UNIVERSITY OF CAMBRIDGE

Faculty of Economics

Cambridge Working Papers in Economics

Cambridge Working Papers in Economics: 1880

THE LOWER REGRESSION FUNCTION AND TESTING EXPECTATION DEPENDENCE DOMINANCE HYPOTHESES

Oliver Linton Yoon Jae Whang Yumin Yen

23 October 2018

We provide an estimator of the lower regression function and provide large sample properties for inference. We also propose a test of the hypothesis of positive expectation dependence and derive its limiting distribution under the null hypothesis and provide consistent critical values. We apply our methodology to several empirical questions.

The Lower Regression Function and Testing Expectation Dependence Dominance Hypotheses

Oliver Linton* — University of Cambridge
Yoon Jae Whang — Seoul National University
Yumin Yen — National College of Commerce

October 23, 2018

Abstract

We provide an estimator of the lower regression function and provide large sample properties for inference. We also propose a test of the hypothesis of positive expectation dependence and derive its limiting distribution under the null hypothesis and provide consistent critical values. We apply our methodology to several empirical questions.

1 Introduction

Suppose that $Y_t \in \mathbb{R}$, $X_t \in \mathbb{R}^d$ be a stationary mixing vector process. We suppose throughout that $E(|Y_t|^r) < \infty$ for some $r > 1$. Let F denote the c.d.f. of X_t. The lower regression function is defined as follows

$$m_<(x) = E\left(Y_t \,|X_t \leq x\right) = \frac{E\left[Y_t 1(X_t \leq x)\right]}{E\left[1(X_t \leq x)\right]} \equiv \frac{R(x)}{F(x)}, \tag{1}$$

for each $x \in \mathbb{R}^d$. Note that as $x \to (\infty, \ldots, \infty)$, $m_<(x) \to E(Y_t)$, the unconditional expectation of Y_t. In the special case that $X_t = Y_t$ and x is the lower α quantile of Y_t, then (1) is known as the expected shortfall. The hypothesis of negative (positive) expectation dependence of a random variable $Y_t \in \mathbb{R}$ on a random variable $X_t \in \mathbb{R}^d$ is that

$$m_<(x) - E\left(Y_t\right) = E\left(Y_t \,|X_t \leq x\right) - E\left(Y_t\right) \geq (\leq) 0 \tag{2}$$

*Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD. Email: obl20@cam.ac.uk.

for all scalar x, denoted $NED(Y|X)$ $(PED(Y|X))$. If the inequality (2) is strict for a set of x with positive probability, we say the expectation dependence is strict.

In this paper we define estimators of $m_<(x)$ and tests of the hypothesis NED and related ones. We obtain their limiting distribution under weak dependence conditions on the sample data, and provide consistent inference tools under general conditions. We apply our procedures to several applications and report a small simulation study.

2 The Lower Regression Function

In this section we provide more discussion about the lower regression function, its properties, and its uses. If Y_t, X_t possess a joint Lebesgue density $f_{Y,X}$, we can write $m_<(x)$ as

$$m_<(x) = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{x} y f_{Y,X}(y, x') dy dx'}{\int_{-\infty}^{\infty} \int_{-\infty}^{x} f_{Y,X}(y, x') dy dx'} = \frac{R(x)}{F(x)}.$$

Note that in this case, $R(x)$ and $F(x)$ are smooth functions, and the ordinary regression function $m(x) = E\left(Y_t \left| X_t = x\right.\right)$ satisfies

$$m(x) = \frac{\nabla R(x)}{\nabla F(x)}, \quad \nabla g(x) = \frac{\partial^d}{\partial x_1 \cdots \partial x_d} g(x). \tag{3}$$

In the scalar case $m(x) = R'(x)/F'(x) = R'(x)/f(x)$, where f is the marginal density of the covariate. Likewise, we can write for all x,

$$m_<(x) = \frac{\int_{-\infty}^{x} m(x') f(x') dx'}{\int_{-\infty}^{x} f(x') dx'}.$$

The lower regression function $\{m_<(x), x \in \mathbb{R}^d\}$ contains essentially the same information as the regression function $\{m(x), x \in \mathbb{R}^d\}$. Escanciano and Hualde (2009) consider the integrated regression function (assume for simplicity that y is centered)

$$IRF(x) = E\left[Y_t 1(X_t \le x)\right] = \int_{-\infty}^{x} m(x') f(x') dx' = \text{cov}(Y_t, 1(X_t \le x)),$$

which is also in one to one relation with the regression function. This has been exploited by Delgado (1992) to provide a test of the equality of two regression functions. Note that $IRF(x)$ does not have the same regression interpretation as $m_<(x)$.

Finally, we can interpret $m_<(x)$ as the minimizer of the no intercept best linear fitting (prediction) of Y_t by the indicator function $1(X_t \le x)$, i.e.,

$$E\left[\left\{Y_t - \beta 1(X_t \le x)\right\}^2\right]$$

with respect to β. This draws out its relation to the regression tree literature that uses multiple indicator functions for regression and classification, Breiman, Friedman, Olshen, and Stone (1984).

2.1 Expectation Dominance and Related Hypotheses

In this section we discuss further the hypothesis (2) and its uses in economics and finance.

Wright (1987) consider the classical portfolio choice problem with two risky assets X_t, Y_t and risk averse preferences expressed through a utility function U. That is, maximize $EU\left(\lambda Y_t + (1-\lambda)X_t\right)$ with respect to $\lambda \in [0,1]$. In the case where $EY_t \geq EX_t$, the sufficient condition that $\lambda^* > 0$ is that the random variable Y is relatively negative expectation dependent on X, denoted $RNED(Y|X)$,

$$E(Y_t - X_t|X_t \leq x) \geq E(Y_t - X_t)$$

on a set x that occurs with probability one, which is equivalent to $NED(Y-X|X)$, that is, Y_t is more negative expectation dependent on X_t than X_t is on itself, $E(Y_t|X_t \leq x) - EY_t \geq E(X_t|X_t \leq x) - E(X_t)$. In the case that $E(X_t) = E(Y_t)$ he shows (Theorem 4.2) that the necessary and sufficient conditions for the optimal λ^* to lie strictly between zero and one (diversification) are that: $E(Y_t - X_t|X_t \leq x) \geq 0$ and $E(X_t - Y_t|Y_t \leq y) \geq 0$ for all x, y and with strict inequality on a set of (x, y) that occurs with probability one. These conditions allow $\mathrm{cov}(X_t, Y_t) > 0$. In the appendix to his paper he considers the case with d assets, lets say $Y_{1t}, \ldots, Y_{dt}$. In this case, he says (p120) that a sufficient condition for all assets with the same mean to be chosen with a positive weight is that

$$E\left(Y_{it} - \theta_{i0}^{\intercal} X_{it}|\theta_{i0}^{\intercal} X_{it} \leq u\right) \geq 0 \tag{4}$$

for all $i = 1, \ldots, d$, where $X_{it} = \{Y_{jt}, j \neq i\}$, and all $u \in \mathbb{R}$. Here, $\theta_{i0} \in \mathbb{R}^{d-1}$ are the optimally chosen non-negative portfolio weights on the other assets X_{it}. A sufficient condition for this to hold is the multivariate dominance condition

$$E\left(Y_{it}|X_{it} \leq x\right) \geq E(Y_{jt}|X_{it} \leq x) \tag{5}$$

for all $j \neq i$ and for all $x \in \mathbb{R}^{d-1}$. Each such condition may be seen as a version of expectation dominance except that the conditioning variable is multivariate.

Li (2011) considers the demand for a risky asset in the presence of a background risk, such as health, and a risk free asset. He shows that the condition (2) determines whether there is a positive demand for the risky asset, depending also on the cross partial of utility between terminal wealth and the background risk. Zhu, Guo, Lin, and Zhu (2014) provide a test of this hypothesis, and of the higher order dominances that were defined in Li (2011). They assumed iid sampling.

Levy and Paroush (1974) considers the bivariate choice problem where utility U is defined over a pair of outcomes, lets say (y, x). They give necessary and sufficient conditions for dominance of one bivariate outcome over another. In the case of "envy", (LP, p141), i.e., $\partial U(y,x)/\partial y \partial x \geq 0$, the necessary and sufficient conditions (using our notation) for $EU(Y_{it}, X_{it}) \geq EU(Y_{jt}, X_{jt})$ are that: (1) $\Pr(Y_{it} \leq y) \leq \Pr(Y_{jt} \leq y)$ for all y and $\Pr(X_{it} \leq x) \leq \Pr(X_{jt} \leq x)$ for all x, and (2) for all x, y :

$$\Pr(Y_{it} \leq y, X_{it} \leq x) - \Pr(Y_{it} \leq y)\Pr(X_{it} \leq x) \geq \Pr(Y_{jt} \leq y, X_{jt} \leq x) - \Pr(Y_{jt} \leq y)\Pr(X_{jt} \leq x), \tag{6}$$

that is, the pair (Y_{it}, X_{it}) are more positively dependent according to the "concordance" measure than (Y_{jt}, X_{jt}). If the utility function is separable, then the cross partial of U is zero, and only condition (1) is needed for bivariate dominance; that is, just ordinary first order stochastic dominance. Note that if the marginal distributions are identical, then this condition $\partial U(y,x)/\partial y \partial x \geq 0$ is sufficient for bivariate dominance. The interpretation of the condition is called correlation aversion. Suppose that we replace condition (6) by the conditional notion of dependence, whereby for all x, y :

$$\frac{\Pr(Y_{it} \leq y, X_{it} \leq x)}{\Pr(X_{it} \leq x)} - \Pr(Y_{it} \leq y) \geq \frac{\Pr(Y_{jt} \leq y, X_{jt} \leq x)}{\Pr(X_{jt} \leq x)} - \Pr(Y_{jt} \leq y). \tag{7}$$

If X_{it} and X_{jt} have the same marginal distributions, then (6) and (7) are equivalent. We may write the ratio of probabilities as a lower regression if we replace Y_{it} by $1(Y_{it} \leq y)$ in (1). If also Y_{it} and Y_{jt} have the same marginal distributions, then this is of the form (8) for the given y. If it holds for all y, then the LP condition is satisfied, but if it only holds for a given y, then the weaker condition of Chiu (2014) is satisfied. Chiu (2014) says that the bivariate distribution $\Pr(Y_{it} \leq y, X_{it} \leq x)$ is a weak correlation increase of $\Pr(Y_{jt} \leq y, X_{jt} \leq x)$ if:

$$E(Y_{it} \,|X_{it} \leq x) - E(Y_{jt} \,|X_{jt} \leq x) \leq 0 \tag{8}$$

for all $x \in \mathcal{X}$, (and $EY_{it} = EY_{jt}$ and $\Pr(X_{it} \leq x) = \Pr(X_{jt} \leq x)$ for all x). This is the condition of more positive expectation.

3 Estimation

We consider explicitly the multisample scalar case, that is, $Y_{jt} \in \mathbb{R}$ and $X_{jt} \in \mathbb{R}$, where $j = 1, \ldots, J$. Let F_j denote the c.d.f. of X_{jt}, and let $m_{<j}(x) = E(Y_{jt} \mid X_{jt} \leq x)$ be the lower regression function for each $x \in \mathbb{R}$ and each $j = 1, \ldots, J$.

Suppose that we have a sample $\{(Y_{jt}, X_{jt}) \in \mathbb{R} \times \mathbb{R} : j = 1, \ldots, J; t = 1, \ldots, T\}$. We consider the following estimator:

$$\hat{m}_{j<}(x) = \frac{\frac{1}{T}\sum_{t=1}^{T} Y_{jt} 1(X_{jt} \leq x)}{\frac{1}{T}\sum_{t=1}^{T} 1(X_{jt} \leq x)} = \frac{\widehat{R}_j(x)}{\widehat{F}_j(x)} \tag{9}$$

for $j = 1, \ldots, J$ and $x \in \mathbb{R}$. Note that there is no bandwidth, and that the function $\widehat{m}_<(x)$ is a step function with jumps at the sample points. The estimator $\widehat{m}_<(x)$ is a ratio of unbiased estimators but is itself biased. In the iid case $E[Y_{:(i)}] = E\left[E\left(Y|X = X_{(r)}\right)\right] = \int m(x)dF_{(r)}(x)$, where $F_{(r)}$ is the c.d.f of the r^{th} order statistic of X, Yang (1977), which allows an exact expression for $E\widehat{m}_{<}(x)$ in that case.

Let $X_{j;(1)} \leq X_{j;(2)} \leq \ldots \leq X_{j;(T)}$ denote the order statistics of the covariate j and let $Y_{j:(i)}$, $i = 1, \ldots, T$ denote the corresponding concomitants. Then it appears that (9) is apparently defined only for $x \in [X_{j;(1)}, \infty)$. We rewrite $\widehat{m}_{j<}(x)$ in the following way and complete the 0/0 issue

$$\widehat{m}_{j<}(x) = \begin{cases} Y_{j:(1)} & \text{for } x \leq X_{j;(1)} \\ \frac{1}{k}\sum_{i=1}^{k} Y_{j:(i)} & \text{for } x \in (X_{j;(k-1)}, X_{j;(k)}] \\ \overline{Y}_j & \text{for } x \geq X_{j;(T)}. \end{cases} \tag{10}$$

This shows that the estimator is well defined throughout the whole real line.

Define the smoothed versions of $\widehat{R}_j$ and $\widehat{F}_j$ as follows:

$$\begin{aligned} \widetilde{R}_{j;h}(x) &= (\widehat{R}_j * K_h)(x) = \int \widehat{R}_j(x')K_h(x - x')dx' \\ \widetilde{F}_{j;h}(x) &= (\widehat{F}_j * K_h)(x) = \int \widehat{F}_j(x')K_h(x - x')dx', \end{aligned}$$

where $K_h(.) = K(./h)/h^d$ and K is a kernel function and h a bandwidth, and let $\widetilde{m}_{j<}(x) = \widetilde{R}_{j;h}(x)/\widetilde{F}_{j;h}(x)$. Then we may show that the ordinary Nadaraya-Watson regression smoother is

$$\widehat{m}_{j;NW}(x) = \frac{\nabla\widetilde{R}_{j;h}(x)}{\nabla\widetilde{F}_{j;h}(x)}.$$

Scaillet (2004,2005) considers smoothed estimators of expected shortfall and conditional expected shortfall.

3.1 Test Statistic

We consider the following general class of hypothesis, $H_0 : \tau \leq 0$, where:

$$\tau = \min_{i \neq j} \sup_{x \in \mathcal{X}} d_{ij}(x), \tag{11}$$

where $d_{ij}(x)$ is a distance measure, for example: $d_{ij}(x) = E\left(Y_{it}\,|X_{it} \leq x\right) - E\left(Y_{jt}\,|X_{jt} \leq x\right)$, $d_{ij}(x) = E(Y_t - X_t) - E(Y_t - X_t|X_t \leq x)$ (also called $d^{RNED}(x)$), or $d_{ij}(x) = E\left(Y_t\,|X_t \leq x\right) - E\left(Y_t\right)$ (also called $d^{NED}(x)$).

Let $\widehat{d}_{ij}(x)$ be the empirical version of $d_{ij}(x)$ for example $\widehat{d}_{ij}(x) = \widehat{m}_{i<}(x) - \widehat{m}_{j<}(x)$. Then let

$$\widehat{\tau}_T = \min_{i \neq j} \sup_{x \in \mathcal{X}} \widehat{d}_{ij}(x). \tag{12}$$

4 Large Sample Properties

We first establish the limit distribution of $\hat{\mathbf{m}}_<(\cdot) = (\hat{m}_{<1}(\cdot), \ldots, \hat{m}_{<J}(\cdot))^\intercal$. Let $\mathcal{X}$ be a compact subset of the union of the supports of $(X_{jt})_{j=1}^J$ such that $\inf_{x\in\mathcal{X}} F_j(x) > 0$ for each j. Define the empirical process $\boldsymbol{\nu}_T(x) = (\nu_{1T}(x), \ldots, \nu_{JT}(x))^\intercal$ for $x \in \mathcal{X}$ to be

$$\mathbf{v}_T(x) = \Delta(x)\Lambda_T(x),$$

where

$$\begin{aligned}
\boldsymbol{\Delta}(x) &= \left(\Delta_1(x)^\intercal, \ldots, \Delta_J(x)^\intercal\right), \ \boldsymbol{\Lambda}_T(x) = (\lambda_{1T}(x), \ldots, \lambda_{JT}(x))^\intercal, \\
\Delta_j(x) &= \left[\frac{1}{F_j(x)}, \ \ -\frac{R_j(x)}{F_j(x)^2}\ \right]^\intercal \\
\lambda_{jT}(x) &= \left[\sqrt{T}\left(\hat{R}_j(x) - R_j(x)\right), \sqrt{T}\left(\hat{F}_j(x) - F_j(x)\right)\right]^\intercal.
\end{aligned}$$

Let $\boldsymbol{\Lambda}(x) = (\lambda_1(x), \ldots, \lambda_J(x))^\intercal$ be a mean zero Gaussian process in $x \in \mathcal{X}$ with covariance function given by

$$C_{\boldsymbol{\Lambda}}(x_1, x_2) = \lim_{T\to\infty} \mathrm{E}\boldsymbol{\Lambda}_T(x_1)\boldsymbol{\Lambda}_T(x_2)^\intercal.$$

We impose the following assumption:

Assumption A. *(i) $\{(Y_{jt}, X_{jt}) : t \geq 1\}$ for $j = 1, \ldots, J$ is a strictly stationary and β-mixing sequence of random variables whose mixing coefficient is of order $O(n^{-b})$ for some $b > r/(r-1)$, where $r > 1$. (ii) $E\left|Y_{jt}\right|^{2(r+\delta)} < \infty$ for some $\delta > 0$ for $j = 1, \ldots, J$. (iii) The distribution of X_{jt} has bounded density with respect to Lebesgue measure for $j = 1, \ldots, J$.*

We require at least second moments for our analysis. Linton and Xiao (2013) develop alternative asymptotics for expected shortfall when only weaker moment conditions are adopted. The large sample properties of $\widetilde{m}_{j<}(x)$ are similar to those of $\widehat{m}_{j<}(x)$ and are not repeated here.

Theorem 1. *Suppose that Assumption B holds. Then,*

$$\sqrt{T}\left(\hat{\mathbf{m}}_<(\cdot) - \mathbf{m}_<(\cdot)\right) \Longrightarrow \mathbf{v}(\cdot).$$

where $\mathbf{v}(x) = (\nu_1(x), \ldots, \nu_J(x))^\intercal$ is a mean zero Gaussian process in $x \in \mathcal{X}$ with covariance function given by

$$C(x_1, x_2) = \boldsymbol{\Delta}(x_1)C_\Lambda(x_1, x_2)\boldsymbol{\Delta}(x_2)^\intercal.$$

We discuss the asymptotic variance in the case $J = 1$. Let

$$u_t(x) = \left(Y_t - m_<(x)1(X_t \leq x)\right)1(X_t \leq x)$$

$$V(x) = \frac{\mathrm{lrvar}(u_t(x))}{F(x)^2},$$

where lrvar is the long run variance. In the iid case we have

$$\begin{aligned} V(x) &= \begin{pmatrix} \frac{1}{F(x)} & -\frac{R_1(x)}{F(x)^2} \end{pmatrix} \begin{pmatrix} R_2(x) - R_1^2(x) & R_1(x)(1-F(x)) \\ R_1(x)(1-F(x)) & F(x)(1-F(x)) \end{pmatrix} \begin{pmatrix} \frac{1}{F(x)} \\ -\frac{R_1(x)}{F(x)^2} \end{pmatrix} \\ &= \frac{1}{F(x)} \left(\frac{R_2(x)}{F(x)} - \frac{R_1^2(x)}{F(x)^2} \right) = \frac{R_2(x)F(x) - R_1^2(x)}{F(x)^3}. \end{aligned}$$

We now turn to the testing issue. Let $\widehat{\tau}_T$ be defined in (12).

COROLLARY 1. *Suppose that the null hypothesis H_0 holds. Then, we have*

$$\widehat{\tau}_T \Rightarrow \begin{cases} \min_{(i,j)\in\mathcal{I}} \sup_{x\in\mathcal{B}_{ij}} \left[\nu_i(x) - \nu_j(x)\right] & \text{if } \tau = 0 \\ -\infty & \text{if } \tau < 0, \end{cases}$$

where $I = \{(i,j) | i \neq j,\ \sup_{x\in\mathcal{X}} (\nu_i(x) - \nu_j(x)) = 0\}$ *and*

$$\mathcal{B}_{ij} = \{x \in \mathcal{X} : \nu_i(x) = \nu_j(x)\}. \tag{13}$$

The proof uses arguments of Linton, Maasoumi, and Whang (2005, Theorem 1).

REMARK. We cannot obtain a FCLT over the whole support of X_t, because the variance $V(x) \to \infty$ as $x \to -\infty$, in fact

$$V(x) \leq \frac{E(Y_t)}{F(x)} \to \infty$$

and $\sqrt{T}$-consistency breaks down as $x \to -\infty$. We may be able to obtain consistency of $\widehat{m}_<(x_T)$ with rates for some sequences $x_T \to -\infty$, but this will not hold for all sequences. Specifically, for the extreme values we may adopt a different approximation based on point process theory, in our case this is about the "concomitant" order statistics. Suppose that $Y_t = m(X_t) + \varepsilon_t$ with ε independent of X and mean zero. In the case where X has compact support with lower bound x_L and m smooth, the asymptotic behaviour of $Y_{:(1)}$ is determined by the distribution of ε, specifically

$$\Pr\left(Y_{:(1)} \leq y\right) \longrightarrow F_\varepsilon(y - m(x_L)),$$

so that $Y_{:(1)}$ converges to a random limit centred at $m(x_L) = m_<(x_L)$. The estimator in that case is asymptotically unbiased but inconsistent. Essentially, we need at least $k \to \infty$ in (10) to obtain consistency to a point value.

4.1 Critical Values and Consistency of Test Statistic

We first define the subsampling procedure. Write $\widehat{\tau}_T = \tau_T(W_1, \ldots, W_T)$ as a function of the data $\{W_t : t = 1, \ldots, T\}$. Let

$$G_T(\cdot) = \Pr\left(\sqrt{T}\tau_T(W_1, \ldots, W_T) \leq \cdot\right) \tag{14}$$

denote the distribution function of $\sqrt{T}\widehat{\tau}_T$. Let $\widehat{\tau}_{T,b,t}$ be equal to the statistic evaluated at the subsample $\{W_t, \ldots, W_{t+b-1}\}$ of size b, i.e.,

$$\widehat{\tau}_{T,b,t} = \tau(W_t, W_{t+1}, \ldots, W_{t+b-1}) \text{ for } t = 1, \ldots, T-b+1.$$

We note that each subsample of size b (taken *without replacement* from the original data) is indeed a sample of size b from the true sampling distribution of the original data. Hence, it is clear that one can approximate the sampling distribution of $\sqrt{T}\widehat{\tau}_T$ using the distribution of the values of $\tau_{T,b,t}$ computed over $T-b+1$ different subsamples of size b. That is, we approximate the sampling distribution G_T of $\sqrt{T}\widehat{\tau}_T$ by

$$\hat{G}_{T,b}(\cdot) = \frac{1}{T-b+1}\sum_{t=1}^{T-b+1} 1\left(\sqrt{b}(\tau_{T,b,t} - \widehat{\tau}_T) \leq \cdot\right).$$

Let $g_{T,b}(1-\alpha)$ denote the $(1-\alpha)$-th sample quantile of $\hat{G}_{T,b}(\cdot)$, i.e.,

$$g_{T,b}(1-\alpha) = \inf\{w : \hat{G}_{T,b}(w) \geq 1-\alpha\}.$$

We call it the *subsample critical value* of significance level α. Thus, we reject the null hypothesis at the significance level α if $\sqrt{T}\widehat{\tau}_T > g_{T,b}(1-\alpha)$. The computation of this critical value is not particularly onerous, although it depends on how big b is. The subsampling method has been proposed in Politis and Romano (1994) and is thoroughly reviewed in Politis, Romano and Wolf (1999). It works in many cases where the standard bootstrap fails: in heavy tailed distributions, in unit root cases, in cases where the parameter is on the boundary of its space, etc.

We now show that our subsampling procedure works under a very weak condition on b. In many practical situations, the choice of b will be data-dependent, see Linton, Maasoumi and Whang (2005, Section 5.2) for some methodology for choosing b. To accommodate such possibilities, we assume that $b = \hat{b}_T$ is a data-dependent sequence satisfying

Assumption B. $\Pr[l_T \leq \hat{b}_T \leq u_T] \to 1$ *where* l_T *and* u_T *are integers satisfying* $1 \leq l_T \leq u_T \leq T$, $l_T \to \infty$ *and* $u_T/T \to 0$ *as* $T \to \infty$.

The following theorem shows that our test based on the subsample critical value has asymptotically correct size.

Theorem 2. *Suppose that Assumptions A and B hold. Then, under the null hypothesis* H_0,

$$\lim_{T\to\infty} \Pr\left(\sqrt{T}\widehat{\tau}_T > g_{T,\hat{b}_T}(1-\alpha)\right) \leq \alpha,$$

with equality holding if $\cup_{i,j}\mathcal{B}_{ij} \neq \emptyset$, *where* $\mathcal{B}_{ij}$ *is defined in (13).*

Theorem 2 shows that our test based on the subsampling critical values has asymptotically valid size under the null hypothesis and has asymptotically exact size on the boundary of the null hypothesis. Under additional regularity conditions, we can extend this pointwise result to establish that our test has asymptotically correct size uniformly over the distributions under the null hypothesis, using the arguments of Andrews and Shi (2013) and Linton, Song and Whang (2010). For brevity, we do not discuss the details of this issue in this paper.

We next establish that the test S_T based on the subsampling critical values is consistent against the fixed alternative H_1.

Theorem 3. *Suppose that Assumptions A and B hold. Then, under the alternative hypothesis H_1,*

$$\lim_{T\to\infty} \Pr\left(\sqrt{T}\widehat{\tau}_T > g_{T,\hat{b}_T}(1-\alpha)\right) = 1.$$

5 Numerical Evidence

In this section we show some empirical results on testing the hypothesis whether a random variable Y is relatively negative expectation dependent on the other random variable X ($RNED\,(Y|X)$). Wright (1987) shows that $RNED(Y|X)$ if and only if $\text{cov}\,(Y-X, f\,(X)) \leq 0$ for every increasing function f for which the covariance is defined. We construct the sample test statistic by using the sample analogue of $d^{RNED}\,(x) = E(Y_t - X_t) - E(Y_t - X_t|X_t \leq x)$. The null hypothesis is that $RNED(Y|X)$ holds: $H_0 : d^{RNED}\,(x) \leq 0$ for all x. Similarly, we construct the sample statistic for testing whether Y is negative expectation dependent on X based on $d^{NED}(x)$. The theoretical properties of these tests follow from Theorems 1,2, and 3.

To compute the statistics (here called $\hat{S}_T^{RNED}$ and $\hat{S}_T^{NED}$), we use a brute-force method: Choose a grid of 500 points x through the empirical quantiles of X_t and find the maximum value of $\hat{d}_T^{NED}(.)$ and $\hat{d}_T^{NED}(.)$ given those x. We use the subsampling scheme introduced in Section 8.1 to construct the empirical distribution of the sample test statistic, and reject the null if the empirical p-value of the sample test statistic is larger than a specified significant level α.

Simulations

We next examine performances of the proposed test statistic by simulations. We focus on testing whether $RNED\,(Y|X)$ holds. We generate samples from a multivariate normal $MV(\mathbf{0}, \Sigma)$ and a multivariate t distributions $t_v(\mathbf{0}, \Sigma)$. The data generating processes for the simulations are as follows:

- DGP 1: $X_t = Z_{2t}$, $Y_t = X_t + Z_{1t}$ and $(Z_{1t}, Z_{2t}) \sim i.i.d.\ MV(\mathbf{0}, \Sigma)$.
- DGP 3: $X_t = Z_{2t}$, $Y_t = X_t + Z_{1t}$ and $(Z_{1t}, Z_{2t}) \sim i.i.d.\ t_v(\mathbf{0}, \Sigma)$ and degrees of freedom $v = 3$.
- DGP 2: $\ln X_t = Z_{2t}$, $Y_t = X_t + Z_{1t}$ and $(Z_{1t}, Z_{2t}) \sim i.i.d.\ MV(\mathbf{0}, \Sigma)$.

- DGP 4: $\ln X_t = Z_{2t}$, $Y_t = X_t + Z_{1t}$ and $(Z_{1t}, Z_{2t}) \sim i.i.d.\ t_v(\mathbf{0}, \Sigma)$ and degrees of freedom $v = 3$.

For the data generating process, we set

$$\Sigma = \begin{pmatrix} 1 & \sigma_{xy} \\ \sigma_{xy} & 1 \end{pmatrix},$$

where $\sigma_{xy} = -0.8$, -0.3, 0, 03, 0, 0.8. Notice that under the multivariate t distribution, $\text{cov}(Z_{1t}, Z_{2t}) = v/(v-2) \times \Sigma$. Finally, $\sigma_{xy} \leq 0$ corresponds to the null and $\sigma_{xy} > 0$ corresponds to the alternative.

We set sample sizes $T = 250$, 500 and 1000 and corresponding subsample sizes $b = 50$, 100 and 200. Each scenario is simulated 1,000 times. Figure 1 shows some examples of empirical distributions of the subsampling test statistics from simulations. We report simulation results in Table 1. Overall, the proposed test statistic performs better under the multivariate normal than under the multivariate t. When the samples are generated from the multivariate normal distribution, the proposed test statistic has a less probability to get wrong rejections when the null is true and a higher probability to get correct rejections when the null is false. Under the multivariate normal distribution, when $\sigma_{xy} = -0.8$ and -0.3, whether X_t is with log transformation or not has a negligible effect on the performance of the proposed test statistic. But when $\sigma_{xy} \geq 0$, the test statistic tends to obtain less rejections with $\ln X_t$ than with X_t. Similar phenomenon also occurs in the case of the multivariate t. The results suggest that log transformation of X_t may damage power of the test statistic but have almost no effect on size distortion when the null becomes strong. Finally, increasing sample size in general seems to have little effect on improving the performance and this might be due to i.i.d. samples.

Applications with Real Data

Optimal Portfolio Choices

We consider two applications with real data. The first application is to test whether an asset should be included in a risk averse investor's portfolio. Wright (1987) provides a sufficient condition associated with the relative negative expectation dependence for justifying this. The condition has been mentioned in previous section. Here we restate it again. Consider a risk averse investor's portfolio optimization with N assets:

$$\max_{\boldsymbol{\theta}} E\left[U\left(\boldsymbol{\theta}^{\intercal}\mathbf{R}\right)\right] \text{ subject to } \boldsymbol{\theta}^{\intercal}\mathbf{1} = 1 \text{ and } \boldsymbol{\theta} \succeq \mathbf{0},$$

where $U' \geq 0$ and $U'' < 0$. $\mathbf{R} = (R_1, \ldots, R_N)^{\intercal}$ is a column vector for asset returns and $\theta = (\theta_1, \ldots, \theta_N)^{\intercal}$ is a column vector for portfolio weights. $\boldsymbol{\theta} \succeq \mathbf{0}$ means $\theta_i \geq 0$ for all $i = 1, \ldots, N$.

Suppose an asset k has the largest expected return among all assets, i.e., $E(R_k) \geq E(R_i)$ for all i, $k \neq i$. Wright (1987) shows that the asset k should be included in the risk averse investor's portfolio ($\theta_k > 0$) if strictly $RNED(R_k|R_i)$ holds for all i, $i \neq k$.

We use two data sets: 10 portfolios formed on size and 10 portfolios formed on stock's Beta. The constituents of these portfolios are U.S. stocks. For the two data sets, the sampling frequency is monthly and sampling period is from July-1963 to June-2015. Both data sets can be downloaded from Kenneth French's website. Table 2 shows summary statistics of returns (in percentage) of these portfolios. For portfolios formed on size, indices 1 to 10 in the first column of the table denote the portfolio of the smallest companies' stocks to the portfolio of the largest companies' stocks. It is the same for portfolios formed on stocks' Beta: Indices 1 to 10 denote the portfolio of the lowest Beta stocks to the portfolio of the highest Beta stocks. From the table, it can be seen that (time series) average returns of portfolios formed on size do not monotonically decreasing with the size, although the portfolios with small companies' stocks often have higher average returns than portfolios with large companies' stocks. It is also similar for average returns of portfolios formed on stocks' Beta: The High Beta portfolios often have higher average returns than does the low Beta portfolios.

For each data sets, we conduct the relative expectation dependence test pairwisely on the portfolio returns and report results in Table 3 (portfolios formed on size) and 4 (portfolio formed on stocks' Beta). In each table we show values of the sample statistic and corresponding empirical p-values obtained from using the subsampling scheme. Notice that for $i \neq j$, strictly $RNED(R_i|R_j)$ does not guarantee strictly $RNED(R_j|R_i)$ and thus the tables are not symmetric. For portfolios formed on size, it can be seen that the hypothesis of $RNED(R_i|R_j)$ is rejected for most $i < j$ but is not rejected for most $i > j$ at the conventional significant level. But for portfolios formed on stocks' Beta, the results are reversed: The hypothesis of $RNED(R_i|R_j)$ is rejected for most $i > j$ but is not rejected for most $i < j$ at the conventional significant level. Notice that portfolios of small size (high Beta) stocks in general have higher average returns than do portfolios of large size (low Beta) stocks. Thus if a risk averse investor considers to form a portfolio from the 10 size (stocks' Beta) portfolios, the test results suggest that assigning positive weights on portfolios of small size (high Beta) stocks may not be optimal.

The test result for portfolios formed on stocks' Beta is consistent with a trading strategy called Betting against Beta (BAB): Buying a low Beta portfolio and selling a high Beta portfolio with appropriate amounts. Such a strategy on average generates a low systematic risk but a high risk-adjusted return. The BAB strategy is based on a historical observation that average returns of portfolios with different Betas do not behave as what the CAPM implies. We discuss the relation of the test result and the strategy as follows. Suppose return of a well diversified portfolio R_i has a form of the CAPM: $R_i = R_f + \beta_i (R_m - R_f) + \varepsilon_i$, where R_f is the risk-free rate, β_i is the portfolio's

Beta and R_m is the market portfolio return. Since the portfolio is well diversified, its idiosyncratic risk $\varepsilon_i \approx 0$ and its expectation is $E(R_i) \approx R_f + \beta_i (E(R_m - R_f))$. In Finance, a line depicts the relation between β_i and $E(R_i)$ is called the security market line (SML). If the CAPM holds, the SML should have a slope of $E(R_m - R_f)$ and an intercept term of R_f.

In Figure 2 we plot the average monthly return (in percentage) against Beta for the 10 portfolios formed on stocks' Beta. The dash line is the hypothetical SML implied by the CAPM (with monthly average R_f about 0.4% and monthly average R_m about 0.9%). The solid line is the actual SML, which is a fitted line of the average returns and Betas of the portfolios. It can be seen that the actual SML deviates substantially from the hypothetical SML.

The deviation indicates that during a very long time (more than 50 years), the low (high) Beta portfolio has a higher (lower) average return than what the CAPM implies and investors put too much (less) money on the high (low) Beta stocks[1]. The long-term deviation suggests that it is safe for an investor to implement the BAB strategy. Notice that the derivation of the BAB strategy is based on the observation that the arbitrage opportunity exists in a very long time. Our test result, however, is based on the framework of the risk averse investor's portfolio optimization. Both our test result and the BAB strategy suggest that it may not optimal for a rational, risk-averse investor to put money on the high Beta stocks.

Growth and Public Debt

The second application is to test whether there is a negative relation between real GDP growth and debt to GDP ratio. Reinhart and Rogoff (2010) shows that in 20 advanced countries, from 1946 to 2009, the relation between real GDP growth and public debt seems relatively insignificant when the countries' debt to GDP ratios are below 90 percent, but median growth rates for the countries with public debt to GDP ratios over 90 percent are about one percent lower than otherwise; average (mean) growth rates are several percent lower. However, Herndon, Ash and Pollin (2013) points out several research flaws in Reinhart and Rogoff (2010). Among them, the most serious one is that in the period from 1946 to 2009, the strong negative relation between real growth and public debt of the 20 advanced countries when their debt to GDP ratios are over 90 percent, is no longer hold.

We use the proposed test of negative expectation dependence (NED) to examine whether there is a negative relation between growth and public debt in the 20 advanced countries in the post-war period. The data we use are compiled by Herndon, Ash and Pollin (2013). Figure 3 is a reproduction of Figure 3 in Herndon, Ash and Pollin (2013) with their R codes. It shows real GDP growth against public debt/GDP for all country-years and an estimated locally smoothed regression function of the

[1] There are several financial theories for explaining this deviation, for example, investors is facing leverage constraints or afraid of risk from using leverage (see, pp. 161 in Pedersen (2015)).

two variables. The estimated locally smoothed regression function suggests that real GDP growth is overall a nonincreasing function of public debt/GDP.

We categorize all country-year observations by their public debt to GDP ratios and report summary statistics and test results in Table 5. Column 7-10 of Table 5 show values of the sample test statistic and the corresponding empirical p-values. To calculate the test statistic, standardized data are used. The empirical p-values are obtained by using the *bootstrap* method since the data are of country-year type. Results in column 7 and 8 indicate that we cannot reject that null that the real GDP growth is negative expectation dependent on the debt to GDP ratio. Results in column 9 and 10 show that the reverse also seems to be true: The null that the debt to GDP ratio is negative expectation dependent on the real GDP growth cannot be rejected. Comparing the test results with the sample correlation coefficients shown in the last column, we find they are only inconsistent in the case when the debt to GDP ratio is 30-60%, in which the sample correlation coefficient is 0.0009. For the other four cases, they are consistent with each other.

6 Conclusions

The lower regression function is simple to compute and to analyze from a statistical point of view. The theory is related to the theory for stochastic dominance, Whang (2019). We applied our techniques to two applications in finance and macroeconomic growth theory.

7 Appendix

Proof of Theorem 1: For $j = 1, \ldots, J$, write

$$\begin{aligned} \sqrt{T}\left[\hat{m}_{<j}(x) - m_{<j}(x)\right] &= \frac{1}{F_j(x)}\sqrt{T}\left[\hat{R}_j(x) - R_j(x)\right] \\ &\quad - \frac{\hat{R}_j(x)}{\hat{F}_j(x)F_j(x)}\sqrt{T}\left[\hat{F}_j(x) - F_j(x)\right]. \end{aligned} \tag{15}$$

Define the following empirical processes indexed by $x \in \mathcal{X}$:

$$\begin{aligned} \alpha_{jT}(x) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[Y_{jt}1(X_{jt} \le x) - EY_{jt}1(X_{jt} \le x)\right] \\ \beta_{jT}(x) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[1(X_{jt} \le x) - E1(X_{jt} \le x)\right]. \end{aligned}$$

We first establish stochastic equicontinuity of $\{\alpha_{jT}(\cdot) : T \geq 1\}$ using the result of Doukhan et. al. (1995, Theorem 1). The class of functions $\mathcal{M} = \{Y_{jt}1(X_{jt} \le x) : x \in \mathcal{X}\}$ is a type IV class (see

Andrews (1994)) that satisfies the L^2-continuity condition, because: $\forall x \in \mathcal{X}$

$$\begin{aligned} & E \sup_{x_1 \in \mathcal{X}: |x_1 - x| < \delta} |Y_{jt} 1(X_{jt} \leq x) - Y_{jt} 1(X_{jt} \leq x_1)|^2 \\ = & \; E |Y_{jt}|^2 1(X_{jt} \in (x - \delta, x + \delta)) \\ \leq & \; C_1 \Pr(X_{jt} \in (x - \delta, x + \delta)) \\ \leq & \; C_2 \delta, \end{aligned}$$

for each $\delta > 0$, where the first and second inequalities hold by Assumptions B(ii) and B(iii), respectively. This implies that the bracketing covering number satisfies $N_2^B(\varepsilon, M) \leq C(1/\varepsilon)^2$, which in turn satisfies the entropy condition of Doukhan et. al. (1995, equation (2.15)). This establishes the stochastic equicontinuity of $\{\alpha_{jT}(\cdot) : T \geq 1\}$ and hence $\{\beta_{jT}(\cdot) : T \geq 1\}$ by taking $Y_{jt} = 1$. The finite dimensional (fidi) convergence holds by the CLT of Herrndorf (1984, Theorem 1) using Assumptions B(i) and (ii) and Cramer-Wold device. Therefore, by Pollard (1990, Sec. 10), we have

$$\Lambda_T(\cdot) = (\lambda_{1T}(\cdot), \ldots, \lambda_{JT}(\cdot))^\intercal \Rightarrow \Lambda(\cdot) \tag{16}$$

where

$$\lambda_{jT}(\cdot) = \left(\alpha_{jT}(\cdot), \beta_{jT}(\cdot)\right)^\intercal$$

and $\Lambda(\cdot)$ is a mean zero Gaussian process on $\mathcal{X}$ with covariance function C_Λ. Finally, the weak convergence results imply that : for all $j = 1, \ldots, J$,

$$\sup_{x \in \mathcal{X}} \left| \hat{F}_j(x) - F_j(x) \right| \xrightarrow{p} 0 \tag{17}$$

$$\sup_{x \in \mathcal{X}} \left| \hat{R}_j(x) - R_j(x) \right| \xrightarrow{p} 0. \tag{18}$$

The results (15) - (18) and the assumption $\inf_{x \in \mathcal{X}} F_j(x) > 0$ for all j establish the desired result. ■

References

[1] Breiman L., Friedman J. H., Olshen R. A., and Stone, C. J. (1984) Classification and Regression Trees. Wadsworth.

[2] Chiu, W.H. (2014). Correlation-Increasing Marginal Risk Increase and Financial Risk Taking in the presence of Non-Financial Background Risks. Working paper, University of Manchester.

[3] Delgado, M. (1993). Testing the equality of nonparametric regression curves. *Statistics and Probability Letters.*

[4] Denuit, M., A. Goderniaux, and O. Scaillet (2007). A Kolmogorov-Smirnov type test for Shortfall Dominance against Parametric Alternatives. *Technometrics* 49, 88-98.

[5] Escanciano, J.C., and J. Hualde (2009). Persistence in Nonlinear Time Series: A Nonparametric Approach (February 6, 2009). CAEPR Working Paper No. 2009-003. Available at SSRN: https://ssrn.com/abstract=1346052 or http://dx.doi.org/10.2139/ssrn.1346052

[6] Herndon, T., M. Ash and Robert Pollin (2013), "Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff," *Cambridge Journal of Economics*, doi:10.1093/cje/bet075.

[7] Levy, H. and J. Paroush (1974). Toward Multivariate Efficiency Criteria. *Journal of Economic Theory* 7, 129-142.

[8] Li, J. (2011). The demand for a risky asset in the presence of a background risk. *Journal of Economic Theory* 146, 372-391.

[9] Linton, O., & Xiao, Z. (2013). Estimation of and inference about the expected hsortfall for time series with infinite variance. *Econometric Theory*, 29(4), 771-807. doi:10.1017/S0266466612000692

[10] Pedersen, L. H. (2015), *Efficiently Inefficient: How Smart Money Invests and Market Prices Are Determined*, Princeton University Press.

[11] Reinhart, C. M. and K. S. Rogoff (2010), "Growth in a Time of Debt," *American Economic Review: Papers and Proceedings* 100, 573-578.

[12] Scaillet, O. (2004) Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance* 14, 115–129.

[13] Scaillet, O. (2005) Nonparametric estimation of conditional expected shortfall. Insurance and Risk *Management Journal* 74, 639–660

[14] Wright, R. (1987), "Expectation Dependence of Random Variables, with an Application in Portfolio Theory," *Theory and Decision* 22, 111-124.

[15] Yang, S.S. (1977) General distribution theory of concomitants of the order statistics. *Annals of Statistics* 5, 996-1002

[16] Zhu, X., X. Guo, L. Lin, and L. Zhu (2014). Testing for positive expectation dependence. *Annals of the Institute of Statistical Mathematics* 68, 135–153.