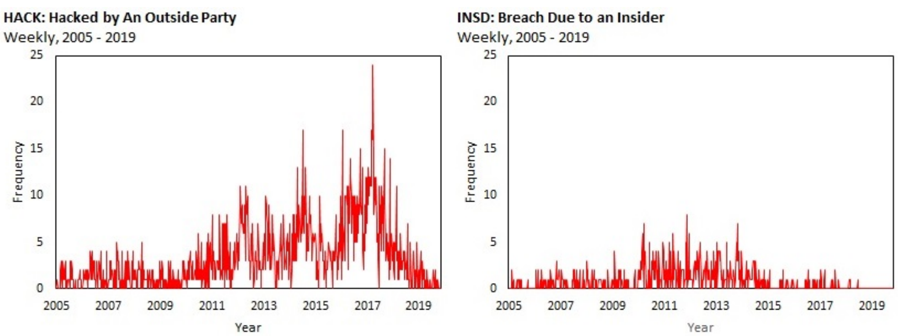


Figure 2: *Time series counts and histograms of HACK and INSD.*

Process	Mean	Variance	Excess Skewness	Excess Kurtosis
HACK	3.28	11.13	1.60	3.31
INSD	0.78	1.46	2.10	5.42

Table 1: Summary statistics for HACK and INSD.

Both series feature a large number of zero counts, with 138 and 440 instances for HACK and INSD, respectively. While both types can be considered breaches with malicious intent, the insider breaches occur much less on average. Indeed, it is much easier for an outsider to gain access without getting caught (such as using an IP spoofer), than an insider to attempt a breach. In each series, we see that the variance is much larger than the mean, which suggests that the count data exhibit overdispersion. There is also positive excess skewness due to the non-negativity of frequency counts. Furthermore, all the series exhibit

[7] A breach is a successful cyber attack where information or data is accessed or stolen by an unauthorized party.

excess kurtosis, so they have relatively fatter tails compared to the normal distribution.

5.2 Bivariate Negative Binomial Autoregressive Model

The INAR model [Al-Osh and Alzaid (1987)] is the basic dynamic model for a series of count data. However, it implies a marginal Poisson distribution, for which the mean is equal to the variance. Therefore, it is not compatible with data on cyber attacks (see Table 1). The INAR model can be extended for more flexibility by introducing stochastic intensity. This leads to the Negative Binomial Autoregressive Process (NBAR) [see Gourieroux and Lu (2019) for a discussion].

We will focus on the joint analysis of the two risks, which has to account for both cross-sectional and serial dependencies between the series. A first insight on the cross-sectional dependence is through the joint stationary distributions. This is illustrated in Figure 3, where on the left hand side we provide a plot of values $(Y_{1,t}, Y_{2,t})$ and on the right hand side a plot of the associated Gaussian ranks ($\pi(Y_{1,t})$,$\pi(Y_{2,t})$), i.e. the estimated copula after the Gaussian transform[8][9]. Such plots are usually given for continuous variables or for count variables taking a large number of values. In our framework, the discrete nature of the variable has to be taken into account in the interpretation. For instance, we observe an overweighting of zero counts for the HACK variable, which can be seen in the left plot of Figure 3, where the joint density rests “against” its axis. The right panel in Figure 3 shows that the two series features some right tail independence.

[8]The raw data are first ranked by increasing order. Rank$(Y_{j,t})$, the rank of $Y_{j,t}$, is valued in $[1, ..., T]$. Then, $\pi(Y_{j,t}) = \Phi^{-1}\,[\text{Rank}(Y_{j,t})/T]$, where Φ is the cumulative distribution function of the standard normal distribution.

[9]This Gaussian transformed unconditional copula is completed on the count of occurrences. It differs from a copula completed from losses that account also for severities [see Eling and Jung (2018)].

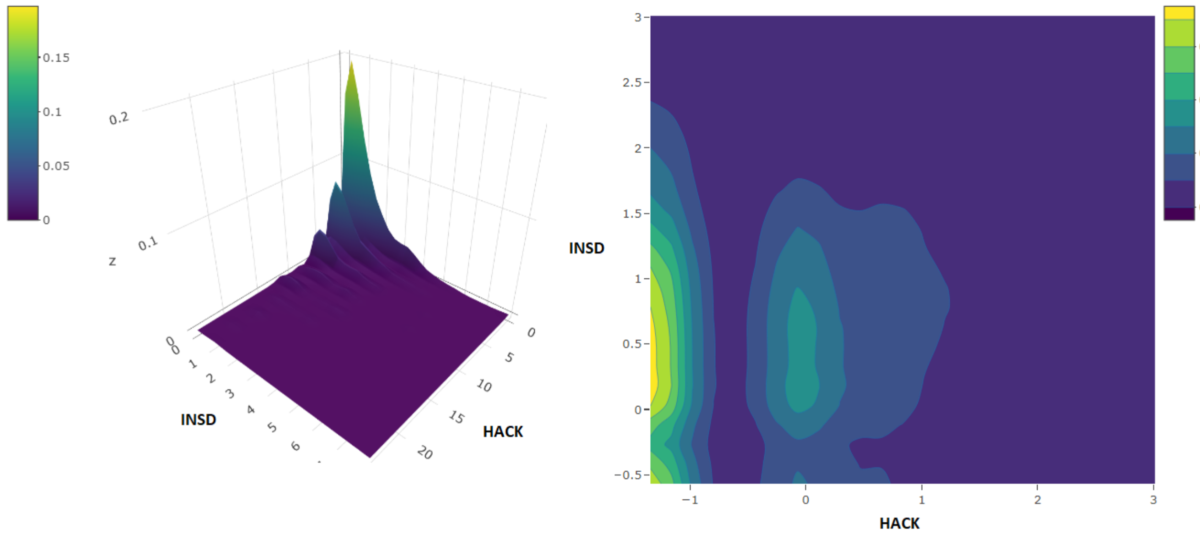


Figure 3: *The joint density plot of HACK and INSD (left) and the contour density plot of their associated Gaussian ranks (right).*

5.2.1 Model

Let us denote Y_{1t}, Y_{2t} the two count series and introduce three state variables interpretable as specific and common stochastic intensities, denoted $X_{1,t}, X_{2,t}$ and Z_t, respectively. The nonlinear state space representation becomes:

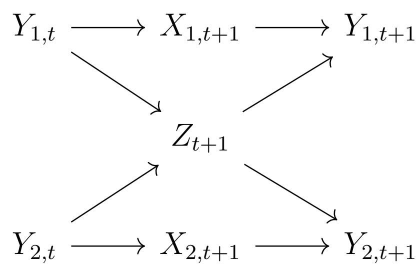
1. **Measurement Equation -** Conditional on $\underline{Y}_t$, $\underline{X}_{t+1}$, $\underline{Z}_{t+1}$, the variables $Y_{1,t+1}$, $Y_{2,t+1}$ are independent $Y_{j,t+1} \sim \mathcal{P}(\alpha_j Z_{t+1} + \beta_j X_{j,t+1})$ for $j = 1, 2$.

2. **Transition Equations -** Conditional on $\underline{Y}_t$, $\underline{X}_t$, $\underline{Z}_t$, the variables $X_{1,t+1}$, $X_{2,t+1}$, Z_{t+1} are independent such that $X_{j,t+1} \sim \gamma(\delta_j + Y_{j,t})$, for $j = 1, 2$, and $Z_{t+1} \sim \gamma(\delta + \sigma_1 Y_{1,t} + \sigma_2 Y_{2,t}, 0, c)$[10].

The dimensionality and the presence of stochastic intensity factors lead to 9 parameters to be estimated: α_1, α_2, β_1, β_2, δ_1, δ_2, σ_1, σ_2, and δ. The process described above corresponds

[10] There is no scale coefficient introduced in the conditional distribution of the idiosyncratic intensities for identification reasons.

to a more complicated causal scheme with one hidden layer and three hidden neurons:



It is easily checked that the bivariate NBAR model is an affine model with a conditional Laplace transform at horizon 1 given by [see Appendix A.3.3]:

$$\Psi(u,1|Y_t) = \mathbb{E}\left[\exp(-u'Y_{t+1})|Y_t\right] = \exp\left[-a_1(u_1,u_2)Y_{1,t} - a_2(u_1,u_2)Y_{1,t} - b(u_1,u_2)\right], \tag{5.1}$$

where:

$$\begin{aligned} a_1(u_1,u_2) =& \log\left[1+\beta_1(1-\exp(-u_1))\right] + \sigma_1 \log[1+\alpha_1(1-\exp(-u_1)) + \alpha_2(1-\exp(-u_2))], \\ a_2(u_1,u_2) =& \log\left[1+\beta_2(1-\exp(-u_2))\right] + \sigma_2 \log[1+\alpha_1(1-\exp(-u_1)) + \alpha_2(1-\exp(-u_2))], \\ b(u_1,u_2) =& \delta_1 \log[1+\beta_1(1-\exp(-u_1))] + \delta_2 \log[1+\beta_2(1-\exp(-u_2))] \\ &+ \delta \log[1+\alpha_1(1-\exp(-u_1)) + \alpha_2(1-\exp(-u_2))]. \end{aligned}$$

We see from the closed form expression of the conditional Laplace transform that all parameters are identifiable.

5.2.2 Estimation

The increase of the parameter dimension and the introduction of a common stochastic intensity lead to a larger number of identifiable parameters to be estimated, equal to 9. A first estimation to consider is by applying a Vector Autoregressive (VAR) representation based on the linear prediction formula [see Appendix A.3.4]:

$$\mathbb{E}\left[Y_t|Y_{t-1}\right] = \begin{bmatrix} \alpha_1\delta + \beta_1\delta \\ \alpha_2\delta + \beta_2\delta \end{bmatrix} + \begin{bmatrix} \alpha_1\sigma_1 + \beta_1 & \alpha_1\sigma_2 \\ \alpha_2\sigma_1 & \alpha_2\sigma_2 + \beta_2 \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix}. \tag{5.2}$$

We can apply OLS to estimate the model $Y_t = C + AY_{t-1} + \varepsilon_t$ (see Table 2 below). The corresponding eigenvalues of $\hat{A} = \begin{bmatrix} 0.638 & 0.053 \\ 0.005 & 0.361 \end{bmatrix}$ are $\lambda_1 = 0.639$ and 0.360, which suggests that the bivariate process is stationary in the conditional mean.

	$Y_{1,t}$	$Y_{2,t}$
$Y_{1,t-1}$	0.638	0.005
	(0.028)	(0.012)
$Y_{2,t-1}$	0.053	0.361
	(0.076)	(0.034)
Constant	1.145	0.486
	(0.140)	(0.062)

Table 2: OLS estimates of the bivariate VAR(1).

More generally, we can apply a Generalized Method of Moments (GMM) procedure based on the unconditional pairwise moment restrictions of the form:

$$\mathbb{E}\left\{\left[\exp(-u'Y_t) - \Psi(u, 1|Y_t)\right]\exp(-v'Y_{t-1})\right\} = 0, \tag{5.3}$$

valued for different u and v, by considering the orthogonality between prediction errors and past values. A global estimation of the 9 parameters can be based on the unconditional pairwise moments given in (5.3) after selecting at least 9 quadruples (u_1, u_2, v_1, v_2) of linearly independent moment restrictions. By fixing u, we select conditional moment restrictions and, by fixing v, we find some instruments to transform them into marginal moment restrictions. The selected values must impose informative restrictions on the parameters that characterize the joint dynamics of tails of the count variables. The information on the tails of a distribution is hidden in the behaviour of the Laplace transform in the neighbourhood of $u = 0$ (for distributions with thin as well as fat tails). This explains the choice of the selected risk aversion values for GMM[11]. We prefer such an approach to the introduction of a huge number of moment restrictions (i.e. of u and v), with the danger of the weak instrument problem and numerical instability when inverting the weighting matrix of the GMM approach. We also include 6 additional moment conditions implied by the OLS first-order conditions. Applying the GMM procedure for the 15 moment conditions, we obtain the following estimates displayed in Table 3.

α_1	α_2	β_1	β_2	δ_1	δ_2	σ_1	σ_2	δ
0.118	-0.067	0.647	0.391	1.20	1.27	-0.075	0.453	1.492
(0.277)	(0.367)	(0.078)	(0.226)	(0.626)	(0.998)	(0.374)	(1.155)	(0.671)

Table 3: Estimated values of the 9 coefficients in the bivariate model of HACK and INSD.

These estimates imply $\hat{\hat{C}} = \begin{bmatrix} 1.143 \\ 0.485 \end{bmatrix}$ $\hat{\hat{A}} = \begin{bmatrix} 0.701 & 0.053 \\ 0.005 & 0.361 \end{bmatrix}$. Hence, our choice of quadruplets not only captures different risk aversion scenarios, but also produce estimates that are compatible with the OLS estimations of A and C. The estimated values suggest that the process is stationary[12].

5.3 FELD Analysis

5.3.1 FELD for Bivariate NBAR

An expression for the conditional Laplace transform at horizon h is difficult to obtain. However, the dynamic affine property of the bivariate NBAR means that it can be derived numerically by means of recursion. In particular, we have:

$$\Psi(u, h|Y_t) = \exp\left[-a_1^{(h)}(u_1, u_2)Y_{1,t} - a_2^{(h)}(u_1, u_2)Y_{1,t} - b^{(h)}(u_1, u_2)\right], \tag{5.4}$$

[11]We choose quadruplets which correspond to a range of different risk aversion scenarios. For instance, the quadruplet (0.41,0.01,0.41,0.01) corresponds to the scenario where there is high risk aversion on only on the series $Y_{1,t}$. Likewise, the quadruplet (0.41,0.41,0.01,0.01) reflects the case where there is high risk aversion for both series, but only at time t.

[12]The stationary conditions are given by: $1 - \alpha_1 - \sigma_1\beta_1 > 0$, $1 - \alpha_2 - \sigma_2\beta_2 > 0$ and $(1 - \alpha_1 - \sigma_1\beta_1)(1 - \alpha_1 - \sigma_1\beta_1) > \sigma_1\sigma_2\beta_1\beta_2$ [Gourieroux and Lu (2019), Proposition 3].

where:

$$
\begin{aligned}
a_1^{(h)}(u_1,u_2) =& a_1(a_1^{(h-1)}(u_1,u_2), a_2^{(h-1)}(u_1,u_2)),\\
a_2^{(h)}(u_1,u_2) =& a_2(a_1^{(h-1)}(u_1,u_2), a_2^{(h-1)}(u_1,u_2)),\\
b^{(h)}(u_1,u_2) =& b(a_1^{(h-1)}(u_1,u_2), a_2^{(h-1)}(u_1,u_2)),\ \forall\ h \geq 2.
\end{aligned}
$$

The estimated FELD for the model can still be obtained by applying the recursion in (5.4) to generate the terms $\Psi(u, h-k|I_{t+k})$ in (3.5) and plugging in the estimated values for the 9 parameters from Table 5.

In this framework, the parameters (u_1, u_2) can be interpreted either as individual risk aversion parameters for HACK and INSD, respectively, with no co-risk aversion between these two types of cyber breaches or weights applied to the two cyber risks with a single common risk aversion parameter. The FELD also depends on the conditioning values $(Y_{1,t}, Y_{2,t})$, which are the observed values of the counts in period t.

5.3.2 Empirical Results

Let us now consider two examples of decomposition analysis using the FELD in this bivariate framework.

Example 1: Risk Aversion and Conditioning Counts are the SAME for both HACK and INSD		
■ Low Risk Aversions, Low Historical Count	$(u_1, u_2) = (0.5, 0.5)$	$(Y_{1,t}, Y_{2,t}) = (0, 0)$
■ High Risk Aversions, Low Historical Count	$(u_1, u_2) = (2, 2)$	$(Y_{1,t}, Y_{2,t}) = (0, 0)$
■ Low Risk Aversions, High Historical Count	$(u_1, u_2) = (0.5, 0.5)$	$(Y_{1,t}, Y_{2,t}) = (5, 5)$
■ High Risk Aversions, High Historical Count	$(u_1, u_2) = (2, 2)$	$(Y_{1,t}, Y_{2,t}) = (5, 5)$

The objective of this first example is to see how the risk aversions and the conditioning counts of the two cyber breaches will influence the total FELD and its decomposition. For now, suppose we are equally risk averse on both cyber risks (i.e. $u_1 = u_2$) and condition on the same values observed at time t (i.e. $Y_{1,t} = Y_{2,t}$). We then compare the four cases: [1] Low Risk Aversions and Low Conditioning Counts. [2] High Risk Aversions and Low

Conditioning Count. [3] Low Risk Aversions and High Conditioning Count. [4] High Risk Aversions and High Conditioning Counts. The results of the FELD are presented in Figure 4 below.

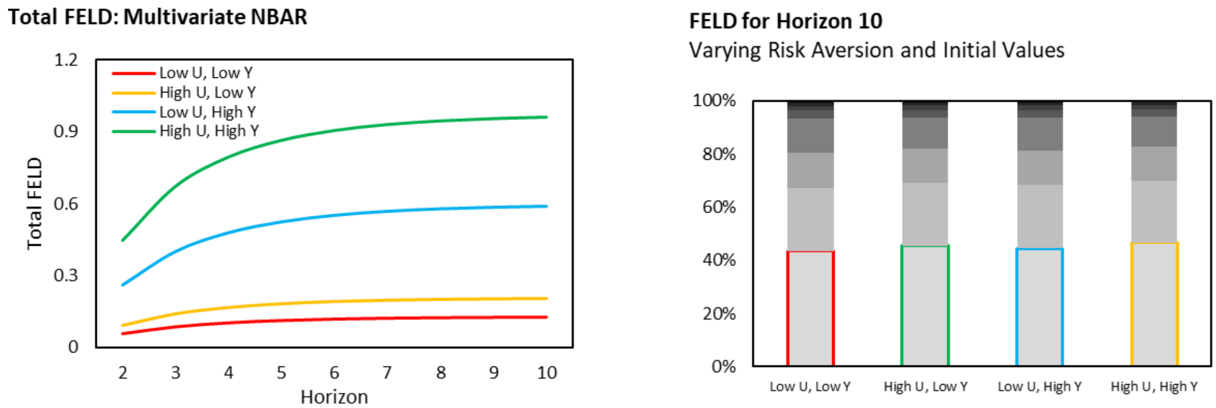


Figure 4: *A comparison of the four scenarios where risk aversion and conditioning counts are the same for both HACK and INSD.*

On the left hand side graph, we plot the total FELD (i.e. the LHS of (3.5)) for the four cases, up to horizon 10. The red and blue lines (corresponding to a risk aversion of 0.5) are lower than the yellow and green lines (corresponding to a risk aversion of 2), respectively. This means that total risk is increasing with the value of the risk aversion parameters (u_1, u_2). Also, the total FELD increases with the conditioning counts $(Y_{1,t}, Y_{2,t})$, since the green and blue lines are higher than the red and yellow ones, respectively. Indeed, if a firm were to observe a higher count of cyber breaches at week t, then the overall risk at future horizons should be increasing.

On the right hand side graph, we show the decomposition of the total FELD at horizon 10. To ease the comparison between the four cases, we express the FELD as a percentage of the total. For instance, the total value of the FELD for "High Risk Aversion, High Historical Count" (the green line) is approximately 0.9, so the value of the terms in the decomposition are taken as a fraction of this total. We have highlighted in red the contribution of risk in updating between horizons 1 and 2 on the FELD at horizon 10. Although each case depends on different risk aversion parameters and conditioning counts, the percentage contribution is roughly 40% in all the cases. Thus, if we are equally risk averse on the two types of cyber risks, and observe the same number of counts at week t, then our decomposition remains the same in proportion.

Exercise 2: Different Values of Risk Aversion and Historical Counts for HACK and INSD		
■ Low $Y_{1,t}$, High u_1, High $Y_{2,t}$, Low u_2	$(u_1, u_2) = (2, 0.5)$	$(Y_{1,t}, Y_{2,t}) = (0, 5)$
■ High $Y_{1,t}$, High u_1, Low $Y_{2,t}$, Low u_2	$(u_1, u_2) = (2, 0.5)$	$(Y_{1,t}, Y_{2,t}) = (5, 0)$
■ Low $Y_{1,t}$, Low u_1, High $Y_{2,t}$, High u_2	$(u_1, u_2) = (0.5, 2)$	$(Y_{1,t}, Y_{2,t}) = (0, 5)$
■ High $Y_{1,t}$, Low u_1, Low $Y_{2,t}$, High u_2	$(u_1, u_2) = (0.5, 2)$	$(Y_{1,t}, Y_{2,t}) = (5, 0)$

Of course, the assumption that the two cyber breaches have the same risk aversions or the same observed conditioning values at time t is an unrealistic one. We consider a second example, with different values for HACK and INSD to see how the FELD can change. We consider four new cases: [1] Low/High Conditioning Count for HACK/INSD, High/Low Risk Aversions for HACK/INSD. [2] High/Low Conditioning Count for HACK/INSD, High/Low Risk Aversions for HACK/INSD. [3] Low/High Conditioning Count for HACK/INSD, Low/High Risk Aversions for HACK/INSD. [4] High/Low Conditioning Count for HACK/INSD, Low/High Risk Aversions for HACK/INSD. The results are presented in Figure 5.

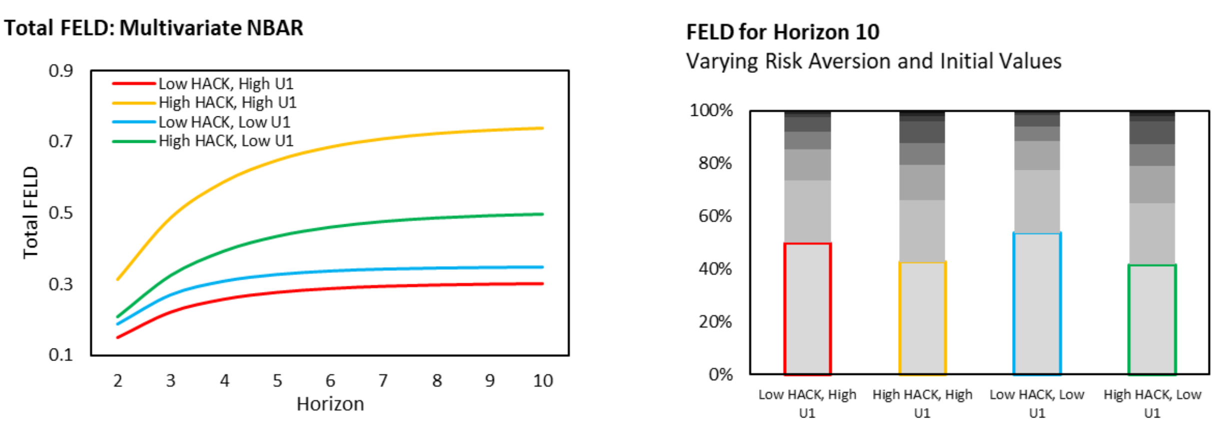


Figure 5: *A comparison of four scenarios where risk aversion and historical counts are different for both HACK and INSD.*

On the left hand side graph of Figure 5, the yellow and green lines correspond to a high observed conditioning value for HACK, but a low observed conditioning value observed for INSD. These lines are much higher than the blue and red lines, which correspond to the opposite case. Intuitively, there is more overall risk associated with observing a high number of outsider breaches at time t compared to insider ones. This result is consistent with our understanding of the risks associated with these cyber breaches. Indeed, when an insider breach is discovered, a company can conduct an exhaustive investigation and likely find most if not all the vulnerabilities that lead to the breach. Then, internal policies can be enacted to prevent such breaches from occurring again in the future. Furthermore, the transmission of insider breaches from one company to another is rather limited. That is, if company A suffers an internal breach, there is no compelling reason to think that company B will also experience something similar. On the other hand, outsider breaches are much harder to insulate against, and sometimes a hack done by an external party is never fully diagnosed. Also, software often undergo regular updates, which introduces more opportunities for bugs that can be exploited by unauthorized parties. Moreover, outsider breaches can be transmitted easily between companies. For instance, a virus can be easily spread between computer systems, and may affect an entire network of companies.

On the right hand side graph of Figure 5, we show the decompositions for horizon 10 in each case. A notable difference is that now, the decompositions are quite different. In particular, when updating from horizon 1 to 2 (outlined in red), each case has a different contribution to the FELD at horizon 10. When the observed conditioning count for HACK is low at time t and INSD is high (that is, the first and third bars in the graph), the risk is front loaded since the risk of updating between horizon 1 and 2 accounts for over 40% of the FELD in these cases. On the other hand, when a high historical count of HACK and a low count of INSD is observed (that is, the second and fourth bars in the graph), the risk has more spread across future horizons. This means that a high count of observed insider breaches implies a more front loaded risk. Again, this is a reasonable conclusion, since outsider breaches should pose more of a long run risk than insider ones for the reasons stated above.

The examples above demonstrate two important factors. Firstly, by considering two series together in a bivariate model, we are able to take advantage of the cross-sectional dependencies between the two types of cyber attacks, even if our decomposition measure is a dynamic separation of effects. Secondly, the FELD proposed in this paper has as many decompositions as values of u and Y_t. In particular, we see that in example 1 and example 2, the decompositions can be very different and are dependent on risk aversions and observed conditioning counts at time t. For an insurance firm or intermediary that specializes in cyber risk, these types of scenario analysis can help price their insurance products, or allocate resources efficiently to hedge against future cyber attacks for a range of customers or insurees with different risk aversions.

5.4 FEKD Analysis

The transition density at horizon h for a NBAR process does not have a simple expression. However, since the NBAR process is also a Markov process, we may represent its transition through an approximation with a Markov chain. Suppose $(Y_t) = (Y_{1,t}, Y_{2,t}) = (i, j)$, where $i = 0, ..., 8$ and $j = 0, ..., 3$, that is (Y_t) is a Markov chain with $9 \times 4 = 36$ states. Values in i represent the counts for HACK and values in j represent the counts for INSD. The final state for each cyber attack represents an overflow bin, that is, 8 denotes "8 or more" and 3 denotes "3 or more" for HACK and INSD respectively. This is a generative model where, for the estimated parameter set in Section 7.2.2, we can compute the transition matrix by simulation. Then, the result of Proposition 5 can be utilized directly.

For exposition, we perform the FEKD conditional on two historical counts: $Y_t = (0, 0)$, corresponding to no observed historical counts of either cyber risk, and $Y_t = (4, 1)$, corresponding to 4 observed counts of HACK and 1 observed count of INSD. The FEKD is then evaluated at different points of the forecast density, namely $y = (0,0), (8,0), (0,3)$ and $(8,3)$. These reflect cases of 0 future counts, extremes for HACK, extremes for INSD, and extremes for both cyber attacks. In Figure 6, we provide the plots of the total FEKD for the first 9 horizons in each scenario. A first insight is to is to compare lines with the same colour. Recall from the definition in (3.2) that the total FEKD is intuitively an expected Kullback divergence (conditional on time t) between the densities $f(y, h|I_t)$ and $f(y, 1|I_{t+h-1})$. A larger value implies a larger difference between these two transitions. The green lines in both graphs represent the FEKD evaluated at the value $y = (8,3)$. This represents an extreme point in the joint predictive distribution, since $y = (8,3)$ is the state where both cyber risks are in their overflow bins. The green line in the left graph is slightly higher (a larger difference) than the one on the right. Indeed, it is more plausible to observe $(8,3)$ given a history of $(4,1)$ than it is given a history of $(0,0)$. Likewise, we see the opposite for the red lines, where the total FEKD for the right graph is higher than the left one. A second insight from these graphs is to compare the total FEKD from the different profiles of y within the same graph. Clearly, observing the extreme state $(8,3)$ is an event of rather low probability, which is reflected by the green line being the highest of the four profiles.

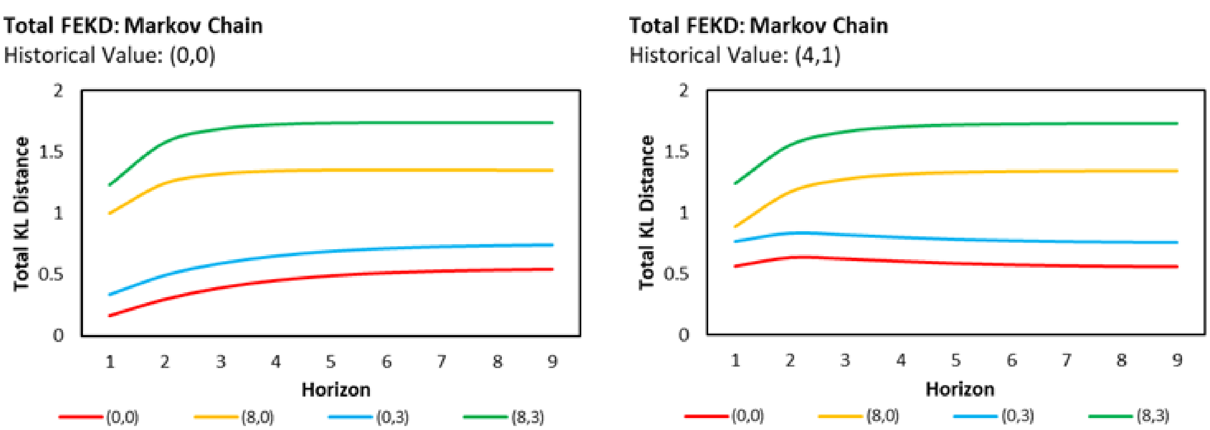


Figure 6: *A comparison of the risks at different values of the predictive distribution.*

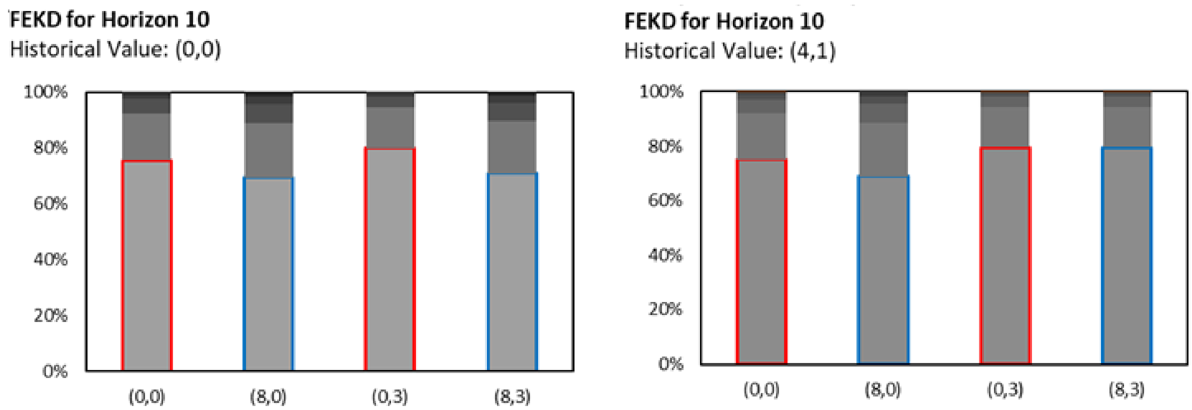


Figure 7: *A comparison of the decomposition at horizon 10 for different values of the predictive distribution.*

In Figure 7, we study the decomposition of risk at horizon 10. Each bar represents the fraction of the total FEKD accounted by the risk at each horizon, with the highlighted bars representing the initial horizon. Red bars correspond to values of y where HACK is low, and blue bars represent values of y where HACK is high. When conditioning on the initial state $Y_t = (0, 0)$, profiles with a high HACK value have less front-loaded risk. This observation is consistent with the discussion of this cyber risk in the application with the FELD. However, when conditioning on the historical value $Y_t = (4, 1)$, the proportion of risk seems to differ for the case $y = (8, 3)$. This reflects the importance of the conditioning value and considering the decomposition at different levels of y.

6 Concluding Remarks

In this paper, we have introduced decomposition formulas for the analysis of global forecast errors in nonlinear dynamic models. These formulas are based on functional measures of nonlinear forecast with respect to either the transition (predictive) densities in the Forecast Error Kullback Decomposition, or the conditional log-Laplace transform in the Forecast Error Laplace Decomposition. Such decompositions could be extended to functional measures of economic or financial interest such as conditional Lorenz curves used in inequality analysis or conditional quantiles (VaR). This is left for future research.

7 References

AL-OSH, M., & ALZAID, A. (1987). *First-order Integer-Valued Autoregressive (INAR (1)) Process.* Journal of Time Series Analysis, 8(3):261-275.

ARROW, K. J. (1965). *Aspects of the theory of risk-bearing.* Yrjo Jahnssonin Saatio.

BARATI, M., & YANKSON, B. (2022). *Predicting the Occurrence of a Data Breach.* International Journal of Information Management Data Insights, 2(2), 100-128.

COX, J., INGERSOLL, J., & ROSS, S. (1985). *A theory of the Term Structure of Interest Rates.* Econometrica, 53(2), 385-407.

DAROLLES, S., GOURIEROUX, C., & JASIAK, J. (2006). *Structural Laplace Transform and Compound Autoregressive Models.* Journal of Time Series Analysis, 27(4), 477-503.

DOOB, J. (1953). *Stochastic Processes*, Wiley.

DUFFIE, D., FILIPOVIC, D., & SCHACHERMAYER, W. (2003).*Affine Processes and Applications in Finance.* The Annals of Applied Probability, 13(3), 984.

ELING, M., IBRAGIMOV, R., & NING, D. (2023). *Time Dynamics of Cyber Risk.* Available at SSRN.

ELING, M., & JUNG, K. (2018). *Copula Approaches for Modeling Cross-Sectional Dependence of Data Breach Losses.* Insurance: Mathematics and Economics, 82, 167-180.

ELING, M., & LOPERFIDO, N. (2017).*Data Breaches: Goodness of Fit, Pricing, and Risk Measurement.* Insurance: Mathematics and Economics, 75, 126-136.

EMBRECHTS, P.(2000). *Actuarial Versus Financial Pricing of Insurance.* The Journal of Risk Finance, 1(4), 17–26.

FELLER, W.(1991). *An Introduction to Probability Theory and its Applications.* Volume 2. John Wiley & Sons.

FISHBURN, P., & VICKSON, R. (1978). *Theoretical Foundations of Stochastic Dominance.* Stochastic Dominance, 39-113.

GALLANT, R., ROSSI, P., AND G., TAUCHEN. (1993). *Nonlinear Dynamic Structures*, Econometrica, 61, 871-907.

GOURIEROUX, C., & JASIAK, J. (2006). *Autoregressive Gamma Processes.* Journal of Forecasting, 25(2), 129–152.

GOURIEROUX, C., & LU, Y. (2019). *Negative Binomial Autoregressive Process with Stochastic Intensity.* Journal of Time Series Analysis, 40(2), 225-247.

ISAKIN, M., & NGO, P. (2020). *Variance Decomposition Analysis for Nonlinear Economic Models.* Oxford Bulletin of Economics and Statistics, 82(6), 1362-1374.

KARNI, E. (1979). *On Multivariate Risk Aversion.* Econometrica, 1391-1401.

KILIAN, L., & LUTKEPOHL, H. (2017). *Structural Vector Autoregressive Analysis.* Cambridge University Press.

LANNE, M., & NYBERG, H. (2016). *Generalized Forecast Error Variance Decomposition for Linear and Nonlinear Multivariate Models.* Oxford Bulletin of Economics and Statistics, 78(4), 595-603.

LEE, Q. (2025). *Nonlinear Forecast Error Variance Decompositions with Hermite Polynomials.* ArXiv:2503.11416.

LU, Y., ZHANG, J., & ZHU, W. (2024). *Cyber Risk Modeling: A Discrete Multivariate Count Process Approach.* Scandinavian Actuarial Journal, 6, 625-655.

MARKOWITZ, H. (1952). *Portfolio Selection.* The Journal of Finance, 7(1), 77-91.

MARKOWITZ, H., & TODD, G. (2000). *Mean-Variance Analysis in Portfolio Choice and Capital Markets.* John Wiley & Sons.

MCKENZIE, E.(1985). *Some Simple Models for Discrete Variate Time Series.* JAWRA Journal of the American Water Resources Association, 21(4), 645-650.

PESARAN, H., & SHIN, Y. (1998). *Generalized Impulse Response Analysis in Linear Multivariate Models.* Economics Letters, 58(1), 17-29.

ROTHSCHILD, M., & STIGLITZ, J. (1970). *Increasing Risk: I. A Definition.* Journal of Economic Theory, 2(3), 225–243.

WHITTLE, P. (1963). *Prediction and Regulation by Linear Least-Square Methods.* University of Minnesota.