Testing for Conditional Mean Independence with Covariates through Martingale Difference Divergence

Ze Jin*
Department of Statistical Science
Cornell University
Ithaca, NY 14850

Xiaohan Yan
Department of Statistical Science
Cornell University
Ithaca, NY 14850

David S. Matteson†
Department of Statistical Science
Cornell University
Ithaca, NY 14850

Abstract

A crucial problem in statistics is to decide whether additional variables are needed in a regression model. We propose a new multivariate test to investigate the conditional mean independence of Y given X conditioning on some known effect Z, i.e., $\mathrm{E}(Y|X,Z) = \mathrm{E}(Y|Z)$. Assuming that $\mathrm{E}(Y|Z)$ and Z are linearly related, we reformulate an equivalent notion of conditional mean independence through transformation, which is approximated in practice. We apply the martingale difference divergence (Shao and Zhang, 2014) to measure conditional mean dependence, and show that the estimation error from approximation is negligible, as it has no impact on the asymptotic distribution of the test statistic under some regularity assumptions. The implementation of our test is demonstrated by both simulations and a financial data example.

1 INTRODUCTION

Testing (conditional) dependence and conditional mean dependence plays an important role in statistics with various applications, including variable selection (Székely and Rizzo, 2014; Park et al., 2015; Zhang et al., 2015; Yan and Bien, 2018), feature screening (Li et al., 2012; Shao and Zhang, 2014; Yan et al., 2017), and graphical models (Gan et al., 2018; Li and McCormick, 2017; Li et al., 2018). Both areas attracted tremendous attention in the last two decades, as datasets have increased in size and dimension. Let $X \in \mathbb{R}^p$, $Y \in \mathbb{R}^q$, $Z \in \mathbb{R}^r$ be the three random vectors of interest, and denote pairwise independence by $\perp\!\!\!\perp$.

Measures of (conditional) dependence have been extensively studied. Székely et al. (2007) proposed distance covariance (dCov) to capture the non-linear and non-monotone pairwise dependence between X and Y, and dCov = 0 if and only if pairwise independence ($X \perp\!\!\!\perp Y$) holds. Jin and Matteson (2017) extended distance covariance to mutual dependence measures (MDMs), which have been applied to independent component analysis in Jin and Matteson (2018). To capture the conditional dependence between X and Y given Z, Székely and Rizzo (2014) generalized distance covariance to partial distance covariance (pdCov), however, pdCov = 0 is not equivalent to conditional independence ($X \perp\!\!\!\perp Y|Z$), and neither one implies the other. Wang et al. (2015) extended distance covariance to conditional distance covariance (CDCov) using kernel estimators, and CDCov = 0 if and only if conditional independence holds. Under a linear assumption between X, Y and Z, Fan et al. (2015) converted testing conditional independence to testing independence, and applied distance covariance to measure the dependence of estimated variables. Moreover, inter-temporal conditional dependence is known as Granger causality in time series analysis. Hiemstra and Jones (1994), Su and White (2007), and Chen and Hong (2012) each introduced non-parametric tests for non-linear Granger causality based on conditional probabilities and characteristic functions.

Likewise, various measures of conditional mean dependence have been broadly developed as well. Testing the conditional mean independence of Y given X, i.e.,

$$H_0 : \mathrm{E}(Y|X) = \mathrm{E}(Y) \quad a.s., \quad H_A : o.w. \tag{1}$$

provides insight on whether X contributes to the conditional mean of Y. Shao and Zhang (2014) generalized distance covariance to martingale difference divergence (MDD), and MDD = 0 if and only if (1) holds. Testing

*Corresponding author. Email address: zj58@cornell.edu.

†Research support from an NSF Award (DMS-1455172), a Xerox PARC Faculty Research Award, and Cornell University Atkinson Center for a Sustainable Future (AVF-2017).

the conditional mean independence of Y given X conditioning on some known effect Z, i.e.,

$$H_0 : \mathrm{E}(Y|X, Z) = \mathrm{E}(Y|Z) \quad a.s., \quad H_A : o.w. \tag{2}$$

sheds light on whether X contributes to the conditional mean of Y when taking known dependence on Z into account. Park et al. (2015) generalized martingale difference divergence to partial martingale difference divergence (pMDD), however, pMDD = 0 is not equivalent to (2). Fan and Li (1996), Lavergne and Vuong (2000), and Aït-Sahalia et al. (2001) each introduced non-parametric tests for (2) using kernel estimators of conditional expectations. Assuming a linear model between Y and (X, Z), Lan et al. (2014) generalized the classical partial F-test (Chatterjee and Hadi, 2015) to a partial covariance-based (pcov) test for (2) in the high-dimensional setting, and Tang et al. (2017) further proposed a hybrid test for (2) through finding the most predictive covariate based on both maximum-type and sum-type statistics. Conditional mean independence conditioning on lagged covariates is known as Granger causality *in mean* in time series analysis. Raïssi et al. (2011) proposed a parametric test for linear Granger causality in mean based on vector autoregressive (VAR) models, and Hong et al. (2009) introduced a non-parametric test for non-linear Granger causality in mean based on cross-correlations.

In this paper, we focus on testing conditional mean independence with covariates and develop a method to test (2) for two main reasons. As Cook and Li (2002) state, regression analysis is mostly concerned with the conditional mean of the response given the predictors, which makes testing conditional mean independence more appealing than testing conditional independence. Further, it is very common in practice that some given covariates Z have been known to affect the conditional mean of Y. In this situation, we aim to determine whether X has marginal effect on the conditional mean of Y in the presence of Z, and decide whether X should be included to model the conditional mean of Y along with Z. In general, testing (2) is more useful than testing (1), but requires more careful handling.

We first simplify testing (2) to testing conditional mean independence through a transformation. Let $V = Y - \mathrm{E}(Y|Z) \in \mathbb{R}^q$, and $U = (X, Z) \in \mathbb{R}^{p+r}$. Then $\mathrm{E}(V) = 0$, and $\mathrm{E}(V|U) = \mathrm{E}(Y|X, Z) - \mathrm{E}(Y|Z)$. As a result, we obtain an equivalent hypothesis test to (2) as

$$H_0 : \mathrm{E}(V|U) = \mathrm{E}(V) = 0 \quad a.s., \quad H_A : o.w. \tag{3}$$

which is conditional mean independence of V given U. Thus, we consider the MDD with U and V to investigate (3). However, there are two problems to solve when we apply MDD to U and V. First, V needs to be estimated since it is unobserved. We will replace V by its estimate $\widehat{V}$ in calculating MDD. Second, we need to confirm that the estimation error of $\widehat{V}$ is negligible, i.e., MDD with $\widehat{V}$ is close enough to that with V, such that $\widehat{V}$ may be used for inference instead of V.

The rest of this paper is organized as follows. In Section 2, we give a brief overview of martingale difference divergence. In Section 3, we estimate V based on the assumption that $\mathrm{E}(Y|Z)$ is a linear function of Z, and prove that the estimation of V does not affect the asymptotic distribution of martingale difference divergence under some regularity conditions. We present simulation results in Section 4, followed by a real data analysis in Section 5[1]. Finally, we summarize our work in Section 6.

The following notation is used throughout this paper. Let $\{(X_i, Y_i, Z_i) : i = 1, \ldots, n\}$ be an i.i.d. sample from the joint distribution $F_{X,Y,Z}$. When A is a matrix, the element of A at row k and column ℓ is denoted by $A(k, \ell)$. When A is a vector, the element of A at index k is denoted by $A(k)$. The Frobenius norm of matrix $A \in \mathbb{R}^{p\times q}$ is denoted by $\|A\|_{\mathrm{F}}$. The Euclidean norm of vector $X \in \mathbb{R}^p$ is denoted by $|X|$. The weighted $\mathcal{L}_2$ norm $\|\cdot\|_w$ of any complex-valued function $\eta(t), t \in \mathbb{R}^p$ is defined by $\|\eta(t)\|_w^2 = \int_{\mathbb{R}^p} |\eta(t)|^2 w(t)\, dt$ where $|\eta(t)|^2 = \eta(t)\overline{\eta(t)}$, $\overline{\eta(t)}$ is the complex conjugate of $\eta(t)$, and $w(t)$ is any positive weight function under which the integral exists. Furthermore, $a.s.$ is an abbreviation of almost surely.

2 MARTINGALE DIFFERENCE DIVERGENCE

Shao and Zhang (2014) proposed martingale difference divergence to capture the conditional mean dependence (in any form) of $Y \in \mathbb{R}^q$ given $X \in \mathbb{R}^p$.

The non-negative martingale difference divergence for X and Y, $\mathrm{MDD}(Y|X)$ is defined by its square

$$\begin{aligned}\mathrm{MDD}^2(Y|X) &= \|\mathrm{E}(Ye^{i\langle s,X\rangle}) - \mathrm{E}(Y)\mathrm{E}(e^{i\langle s,X\rangle})\|_{w_p}^2 \\ &\triangleq \int_{\mathbb{R}^p} |\mathrm{E}(Ye^{i\langle s,X\rangle}) - \mathrm{E}(Y)\mathrm{E}(e^{i\langle s,X\rangle})|^2 w_p(s)\, ds,\end{aligned}$$

where the weight $w_p(s) = c_p|s|^{1+p}$, with $c_p = \frac{\pi^{(1+p)/2}}{\Gamma((1+p)/2)}$, and Γ is the gamma function. If $\mathrm{E}(|X|^2 + |Y|^2) < \infty$, then $\mathrm{MDD}(Y|X) = 0$ if and only if $\mathrm{E}(Y|X) = \mathrm{E}(Y)$ holds $a.s.$

The non-negative empirical martingale difference diver-

[1]See CRAN for an accompanying R package `EDMeasure` (Jin et al., 2018).

gence $\mathrm{MDD}_n(Y|X)$ is analogously defined by

$$\mathrm{MDD}_n^2(Y|X) = \frac{1}{n^2}\sum_{i,j=1}^{n} A_{ij}B_{ij},$$

where $A_{ij} = a_{ij} - \bar{a}_{i\cdot} - \bar{a}_{\cdot j} + \bar{a}_{\cdot\cdot}$, $\bar{a}_{i\cdot} = \frac{1}{n}\sum_{j=1}^{n} a_{ij}$, $\bar{a}_{\cdot j} = \frac{1}{n}\sum_{i=1}^{n} a_{ij}$, $\bar{a}_{\cdot\cdot} = \frac{1}{n^2}\sum_{i,j=1}^{n} a_{ij}$, $a_{ij} = |X_i - X_j|$, and similarly for B_{ij} with $b_{ij} = \frac{1}{2}|Y_i - Y_j|^2$.

The consistency and weak convergence of $\mathrm{MDD}_n(Y|X)$ are derived as follows. If $\mathrm{E}(|X| + |Y|^2) < \infty$, we have (i) $\mathrm{MDD}_n(Y|X) \xrightarrow[n\to\infty]{a.s.} \mathrm{MDD}(Y|X)$; (ii) under H_0 : $\mathrm{E}(Y|X) = \mathrm{E}(Y)\ a.s.$, $n\mathrm{MDD}_n^2(Y|X) \xrightarrow[n\to\infty]{\mathcal{D}} \|\zeta(s)\|_{w_p}^2$, where $\zeta(\cdot)$ is a complex-valued zero-mean Gaussian process whose covariance function depends on $F_{X,Y}$; (iii) under $H_A : o.w.$, $n\mathrm{MDD}_n^2(Y|X) \xrightarrow[n\to\infty]{a.s.} \infty$. Utilizing the nice properties of MDD, we next propose our test for (3).

3 METHODOLOGY

Inspired by the linear assumption to simplify the conditional dependence structure in Fan et al. (2015), we assume that the conditional expectation $\mathrm{E}(Y|Z)$ is a linear function of Z, simplifying the conditional mean dependence structure. As a result, we can decompose Y into the conditional expectation and reminder as

$$Y = \mathrm{E}(Y|Z) + [Y - \mathrm{E}(Y|Z)] \triangleq BZ + V,$$

where $B \in \mathbb{R}^{q\times r}$, $V \in \mathbb{R}^q$. Then we have $\mathrm{E}(V|Z) = 0$, and $\mathrm{E}(V) = 0$. Similarly, the ith sample counterpart is $Y_i = \mathrm{E}(Y_i|Z_i) + V_i \triangleq BZ_i + V_i, i = 1, \ldots, n$.

Suppose $\widehat{B}$ is the ordinary least squares (OLS) estimator of B when regressing Y on Z. We will then replace B with $\widehat{B}$ to estimate $\mathrm{E}(Y_i|Z_i)$ as $\widehat{\mathrm{E}}(Y_i|Z_i) = \widehat{B}Z_i$, and V_i as $\widehat{V}_i = Y_i - \widehat{B}Z_i = (B - \widehat{B})Z_i + V_i$. When estimating B via the OLS, Z is implicitly assumed to have full column rank. In case Z is high-dimensional, i.e., $r > n$, we can estimate B by the penalized least squares (PLS) similar to Fan et al. (2015), including ridge (Hoerl and Kennard, 1970) and lasso (Tibshirani, 1996).

We now construct a test for (3) based on $\mathrm{MDD}_n^2(\widehat{V}|U)$ and its counterparts using permutation samples, then calculate the empirical p-value following the permutation in Park et al. (2015). Because the samples are independent, but with an unspecified distribution, permutation tests are a convenient tool for inference. We will later show in Theorem 2 that the asymptotic distribution of $n\mathrm{MDD}_n^2(\widehat{V}|U)$ depends on an unknown underlying distribution, which justifies the use of permutation tests. To measure the conditional mean dependence of V given U, we first compute the test statistic $\mathrm{MDD}_n^2(\widehat{V}|U)$ from the sample $\{(\widehat{V}_i, U_i) : i = 1, \ldots, n\}$, where $U_i = (X_i, Z_i)$. That is, $\mathrm{MDD}_n^2(\widehat{V}|U)$ depends on the i.i.d. sample $\{(X_i, Y_i, Z_i) : i = 1, \ldots, n\}$. Next we draw B permutation samples of size n as $\{(X_i^*, Y_i, Z_i) : i = 1, \cdots, n\}$, where only the sample of X is permuted in order to approximate the sampling distribution. For each permutation sample, we calculate the test statistic $\mathrm{MDD}_{n,b}^2(\widehat{V}|U)$, $b = 1, \cdots, B$. Then the empirical p-value is given by

$$\widehat{p} = \frac{\sum_{b=1}^{B} \mathbf{1}\left\{\mathrm{MDD}_{n,b}^2(\widehat{V}|U) \geq \mathrm{MDD}_n^2(\widehat{V}|U)\right\}}{B}.$$

When H_0 is false, $\mathrm{MDD}_n^2(\widehat{V}|U)$ tends to be large while $\mathrm{MDD}_{n,b}^2(\widehat{V}|U)$ tends to be small. As a result, the empirical p-value is expected to be very small, leading to a rejection of H_0. We name the proposed test linear martingale difference divergence (LinMDD). To justify our LinMDD test, it remains to validate that $\mathrm{MDD}_n^2(\widehat{V}|U)$ is close enough to $\mathrm{MDD}_n^2(V|U)$, i.e., the estimation error in $\widehat{V}$ is negligible for the sampling distribution of the test statistic, focusing on the asymptotic case. To begin with, we introduce some regularity conditions to derive the asymptotic distribution of $\mathrm{MDD}_n^2(\widehat{V}|U)$.

Condition 1. *There exist constants* $0 < c_1, c_2, c_3 < \infty$, *such that* $E(|U_i - U_j|^2) = c_1$, $i \neq j$; $E(|U_i - U_j||U_i - U_k|) = c_2$, $i \neq j \neq k$; $E(|U_i - U_j||U_k - U_\ell|) = c_3$, $i \neq j \neq k \neq \ell$.

Condition 2. *There exists constant* $0 < c_4 < \infty$, *such that* $E[(Z_i(t) - Z_j(t))^2(Z_i(s) - Z_j(s))^2] \leq c_4$, $i \neq j$, $\forall t, s$.

Condition 3. *There exists constant* $0 < c_5 < \infty$, *such that* $E[(Z_i(t) - Z_j(t))^2(V_i(s) - V_j(s))^2] \leq c_5$, $i \neq j$, $\forall t, s$.

Condition 4. $\|\widehat{B} - B\|_F = O_p(n^{-1/2})$.

Remark. Condition 4 can be derived from the bounded density of $|V_i - V_j|$ and non-heavy tails of $Z_i(t)$ and $V_i(t)$ according to Fan et al. (2015) and Fan et al. (2011).

Through a similar derivation to Theorem 2 of Fan et al. (2015), we justify the choice of using $\mathrm{MDD}_n^2(\widehat{V}|U)$ in place of $\mathrm{MDD}_n^2(V|U)$ by the following lemma and theorems. Lemma 1 shows that the difference between $\mathrm{MDD}_n^2(\widehat{V}|U)$ and $\mathrm{MDD}_n^2(V|U)$ is negligible as the sample size increases. The proof of Lemma 1 can be found in Appendix 6.

Lemma 1. *If* $Y = BZ + V$ *and Conditions 1-4 hold, we have*

$$MDD_n^2(\widehat{V}|U) - MDD_n^2(V|U) = O_p(n^{-3/2}).$$

Consequently, the consistency and weak convergence of $\mathrm{MDD}_n(\widehat{V}|U)$ follow from Lemma 1 and are summarized in Theorem 1 and 2 below.

Theorem 1 (Consistency)**.** *If $Y = BZ + V$ and Conditions 1-4 hold, we have*

$$MDD_n(\widehat{V}|U) \xrightarrow[n\to\infty]{\mathcal{P}} MDD(V|U).$$

Theorem 2 (Weak convergence)**.** *If $Y = BZ + V$ and Conditions 1-4 hold, under H_0, we have*

$$nMDD_n^2(\widehat{V}|U) \xrightarrow[n\to\infty]{\mathcal{D}} \|\zeta(s)\|_{w_p}^2,$$

where $\zeta(\cdot)$ denotes the complex-valued Gaussian random process corresponding to the asymptotic distribution of $nMDD_n^2(V|U)$. Under H_A, we have

$$nMDD_n^2(\widehat{V}|U) \xrightarrow[n\to\infty]{\mathcal{P}} \infty.$$

According to Theorem 1, $\mathrm{MDD}_n(\widehat{V}|U)$ converges to the same population statistic $\mathrm{MDD}(V|U)$ as $\mathrm{MDD}_n(V|U)$, and thus it can serve to measure the conditional mean dependence of V given U. In addition, $n\mathrm{MDD}_n^2(\widehat{V}|U)$ and $n\mathrm{MDD}_n^2(V|U)$ have the same asymptotic distribution stated in Theorem 2, which establishes the effectiveness of LinMDD test, as we approximate the limiting distribution of $n\mathrm{MDD}_n^2(V|U)$ using $n\mathrm{MDD}_n^2(\widehat{V}|U)$ in LinMDD test. In Section 4 and Section 5, we will present the finite-sample performance of our LinMDD test through simulations and a real data example, respectively.

4 SIMULATION STUDIES

To evaluate the performance of our LinMDD test, we adopt the simulation setup in Lavergne and Vuong (2000), and compare our test to the pMDD test (Park et al., 2015), pdCov test (Székely and Rizzo, 2014), and pcov test (Lan et al., 2014) as benchmarks. All tests are implemented as permutation tests with permutation size $B = 500$, in which we only permute the sample of X to approximate the distribution of the test statistic.

We generate data from the underlying model

$$Y = -Z + b \cdot Z^3 + f(X) + \epsilon,$$

where $Z \sim N(0,1)$, $X \sim \mathcal{N}(0,1)$, $\epsilon \sim \mathcal{N}(0,4)$, and Z, X, ϵ are independent. We test the null hypothesis H_0 : $\mathrm{E}(Y|X,Z) = \mathrm{E}(Y|Z)$ $a.s.$ with significance level $\alpha \in \{0.05, 0.1\}$, and examine the empirical size and power of each test. We run 1000 replications with sample size $n \in \{20, 30, 50, 70, 100\}$ for each specific model.

Model 1 (Linear Z, linear X)**.** $b = 0$, $f(X) = cX$ where $c \in \{0, \frac{2}{3}, 1, \frac{3}{2}\}$.

Model 2 (Linear Z, non-linear X)**.** $b = 0$, $f(X) = \sin(c\pi X)$ where $c \in \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}\}$. We omit $c = 0$ as it is exactly the same as $c = 0$ in Model 1.

From Figure 1, the empirical size of all tests is around 0.05 (0.1). The empirical power of all tests increases as n increases. For the linear X case, the empirical power of all tests is higher when c is larger, since the signal-to-noise ratio increases. Moreover, the empirical power of the LinMDD and pcov tests is consistently higher than that of the other tests, because the linear assumption is valid, and only LinMDD and pcov tests are designed for linear Z. For the non-linear X case, the LinMDD test still outperforms the other tests, while the performance of the pcov test degrades as c increases, because the LinMDD test is designed for non-linear X while pcov test is suitable only for linear X.

Model 3 (Nonlinear Z, linear X)**.** $b = 1$, $f(X) = cX$ where $c \in \{0, \frac{2}{3}, 1, \frac{3}{2}\}$.

Model 4 (Nonlinear Z, non-linear X)**.** $b = 1$, $f(X) = \sin(c\pi X)$ where $c \in \{\frac{1}{4}, \frac{1}{3}, \frac{1}{2}\}$. We omit $c = 0$ as it is exactly the same as $c = 0$ in Model 3.

From Figure 2, the empirical size of all tests is around 0.05 (0.1). For the linear X case, the empirical power of the LinMDD and pcov tests is competitive with but not always higher than that of the other tests. The reason is that the linear dependence of Y on Z is violated while the other tests do not rely it. For the non-linear X case, we similarly find that the performance of the pcov test degrades as c increases. The simulation results show that our LinMDD test achieves competitive and often better performance than the others in these situations. Next, we apply the proposed LinMDD test on a real dataset.

5 FINANCIAL DATA APPLICATION

In finance, the capital asset pricing model (CAPM) was proposed by Sharpe (1964), Lintner (1965), and Mossin (1966) to describe the stock returns through the market risk as

$$r_t = \alpha + \beta_1 m_t,$$

where r_t is the excess stock return (in excess the risk-free return), and m_t is the excess market return at time t. Fama and French (1993) added size and value factors to the CAPM, and proposed the Fama−French three-factor model as

$$r_t = \alpha + \beta_1 m_t + \beta_2\,\mathrm{SMB}_t + \beta_3\,\mathrm{HML}_t,$$

where SMB (small minus big) and HML (high minus low) account for stocks with small/big market capitalization and high/low book-to-market ratio, respectively. Fama and French (2015) further added profitability and investment factors to the three-factor model, and extended it to the Fama−French five-factor model as

$$\begin{aligned} r_t = \alpha + \beta_1 m_t + \beta_2\,\mathrm{SMB}_t + \beta_3\,\mathrm{HML}_t \\ + \beta_4\,\mathrm{RMW}_t + \beta_5\,\mathrm{CMA}_t, \end{aligned}$$

where RMW (robust minus weak) and CMA (conservative minus aggressive) further account for stocks with robust/weak operating profitability and conservative/aggressive investment, respectively.

We collect the annual risk-free returns and Fama−French five factors[2], and the annual returns of Boeing (BA) stock[3] in the past 53 years between 1964 and 2016. The time series and histograms of excessive BA stock returns and Fama−French five factors are depicted in Figure 3.

5.1 CAPM VS. FAMA−FRENCH THREE-FACTOR MODEL

First, we are curious whether the size and value factors should be added to the CAPM, i.e., whether SMB and HML in the Fama−French three-factor model contribute to the expectation of excess stock returns given the market risk. Thus, we test H_0 : $\mathrm{E}(Y|X, Z) = \mathrm{E}(Y|Z)\ a.s.$, where $X_t = (\mathrm{SMB}_t, \mathrm{HML}_t)$, $Y_t = r_t$, and $Z_t = (1, m_t)$.

We apply our LinMDD test to the data with $n = 53$ and $B = 500$. Our p-value is 0.072, while the p-values are 0.012 (pMDD), 0.092 (pdCov) and 0.096 (pcov) using competing tests. As a result, we reject H_0 with significance level $\alpha = 0.1$, and conclude that SMB and HML help determine the excess returns of BA stock in the presence of the market risk. Our results align with the research in finance that the Fama−French three-factor model remarkably outperforms the CAPM in explaining excess stock returns.

5.2 FAMA−FRENCH THREE-FACTOR MODEL VS. FIVE-FACTOR MODEL

Similarly, we are interested in whether the profitability and investment factors should be further added to the Fama−French three-factor model, i.e., whether RMW and CMA in the Fama−French five-factor model contribute to the description of excess stock returns given the other three factors. Hence, we test H_0 : $\mathrm{E}(Y|X, Z) = \mathrm{E}(Y|Z)\ a.s.$, in which $X_t = (\mathrm{RMW}_t, \mathrm{CMA}_t)$, $Y_t = r_t$, and $Z_t = (1, m_t, \mathrm{SMB}_t, \mathrm{HML}_t)$.

We apply our LinMDD test to the data with $n = 53$ and $B = 500$, and its p-value is 0.360, while the p-values are 0.358 (pMDD), 0.878 (pdCov) and 0.768 (pcov) using competing tests. As a result, we fail to reject H_0 with significance level $\alpha = 0.1$, and conclude that RMW and CMA are unable to help determine the excess returns of BA stock in the presence of the other three factors. Our results align with the research in finance that the Fama−French five-factor model has yet to be proven as a significant improvement over the three-factor model in describing excess stock returns.

5.3 FAMA−FRENCH FOUR-FACTOR MODEL VS. FIVE-FACTOR MODEL

Fama and French (2015) showed that the value factor HML becomes redundant when profitability and investment factors are added to the Fama−French three-factor model, because HML is fully captured by its exposures to the other four factors, especially RMW and CMA. To validate this argument, we test H_0 : $\mathrm{E}(Y|X, Z) = \mathrm{E}(Y|Z)\ a.s.$, where $X_t = \mathrm{HML}_t$, $Y_t = r_t$, and $Z_t = (1, m_t, \mathrm{SMB}_t, \mathrm{RMW}_t, \mathrm{CMA}_t)$.

We apply our LinMDD test to the data with $n = 53$ and $B = 500$. Our p-value is 0.218, while the p-values are 0.438 (pMDD), 0.540 (pdCov) and 0.858 (pcov) using competing tests. As a result, we fail to reject H_0 with significance level $\alpha = 0.1$, and conclude that HML cannot help explain the excess returns of BA stock in the presence of the other four factors. Our results demonstrate that HML is redundant for describing excess stock returns in the Fama−French five-factor model.

6 CONCLUSION

In this paper, we propose a new test, LinMDD, for the null hypothesis H_0 : $\mathrm{E}(Y|X, Z) = \mathrm{E}(Y|Z)\ a.s.$ by investigating an equivalent one H_0 : $\mathrm{E}(V|U) = \mathrm{E}(V) = 0\ a.s.$, derived from a transformation involving the conditional expectation. When applying martingale difference divergence (Shao and Zhang, 2014) to test H_0 : $\mathrm{E}(V|U) = \mathrm{E}(V) = 0\ a.s.$, we make two major contributions.

(1) Since V is unobservable, we estimate V based on the assumption that $\mathrm{E}(Y|Z)$ is a linear function of Z, simplifying the conditional mean dependence structure.

(2) We prove that the estimation error in $\widehat{V}$ is negligible for the asymptotic distribution of the test statistic. Thus, we can replace V with $\widehat{V}$ in the test statistic for inference in large samples.

We implement the LinMDD test as a permutation test following Park et al. (2015), and compare it with existing tests in various simulation studies. The LinMDD test consistently outperforms existing tests when its linear assumption is valid, and it achieves competitive results with existing tests even when its linear assumption is violated.

[2]Download data at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

[3]Download data using `get.hist.quote` in the R package `tseries` (Trapletti and Hornik, 2017).

To illustrate the practical value of the LinMDD test, we compare the CAPM, the Fama−French three-factor and five-factor models by applying LinMDD test to the financial data. We find that the Fama−French three-factor outperforms the CAPM, while the Fama−French five-factor is not a significant improvement over the three-factor model when explaining the excess annual returns of a major stock. Moreover, we validate the statement that the value factor is redundant in the Fama−French five-factor model (Fama and French, 2015) using the LinMDD test.

The relaxation of the linear assumption is an important topic for future research. Our method will become more general if the linear assumption of conditional mean dependence can be generalized to a non-linear one, using non-parametric regression (local regression, splines) instead of linear regression in the estimation of conditional mean. In addition, the high-dimensional setting regarding Z where $r > n$ is an interesting direction to consider as well.

APPENDIX

PROOF OF LEMMA 1

Proof. We define T

$$\begin{aligned}
&= n\mathrm{MDD}_n^2(\widehat{V}|U) - n\mathrm{MDD}_n^2(V|U) \\
&= \frac{1}{2n}\sum_{i,j}[(F_{ij} - \frac{1}{n}\sum_k F_{kj} - \frac{1}{n}\sum_k F_{ik} + \frac{1}{n^2}\sum_{k,\ell} F_{k\ell}) \\
&\times\ (E_{ij} - \frac{1}{n}\sum_k E_{kj} - \frac{1}{n}\sum_k E_{ik} + \frac{1}{n^2}\sum_{k,\ell} E_{k\ell})],
\end{aligned}$$

where $F_{ij} = |\widehat{V}_i - \widehat{V}_j|^2 - |V_i - V_j|^2$, $E_{ij} = |U_i - U_j|$.

We apply Taylor expansion to $|\widehat{V}_t - \widehat{V}_s|^2$ at $V_t - V_s$ in terms of $f(x) = x^T x$, $f'(x) = 2x^T$, then there exists $\lambda \in (0, 1)$, such that F_{ij}

$$\begin{aligned}
&= 2[\lambda(\widehat{V}_i - \widehat{V}_j) + (1-\lambda)(V_i - V_j)]^T(\widehat{V}_i - \widehat{V}_j - V_i + V_j) \\
&= 2[\lambda(Z_i - Z_j)^T(B - \widehat{B})^T(B - \widehat{B})(Z_i - Z_j) \\
&+\ (V_i - V_j)^T(B - \widehat{B})(Z_i - Z_j)].
\end{aligned}$$

Thus, we have $T = T_1 + T_2$, where T_1

$$\begin{aligned}
&= \frac{\lambda}{n}\sum_{i,j}[(G_{ij} - \frac{1}{n}\sum_k G_{kj} - \frac{1}{n}\sum_k G_{ik} + \frac{1}{n^2}\sum_{k,\ell} G_{k\ell}) \\
&\times\ (E_{ij} - \frac{1}{n}\sum_k E_{kj} - \frac{1}{n}\sum_k E_{ik} + \frac{1}{n^2}\sum_{k,\ell} E_{k\ell})],
\end{aligned}$$

$$G_{ij} = (Z_i - Z_j)^T(B - \widehat{B})^T(B - \widehat{B})(Z_i - Z_j),$$

and T_2

$$\begin{aligned}
&= \frac{1}{n}\sum_{i,j}[(H_{ij} - \frac{1}{n}\sum_k H_{kj} - \frac{1}{n}\sum_k H_{ik}\frac{1}{n^2}\sum_{k,\ell} H_{k\ell}) \\
&\times\ (E_{ij} - \frac{1}{n}\sum_k E_{kj} - \frac{1}{n}\sum_k E_{ik} + \frac{1}{n^2}\sum_{k,\ell} E_{k\ell})],
\end{aligned}$$

$$H_{ij} = (V_i - V_j)^T(B - \widehat{B})(Z_i - Z_j).$$

First, we will show (i) $T_1 = O_p(n^{-1})$.

After a simple calculation, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i,j}(G_{ij} - \frac{1}{n}\sum_k G_{kj} - \frac{1}{n}\sum_k G_{ik} + \frac{1}{n^2}\sum_{k,\ell} G_{k\ell})E_{ij} \\
&= tr[\frac{1}{n}\sum_{i,j}|U_i - U_j|(G_{ij} - \frac{1}{n}\sum_k G_{kj} - \frac{1}{n}\sum_k G_{ik} \\
&+\ \frac{1}{n^2}\sum_{k,\ell} G_{k\ell})] \\
&= tr[(B - \widehat{B})^T(B - \widehat{B})M],
\end{aligned}$$

where $M = \frac{1}{n}\sum_{i,j}|U_i - U_j|S_{ij}$, and

$$S_{ij} = R_{ij} - \frac{1}{n}\sum_k R_{kj} - \frac{1}{n}\sum_k R_{ik} + \frac{1}{n^2}\sum_{k,\ell} R_{k\ell},$$

$R_{ij} = (Z_i - Z_j)(Z_i - Z_j)^T$, $R_{ij} = R_{ji}$, $S_{ij} = S_{ji}$, then

$$\begin{aligned}
&\mathrm{E}[(M(t,s))^2] \\
&= \mathrm{E}[\frac{1}{n^2}(\sum_{i,j}|U_i - U_j|S_{ij}(t,s))^2] \\
&= \mathrm{E}\{\mathrm{E}[\frac{1}{n^2}(\sum_{i,j}|U_i - U_j|S_{ij}(t,s))^2|U_i, \forall i]\} \\
&= \mathrm{E}[\frac{2c_1}{n^2}\sum_{i\neq j}(S_{ij}(t,s))^2 \\
&+\ \frac{2c_2}{n^2}\sum_{i\neq j\neq k}(S_{ij}(t,s)S_{ik}(t,s) + S_{ij}(t,s)S_{kj}(t,s)) \\
&+\ \frac{c_3}{n^2}\sum_{i\neq j\neq k\neq \ell} S_{ij}(t,s)S_{k\ell}(t,s)],
\end{aligned}$$

where $c_1 = \mathrm{E}(|U_i - U_j|^2)$, $i \neq j$; $c_2 = \mathrm{E}(|U_i - U_j||U_i - U_k|)$, $i \neq j \neq k$; $c_3 = \mathrm{E}(|U_i - U_j||U_k - U_\ell|)$, $i \neq j \neq k \neq \ell$.

Considering that $\mathrm{E}[(R_{ij}(t,s))^2] = \mathrm{E}[(Z_i - Z_j)_t^2(Z_i - Z_j)_s^2] \leq c_4$, $i \neq j$, $\forall t, s$, we have $\mathrm{E}[(R_{ij}(t,s))^2] = O(1)$, which implies $\mathrm{E}[(S_{ij}(t,s))^2] = O(1)$, and thus $\mathrm{E}[\frac{1}{n^2}\sum_{i\neq j}(S_{ij}(t,s))^2] = O(1)$.

After a simple calculation, we have $\sum_i S_{ij}(t,s) = 0$, $\sum_j S_{ij}(t,s) = 0$, $\sum_i \sum_j S_{ij}(t,s) = 0$, and

$$\begin{aligned}
&\sum_{i\neq j\neq k} S_{ij}(t,s)S_{ik}(t,s)\\
&= \sum_i (S_{ii}(t,s))^2 - \sum_{i\neq j}(S_{ij}(t,s))^2,\\
&\sum_{i\neq j\neq k} S_{ii}(t,s)S_{jk}(t,s)\\
&= \sum_i (S_{ii}(t,s))^2 - \sum_{i\neq j} S_{ii}(t,s)S_{jj}(t,s),\\
&\sum_{i\neq j\neq k\neq \ell} S_{ij}(t,s)S_{k\ell}(t,s)\\
&= -2\sum_{i\neq j\neq k}[S_{ii}(t,s)S_{jk}(t,s) + S_{ij}(t,s)S_{ik}(t,s)\\
&+ S_{ij}(t,s)S_{kj}(t,s)]\\
&- \sum_{i\neq j}[4S_{ii}(t,s)S_{ij}(t,s) + S_{ii}(t,s)S_{jj}(t,s)\\
&+ 2(S_{ij}(t,s))^2] - \sum_i (S_{ii}(t,s))^2,
\end{aligned}$$

we have

$$\begin{aligned}
\mathrm{E}[\frac{1}{n^2}\sum_{i\neq j\neq k} S_{ij}(t,s)S_{ik}(t,s)] &= O(1),\\
\mathrm{E}[\frac{1}{n^2}\sum_{i\neq j\neq k} S_{ij}(t,s)S_{kj}(t,s)] &= O(1),\\
\mathrm{E}[\frac{1}{n^2}\sum_{i\neq j\neq k} S_{ii}(t,s)S_{jk}(t,s)] &= O(1),\\
\mathrm{E}[\frac{1}{n^2}\sum_{i\neq j\neq k\neq \ell} S_{ij}(t,s)S_{k\ell}(t,s)] &= O(1).
\end{aligned}$$

Therefore, $\mathrm{E}[(M(t,s))^2] = O(1)$.

Applying Chebyshev's inequality to $M(t,s)$, we have

$$P(|M(t,s) - \mu| \geq k\sigma) \leq 1/k^2,$$

where $\mu = \mathrm{E}[M(t,s)]$, $\sigma^2 = \mathrm{Var}[M(t,s)]$. As a result, $M(t,s) = O_p(1)$.

Given that $\|\widehat{B} - B\|_F = O_p(n^{-1/2})$, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i,j}(G_{ij} - \frac{1}{n}\sum_k G_{kj} - \frac{1}{n}\sum_k G_{ik} + \frac{1}{n^2}\sum_{k,\ell} G_{k\ell})E_{ij}\\
&= tr[(B-\widehat{B})^T(B-\widehat{B})M]\\
&= pq^2 O_p(n^{-1})O_p(1)\\
&= O_p(n^{-1}).
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i,j}(G_{ij} - \frac{1}{n}\sum_k G_{kj} - \frac{1}{n}\sum_k G_{ik} + \frac{1}{n^2}\sum_{k,\ell} G_{k\ell})E_{kj},\\
&\frac{1}{n}\sum_{i,j}(G_{ij} - \frac{1}{n}\sum_k G_{kj} - \frac{1}{n}\sum_k G_{ik} + \frac{1}{n^2}\sum_{k,\ell} G_{k\ell})E_{ik},\\
&\frac{1}{n}\sum_{i,j}(G_{ij} - \frac{1}{n}\sum_k G_{kj} - \frac{1}{n}\sum_k G_{ik} + \frac{1}{n^2}\sum_{k,\ell} G_{k\ell})E_{k\ell}
\end{aligned}$$

are all $O_p(n^{-1})$. Therefore, $T_1 = O_p(n^{-1})$.

Analogous to (i), we can show (ii) $T_2 = O_p(n^{-1/2})$. The only differences are

$$\begin{aligned}
&\frac{1}{n}\sum_{i,j}(H_{ij} - \frac{1}{n}\sum_k H_{kj} - \frac{1}{n}\sum_k H_{ik} + \frac{1}{n^2}\sum_{k,\ell} H_{k\ell})E_{ij}\\
&= tr[(B-\widehat{B})M],
\end{aligned}$$

where M is defined similarly with $R_{ij} = (Z_i - Z_j)(V_i - V_j)^T$, and $\mathrm{E}[(R_{ij}(t,s))^2] = \mathrm{E}[(Z_i - Z_j)_t^2(V_i - V_j)_s^2] \leq c_5$, $i \neq j$, $\forall t, s$, and

$$\begin{aligned}
&\frac{1}{n}\sum_{i,j}(H_{ij} - \frac{1}{n}\sum_k H_{kj} - \frac{1}{n}\sum_k H_{ik} + \frac{1}{n^2}\sum_{k,\ell} H_{k\ell})E_{ij}\\
&= tr[(B-\widehat{B})M]\\
&= pqO_p(n^{-1/2})O_p(1)\\
&= O_p(n^{-1/2}),
\end{aligned}$$

and therefore $T_2 = O_p(n^{-1/2})$.

As a conclusion, $T = T_1 + T_2 = O_p(n^{-1/2})$. $\square$

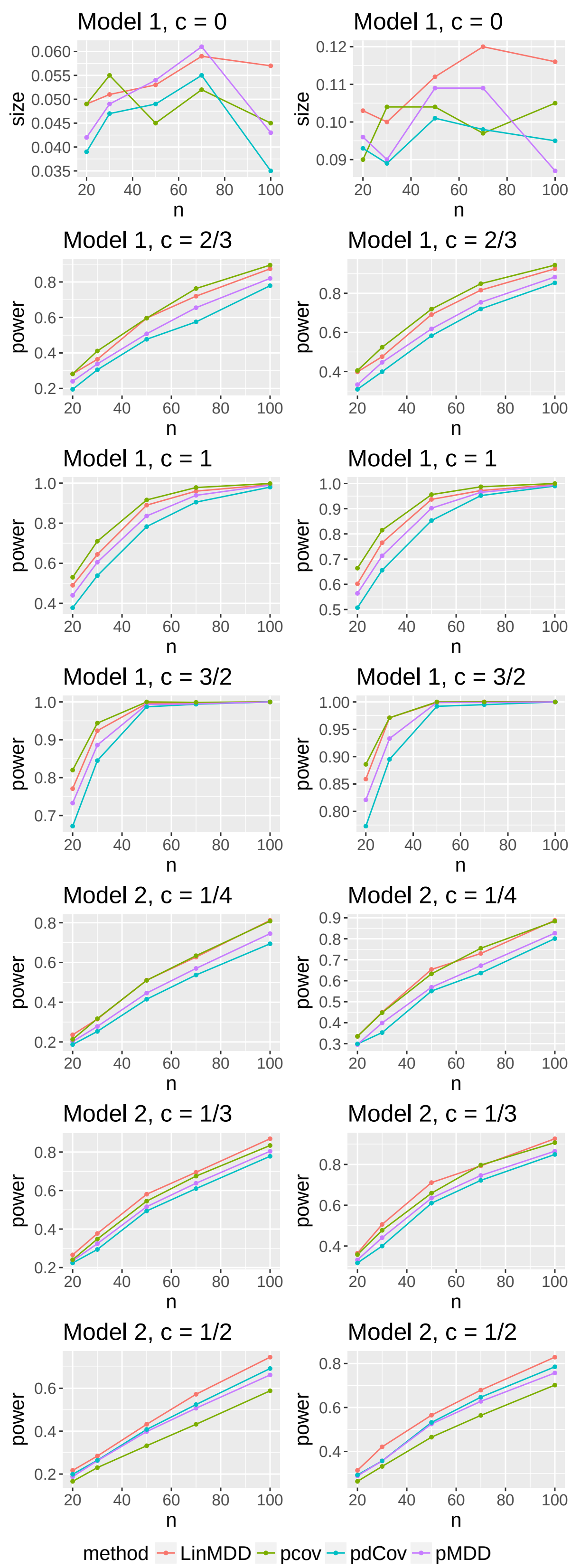

Figure 1: Empirical size and power of 1000 replications with $B = 500$ for Model 1 & 2.

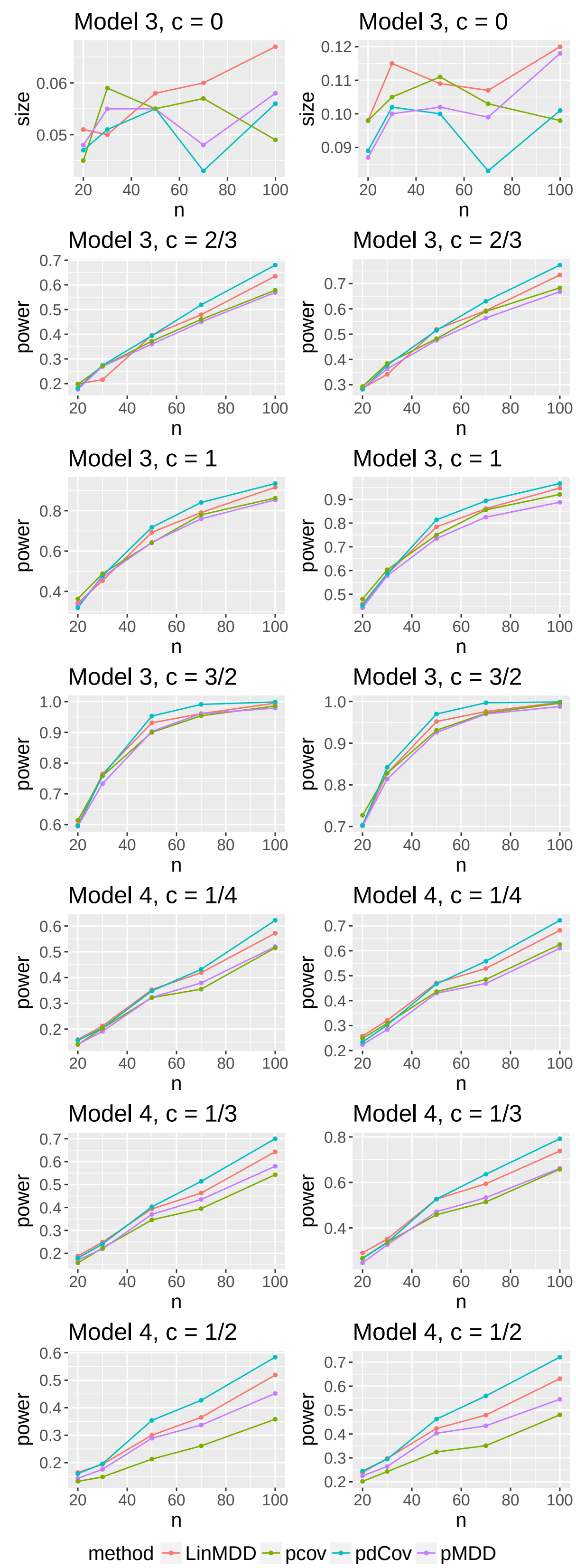

Figure 2: Empirical size and power of 1000 replications with $B = 500$ for Model 3 & 4.

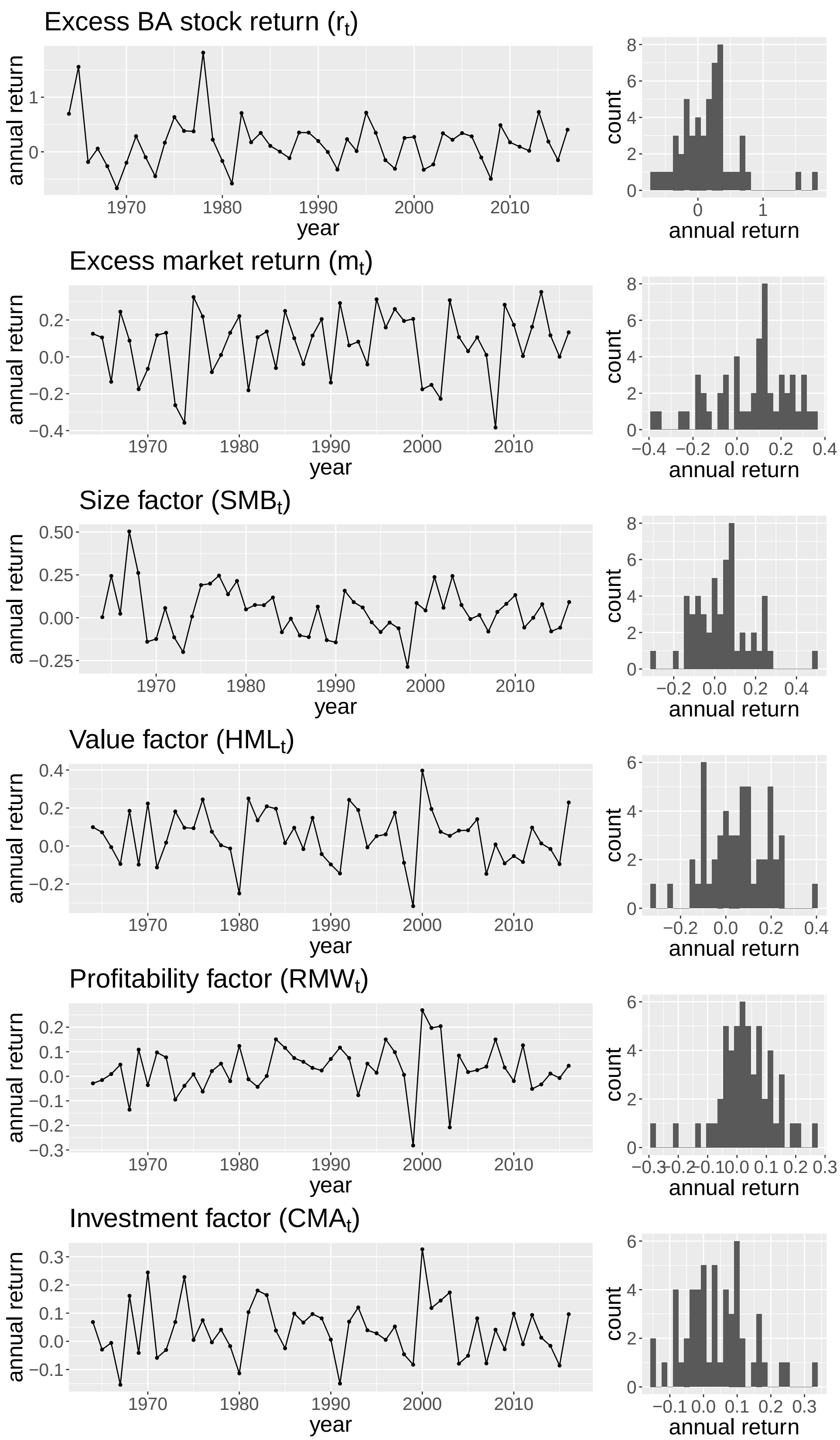

Figure 3: Time series and histograms of excess BA stock returns (r_t), excess market returns (m_t), size factors (SMB$_t$), value factors (HML$_t$), profitability factors (RMW$_t$), and investment factors (CMA$_t$) between 1964 and 2016.

References

Y. Aït-Sahalia, P. J. Bickel, and T. M. Stoker. Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of Econometrics*, 105(2):363–412, 2001.

S. Chatterjee and A. S. Hadi. *Regression analysis by example*. John Wiley & Sons, 2015.

B. Chen and Y. Hong. Testing for the markov property in time series. *Econometric Theory*, 28(1):130–178, 2012.

R. D. Cook and B. Li. Dimension reduction for conditional mean in regression. *Annals of Statistics*, pages 455–474, 2002.

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22, 2015.

J. Fan, Y. Liao, and M. Mincheva. High dimensional covariance matrix estimation in approximate factor models. *Annals of statistics*, 39(6):3320, 2011.

J. Fan, Y. Feng, and L. Xia. A conditional dependence measure with applications to undirected graphical models. *arXiv preprint arXiv:1501.01617*, 2015.

Y. Fan and Q. Li. Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica: Journal of the econometric society*, pages 865–890, 1996.

L. Gan, N. N. Narisetty, and F. Liang. Bayesian regularization for graphical models with unequal shrinkage. *Journal of the American Statistical Association*, (just-accepted), 2018.

C. Hiemstra and J. D. Jones. Testing for linear and nonlinear granger causality in the stock price-volume relation. *The Journal of Finance*, 49(5):1639–1664, 1994.

A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

Y. Hong, Y. Liu, and S. Wang. Granger causality in risk and detection of extreme risk spillover between financial markets. *Journal of Econometrics*, 150(2):271–287, 2009.

Z. Jin and D. S. Matteson. Generalizing distance covariance to measure and test multivariate mutual dependence. *arXiv preprint arXiv:1709.02532*, 2017.

Z. Jin and D. S. Matteson. Independent component analysis via energy-based and kernel-based mutual dependence measures. *arXiv preprint arXiv:1805.06639*, 2018.

Z. Jin, S. Yao, D. S. Matteson, and X. Shao. *EDMeasure: Energy-Based Dependence Measures*, 2018. R package version 1.2.

W. Lan, H. Wang, and C.-L. Tsai. Testing covariates in high-dimensional regression. *Annals of the Institute of Statistical Mathematics*, 66(2):279–301, 2014.

P. Lavergne and Q. Vuong. Nonparametric significance testing. *Econometric Theory*, 16(4):576–601, 2000.

R. Li, W. Zhong, and L. Zhu. Feature screening via distance correlation learning. *Journal of the American Statistical Association*, 107(499):1129–1139, 2012.

Z. Li and T. H. McCormick. An expectation conditional maximization approach for gaussian graphical models. *arXiv preprint arXiv:1709.06970*, 2017.

Z. R. Li, T. H. McCormick, and S. J. Clark. Bayesian joint spike-and-slab graphical lasso. *arXiv preprint arXiv:1805.07051*, 2018.

J. Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The review of economics and statistics*, pages 13–37, 1965.

J. Mossin. Equilibrium in a capital asset market. *Econometrica: Journal of the econometric society*, pages 768–783, 1966.

T. Park, X. Shao, and S. Yao. Partial martingale difference correlation. *Electronic Journal of Statistics*, 9(1): 1492–1517, 2015.

H. Raïssi et al. Testing linear causality in mean when the number of estimated parameters is high. *Electronic journal of statistics*, 5:507–533, 2011.

X. Shao and J. Zhang. Martingale difference correlation and its use in high-dimensional variable screening. *Journal of the American Statistical Association*, 109(507):1302–1318, 2014.

W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.

L. Su and H. White. A consistent characteristic function-based test for conditional independence. *Journal of Econometrics*, 141(2):807–834, 2007.

G. J. Székely and M. L. Rizzo. Partial distance correlation with methods for dissimilarities. *The Annals of Statistics*, 42(6):2382–2412, 2014.

G. J. Székely, M. L. Rizzo, and N. K. Bakirov. Measuring and testing dependence by correlation of distances. *The annals of statistics*, 35(6):2769–2794, 2007.

Y. Tang, H. J. Wang, and E. Barut. Testing for the presence of significant covariates through conditional marginal regression. *Biometrika*, 2017.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

A. Trapletti and K. Hornik. *tseries: Time Series Analysis and Computational Finance*, 2017. R package version 0.10-42.

X. Wang, W. Pan, W. Hu, Y. Tian, and H. Zhang. Conditional distance correlation. *Journal of the American Statistical Association*, 110(512):1726–1734, 2015.

X. Yan and J. Bien. Rare feature selection in high dimensions. *arXiv preprint arXiv:1803.06675*, 2018.

X. Yan, J. Bien, et al. Hierarchical sparse modeling: A choice of two group lasso formulations. *Statistical Science*, 32(4):531–560, 2017.

B. Zhang, N. Mohammed, V. Dave, and M. A. Hasan. Feature selection for classification under anonymity constraint. *arXiv preprint arXiv:1512.07158*, 2015.