Article

Detecting and Measuring Nonlinearity

Rachidi Kotchoni

EconomiX-CNRS (UMR7235), Bureau G-517, Université Paris Nanterre, 92000 Nanterre, France;
rachidi.kotchoni@parisnanterre.fr; Tel.: +33-(0)1-40-97-59-47

Abstract: This paper proposes an approach to measure the extent of nonlinearity of the exposure of a financial asset to a given risk factor. The proposed measure exploits the decomposition of a conditional expectation into its linear and nonlinear components. We illustrate the method with the measurement of the degree of nonlinearity of a European style option with respect to the underlying asset. Next, we use the method to identify the empirical patterns of the return-risk trade-off on the SP500. The results are strongly supportive of a nonlinear relationship between expected return and expected volatility. The data seem to be driven by two regimes: one regime with a positive return-risk trade-off and one with a negative trade-off.

Keywords: conditional expectation; nonlinearity; orthogonal polynomials; return-risk trade-off

JEL Classification: C10; G10

1. Introduction

Economic theories are often operationalized under linearity assumptions on the relationships between the underlying variables or joint normality assumptions on their distributions. Famous examples include the Capital Asset Pricing Model (CAPM) and Value at Risk models. Often, linearity and normality assumptions are needed in order to obtain analytical formulas and elegant characterizations of the phenomena of interest. However, such assumptions may lead to wrong conclusions when they are not valid. Recognizing these limits and observing the relatively high frequency of extreme economic events (e.g., financial crises and recessions), a body of the empirical literature in finance emphasizes the distributional characteristics of assets returns that do not reflect normality, in particular asymmetry and fat tails.

For example, Harvey and Siddique (1999, 2000) propose a model to estimate the conditional skewness and highlight the importance of taking this into account when analyzing the cross sectional properties of assets prices. Christoffersen et al. (2006) propose a framework to price options in the presence of conditional skewness. Feunou and Tedongap (2012) propose a stochastic volatility model with conditional skewness for assets prices and show that these distributional aspects are very important to explain how investors value options. Gabaix (2009, 2016) argues that stable laws approximate the distribution of many economic and financial variables fairly well while Gabaix (2011) proves that macroeconomic fluctuations can have granular origins. Indeed, Gabaix's work draws attention on a more general issue, namely, the fact that the aggregation of independent phenomena does not always lead to the normal distribution as stipulated by the central limit theorem.

If non-normality is now well entrenched in the mind of academic researchers and financial risk managers, non-linearity has received much less attention in the literature. The joint normality of two random variables X and Y implies that $E(Y|X)$ is a linear function of X and that conversely $E(X|Y)$ is linear in Y. While a linear relation can still exist between X and Y without joint normality, a nonlinear relationship precludes joint normality. This shows that nonlinearity and non-normality are distinct

concepts. Despite the availability of more and more complex computer solutions, linear relationships remain largely advocated for empirical inquiries in economics and finance. Prominent models that are based on linear relationships include the Arbitrage Pricing Theory (APT), Taylor rules and Keynesian consumption functions. Unfortunately, a linearity assumption may lead to wrongly falsifying a theory when the true unknown relationship of interest is nonlinear. Even after acknowledging that the relationship between two variables is nonlinear, finding the functional form that works best for the situation of interest may still be difficult.

In this paper, I propose an approach to measure the degree of nonlinearity of the relationship between two variables. For any pair (X, Y), I define the exposure of Y to X as the expectation of Y given X. The relationship between X and Y is then said to be nonlinear if either Y is nonlinearly exposed to X or X is nonlinearly exposed to Y. Said differently, the relationship between X and Y is linear if and only if Y is linearly exposed to X and in turn X is linearly exposed to Y. Indeed, it is possible that $E(Y|X)$ be a linear function of X without $E(X|Y)$ being linear in Y. In this case, the linearity of $E(Y|X)$ is spurious as it is misleading about the true relationship between the two variables. Please note that the relationship between variables is approached in this paper from a predictive point of view rather than from a causal perspective. Knowing that the predictive relationship between X and Y is nonlinear is a good starting point for the search of the true underlying causal relationship.

The proposed measure for the degree of nonlinearity of the function $E(Y|X)$, denoted κ, exploits the relative importance of the norms of the linear and nonlinear parts of $E(Y|X)$ in a functional space. The decomposition of $E(Y|X)$ into its linear and nonlinear part is done via a functional projection of $E(Y|X)$ onto a basis of orthogonal polynomials $\left\{P_j(X)\right\}_{j=0}^{\infty}$, where $P_j(X)$ is a polynomial of order j and the orthogonality is defined with respect to a metrics $m(x)$. Upon observing that linear functions of X are loaded only on $P_0(X)$ and $P_1(X)$, a function is said to be purely nonlinear when it is entirely loaded on the higher order polynomials $\left\{P_j(X)\right\}_{j=2}^{n}$. For any function of X, the value of κ always lies between 0 and 1, with 0 meaning that $E(Y|X)$ is linear in X and 1 meaning that $E(Y|X)$ is purely nonlinear as per the previous definition. The index κ is invariant to linear transformations of Y as well as to the addition of an independent noise to Y. It is not invariant to the choice of the metrics $m(x)$ and it is therefore sensitive to transformations of X.

Unlike our proposed measure of nonlinearity, the Pearson *linear correlation coefficient* measures the propensity of a random variable Y of being replicated by a linear function of another variable X. It always lies between −1 and +1, with 1 meaning a perfectly linear and positive relationship, 0 the absence of linear relationship and −1 a perfectly linear and negative relationship. A linear correlation coefficient lying strictly between −1 and 1 indicates that a fit of Y by a linear function of X will not be perfect. This imperfection arises either from the dependence of Y on random factors other than X or from nonlinearity in the relationship between Y and X (or both). The linear correlation coefficient is smaller than 1 in absolute value when X and Y are bound by a deterministic but nonlinear relationship. In an effort to repair the limitations of the linear correlation, measures of nonlinear association (typically, based on ranks) have been proposed. Notably, we have the rank correlation coefficient of Spearman (1904), Kendall's tau (Kendall 1938, 1970), Goodman and Kruskal's gamma (Goodman and Kruskal 1954, 1959) and the quadrant count ratio (see Holmes 2001). All rank correlation measures are designed to detect the strength of possibly nonlinear but monotonic relationships between Y and X. Unlike the linear correlation, a rank correlation coefficient equals 1 if for instance $Y = \log X$ and −1 if $Y = 1/X$ (assuming $X > 0$). However, like the linear correlation, a rank correlation coefficient can be non significant if the relationship between X and Y is non-monotonic.

The remainder of the paper is organized as follows. Section 2 motivates the use of the nonlinearity index (κ) proposed in this paper. Section 3 presents the derivation of κ. Section 4 discusses the choice of the metrics $m(x)$ used to calculate κ. Section 5 examines the invariance properties of κ. Section 6 illustrates the calculation of κ for simple functions and warns against coarse errors when choosing the metrics $m(x)$. Section 7 proposes a feasible estimator for κ and shows its consistency. Section 8 proposes three applications of our methodology. In the first application, I compute the nonlinearity

index of a European option relatively to the underlying asset. It is found that the degree of nonlinearity of the option depends on its maturity and strike as well as on the volatility of the underlying asset. The second application underscores the importance of performing a nonlinearity diagnosis prior to designing a hedging strategy for a portfolio. In the third application, I analyze the nature of the relationship between the returns on the SP500 index and the associated risk as measured by the realized variance (RV). The empirical results are supportive of the existence of a nonlinear relationship between the expected return and the expected risk. The SP500 seems to be driven by two regimes, one regime in which the expected return is increasing in the expected risk and another regime in which the trade-off is negative. Within each regime, the return-risk trade-off is approximately linear. Section 9 concludes. The mathematical proofs are gathered in Appendix A.

2. Motivation

This section starts with an empirical example which shows that $E(Y|X)$ can be linear with $E(X|Y)$ being concomitantly nonlinear. The second subsection presents a theoretical explanation for the empirical example. The third subsection presents a situation where ignoring the presence of nonlinearity in the data can be harmful.

2.1. Pitfalls in the Linearity of Conditional Expectations

It is possible to have two random variables X and Y such that $E(Y|X)$ is linear in X while $E(X|Y)$ is nonlinear in Y. To support this point by empirical arguments, I downloaded daily observations on the SP500 from Yahoo Finance covering the period from 1 February 1959 to 30 April 2013 (14,173 days). The daily data are used to generate monthly returns (R_t) and log-returns denoted $r_t = \log R_t$ (652 months). The realized volatility (RV_t) is computed as the sum of squared daily log-returns within Month t. Figure 1a shows the scatter plot of the r_t on the y-axis against RV_t on the x-axis. The relationship between r_t and V_t is not visible on this plot. Figure 1b,c shows the scatter plots of r_t against $v_t = \log RV_t$ and v_{t-1} respectively. Taking the log of RV_t zooms the pictures out. The shapes of the two graphs are quite similar.


(a)　　　　　　　　　　　　　　　　　　　　　　(b)

Figure 1. *Cont.*

(c)
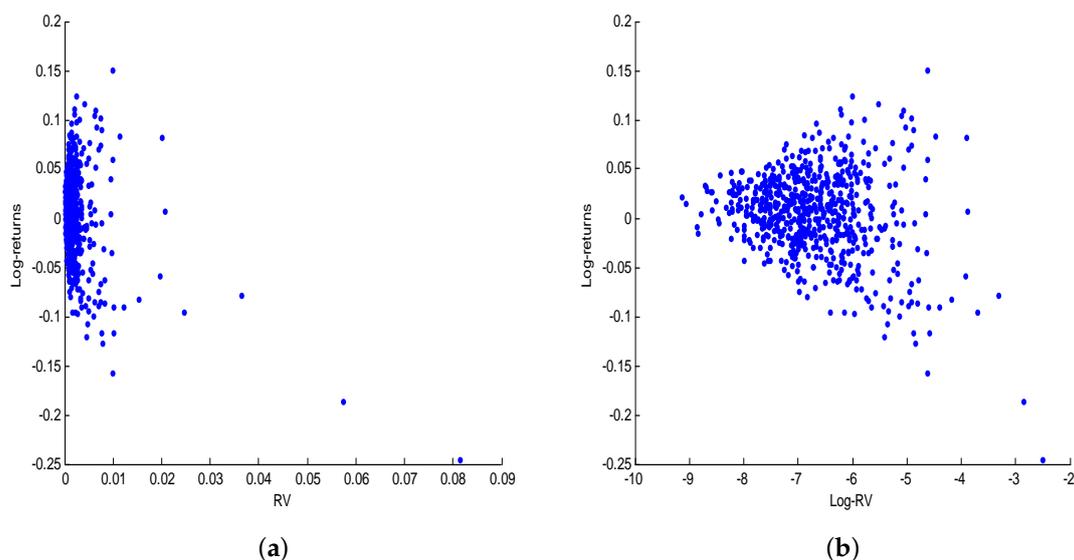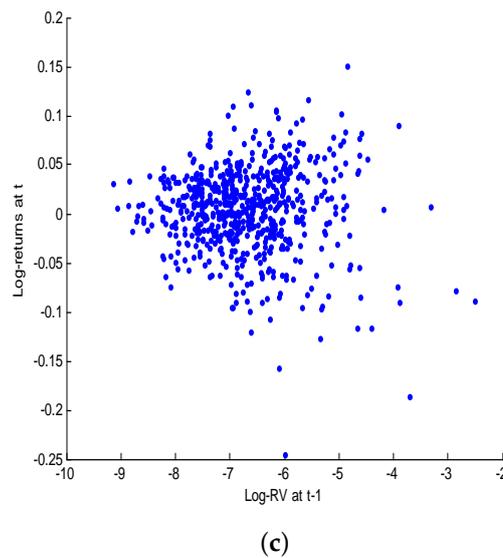
Figure 1. Scatter plots of daily log-returns against RV and log-RV. (**a**) r_t against RV_t; (**b**) r_t against v_t; (**c**) r_t against v_{t-1}.

Figure 2 shows the nonparametric estimators of $E(r_t|v_t)$ and $E(v_t|r_t)$. For any pair (X, Y), the expectation of Y conditional on $X = x$ is estimated as:

$$\widehat{G}(x) = \frac{\sum_{t=1}^{T} y_t \varphi\left(\frac{x-x_t}{h}\right)}{\sum_{t=1}^{T} \varphi\left(\frac{x-x_t}{h}\right)},$$

where $\{(x_1, y_1), ..., (x_T, y_T)\}$ is the observed sample, $\varphi()$ is the Gaussian kernel, $h = \frac{1}{3T}\sum_{t=1}^{T}(x_t - \overline{x})^2$, and $\overline{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$. The curve of $E(r_t|v_t)$ shown on Figure 2a suggests that the relationship between r_t and v_t is more or less linear. However, the curve of $E(v_t|r_t)$ shows on Figure 2b is clearly nonlinear.


(a)　　　　　　　　　　　　　　　　　　　　　　　　**(b)**
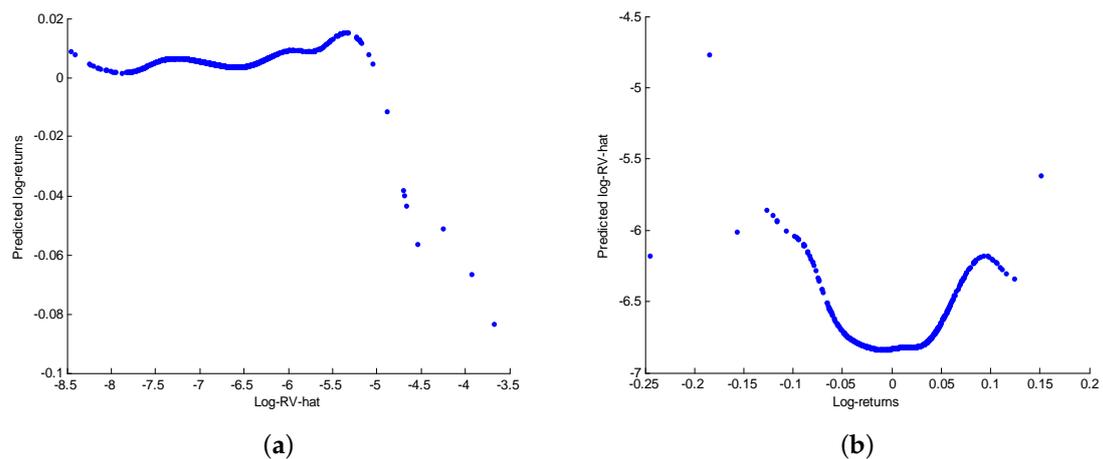
Figure 2. Nonparametric estimation of $E(r_t|v_t)$ and $E(v_t|r_t)$. (**a**) $\widehat{G}(v_t) \equiv E(r_t|v_t)$; (**b**) $\widehat{G}(r_t) \equiv E(v_t|r_t)$.

The shape of $E(v_t|r_t)$ is consistent with the existence of two regimes in the joint dynamics of (v_t, r_t), as found by Ghysels et al. (2014). The first regime is a vicious circle in which higher levels of

expected risk are associated with higher losses (financial crises occur during this regime)[1]. By contrast, the second regime is a virtuous circle where higher levels of expected risk are associated with higher gains. The nonlinearity index proposed in this paper can be used to identify the two regimes.

2.2. Spurious Linearity

The expression *"spurious linearity"* may be used to describe a situation where the exposure of X to Y is linear while the reverse exposure is nonlinear, as in the previous example. To illustrate this concept theoretically, let us consider a deterministic mapping of the following form:

$$Y = \begin{cases} 2X, & \text{if } X \in [0,1) \\ 3 - X, & \text{if } X \in [1,3]. \end{cases} \tag{1}$$

The relationship described by Equation (1) is plotted in Figure 3a. In this equation, Y is a deterministic and nonlinear transformation of X, which means that $Y = E(Y|X)$ and that the pseudo-R^2 of the nonlinear regression of Y onto X is $\rho^2(Y, E(Y|X)) = 1$.[2]

However the reciprocal mapping from Y to X is not deterministic. To see this, assume that X follows a Uniform distribution on $[0,3]$. Figure 3b plots X on the vertical axis against Y on the horizontal axis. Each value of Y is related to two possible values of X. Therefore, the knowledge of Y does not permit to identify with certainty the value of X to which it is associated. For each Y, the possible values of X are $\frac{Y}{2}$ and $3 - Y$.


(a)   **(b)**
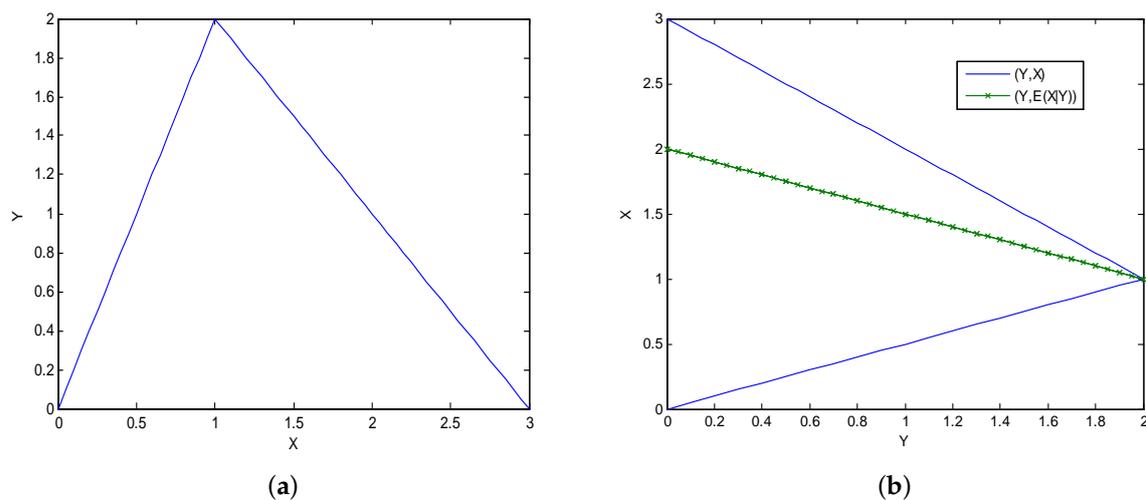
Figure 3. Spurious linearity. (**a**) Y is nonlinear deterministic function of X; (**b**) Expectation of X given Y is linear function of Y.

---

[1]   Please note that outliers have not been removed from the sample.

[2]   As $Y = E(Y|X) + \varepsilon$ with $E(\varepsilon|X) = 0$, we have $\dfrac{Cov(Y, E(Y|X))}{Var[E(Y|X)]} = 1$. Hence:

$$\rho^2(Y, E(Y|X)) = \left(\frac{Cov(Y, E(Y|X))}{\sqrt{Var(Y) \, Var[E(Y|X)]}} \right)^2 = \frac{Var[E(Y|X)]}{Var(Y)}.$$

Please note that $\dfrac{Var[E(Y|X)]}{Var(Y)}$ and $\rho^2(Y, E(Y|X))$ coincide with the R^2 in linear regression models. In nonlinear regressions, the empirical counterpart of $\frac{Var[E(Y|X)]}{Var(Y)}$ can be larger than one in finite sample and depending on the choice of functional form for $E(Y|X)$. However, the empirical version of $\rho^2(Y, E(Y|X))$ always lies between 0 and 1. Hence, $\rho^2(Y, E(Y|X))$ is used here as Pseudo-R^2 for nonlinear regression models.

To find $E(X|Y)$, it helps to think of the joint distribution of (X, Y) as being generated by two regimes. Under Regime $j = 1$, $Y = 2X$ and $E(X|Y, j = 1) = \frac{Y}{2}$ whereas under Regime $j = 2$, we have $Y = 3 - X$ and $E(X|Y, j = 2) = 3 - Y$. Hence:

$$
\begin{aligned}
E(X|Y) &= E(X|Y, j = 1) \Pr(j = 1) + E(X|Y, j = 2) \Pr(j = 2) \\
&= E(X|Y, j = 1) \Pr(0 \leq X \leq 1) + E(X|Y, j = 2) \Pr(1 \leq X \leq 3) \\
&= \frac{1}{3} \frac{Y}{2} + \frac{2}{3}(3 - Y) = 2 - \frac{Y}{2}.
\end{aligned}
$$

This shows that $E(X|Y)$ is linear in Y while $E(Y|X)$ is nonlinear in Y. The pseudo-R^2 of the nonlinear regression of X onto Y is given by:

$$
\rho(X, E(X|Y)) = \frac{Cov(X, E(X|Y))}{\sqrt{Var(X)Var(E(X|Y))}} = \frac{-\frac{1}{2}Cov(X, Y)}{\frac{1}{2}\sqrt{Var(X)Var(Y)}} = -1/3.
$$

In summary, the exposure of Y to X is nonlinear and strong while the exposure of X to Y is linear and weak. This simply reflects the fact that X is a richer conditioning information set than Y. Indeed, the values of X implicitly determine the regimes so that:

$$
E(Y|X = x) = E[Y|X = x, j = k],
$$

where $k \in \{1, 2\}$. Consequently, more information is gleaned about the type of relationship between X and Y by examining the expectation of Y given X.

The methodology proposed in this paper can be used to compute and compare the degrees of nonlinearity of $E(X|Y)$ and $E(Y|X)$. If one of the two conditional expectations is far more nonlinear than the other, this would be suggestive that the relationship between X and Y is non-monotonic. In the empirical example of Figure 2b and the theoretical illustration of Figure 3a, an examination of the degree of nonlinearity of $E(Y|X)$ on increasing subsets of type $(-\infty, x]$ of the support of X can be used to identify a candidate threshold x^* for a piecewise linear approximation to the true model.

2.3. Disentangling Nonlinearity from Non-Normality

Models of equity premium prediction are typically specified as:

$$
y_t = \alpha + \beta x_t + \varepsilon_t, \tag{2}
$$

where y_t is an excess return process, x_t is a vector of risk factors and ε_t is an error term. The Capital Asset Pricing Model (CAPM) is a one factor model which takes x_t to be the excess return on the market portfolio (See e.g., Sharpe 1964). Merton (1973) proposed an Intertemporal Capital Asset Pricing Model (ICAPM) where x_t is the lagged realization of the asset's volatility. Merton (1980) further insisted on the fact that empirical predictions of expected returns must be related to changes in expected future market risk in order to be consistent with equilibria model.

Following this idea, French et al. (1987) conducted an empirical study where y_t and x_t are respectively the stock market excess return and expected volatility. They found that expected return is positively related to expected risk while unexpected return is negatively related to unexpected risk (so-called "leverage effect"). Attempts have also been made to predict the equity premium using valuation ratios, such as companies' sizes, earnings-to-price ratios, cash flow-to-price ratios, book-to-market equity, sales growth, etc. See for instance Fama and French (1996) and their famous three-factor model.

Another strand of this literature focuses on the estimation of long run return-risk trade-off. Using long horizon regression models, Bandi and Perron (2008) find that past realized market variance is a good predictor of future excess returns. Jacquier and Okou (2014) reappraise this result by

separating the market realized variance into its continuous part and its jump component. They find that the power of past realized volatility at predicting the future risk premium is attributable to its continuous part. Okou and Jacquier (2016) refined their own results by performing statistical inferences on the term structure of the return-risk trade off at long horizon. Their main finding is that the results depend much on whether an intercept is included in the regressions or not, and that there is an horizon effect in the risk return trade-off. They argued that differences in the data frequency may be responsible for the conflicting empirical conclusions on the risk-return trade-off.[3] I argue that in the presence of unsuspected nonlinearity, a linear regression can lead to quite misleading conclusions. In fact, the slope of the regression (2) can be insignificant as a result of $E(y_t|x_t)$ being nonlinear in x_t. Moreover, the estimate of β that comes out of this regression is meaningless in the presence of nonlinearity.

Nonparametric regressions are sometimes advocated by empirical researchers on the ground that excess returns are non-normal (See e.g., Harvey 2001). However, it is easy to conceive a linear relationship between y_t and x_t with non-normal residuals. Likewise, $E(y_t|x_t)$ can be nonlinear while ε_t is Gaussian. In the latter case, nonlinearity can cause the residuals of the linear regression of y_t onto x_t to be skewed or fat-tailed. The κ index proposed in this paper can help suspect whether the behavior of the model is driven by nonlinearity or non-normality.

3. Measuring Nonlinearity

Let the exposure of Y to X be given by:

$$G(X) = E(Y|X). \tag{3}$$

where X and Y are scalar random variables and $G(X)$ is a possibly nonlinear function of X. In assets pricing, Y could be the risk premium on an asset and X a measure of the risk for bearing that asset, in which case $G(X)$ describes a nonlinear risk-return trade-off. Alternatively, Y can be viewed as an investors portfolio and X the traditional market index. In the latter case, $G(X)$ would be reflecting a nonlinear exposure to the market stemming from a non-directional investment strategy. In macroeconomics, Y could be the inflation rate and X the unemployment rate, in which case $G(X)$ features a nonlinear Phillips curve. Finally, $Y = Y_t$ could be an arbitrary process and $X = Y_{t-1}$ its lagged value in a nonlinear time series model.

My objective is to assess the extent of nonlinearity of $G(X)$. Let the population linear regression of Y onto X be denoted by:

$$EL(X) = \alpha + \beta X, \tag{4}$$

where α and β are real numbers. Please note that $EL(X)$ coincides with the linear regression of $G(X)$ onto X and that $EL(X)$ coincides with $G(X)$ if and only if Y can be represented as:

$$Y = \alpha + \beta X + \varepsilon,$$

where ε is linearly uncorrelated with X. In the particular case where (X, Y) is bivariate Gaussian, $G(X)$ and $EL(X)$ are necessarily identical and ε is normally distributed as well. That is, the joint Gaussianity of (X, Y) is sufficient but not necessary for linearity.

With no loss of generality, I assume that $G(X)$ admits a series representation of the following form:

$$G(X) = \sum_{j=0}^{\infty} \gamma_j P_j(X), \tag{5}$$

---

[3] Two empirical studies using different databases recorded at the same frequency or two studies based on the same dataset but using different methods, may still lead to conflicting conclusion. See the introduction of Ghysels et al. (2005).

where $\{P_j(X)\}_{j=0}^{\infty}$ is a complete sequence of orthogonal polynomials over the support of X under some metrics $m(x)$, that is:

$$\int_{D(X)} P_i(x) P_j(x) m(x) dx = 0 \text{ if } i \neq j,$$

and $D(X)$ is the *tight* range of X, that is, the domain over which the values of X are meaningful. For instance, $D(X)$ is a priori $(0, \infty)$ for a price process and $\mathbb{R}$ for a log-return process. $P_j(x)$ is a polynomial of order j and $P_0(x) = 1$. Please note that $G(x)$ satisfies (5) if and only if:

$$\int_{D(X)} G(x)^2 m(x) dx < \infty, \tag{6}$$

See Carrasco et al. (2007) and the references therein. Upon knowing $G(X)$, the metrics $m(x)$ can always be selected to meet the condition (6).

Let the projection of $G(X)$ onto $[P_0(.), P_1(.)]$ under the metrics $m(x)$ be given by:

$$LP(X) = \gamma_0 P_0(X) + \gamma_1 P_1(X), \tag{7}$$

where:

$$\gamma_0 \equiv \frac{\langle G(.), P_0(.) \rangle_m}{\langle P_0(.), P_0(.) \rangle_m} = \frac{\int_{-\infty}^{\infty} P_0(x) G(x) m(x) dx}{\int_{-\infty}^{\infty} P_0(x)^2 m(x) dx} \text{ and} \tag{8}$$

$$\gamma_1 \equiv \frac{\langle G(.), P_1(.) \rangle_m}{\langle P_1(.), P_1(.) \rangle_m} = \frac{\int_{-\infty}^{\infty} P_1(x) G(x) m(x) dx}{\int_{-\infty}^{\infty} P_1(x)^2 m(x) dx}, \tag{9}$$

and $\langle ., . \rangle_m$ denotes the scalar product under $m(x)$. The function $G(X)$ is linear if and only if it is loaded only on the first two basis functions, that is:

$$G(X) = LP(X) = EL(X).$$

where the first equality is deduced from (5) and the last equality stems from (4).

$G(X)$ is nonlinear if and only if the residual of the projection of $G(X)$ onto $[P_0(.), P_1(.)]$, i.e., $G(X) - LP(X)$, is not identically null. Therefore, the nonlinear part of $G(X)$ may be isolated as:

$$G(X) - LP(X) = \sum_{j=2}^{\infty} \gamma_j P_j(X). \tag{10}$$

Based of this observation, a measure of the degree of nonlinearity of $G(X)$ is given by:

$$\kappa = \sqrt{1 - \frac{\|LP(X) - \gamma_0\|_m^2}{\|G(X) - \gamma_0\|_m^2}}. \tag{11}$$

By construction, $\kappa = 0$ if $G(X)$ is perfectly linear and $\kappa = 1$ if $G(X)$ is fully loaded on the nonlinear basis functions. Hence, κ always lies between 0 and 1 and is decreasing in the degree of linearity of $G(X)$ as measured by the ratio of the norms of $LP(X) - \gamma_0 = \gamma_1 P_1(X)$ and $G(X) - \gamma_0 = \sum_{j=1}^{\infty} \gamma_j P_j(X)$. Equivalent expressions of κ are therefore given by:

$$\kappa^2 = \frac{\sum_{j=2}^{\infty} \gamma_j^2 \|P_j\|_m^2}{\sum_{j=1}^{\infty} \gamma_j^2 \|P_j\|_m^2} = 1 - \frac{\gamma_1^2 \|P_1\|_m^2}{\sum_{j=1}^{\infty} \gamma_j^2 \|P_j\|_m^2}. \tag{12}$$

Please note that κ is not defined when $G(X) = \gamma_0$, just as the linear correlation coefficient between Y and a constant does not exist.

The choice of the metrics $m(x)$ is a crucial step of the methodology presented above. Indeed, the value of κ depends on the metrics used and a bad choice of metrics may lead to spuriously detect nonlinearity. This issue is discussed in the next section.

4. The Conditioning Information Set and the Suitable Choice of Metrics

Let $D(X)$ denote the conditioning information set, that is, the support of X. A probability measure $f(x)$ on $D(X)$ is said to belong to Pearson's family if and only if it satisfies:

$$\frac{f'(x)}{f(x)} = \frac{a_1 x + a_2}{1 + c_1 x + c_2 x^2}, \, a_1 \leq 0, a_2, c_1, c_2 \in \mathbb{R}. \tag{13}$$

For instance, letting $a_1 = -1$ and $a_2 = c_1 = c_2 = 0$ leads to:

$$\frac{f'(x)}{f(x)} = \frac{d \log f(x)}{dx} = -x \Leftrightarrow f(x) \propto \exp\left(-\frac{1}{2}x^2\right),$$

which is the Gaussian probability distribution function on $(-\infty, +\infty)$. Likewise, letting $a_1 = a_2 = 0$ yields the uniform distribution on $[\underline{b}, \overline{b}]$ while setting $c_1 = \frac{a_1}{a_2}$ ($a_1 \neq 0, a_2 < 0$) and $c_2 = 0$ yields an exponential distribution on $[\underline{b}, +\infty)$. The Student, Gamma, Beta distributions are also special members of the Pearson family. See Johnson et al. (1994, pp. 15–25) and Bontemps and Meddahi (2012) for more details.

If a probability measure $f(x)$ defined on $[\underline{b}, \overline{b}]$ belongs to the Pearson family, the sequence of orthogonal polynomials under $f(x)$ are given by Rodrigues' formula (see Askey 2005):

$$P_j(x) = e_j \frac{1}{f(x)} \frac{d^{(j)}}{dx^j} \left[\left(1 + c_1 x + c_2 x^2\right)^j f(x)\right], j = 0, 1, ..., \infty \tag{14}$$

where $\frac{d^{(j)}}{dx^j}$ is the nth order differentiation operator and e_j is a sequence of normalization factors that could be chosen so as to achieve specific purposes. That is, any sequence $P_j(x)$ given by (14) satisfies:

$$\int_{\underline{b}}^{\overline{b}} P_j(x) P_i(x) f(x) dx = 0.$$

Subsequently, we consider five cases that are representative of the situations that researchers will often face in practice.

Case 1: When $D(X) = \mathbb{R}$, a suitable choice of metrics is $m(x) = e^{-x^2}$. The corresponding orthogonal basis is given by Hermite polynomials:

$$\begin{aligned} H_0(x) &= 1; \, H_1(x) = 2x \text{ and} \\ H_{j+1}(x) &= 2x H_j(x) - H_j'(x) \text{ for all } j \geq 2. \end{aligned}$$

where $H_j'(x)$ is the derivative of $H_j(x)$ with respect to x. In this case, the nonlinearity of $G(X)$ is given by:

$$\kappa = \sqrt{1 - \frac{2\sqrt{\pi}\gamma_1^2}{\int_{-\infty}^{\infty} G(x)^2 e^{-x^2} dx - \gamma_0^2 \sqrt{\pi}}} \tag{15}$$

where:

$$\gamma_0 \equiv \frac{\langle G(.), H_0(.)\rangle_m}{\langle H_0(.), H_0(.)\rangle_m} = \frac{\int_{-\infty}^{\infty} G(x)e^{-x^2}dx}{\int_{-\infty}^{\infty} e^{-x^2}dx} = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} G(x)e^{-x^2}dx \text{ and} \tag{16}$$

$$\gamma_1 \equiv \frac{\langle G(.), H_1(.)\rangle_m}{\langle H_1(.), H_1(.)\rangle_m} = \frac{\int_{-\infty}^{\infty} 2xG(x)e^{-x^2}dx}{\int_{-\infty}^{\infty} (2x)^2 e^{-x^2}dx} = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} xG(x)e^{-x^2}dx, \tag{17}$$

Case 2: When $D(X) = [0, +\infty)$, one may use $m(x) = e^{-x}$ along with the corresponding orthogonal basis formed by Laguerre polynomials:

$$L_0(x) = 1; \ L_1(x) = -x + 1 \text{ and}$$
$$L_j(x) = \left(2 - \frac{1+x}{j}\right)L_{j-1}(x) - \left(1 - \frac{1}{j}\right)L_{j-2}(x); \text{ for all } j \geq 2.$$

The nonlinearity of $G(X)$ is then given by:

$$\kappa = \sqrt{1 - \frac{\delta_1^2}{\int_0^{\infty} G(x)^2 e^{-x}dx - \delta_0^2}} \tag{18}$$

where:

$$\delta_0 \equiv \frac{\langle G(.), L_0(.)\rangle_m}{\langle L_0(.), L_0(.)\rangle_m} = \frac{\int_0^{\infty} G(x)e^{-x}dx}{\int_0^{\infty} e^{-x}dx} = \int_0^{\infty} G(x)e^{-x}dx \text{ and} \tag{19}$$

$$\delta_1 \equiv \frac{\langle G(.), L_1(.)\rangle_m}{\langle L_1(.), L_1(.)\rangle_m} = \frac{\int_0^{\infty} (1-x)G(x)e^{-x}dx}{\int_0^{\infty} (1-x)^2 e^{-x}dx} = \int_0^{\infty} (1-x)G(x)e^{-x}dx. \tag{20}$$

Case 3: If X can realistically not fall below a given threshold $\underline{b}$, one may consider defining the domain of X as $D(X) = [\underline{b}, \infty)$. The corresponding Laguerre polynomials are obtained by noting that:

$$\int_0^{\infty} L_i(u)L_j(u)e^{-u}du = 0 \Leftrightarrow \int_{\underline{b}}^{\infty} L_i(x - \underline{b})L_j(x - \underline{b})e^{-x}dx = 0.$$

Hence for $i \neq j$, $L_i(x - \underline{b})$ is orthogonal to $L_j(x - \underline{b})$ with respect to $m(x) = e^{-x}$ on $[\underline{b}, \infty)$.

Case 4: When $D(X) = [-1, 1]$, the simplest possible choice of metrics is uniform weighting function $m(x) = 1$. The corresponding orthogonal basis of functions consists of Legendre's polynomials:[4]

$$P_0(x) = 1; \ P_1(x) = x$$
$$P_{j+1}(x) = \frac{2n+1}{n+1}xP_j(x) - \frac{n}{n+1}P_{j-1}(x).$$

Case 5: For an arbitrary bounded domain $D(X) = [\underline{b}, \overline{b}]$, we simply note that:

$$\underline{b} \leq X \leq \overline{b} \Leftrightarrow -1 \leq \frac{2X - \underline{b} - \overline{b}}{\overline{b} - \underline{b}} \leq 1.$$

Hence, by letting $u = \frac{2X - \underline{b} - \overline{b}}{\overline{b} - \underline{b}}$, it is straightforward to show that:

4　Alternative choices of basis functions are Jacobi, Gegenbauer or Tchebychev's polynomials along with the suitable metrics.

$$\int_{-1}^{1} P_i(u) P_j(u) du = 0 \Leftrightarrow$$

$$\int_{\underline{b}}^{\overline{b}} P_i\left(\frac{2x - \underline{b} - \overline{b}}{\overline{b} - \underline{b}}\right) P_j\left(\frac{2x - \underline{b} - \overline{b}}{\overline{b} - \underline{b}}\right) dx = 0.$$

Hence, a suitable choice of basis functions when $X \in [\underline{b}, \overline{b}]$ is given by the sequence $\left\{P_j\left(\frac{2x - \underline{b} - \overline{b}}{\overline{b} - \underline{b}}\right)\right\}_{j=0}^{\infty}$.
The measure of nonlinearity for this case is:

$$\kappa = \sqrt{1 - \frac{\gamma_1^2\left(\overline{b} - \underline{b}\right)/3}{\int_{\underline{b}}^{\overline{b}} G(x)^2 dx - \gamma_0^2\left(\overline{b} - \underline{b}\right)}} \tag{21}$$

where:

$$\gamma_0 \equiv \frac{\langle G(.), P_0(.)\rangle_m}{\langle P_0(.), P_0(.)\rangle_m} = \frac{\int_{\underline{b}}^{\overline{b}} G(x) dx}{\int_{\underline{b}}^{\overline{b}} dx} = \frac{1}{\overline{b} - \underline{b}} \int_{\underline{b}}^{\overline{b}} G(x) dx \text{ and} \tag{22}$$

$$\gamma_1 \equiv \frac{\langle G(.), P_1(.)\rangle_m}{\langle P_1(.), P_1(.)\rangle_m} = \frac{\int_{\underline{b}}^{\overline{b}}\left(\frac{2x - \underline{b} - \overline{b}}{\overline{b} - \underline{b}}\right) G(x) dx}{\int_{\underline{b}}^{\overline{b}}\left(\frac{2x - \underline{b} - \overline{b}}{\overline{b} - \underline{b}}\right)^2 dx} = \frac{3}{\overline{b} - \underline{b}} \int_{\underline{b}}^{\overline{b}}\left(\frac{2x - \underline{b} - \overline{b}}{\overline{b} - \underline{b}}\right) G(x) dx. \tag{23}$$

The latter set up best suits for measuring nonlinearity on segments of the support of X.

Any metrics that follows the guideline described above will delivers a measure of nonlinearity that is reliable. This means that for an appropriately chosen metrics, the function under consideration is nonlinear as soon as κ is strictly positive. However, the interpretation of the result is "*metrics specific*", meaning that κ is a relative measure of nonlinearity. While the degrees of nonlinearity of different functions obtained under different metrics cannot be compared, different functions sharing the same support can be compared under the same metrics.

5. Invariance Properties

Observe that $1 - \kappa^2$ has the flavor of the R^2 of a linear regression as it measures the "goodness-of-fit" of the functional projection of $G(X)$ onto $[P_0(.), P_1(.)]$. Based on this observation, one is tempted to claim that κ shares all the invariance properties as an R^2. However, such a statement is only partially true because of the dependence of κ on the metrics $m(x)$. The invariance properties of κ are discussed below.

Proposition 1. *κ is invariant to a linear transformation of Y.*

Proposition 1 establishes that the amount of nonlinearity remains the same under drifting and scaling of Y. Applied to a portfolio of financial assets, this property means that leverage does not affect the nonlinearity of a financial position. Another property shared by the R^2 is stated below.

Proposition 2. *κ is invariant to the addition of a randomness ε to Y provided that ε is independent of X.*

This property is rather interesting as it implies that κ may be used to diagnose linear models with additive error terms. Let us assume that X is the return on the market index and Y the return on the portfolio of an investor such that $E(Y|X) = \alpha + \beta X$ (i.e., we are assuming that the exposure of Y to X is perfectly linear). The strength of the exposure of Y to X is given by the linear correlation between Y and $\alpha + \beta X$, that is, the square root of R^2 of the regression of Y onto X. Suppose the investor decides to

implement a non-directional (i.e., market neutral) diversification strategy by changing Y into $\frac{1}{2}Y + \frac{1}{2}\varepsilon$, where ε is independent of X. The exposure of the new position is given by $E\left(\frac{1}{2}Y + \frac{1}{2}\varepsilon|X\right) = \widetilde{\alpha} + \widetilde{\beta}X$, where $\widetilde{\alpha} = \alpha + \frac{1}{2}E\left(\varepsilon\right)$ and $\widetilde{\beta} = \frac{1}{2}\beta$. The alpha of the new position may have increased of decreased depending on the value of $E\left(\varepsilon\right)$ and its beta is reduced by half. However, the nonlinearity of the new position as measured by κ is unaltered since $E\left(\frac{1}{2}Y + \frac{1}{2}\varepsilon|X\right)$ remains linear in X.

The next proposition examines the consequence of an addition of a linear function of X to Y.

Proposition 3. *If Y is nonlinearly exposed to X, Adding a linear function of X to Y does not necessarily decrease its degree of nonlinearity.*

Indeed, κ is not invariant to the addition of a linear function of X to Y. To understand this result in the context of the proposition (see the proof in Appendix A for more details), suppose γ_1 is positive so that $-4\gamma_1 < 0$. Then adding $Z = aX + b$ to Y exacerbates its nonlinearity if a is negative and lies within the range $[-4\gamma_1, 0]$. The degree of nonlinearity decreases only if $a < -4\gamma_1$ or $a > 0$. Alternatively, suppose γ_1 is negative so that $-4\gamma_1 > 0$. Then adding $Z = aX + b$ to Y exacerbates its nonlinearity if a is positive and lies within the range $[0, -4\gamma_1]$. Otherwise, the nonlinearity of Y decreases. Applied to portfolio choice, Proposition 3 implies that an asset that is linearly exposed to X can be used to increase the nonlinearity of an already nonlinear position Y. A sufficient condition for the addition of $Z = aX + b$ to reduce the nonlinearity of Y is that a be of the same sign as γ_1.

The property of κ highlighted by Proposition 3 is also shared by the linear correlation coefficient. Unlike the correlation coefficient however, the value of κ is sensitive to drifting and scaling of X.

Proposition 4. *Let $Z = aX + b$ where a and b are some constants. Then the degree of nonlinearity of Y with respect to X under $m(x)$ is equal to its degree of nonlinearity with respect to Z under $\widetilde{m}(z) = \frac{1}{a}m\left(\frac{z-b}{a}\right)$.*

The result of Proposition 4 stems from the fact that any transformation of X alters the metrics $m(x)$, which in turn invalidates the orthogonality of $\left\{P_j(X)\right\}_{j=0}^{\infty}$. It implies that the values of κ are not directly comparable across different choices of metrics, which is a drawback of the proposed methodology. However, this drawback is a minor one if $m(x)$ is continuous and puts zero weights outside the support of X. Also, functions that are defined on the same domain $D(X)$ may be compared under the same metrics provided that they all have finite norms.

6. Spurious Nonlinearity

This section illustrates how to compute κ when $G(X)$ is known and in passing, underscores the importance of selecting the metrics $m(x)$ wisely. Indeed, *spurious nonlinearity* may arise from a bad choice of metrics. To see this, let us consider the exponential function $G(X) = e^X$, which also has the following representation:

$$e^X = \sum_{j=1}^{\infty} \frac{X^k}{k!} = 1 + X + \frac{X^2}{2} + ..., \text{ for } X \in \mathbb{R}. \tag{24}$$

It is tempting to claim based on (24) that $\gamma_0 = \gamma_1 = 1$. However, such a claim would be false since γ_0 and γ_1 have been defined as the coordinates of $G(X)$ in the basis formed by the orthogonal polynomials $\left\{P_j(X)\right\}_{j=0}^{\infty}$. By noting that $D(X) = \mathbb{R}$ for the exponential function, we let $m(X) = e^{-X^2}$ so that $G(X) = \sum_{j=1}^{\infty} \gamma_j H_j(X)$, where $\left\{H_j(X)\right\}_{j=0}^{\infty}$ are Hermite polynomials and:

$$\gamma_0 = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^x \exp\left(-x^2\right) dx = e^{\frac{1}{4}},$$

$$\gamma_1 = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} xe^x \exp\left(-x^2\right) dx = \frac{1}{2}e^{\frac{1}{4}} \text{ and}$$

$$\|G\|_m^2 \quad = \quad \int_{-\infty}^{\infty} e^{2x} \exp\left(-x^2\right) dx = \sqrt{\pi} e.$$

This yields the following measure of nonlinearity:

$$\kappa = \sqrt{1 - \frac{2\sqrt{\pi}\left(\frac{1}{2}e^{\frac{1}{4}}\right)^2}{\sqrt{\pi}e - \left(e^{\frac{1}{4}}\right)^2 \sqrt{\pi}}} = 0.479. \tag{25}$$

Let us now consider $G(X) = \log X$. This function is defined only for $x > 0$ and hence, its nonlinearity *should not be* measured as though x lies on the whole real line. For illustration purposes, let us ignore this warning by letting $G(X) = \sum_{j=1}^{\infty} \gamma_j H_j(X)$. This leads to:

$$\gamma_0 \quad = \quad \frac{1}{\sqrt{\pi}} \int_0^{\infty} (\log x) \exp\left(-x^2\right) dx = \frac{-0.870}{\sqrt{\pi}},$$

$$\gamma_1 \quad = \quad \frac{1}{\sqrt{\pi}} \int_0^{\infty} x(\log x) \exp\left(-x^2\right) dx = \frac{-0.144}{\sqrt{\pi}} \text{ and}$$

$$\|G\|_m^2 \quad = \quad \int_0^{\infty} (\log x)^2 \exp\left(-x^2\right) dx = 1.947.$$

The measure of nonlinearity that results is:

$$\kappa = \sqrt{1 - \frac{2\sqrt{\pi}\left(\frac{-0.144}{\sqrt{\pi}}\right)^2}{1.947 - \left(\frac{-0.870}{\sqrt{\pi}}\right)^2 \sqrt{\pi}}} = 0.992, \tag{26}$$

which is quite excessive compared to what is obtained for the exponential. In reality, the nonlinearity calculated above is for the function given by:

$$\widetilde{G}(X) = \begin{cases} \log X \text{ if } X > 0 \\ 0 \text{ if } X \leq 0, \end{cases}$$

which is distinct from $G(X) = \log x$ on $(0, +\infty)$. The domain of $\widetilde{G}(X)$ is the whole real line whereas the *tight* domain of $G(X)$ is $D(X) = [0, +\infty)$. This explains why we obtain a spuriously high value of κ.

Let us now account for the fact that the domain of $\log X$ is $(0, +\infty)$ by letting $G(X) = \sum_{j=1}^{\infty} \gamma_j L_j(X)$, where $\{L_j(X)\}_{j=0}^{\infty}$ are Laguerre polynomials. We obtain:

$$\delta_0 \quad = \quad \int_0^{\infty} (\log x) \exp\left(-x\right) dx = -0.577,$$

$$\delta_1 \quad = \quad \int_0^{\infty} (1-x)(\log x) \exp\left(-x\right) dx = -1 \text{ and}$$

$$\|G\|_m^2 \quad = \quad \int_0^{\infty} (\log x)^2 \exp\left(-x\right) dx = 1.978.$$

This yields:

$$\kappa = \sqrt{1 - \frac{(-1)^2}{1.978 \, 1 - (-0.57722)^2}} = 0.626, \tag{27}$$

which is more reasonable than previously.

7. Feasible Estimators

Upon observing a sample $\{(x_1, y_1), ..., (x_T, y_T)\}$ of size T from the joint distribution of (X, Y), the conditional expectation $E(Y|X = x) = G(x)$ can be estimated by the nonparametric method of Nadaraya (1964) and Watson (1964):

$$\widehat{G}(x) = \frac{\sum_{t=1}^{T} y_t K\left(\frac{x-x_t}{h}\right)}{\sum_{t=1}^{T} K\left(\frac{x-x_t}{h}\right)}. \tag{28}$$

where $K(z)$ is a kernel function and h is a bandwidth.

With the estimator $\widehat{G}(x)$ above in hand, the sample counterpart of κ is:

$$\widehat{\kappa} = \sqrt{1 - \frac{\widehat{\gamma}_1^2 \|P_1\|_m^2}{\|\widehat{G}\|_m^2 - \widehat{\gamma}_0^2 \|P_0\|_m^2}} \tag{29}$$

where

$$\|\widehat{G}\|_m^2 = \int_{D(X)} \widehat{G}(x)^2 m(x) dx, \tag{30}$$

$$\widehat{\gamma}_0 = \frac{1}{\|P_0\|_m^2} \int_{D(X)} \widehat{G}(x) m(x) dx \text{ and} \tag{31}$$

$$\widehat{\gamma}_1 = \frac{1}{\|P_1\|_m^2} \int_{D(X)} \widehat{G}(x) P_1(x) m(x) dx, \tag{32}$$

with $m(x) = e^{-x^2}$ if $D(X) = \mathbb{R}$, $m(x) = e^{-x}$ if $D(X) = (0, +\infty)$ and $m(x) = 1$ if $D(X) = [\underline{b}, \overline{b}]$. Given these choices of metrics, integrals of the form $\int_{D(X)} f(x) m(x) dx$ can be solved analytically when $f(x)$ is a polynomial. For more complicated functions however, numerical quadratures should be used.

When $m(x) = e^{-x^2}$, the quadrature rule yields:

$$\|\widehat{G}\|_m^2 \simeq \sum_{k=1}^{N} \omega_k \widehat{G}(x_k)^2, \tag{33}$$

$$\widehat{\gamma}_0 = \frac{1}{\sqrt{\pi}} \sum_{k=1}^{N} \omega_k \widehat{G}(x_k) \text{ and} \tag{34}$$

$$\widehat{\gamma}_1 = \frac{1}{\sqrt{\pi}} \sum_{k=1}^{N} \omega_k x_k \widehat{G}(x_k), \tag{35}$$

where $\{x_k\}_{k=1}^{N}$ are Gauss-Hermite quadrature points associated with the weights $\{\omega_k\}_{k=1}^{N}$.

For the alternative metrics $m(x) = e^{-x}$, the quadrature rule becomes:

$$\|\widehat{G}\|_m^2 \simeq \sum_{k=1}^{N} \omega_k \widehat{G}(x_k)^2, \tag{36}$$

$$\widehat{\gamma}_0 = \sum_{k=1}^{N} \omega_k \widehat{G}(x_k) \text{ and} \tag{37}$$

$$\widehat{\gamma}_1 = \sum_{k=1}^{N} \omega_k (1 - x_k) \widehat{G}(x_k), \tag{38}$$

where $\{x_k\}_{k=1}^{N}$ are Gauss-Laguerre quadrature points associated with the weights $\{\omega_k\}_{k=1}^{N}$.

Finally, for the natural metrics $m(x) = 1$, we have:

$$\left\| \widehat{G} \right\|_m^2 \simeq \sum_{k=1}^N \omega_k \widehat{G}(x_k)^2, \tag{39}$$

$$\widehat{\gamma}_0 = \frac{1}{\overline{b} - \underline{b}} \sum_{k=1}^N \omega_k \widehat{G}(x_k) \text{ and} \tag{40}$$

$$\widehat{\gamma}_1 = \frac{3}{\overline{b} - \underline{b}} \sum_{k=1}^N \omega_k \left(\frac{2x^{(k)} - \underline{b} - \overline{b}}{\overline{b} - \underline{b}} \right) \widehat{G}(x_k), \tag{41}$$

where $\{x_k\}_{k=1}^N$ are Gauss-Legendre quadrature points associated with the weights $\{\omega_k\}_{k=1}^N$.

The quadrature rules above are designed such that they are exact when the integrand is a polynomial of order $n \le 2N - 1$. That is:

$$\sum_{k=1}^N \omega_k P_j(x_k) = \int_{D(X)} P_j(x) m(x) dx \text{ for } j = 0, ..., 2N - 1.$$

Thus, it is straightforward to show that:

$$\sum_{k=1}^N \omega_k \widehat{G}(x_k) P_j(x_k) = \widehat{\gamma}_j \left\| P_j \right\|^2 I(N > j) + \sum_{j=2N-j}^\infty \widehat{\gamma}_j \sum_{k=1}^N \omega_k P_l(x_k) P_j(x_k)$$

where $\widehat{\gamma}_j = \frac{1}{\left\| P_j \right\|_m^2} \int_{D(X)} \widehat{G}(x) P_j(x) m(x) dx$. If $N > j$, the approximation error of $\widehat{\gamma}_j$ by a quadrature rule is:

$$\frac{1}{\left\| P_j \right\|^2} \sum_{k=1}^N \omega_k \widehat{G}(x_k) P_j(x_k) - \widehat{\gamma}_j = \frac{1}{\left\| P_j \right\|} \sum_{l=2N-j}^\infty (\widehat{\gamma}_l \left\| P_l \right\|) \left(\sum_{k=1}^N \omega_k \frac{P_l(x_k)}{\left\| P_l \right\|} \frac{P_j(x_k)}{\left\| P_j \right\|} \right) \tag{42}$$

Please note that the finiteness of the norm of $\widehat{G}(x)$ imposes that $\widehat{\gamma}_l \left\| P_l \right\| \to 0$. Furthermore, $\sum_{k=1}^N \omega_k \frac{P_l(x_k)}{\left\| P_l \right\|} \frac{P_j(x_k)}{\left\| P_j \right\|}$ is an approximation of $\frac{1}{\left\| P_l \right\| \left\| P_j \right\|} \int P_l(x) P_j(x) dx = 0$. Consequently, the approximation error (42) converges to zero fast as N increases (especially for $j = 0$ and $j = 1$). This is a good news given that the expression of $\widehat{\kappa}$ only requires on $\widehat{\gamma}_0$ and $\widehat{\gamma}_1$.

The next proposition states a consistency result for $\widehat{\kappa}$ by assuming that its numerical approximation error is negligible.

Proposition 5. *Assume that* $\left\| \widehat{G} - G \right\| = O_p\left(h_T^{-\beta/2} T^{-1/2} \right)$, *where* h_T *is a sequence of bandwidth satisfying* $h_T^{-\beta} T^{-1} \to 0$ *as* $T \to \infty$. *Then we have:*

$$|\widehat{\kappa} - \kappa| = O_p\left(h_T^{-\beta} T^{-1} \right) \text{ as } T \to \infty.$$

According to Proposition 5, any consistent estimator $\widehat{G}(X)$ of $G(X) = E(Y|X)$ may be plugged into (30) to (32) to obtain a consistent estimator of κ. Nonparametric estimators of type (28) have been shown to be consistent under quite general settings (Bierens 1987).

8. Applications

This section presents three applications of κ. The first subsection discusses the degree of nonlinearity of a European option with respect to the underlying asset. The second subsection discusses the optimal hedge ratio in the presence of nonlinearity. The third subsection presents an empirical example where the relationship between the risk and the returns on the SP500 index is analyzed.

8.1. How Nonlinear Are Put and Call Options?

This section examine the nonlinearity of the price of a European style option with respect to the underlying asset. This exercise is trivial in a sense as options generate nonlinear payoffs by construction. However, it is nevertheless useful as it gives us a pretext to compare the degree of nonlinearity of several functions under the same metrics and to further illustrate the importance of correctly selecting the metrics $m(x)$.

The payoff a European Call option at maturity is given by $G(X) = \max(X - K, 0)$, where X is the price of the underlying asset and K is the strike.

If one ignores the fact that the support of X is $[0, \infty)$ and use $m(x) = e^{-x^2}$ to compute the nonlinearity of $G(X)$, then one obtains the following expressions:

$$
\begin{aligned}
\|G\|_m^2 &= \int_K^\infty (x - K)^2 \exp\left(-x^2\right) dx \\
&= \frac{1}{2}\sqrt{\pi} - \frac{1}{2}\sqrt{\pi}\Phi\left(\sqrt{2}K\right) + \sqrt{\pi}K^2 - \sqrt{\pi}K^2\Phi\left(\sqrt{2}K\right) - \frac{1}{2}Ke^{-K^2},
\end{aligned}
$$

$$
\begin{aligned}
\gamma_0 &= \frac{1}{\sqrt{\pi}}\int_K^\infty (x - K)\exp\left(-x^2\right) dx = \frac{1}{2\sqrt{\pi}}e^{-K^2} - K + K\Phi\left(\sqrt{2}K\right) \text{ and} \\
\gamma_1 &= \frac{1}{\sqrt{\pi}}\int_K^\infty x(x - K)\exp\left(-x^2\right) dx = \frac{1}{2} - \frac{1}{2}\Phi\left(\sqrt{2}K\right),
\end{aligned}
$$

where $\Phi(x)$ is the standard normal cumulative distribution function. This suggests that $\kappa_C(K) = \sqrt{1 - \frac{2\sqrt{\pi}\gamma_1^2}{\|G\|_m^2 - \sqrt{\pi}\gamma_0^2}}$ where γ_0, γ_1 and $\|G\|_m^2$ are given above. Evaluating this formula at $K = 0$ yields:

$$
\kappa_C(0) = \sqrt{1 - \frac{2\sqrt{\pi}\left(\frac{1}{4}\right)^2}{\frac{1}{4}\sqrt{\pi} - \sqrt{\pi}\left(\frac{1}{2\sqrt{\pi}}\right)^2}} = 0.516.
$$

The value of $\kappa_C(0)$ is clearly misleading since $G(X)$ is linear when $K = 0$.

To avoid spurious nonlinearity, one uses $m(x) = e^{-x}$. We have:

$$
\|G\|_m^2 = \int_K^\infty (x - K)^2 \exp(-x) dx = 2e^{-K}, \tag{43}
$$

$$
\delta_0 = \int_K^\infty (x - K)\exp(-x) dx = e^{-K} \text{ and} \tag{44}
$$

$$
\delta_1 = \int_K^\infty (1 - x)(x - K)\exp(-x) dx = -e^{-K}(K + 1), \tag{45}
$$

Hence the nonlinearity of the payoff of an European Call is given by:

$$
\kappa_C(K) = \sqrt{1 - \frac{(K + 1)^2}{2e^K - 1}}. \tag{46}
$$

We see that the formula above now implies that $\kappa_C(0) = 0$, consistently with the fact that $G(X) = X$ when $K = 0$.

I now consider the payoff a Put option with strike K, given by $G(X) = \max(K - X, 0)$. Having learned from the previous example, I set $m(x) = e^{-x}$ and obtain:

$$
\|G\|_m^2 = \int_0^K (K - x)^2 \exp(-x) dx = K^2 - 2e^{-K} - 2K + 2, \tag{47}
$$

$$
\delta_0 = \int_0^K (K - x)\exp(-x) dx = K + e^{-K} - 1 \text{ and} \tag{48}
$$

$$\delta_1 = \int_0^K (1-x)(K-x)\exp(-x)\,dx = 1 - Ke^{-K} - e^{-K}. \tag{49}$$

The nonlinearity of the payoff of an European Put is given by:

$$\kappa_P(K) = \sqrt{1 - \frac{\left(1 - Ke^{-K} - e^{-K}\right)^2}{1 - 2Ke^{-K} - e^{-2K}}}.$$

Please note that for a Put, $G(X) \to K - X$ when $K \to \infty$. Hence, we expect to see $\kappa_P(K) \to 0$ as $K \to \infty$. Indeed, we have:

$$\lim_{K \to \infty} \kappa_P(K) = \sqrt{1 - \frac{(1-0-0)^2}{1-0-0}} = 0.$$

Figure 4 compares the nonlinearity of a the payoffs of a Call and a Put under the metrics $m(x) = e^{-x}$. As K increases to infinity, $\kappa_P(K)$ starts at $\kappa_P(0) = 0$ and converges to 1 whereas $\kappa_C(K)$ at $\kappa_C(0) = 1$ and converges to zero $K \to \infty$.[5]
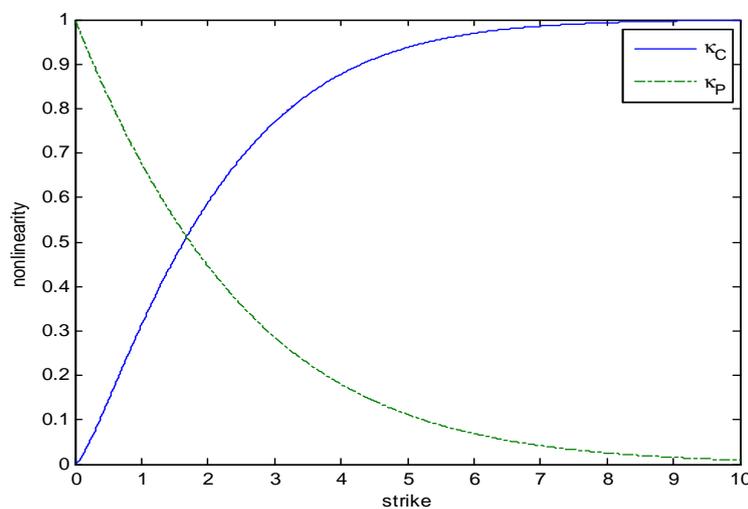
Figure 4. Comparing the Nonlinearity of a Call and Put with Same Strike.

Over the course of its lifetime, the price of a Call as given by the Black-Scholes formula is:

$$C(X,K,\sigma,t) = X\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2), \tag{50}$$

where:

$$d_1 = \frac{1}{\sigma\sqrt{T-t}}\left(\log(X/K) + \left(r + \frac{\sigma^2}{2}\right)(T-t)\right),$$

$$d_1 = \frac{1}{\sigma\sqrt{T-t}}\left(\log(X/K) + \left(r - \frac{\sigma^2}{2}\right)(T-t)\right),$$

X is the price of the underlying asset, K is the strike, t is the current date, T is the maturity and σ is the spot volatility of X.

The following quantities are needed in order to evaluate the nonlinearity of $C(X,K,\sigma,t)$ as K, t and σ vary:

---

[5] The crossing point where $\kappa_C(K) = \kappa_P(K)$ is approximately $K \simeq 1.67$.

$$\gamma_{C,0}(K,\sigma,t) = \int_0^\infty C(x,K,\sigma,t)\exp(-x)\,dx,$$

$$\gamma_{C,1}(K,\sigma,t) = \int_0^\infty (1-x)\,C(x,K,\sigma,t)\exp(-x)\,dx \text{ and}$$

$$\|C(.,K,\sigma,t)\|_m^2 = \int_0^\infty C(x,K,\sigma,t)^2\exp(-x)\,dx.$$

These integrals cannot be computed in closed form. A numerical approximation based on Gauss-Laguerre rule yields:

$$\gamma_{C,0}(K,\sigma,t) \simeq \sum_{l=1}^N \omega_l C(x_l,K,\sigma,t),$$

$$\gamma_{C,1}(K,\sigma,t) \simeq \sum_{l=1}^N \omega_l(1-x_l)C(x_l,K,\sigma,t) \text{ and}$$

$$\|C(.,K,\sigma,t)\|_m^2 \simeq \sum_{l=1}^N \omega_l C(x_l,K,\sigma,t)^2.$$

where $x_l, l = 1, ..., N$ are quadrature points associated with weights $\omega_l, l = 1, ..., N$. Hence,

$$\kappa_C(K,\sigma,t) = \sqrt{1 - \frac{\left(\sum_{l=1}^N \omega_l(1-x_l)C(x_l,K,\sigma,t)\right)^2}{\sum_{l=1}^N \omega_l C(x_l,K,\sigma,t)^2 - \left(\sum_{l=1}^N \omega_l C(x_l,K,\sigma,t)\right)^2}}. \tag{51}$$

For an European Put option, the price evolves according to:

$$P(X,K,\sigma,t) = Ke^{-r(T-t)}\Phi(-d_2) - X\Phi(-d_1). \tag{52}$$

The nonlinearity of this price with respect to the underlying asset is:

$$\kappa_P(K,\sigma,t) = \sqrt{1 - \frac{\left(\sum_{l=1}^N \omega_l(1-x_l)P(x_l,K,\sigma,t)\right)^2}{\sum_{l=1}^N \omega_l P(x_l,K,\sigma,t)^2 - \left(\sum_{l=1}^N \omega_l P(x_l,K,\sigma,t)\right)^2}}. \tag{53}$$

Figure 5 is drawn by assuming that $T = 1$ year, $K \in [0,5]$, $\sigma \in [0,80\%]$ and $r = 3\%$. For a Call option (resp., Put options), nonlinearity increases (resp., decreases) in the strike K for all values of volatility and time to maturity. For both types of options, the nonlinearity is decreasing in the volatility and time to maturity. The nonlinearity appears to be more sensitive in the volatility and time to maturity for a Call than for a Put.
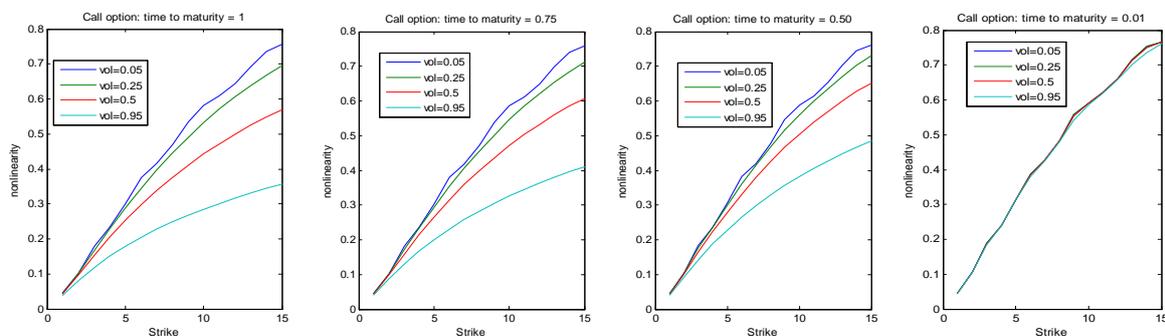

Figure 5. *Cont.*
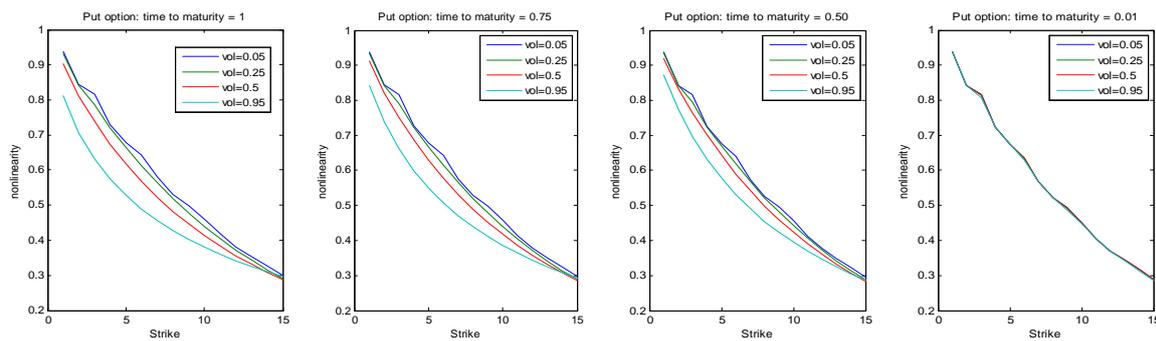
Figure 5. Nonlinearity of the Prices of Calls and Puts w.r.t the Underlying Asset.

8.2. Nonlinearity and Optimal Hedging

Assume that an investor wants to hold $-\beta_0$ units of a risk free asset (B_t) and $-\beta_1$ units of a risky asset (X_t) so as to hedge against the volatility of an asset Y_t. This investor would form a hedge portfolio whose value is $Y_t - \beta_0 B_{f,t} - \beta_1 X_t$. The hedging error is given by:

$$e_{t+1} = \omega_{y,t} y_{t+1} + \left(1 - \omega_{y,t} - \omega_{x,t}\right) b_{t+1} + \omega_{x,t} x_{t+1},$$

where

$$\omega_{y,t} = \frac{Y_t}{Y_t - \beta_0 B_t - \beta_1 X_t}; \omega_{x,t} = \frac{-\beta_1 X_t}{Y_t - \beta_0 B_t - \beta_1 X_t};$$

$$y_{t+1} = \frac{Y_{t+1} - Y_t}{Y_t}; x_{t+1} = \frac{X_{t+1} - X_t}{X_t} \text{ and } b_{t+1} = \frac{B_{t+1} - B_t}{B_t}$$

In practice, a perfect hedge that results in $e_{t+1} = 0$ can rarely be achieved. Therefore, the best solution often consists of minimizing the variance of the hedging error e_{t+1} with respect to the "hedge ratio" $\frac{\omega_{x,t}}{\omega_{y,t}} = k_t$, which is the number of units of Asset X_t to hold per unit of Asset Y_t. When Y_t is the payoff of an option and X_t the price of the underlying asset, k_t corresponds to the "delta" of the option (hence the expression "Delta Hedging").

The optimal hedging problem boils down to the following minimization:

$$\frac{Var_t\left(e_{t+1}\right)}{\omega_{y,t}^2} = Var_t\left(y_{t+1}\right) + 2k_t Cov\left(y_{t+1}, x_{t+1}\right) + k_t^2 Var_t\left(x_{t+1}\right).$$

The optimal hedge ratio is given by $k_t^* = -\frac{Cov_t(y_{t+1}, x_{t+1})}{Var_t(x_{t+1})}$. However, $Cov_t\left(y_{t+1}, x_{t+1}\right)$ can be zero if y_{t+1} is nonlinearly exposed to x_{t+1}. In this case, one would wrongly conclude that Asset X_t is of no help for hedging against the fluctuations of Y_t.

Now, suppose we have detected that $E\left(y_t | x_t\right) = g\left(x_t\right)$ is nonlinear using the methodology proposed in this paper. A reasonable hedging strategy would therefore consist of first using state-of-the-art models to predict x_{t+1} as $\overline{x}_{t+1}$ and next linearizing $g\left(x_t\right)$ around this prediction to obtain:

$$g\left(x_{t+1}\right) \simeq g\left(\overline{x}_{t+1}\right) + g'\left(\overline{x}_{t+1}\right)\left(x_{t+1} - \overline{x}_{t+1}\right).$$

An approximately optimal hedge ratio would then be given by:

$$k_t^* \simeq -\frac{Cov_t\left(g\left(\overline{x}_{t+1}\right) + g'\left(\overline{x}_{t+1}\right)\left(x_{t+1} - \overline{x}_{t+1}\right), x_{t+1}\right)}{Var_t\left(x_{t+1}\right)} = -g'\left(\overline{x}_{t+1}\right).$$

This example underscores the importance of being aware of the presence of nonlinearity in assets returns for a sound portfolio risk management.

8.3. Empirical Application: Return-Risk Trade-Off on the SP500

In this section, I illustrate an empirical use of the measure of nonlinearity κ by performing an analysis of the return-risk trade-off based on the Merton's (1973) intertemporal capital asset pricing model (ICAPM). This model posits the following relation between the conditional expected return on the market index $E_t(R_{t+1})$ and the conditional expected variance $Var_t(R_{t+1})$:

$$E_t(R_{t+1}) = R_{f,t+1} + \theta Var_t(R_{t+1}),$$

where $R_{f,t+1}$ is the risk free rate and θ is the relative risk aversion coefficient of the representative agent. A multivariate version of the ICAPM is proposed in Bollerslev et al. (1988).

Ghysels et al. (2005) employed a MIxed DAta Sampling (MIDAS) methodology to the CRSP value-weighted portfolio and concluded that "*there is a [positive] risk-return trade-off after all.*" Previous studies who found a positive relationship between risk premium and expected risk include French et al. (1987) and Campbell and Hentschel (1992), in contrast with Nelson (1991) who find a negative relationship or Glosten et al. (1993) and Harvey (2001) whose conclusions are mixed. Ghysels et al. (2005) attributes the conflicting conclusions to differences in the models posited for the conditional variance. Jacquier and Okou (2014) suspect the differences in data frequencies to be responsible of the inconsistencies across findings. The empirical results of the current paper suggest that the controversy on the nature of the return-risk trade-off is mainly due to nonlinearity.

I consider estimating the following nonlinear version of the ICAPM:

$$r_{t+1} = \mu + \theta E_t(v_{t+1}) + u_{t+1} \text{ and} \tag{54}$$

$$v_{t+1} = c_0 + c_1 v_t + \varepsilon_{t+1}. \tag{55}$$

If one assume that u_t and ε_t are IID and Gaussian, Equations (54) and (55) imply that:

$$E_t(R_{t+1}) = \exp\left(\mu + \frac{1}{2}Var(u_{t+1}) - \frac{\theta}{2}Var(\varepsilon_{t+1})\right)E_t(RV_{t+1})^\theta. \tag{56}$$

As in the ICAPM, the conditional expected return $E_t(R_{t+1})$ is increasing in conditional expected variance $E_t(RV_{t+1})$. Unlike in the ICAPM, the relationship between the conditional expected return and the innovation on the log-variance process (ε_{t+1}) is explicitly characterized. Namely, the conditional expected return is decreasing in the variance of ε_{t+1}, as observed empirically by French et al. (1987). Finally, the market price of risk is a nonlinear function of the conditional expected variance.

I estimate the AR(1) $v_t = c_0 + c_1 v_{t-1} + \varepsilon_t$ for the log-RV and compute the fitted values as $\widehat{v}_t = \widehat{c}_0 + \widehat{c}_1 v_{t-1}$. These fitted values are the expected risk at time t as perceived by investors at time $t-1$. The estimated coefficients are $\widehat{c}_0 = -1.8683$ (significant at 5% level) and $\widehat{c}_1 = 0.7209$ (significant at 5% level). The R^2 of this regression is 52% and the estimated error variance is 0.4193. Figure 5a shows the scatter plot of v_t against $\widehat{v}_t$. Next, I regress the log-return r_t on $\widehat{v}_t$ and a constant and compute the fitted values as $\widehat{r}_t = \widehat{\mu}_0 + \widehat{\mu}_1 \widehat{v}_t$. This yields $\widehat{\mu}_0 = -0.0056$ and $\widehat{\mu}_1 = -0.0016$. Neither of these coefficients is significant at 5% level and not surprisingly, the R^2 of the regression is less than 1%. The poor fit provided but the linear regression of r_t onto $\widehat{v}_t$ suggests that the relationship between these variable is nonlinear. This is confirmed by the nonparametric estimator of $E(\widehat{v}_t|r_t)$ shown by Figure 2.

Figure 6a,b respectively show the estimated nonlinearity of $E(r_t|\widehat{v}_t)$ and $E(\widehat{v}_t|r_t)$ on increasing segments of the support of $\widehat{v}_t$ and r_t. The scaling of the x-axis corresponds to the quantiles of the conditioning variable. For any pair (X, Y), the nonlinearity of $\widehat{G}(x)$ is computed on the segments $[\underline{x}, x_k]$ using the metrics $m(x) = 1$ along with Gauss-Legendre's polynomials, where:

$$
\begin{aligned}
x_k &= \underline{x} + k\left(\overline{x} - \underline{x}\right)/20, \, k = 0, ..., 20, \\
\underline{x} &= \min\left(x_1, ..., x_T\right) \text{ and} \\
\overline{x} &= \max\left(x_1, ..., x_T\right).
\end{aligned}
$$

For each x_k, $\widehat{G}\left(x_k\right)$ is obtained by spline interpolation based on $\left\{\left(x_t, \widehat{G}\left(x_t\right)\right)\right\}_{t=1}^{T}$.
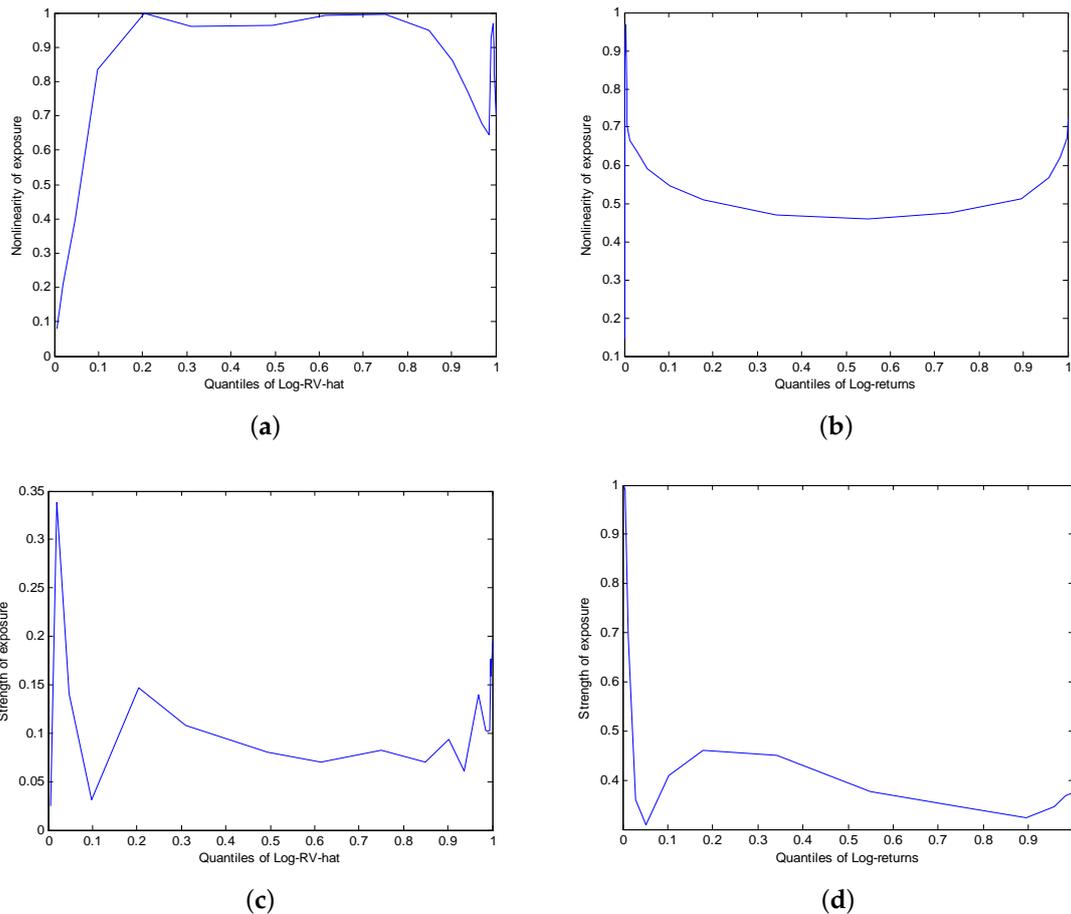
Figure 6. Nonlinearity and strength of exposure. (**a**) Nonlinearity of exposure of r_t to $\widehat{v}_t$; (**b**) Nonlinearity of exposure of $\widehat{v}_t$ to r_t; (**c**) Strength of exposure of r_t to $\widehat{v}_t$; (**d**) Strength of exposure of $\widehat{v}_t$ to r_t.

We see that the nonlinearity of $E\left(r_t|\widehat{v}_t\right)$ increases fast and remains high after the 20th percentile of $\widehat{v}_t$. By contrast, the nonlinearity of $E\left(\widehat{v}_t|r_t\right)$ decreased on the first portion of the support of r_t and increases steadily after the median. The point at which the nonlinearity of $E\left(\widehat{v}_t|r_t\right)$ is minimized, $r^* = -0.0075$, is a good candidate for the threshold of a piecewise linear model.

Figure 6c,d respectively show the strength of the exposure of r_t to $\widehat{v}_t$ and $\widehat{v}_t$ to r_t on increasing subsample of type $\{(x_t, y_t) : x_t \le x_k\}$. The correlation of r_t and $E\left(r_t|\widehat{v}_t\right)$ approximately equals 0.1 on a large portion of the support of $\widehat{v}_t$ while that of $\widehat{v}_t$ and $E\left(\widehat{v}_t|r_t\right)$ is on average equal to 0.4. This means that $E\left(\widehat{v}_t|r_t\right)$ fits $\widehat{v}_t$ better than $E\left(r_t|\widehat{v}_t\right)$ fits r_t.

Acting on the fact that the *linearity* of $E(\widehat{v}_t|r_t)$ reaches its *maximum* at $r^* = -0.0075$, I estimate the following piecewise linear regression:

$$r_t = \mu_0^{(1)} + \mu_1^{(1)} v_t + u_{1,t}, \, r_t \le r^* \text{ and}$$
$$r_t = \mu_0^{(2)} + \mu_1^{(2)} v_t + u_{2,1}, \, r_t > r^*.$$

Let $\widehat{r}_t$ denote the fitted values obtained from this piecewise linear regression.

Figure 7a shows the fitted regression lines while Figure 7b plots the linear correlation between r_t and $\widehat{r}_t$ against the quantiles of $\widehat{v}_t$. The correlation between r_t and $\widehat{r}_t$ reaches 50% at the 10th percentile and remains above 75% thereafter. This strong nonlinear relationship between log-returns and log-realized volatility is completely missed by the naive linear regression of r_t onto $\widehat{v}_t$.


(a)

(b)

(c)
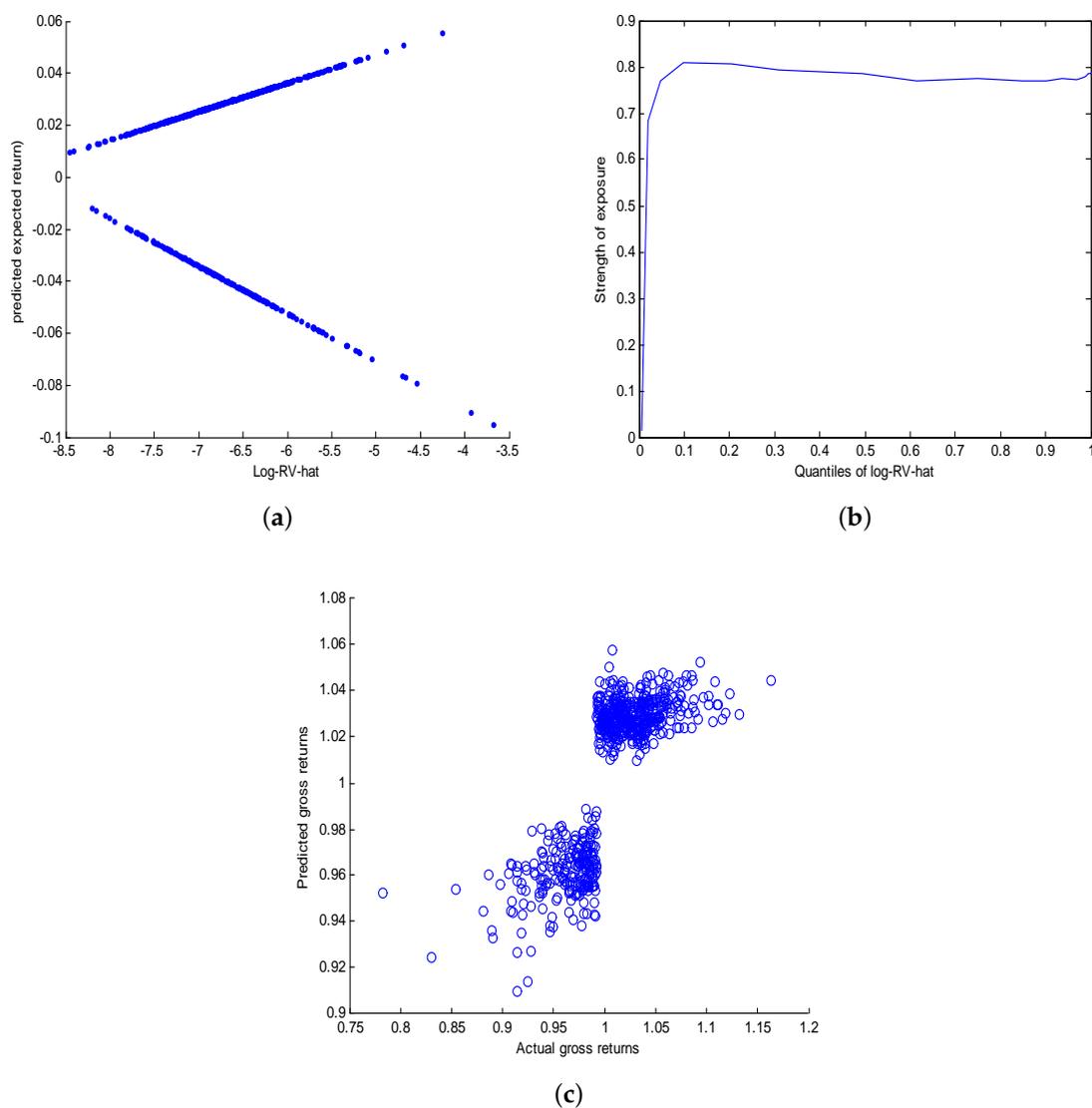
Figure 7. Piecewise linear regression of $\widehat{r}_t$ on $\widehat{v}_t$. (**a**) Fitted log-returns; (**b**) Strength of the fit; (**c**) Predicted gross-returns.

The estimated regression lines are:

$$\widehat{r}_t \;=\; -0.1622 - 0.0183\widehat{v}_t \text{ for } r_t \leq r^* \text{ and}$$
$$\widehat{r}_t \;=\; 0.1022 + 0.011\widehat{v}_t \text{ for } r_t > r^*.$$

The estimated error variances are respectively $\widehat{Var}\left(u_{1,t}\right) = 0.00062$ for the first regime and $\widehat{Var}\left(u_{2,t}\right) = 0.00089$ for the second regime. Finally, the estimated nonlinear return-risk trade-offs are:

$$E_t\left(R_{t+1}\right) \;=\; 1.1054 E_t\left(RV_{t+1}\right)^{0.011} \text{ if } r_t \leq r^* \text{ and}$$
$$E_t\left(R_{t+1}\right) \;=\; 0.8539 E_t\left(RV_{t+1}\right)^{-0.0183} \text{ if } r_t > r^*$$

Based on the equation, the unconditional correlation between the actual gross-returns and their predictions shown on Figure 7c is 0.7861.

A natural implication of these results is that option pricing models should at least account for the presence of regimes or parameter uncertainty in the distribution of the returns on the underlying asset. Our findings provide a strong empirical support for regime switching models in which volatility influences returns (as in Duan et al. 2002), heteroskedastic mixture models or Bayesian models that naturally account for parameter uncertainty (e.g., Rombouts and Stentoft 2014, 2015).

9. Conclusions

This paper proposes an approach to measure the degree of nonlinearity of the exposure of a variable Y to the movements of another variable X. The exposure is defined in terms of the expectation of Y conditional on X, denoted $G(X)$. The proposed measure of nonlinearity, denoted κ, is constructed by exploiting the ratio of the norms of the linear and nonlinear parts of $G(X)$. The separation of $G(X)$ into its linear and nonlinear parts is done via its projection onto an orthogonal basis of polynomials with respect to some metrics $m(x)$. The invariance properties of κ are studied. For cases where the exposure function $G(X)$ is unknown, an estimator $\widehat{\kappa}$ of κ that exploits a consistent nonparametric estimator $\widehat{G}(X)$ is proposed. It is shown that $\widehat{\kappa}$ inherits the consistency of $\widehat{G}(X)$.

Three fields of application are proposed. The first application concerns the measurement of the nonlinearity of the price of a European style option with respect to the underlying asset. I find that the nonlinearity of a Call increases with the strike but decreases with the volatility and time to maturity. For a Put, the nonlinearity decreased with the strike, volatility and time to maturity. The second application attempts to motivate the use of κ for portfolio risk management. Basically, I argue that hedge ratios are misleading in the presence of nonlinearity and therefore, that a diagnosis of nonlinearity should be performed prior to designing the optimal hedge strategy. The third application is empirical and concerns the relationship between the return and realized volatility of the SP500 index. The empirical results provide supportive evidence that the return-risk trade-off on the SP500 is nonlinear and governed by regimes.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Proofs of Propositions

Proof of Proposition 1. Let $Z = aY + b$ where a and b are constants. We have:

$$
\begin{aligned}
E\left(Z|X\right) &= aG(X) + b \\
&= a\sum_{j=0}^{\infty}\gamma_j P_j\left(X\right) + b \equiv \sum_{j=0}^{\infty}\delta_j P_j\left(X\right),
\end{aligned}
$$

where $\delta_0 = a\gamma_0 + b$ and $\delta_j = a\gamma_j$ for all $j \geq 1$. Hence:

$$\kappa_Z^2 = 1 - \frac{a^2 \gamma_1^2 \|P_1\|_m^2}{\|E(Z|X)\|_m^2 - (a\gamma_0 + b)^2 \|P_0\|_m^2}$$

$$= 1 - \frac{a^2 \gamma_1^2 \|P_1\|_m^2}{a^2 \sum_{j=1}^{\infty} \gamma_j^2 \|P_j\|_m^2} = 1 - \frac{\gamma_1^2 \|P_1\|_m^2}{\|G\|_m^2 - \gamma_0^2 \|P_0\|_m^2} = \kappa_Y^2.$$

This shows that $Z = aY + b$ and Y have the same amount of nonlinearity. $\square$

Proof of Proposition 2. Please note that $E(Y + \varepsilon|X) = E(Y|X) + E(\varepsilon|X) = G(X) + E(\varepsilon)$, which has the same amount of nonlinearity as $G(X)$ according to Proposition 1. $\square$

Proof of Proposition 3. $Z = aY + b$ implies $E(Y + Z|X) = \sum_{j=0}^{\infty} \delta_j P_j(X)$, where $\delta_0 = \gamma_0 + b$, $\delta_1 = \gamma_1 + a/2$ and $\delta_j = \gamma_j$ for all $j \geq 2$. Also, "Y is nonlinearly exposed to X" means that $\gamma_2 \neq 0$. We have,

$$\kappa_{Y+Z}^2 = 1 - \frac{\delta_1^2 \|P_1\|_m^2}{\left\|\sum_{j=0}^{\infty} \delta_j P_j\right\|_m^2 - \delta_0^2 \|P_0\|_m^2}$$

$$= 1 - \frac{(\gamma_1 + a/2)^2 \|P_1\|_m^2}{(\gamma_1 + a/2)^2 \|P_1\|_m^2 + \sum_{j=3}^{\infty} \gamma_j^2 \|P_j\|_m^2} = \kappa_{Y+Z}^2(a)$$

where we note that $\kappa_{Y+Z}^2(0) = \kappa_Y^2$. Taking the derivative of $\kappa_{Y+Z}^2(a)$ with respect to a yields:

$$\frac{\partial \kappa_{Y+Z}^2(a)}{\partial a} = \frac{-(\gamma_1 + a/2) \|P_1\|_m^2 \sum_{j=3}^{\infty} \gamma_j^2 \|P_j\|_m^2}{\left((\gamma_1 + a/2)^2 \|P_1\|_m^2 + \sum_{j=3}^{\infty} \gamma_j^2 \|P_j\|_m^2\right)^2}$$

Hence, $\kappa_{Y+Z}^2(a)$ is increasing in the region $a < -2\gamma_1$, and decreasing in the region $a > -2\gamma_1$. Furthermore, $\kappa_{Y+Z}^2(a)$ is symmetric around $a = -2\gamma_1$.[6] Consequently:

$$\begin{aligned} \kappa_{Y+Z}^2(0) &= \kappa_{Y+Z}^2(-4\gamma_1) = \kappa_Y^2, \\ \kappa_{Y+Z}^2(a) &> \kappa_Y^2 \text{ if } a \in [lb, ub] \text{ and.} \\ \kappa_{Y+Z}^2(a) &< \kappa_Y^2 \text{ if } a \in (-\infty, lb] \cup [ub, +\infty). \end{aligned}$$

where $lb = \min(0, -4\gamma_1)$ and $ub = \max(0, -4\gamma_1)$. $\square$

Proof of Proposition 4. First, note that:

$$G(X) = G\left(\frac{Z - b}{a}\right) = \sum_{j=0}^{\infty} \gamma_j P_j\left(\frac{Z - b}{a}\right) \equiv E(Y|Z).$$

Next, observe that $\{P_j(X)\}_{j=0}^{\infty}$ is an orthogonal basis under $m(x)$. That is, for $i \neq j$:

$$\int P_i(x) P_j(x) m(x) dx = 0 \Leftrightarrow$$

$$\int P_i\left(\frac{z - b}{a}\right) P_j\left(\frac{z - b}{a}\right) \frac{1}{a} m\left(\frac{z - b}{a}\right) dz = 0.$$

---

[6] This can be seen by solving $\kappa_{Y+Z}^2(a) = \bar{\kappa}$ for given $\bar{\kappa}$. When $\bar{\kappa} = \kappa_Y^2$, the solutions are $a = 0$ and $a = 0 - 4\gamma_1$.

This shows that $\left\{P_j\left(\frac{Z-b}{a}\right)\right\}_{j=0}^{\infty}$ is an orthogonal basis under $\widetilde{m}(z) = \frac{1}{a}m(\frac{z-b}{a})$. The coordinates of $E(Y|Z)$ in this basis is:

$$\frac{\left\langle E(Y|Z), P_0\left(\frac{Z-b}{a}\right)\right\rangle_{\widetilde{m}}}{\left\langle P_0\left(\frac{Z-b}{a}\right), P_0\left(\frac{Z-b}{a}\right)\right\rangle} = \frac{\left\langle \sum_{j=0}^{\infty}\gamma_j P_j\left(\frac{Z-b}{a}\right), P_0\left(\frac{Z-b}{a}\right)\right\rangle_{\widetilde{m}}}{\left\langle P_0\left(\frac{Z-b}{a}\right), P_0\left(\frac{Z-b}{a}\right)\right\rangle_{\widetilde{m}}} = \gamma_0 \text{ and}$$

$$\frac{\left\langle E(Y|Z), P_0\left(\frac{Z-b}{a}\right)\right\rangle_{\widetilde{m}}}{\left\langle P_1\left(\frac{Z-b}{a}\right), P_1\left(\frac{Z-b}{a}\right)\right\rangle_{\widetilde{m}}} = \frac{\left\langle \sum_{j=0}^{\infty}\gamma_j P_j\left(\frac{Z-b}{a}\right), P_1\left(\frac{Z-b}{a}\right)\right\rangle}{\left\langle P_1\left(\frac{Z-b}{a}\right), P_1\left(\frac{Z-b}{a}\right)\right\rangle_{\widetilde{m}}} = \gamma_1$$

Furthermore, $\|E(Y|Z = .)\|_{\widetilde{m}}^2 = \|G\|_m^2$ and hence:

$$\sqrt{1 - \frac{\gamma_1^2\|P_1\|_m^2}{\|G\|_m^2 - \gamma_0^2\|P_0(X)\|_m^2}} = \sqrt{1 - \frac{\gamma_1^2\left\|P_1\left(\frac{Z-b}{a}\right)\right\|_{\widetilde{m}}^2}{\|E(Y|Z = .)\|_{\widetilde{m}}^2 - \gamma_0^2\left\|P_0\left(\frac{Z-b}{a}\right)\right\|_{\widetilde{m}}^2}}.$$

The L.H.S. is the degree of nonlinearity of Y with respect to X under $m(x)$ while the R.H.S. is its degree of nonlinearity with respect to Z under $\widetilde{m}(z) = \frac{1}{a}m(\frac{z-b}{a})$. $\square$

Proof of Proposition 5. Please note that $\widehat{G}(x) = \sum_{j=0}^{\infty}\widehat{\gamma}_j P_j(x)$ where $\widehat{\gamma}_j = \frac{\langle \widehat{G}, P_j\rangle}{\|P_j\|_m^2}$ for all j. Hence:

$$\widehat{\kappa} = \sqrt{1 - \frac{\widehat{\gamma}_1^2\|P_1\|_m^2}{\|\widehat{G}\|_m^2 - \widehat{\gamma}_0^2\|P_0\|_m^2}} = \sqrt{1 - \frac{\widehat{\alpha}_1^2}{1 - \widehat{\alpha}_0^2}} \equiv \kappa\left(\widehat{\alpha}_0^2, \widehat{\alpha}_1^2\right),$$

where $\widehat{\alpha}_j^2 \equiv \frac{\widehat{\gamma}_j^2\|P_j\|_m^2}{\|\widehat{G}\|_m^2} = \frac{\widehat{\gamma}_j^2\|P_j\|_m^2}{\sum_{j=0}^{\infty}\widehat{\gamma}_j^2\|P_j\|_m^2} \in [0,1]$ for all j. Let $\alpha_j^2 \equiv \frac{\gamma_j^2\|P_j\|_m^2}{\|G\|_m^2}$ so that $\kappa \equiv \kappa\left(\alpha_0^2, \alpha_1^2\right)$. We have:

$$\widehat{\alpha}_j^2 - \alpha_j^2 = \alpha_j^2\left(\left(\frac{\|\widehat{G}\|_m^2}{\|G\|_m^2}\right)^{-1}\left(\frac{\widehat{\gamma}_j^2}{\gamma_j^2}\right) - 1\right)$$

Now, note that $\|\widehat{G}\|_m^2 = \|\widehat{G} - G\|_m^2 + 2\langle \widehat{G} - G, G\rangle + \|G\|_m^2$. Hence

$$\frac{\|\widehat{G}\|_m^2}{\|G\|_m^2} - 1 = \frac{\|\widehat{G} - G\|_m^2}{\|G\|_m^2} + 2\frac{\langle \widehat{G} - G, G\rangle}{\|G\|_m^2}$$

The dominant term of this expression is $\frac{\langle \widehat{G} - G, G\rangle}{\|G\|_m^2}$ and it satisfies:

$$\frac{|\langle \widehat{G} - G, G\rangle|}{\|G\|_m^2} \leq \frac{\|\widehat{G} - G\|_m}{\|G\|_m} = O_p\left(h_T^{-\beta/2}T^{-1/2}\right)$$

This shows that $\frac{\|\widehat{G}\|_m^2}{\|G\|_m^2} = 1 + O_p\left(h_T^{-\beta/2}T^{-1/2}\right)$. The same rate of convergence is obtained for $\left(\frac{\|\widehat{G}\|_m^2}{\|G\|_m^2}\right)^{-1}$ by using Slutsky's Theorem. For the term $\widehat{\gamma}_j^2/\gamma_j^2$, we have:

$$\frac{\widehat{\gamma}_j}{\gamma_j} = \frac{1}{\gamma_j} \frac{\langle \widehat{G}, P_j \rangle_m}{\|P_j\|_m^2} = \frac{1}{\gamma_j} \frac{\langle \widehat{G} - G, P_j \rangle_m}{\|P_j\|_m^2} + \frac{1}{\gamma_j} \frac{\langle G, P_j \rangle}{\|P_j\|_m^2}$$

$$\leq \frac{1}{\gamma_j \|P_j\|_m} \|\widehat{G} - G\|_m + 1 = 1 + O_p \left(h_T^{-\beta/2} T^{-1/2} \right)$$

so that

$$\widehat{\alpha}_j^2 - \alpha_j^2 = \left(1 + O_p \left(h_T^{-\beta/2} T^{-1/2} \right) \right) \left(1 + O_p \left(h_T^{-\beta/2} T^{-1/2} \right) \right) - 1 = O_p \left(h_T^{-\beta/2} T^{-1/2} \right).$$

Finally, $\widehat{\kappa} - \kappa = \kappa \left(\widehat{\alpha}_0^2, \widehat{\alpha}_1^2 \right) - \kappa \left(\alpha_0^2, \alpha_1^2 \right) = O_p \left(h_T^{-\beta/2} T^{-1/2} \right)$ by Slutsky's Theorem. □

References

Askey, Richard. 2005. The 1839 paper on permutations: Its relation to the Rodrigues formula and further developments. In *Mathematics and Social Utopias in France: Olinde Rodrigues and His Times.* Edited by Simon Altmann and Eduardo. L. Ortiz. History and Mathematics 28. Providence: American Mathematical Society, pp. 105–18.

Bandi, Federico M., and Benoit Perron. 2008. Long-Run Risk-Returns Trade-Offs. *Journal of Econometrics* 143: 349–74. [CrossRef]

Bierens, Herman J. 1987. Kernel estimators of regression functions. In *Advances in Econometrics 5th World Congress*. Edited by Truman F. Bewley. Cambridge: Cambridge University Press, pp. 99–144.

Bollerslev, Tim, Robert F. Engle, and Jeffrey M. Wooldridge. 1988. A Capital Asset Pricing Model with Time-Varying Covariances. *Journal of Political Economy* 96: 116–31. [CrossRef]

Bontemps, Christian, and Nour Meddahi. 2012. Testing Distributional Assumptions: A GMM Approach. *Journal of Applied Econometrics* 27: 978–1012. [CrossRef]

Campbell, John. Y., and Ludger Hentschel. 1992. No news is goodnews: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics* 31: 281–318. [CrossRef]

Carrasco, Marine, Jean-Pierre Florens, and Eric Renault. 2007. Linear inverse problems and structural econometrics: Estimation based on spectral decomposition and regularization. *Handbook of Econometrics* 6: 5633–751.

Christoffersen, Peter, Steve Heston, and Kris Jacobs 2006. Option valuation with conditional skewness. *Journal of Econometrics* 131: 253–84.

Duan, Jin-Chuan, Ivilina Popova, and Peter Ritchken. 2002. Option pricing under regime switching. *Quantitative Finance* 2: 1–17. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 1996. Multifactor Explanations of Asset Pricing Anomalies. *The Journal of Finance* 51: 55–84. [CrossRef]

Feunou, Bruno, and Romeo Tedongap. 2012. A Stochastic Volatility Model with Conditional Skewness. *The Journal of Business and Economic Statistics* 30: 576–91. [CrossRef]

French, Kenneth R., G. William Schwert, and Robert F. Stambaugh. 1987. Expected stock returns and volatility. *Journal of Financial Economics* 19: 3–29. [CrossRef]

Gabaix, Xavier. 2016. Power Laws in Economics: An Introduction. *Journal of Economic Perspectives* 30: 185–206. [CrossRef]

Gabaix, Xavier. 2011. The Granular Origins of Aggregate Fluctuations. *Econometrica* 79: 733–72.

Gabaix, Xavier. 2009. Power Laws in Economics and Finance. *Annual Review of Economics* 1: 255–93. [CrossRef]

Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov. 2005. There is a risk-return trade-off after all. *Journal of Financial Economics* 76: 509–48. [CrossRef]

Ghysels, Eric, Pierre Guérin, and Massimiliano Marcellino. 2014. Regime switches in the risk–return trade-off. *Journal of Empirical Finance* 28: 118–38. [CrossRef]

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48: 1779–801. [CrossRef]

Goodman, Leo A., and William H. Kruskal. 1954. Measures of Association for Cross Classifications. *Journal of the American Statistical Association* 49: 732–64.

Goodman, Leo A., and William H. Kruskal. 1959. Measures of Association for Cross Classifications. II: Further Discussion and References. *Journal of the American Statistical Association* 54: 123–63. [CrossRef]

Harvey, Campbell R. 2001. The specification of conditional expectations. *Journal of Empirical Finance* 8: 573–637. [CrossRef]

Harvey, Campbell R., and Akhtar Siddique. 1999. Autoregressive Conditional Skewness. *The Journal of Financial and Quantitative Analysis* 34: 465–87. [CrossRef]

Harvey, Campbell R., and Akhtar Siddique. 2000. Conditional Skewness in Asset Pricing Tests. *The Journal of Finance* 55: 1263–95. [CrossRef]

Holmes, Peter. 2001. Correlation: From Picture to Formula. *Teaching Statistics* 23: 67–71. [CrossRef]

Jacquier, Eric, and Cédric Okou. 2014. Disentangling Continuous Volatility from Jumps in Long-Run Risk–Return Relationships. *Journal of Financial Econometrics* 12: 544–83. [CrossRef]

Johnson, Norman L., Samuel Kotz, and Narayanaswam Balakrishnan. 1994. *Continuous Univariate Distributions*. Chichester: Wiley.

Kendall, Maurice G. 1938. A New Measure of Rank Correlation. *Biometrika* 30: 81–89. [CrossRef]

Kendall, Maurice G. 1970. *Rank Correlation Methods*, 4th ed. London: Griffin. ISBN 978-0-852-6419-96.

Merton, Robert C. 1973. An intertemporal capital asset pricing model. *Econometrica* 41: 867–87. [CrossRef]

Merton, Robert C. 1980. On Estimating the Expected Return on the Market: An Exploratory Investigation. *Journal of Financial Economics* 8: 323–61. [CrossRef]

Nadaraya, E. A. 1964. On Estimating Regression. *Theory of Probability and its Applications* 9: 141–42. [CrossRef]

Nelson, Daniel B. 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59: 347–70. [CrossRef]

Okou, Cédric, and Eric Jacquier. 2016. Horizon effect in the term structure of long-run risk-return trade-offs. *Computational Statistics and Data Analysis* 100: 445–66. [CrossRef]

Rombouts, Jeroen V. K., and Lars Stentoft. 2014. Bayesian option pricing using mixed normal heteroskedasticity models. *Computational Statistics and Data Analysis* 76: 588–605. [CrossRef]

Rombouts, Jeroen V. K., and Lars Stentoft. 2015. Option pricing with asymmetric heteroskedastic normal mixture models. *Computational Statistics and Data Analysis* 31: 635–50. [CrossRef]

Sharpe, William F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–42.

Spearman, Charles. 1904. The proof and measurement of association between two things. *American Journal of Psychology* 15: 72–101. [CrossRef]

Watson, Geoffrey S. 1964. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics Series A* 26: 359–72.