Autoregressive Conditional Models for Interval-Valued Time Series Data

(Preliminary Version)

Ai Han
Chinese Academy of Sciences

Yongmiao Hong
Cornell University

Shouyang Wang
Chinese Academy of Sciences

September 2012

ABSTRACT

An interval-valued observation in a time period contains more information than a point-valued observation in the same time period. Examples of interval data include the maximum and minimum temperatures in a day, the maximum and minimum GDP growth rates in a year, the maximum and minimum asset prices in a trading day, the bid and ask prices in a trading period, the long term and short term interests, and the 90%-tile and 10%-tile incomes of a cohort in a year, etc. Interval forecasts may be of direct interest in practice, as it contains information on the range of variation and the level or trend of economic processes. Moreover, the informational advantage of interval data can be exploited for more efficient econometric estimation and inference.

We propose a new class of autoregressive conditional interval (ACI) models for interval-valued time series data. A minimum distance estimation method is proposed to estimate the parameters of an ACI model, and the consistency, asymptotic normality and asymptotic efficiency of the proposed estimator are established. It is shown that a two-stage minimum distance estimator is asymptotically most efficient among a class of minimum distance estimators, and it achieves the Cramer-Rao lower bound when the left and right bounds of the interval innovation process follow a bivariate normal distribution. Simulation studies show that the two-stage minimum distance estimator outperform conditional least squares estimators based on the ranges and/or midpoints of the interval sample, as well as the conditional quasi-maximum likelihood estimator based on the bivariate left and right bound information of the interval sample. In an empirical study on asset pricing, we document that when return interval data is used, some bond market factors, particularly the default risk factor, are significant in explaining excess stock returns, even after the stock market factors are controlled in regressions. This differs from the previous findings (e.g., Fama and French (1993)) in the literature.

Key Words: Asymptotic normality, Asset Pricing, Autoregressive conditional interval models, Interval time series, Mean squared error, Minimum distance estimation

JEL NO: C4, C2

1. Introduction

Time series analysis has been concerned with modelling the dynamics of a stochastic point-valued time series process. This paper is perhaps a first attempt to model the dynamics of a stochastic interval-valued time series which exhibits both 'range' and 'level' characteristics of the underlying process. A regular real-valued interval is a set of ordered real numbers defined by $\mathbf{y} = [a, b] = \{y \in \mathbf{R} | \quad a \leq y \leq b, \text{ where } a, b \in \mathbf{R}\}$. More generally, one can represent a certain region in the n-dimensional Euclidean space by an interval vector, that is, a n-tuple of intervals; see Moore, Kearfott and Cloud (2009). A stochastic interval time series is a sequence of interval-valued random variables indexed by time t.

There exists a relatively large body of evidence of interval-valued data in economics and finance. In microeconomics, interval-valued observations are often used to provide rigorous enclosures of the actual point data due to incomplete information (e.g., Manski 2007). In time series analysis, however, interval data in a time period often contain richer information than point-based observations in the same period since an interval number captures both the 'range' (or 'volatility') and 'level' (or 'trend') characteristics of the underlying process. A well-known example of interval-valued time series processes is the daily temperatures, e.g., $[Y_{L,t}, Y_{R,t}]$, where the left and right bounds denote the minimum and maximum temperatures in day t respectively. In macroeconomics, the minimum and maximum annualized monthly GDP growth rates form an annual interval-valued GDP growth rate data that indicates the range within which it varies in a given year. In finance, an interval can be an alternative volatility measure, due to its dual natures in assessing the fluctuating range as well as the level of an asset price during a trading period, e.g., $P_t = [P_{L,t}, P_{R,t}]$. In the study of the dynamics of bid-ask price spread of an asset, one can construct an interval data $[Y_{L,t}, Y_{R,t}]$ to present the bid-ask price spread, where $Y_{L,t}$ and $Y_{R,t}$ are the ask and bid prices of the asset at time t. In asset pricing modelling, $Y_{L,t}$ and $Y_{R,t}$ denote the risk-free and equity returns, respectively. Besides the interval-valued observations formed by the minimum and maximum point observations, quantile-based data are also informative. In study of income inequality, for example, the lowest 10% and highest 10% quantiles of the incomes of a cohort can be used as a robust measure of income inequality.

Interval forecasts may be of direct interest in practice because, compared to point forecasts, intervals contain rich information about the range of variation and the level of economic processes. Engle and Russell (2009) argued that intraday financial time series reveal subtle characteristics, e.g., irregular temporal spacing, strong diurnal patterns and complex dependence that present obstacles for traditional forecasting methods. In addition, it is rather difficult to accurately forecast the entire sequence of intraday prices for one day ahead. Thus, interval modelling may be an alternate away to analyze intraday time series. Examples are interval forecasts of temperatures, GDP growth rates, inflation rates, bid and ask prices, as well as long-term and short term interest rates in a given time period.

Since an interval observation in a time period provides more information than a point-valued observation in the same time period, this informational advantage can be exploited for more efficient estimation and inference in econometrics. To elaborate this, let us consider volatility modelling as an example, which has been a central theme in financial econometrics. Most studies on volatility modelling employ point-based data, e.g., the daily closing price of an asset rather than the interval data consisting of the maximum and minimum prices in a trading day. This is the case for the popular GARCH and Stochastic Volatility (SV) models in the literature. Although GARCH and SV models aim to study the dynamics of volatility of an asset price, the closing price observations fail to capture the 'fluctuation' information within a time period. A development in the literature that improves upon GARCH and SV models is to use range observations, based on the difference between the maximum and minimum asset prices in a time period, which are more informative than returns based on closing prices. Early models of this class include Parkinson (1980) and Beckers (1983). Recently, Alizadeh, Brandt and Diebold (2002) have used range observations of stock prices to obtain more efficient estimation for SV models. See also Diebold and Yilmaz (2009) for the use of range observations as measures for volatility. Chou (2005), on the other hand, develops a class of Conditional Autoregressive Range (CARR) models to capture the dynamics of the range of an asset price. Chou (2005) documents that CARR models have better forecasts of volatility than GARCH models, indicating the gain of utilizing range data over point-valued closing price data. However, an inherent problem of the CARR models is that using range as a volatility measure is unable to simultaneously capture the dual empirical features, i.e., 'range' and 'level'. For example, the same range observations in different time periods yield the same information for range, yet possible distinct price levels are ignored.

It is possible to capture the dual features of range and level by a bivariate point-valued model for the left and right bounds of an interval process. Existing methods include modelling the two univariate point-valued processes separately or joint modelling with vector autogression; see Maia et al (2008), Arroyo (2010), Neto et al. (2008), Neto and Carvalho (2010) and the references therein. However, a bivariate point-valued sample may not efficiently make use of the information of the underlying interval process; see Blanco-Ferández et al (2011). Furthermore, a certain region which an interval vector presents, e.g., a squared box which a bivariate interval vector presents, contains at least twice simultaneous equations as a single interval model, which may involve a large number of unknown parameters.

To capture the dynamics of an interval process, to forecast an interval and to explore the potential gain of using interval time series data over using point-valued time series data, we propose a new class of autoregressive conditional interval (ACIX thereof) models for interval-valued time series processes, possibly with exogenous explanatory variables. We develop an asymptotic theory for estimation, testing and inference. In addition to direct interest in interval forecasts by policy makers and practitioners, the advantages of ACIX models over the existing volatility and range models are at least twofold. First, it utilizes the information of both range and level contained in interval data, and thus it is expected to yield more efficient estimation and inference than point-valued data. Consider a case in which the interest is to model the conditional range of the daily price of some asset but there are more variations in the level sample than in the range sample. Because range and level are generally correlated, it may not be efficient to estimate parameters in a range model by using the range information alone. Instead, one may obtain more efficient parameter estimation for an ACIX model with an interval sample, thus providing more accurate forecasts for range.

A parsimonious ACIX model provides a simple and convenient unified framework to infer the dynamics of the interval population, which can also be used to derive some important point-based time series models as special cases. For example, when interval data are transformed to the point-valued 'range', the ACIX model then yields an ARMAX-type range model, which is an alternative to Chou's (2005) CARR model. Because our approach is based on the concept of extended interval for which the left bound needs not to be smaller than the right bound, the aforementioned advantages of our methodology also carry over to a large class of point-valued regression models, where the regressand and regressors are defined as differences between economic variables. See Section 7 for an example of capital asset pricing modelling (Fama and French (1993)).

The remainder of this paper is organized as follows. Section 2 introduces basic algebra of intervals, interval time series, and the class of ACIX models. In Section 3, we propose a minimum distance estimation method and establish the asymptotic theory of consistency and normality of the proposed estimators. We also show how various estimators for the point-based models can be derived as special cases of the proposed minimum distance estimator. Section 4 derives the optimal kernel function that yields the asymptotic most efficient minimum distance estimator, and proposes a feasible asymptotically most efficient two-stage minimum distance estimator. Section 5 develops a Lagrange Multiplier test and a Wald test for the hypotheses on model parameters. Section 6 presents a simulation study, comparing the performance of the proposed two-stage minimum distance estimator with various parameter estimators in finite samples. It is confirmed that more efficient parameter estimation can be obtained when interval data rather than point-valued data are utilized, and the proposed two-stage minimum distance estimator perform the best. Section 7 is an empirical study of Fama-French's (1993) asset pricing model, comparing the OLS estimator and the proposed two stage interval-based minimum distance estimator. We document that the use of interval risk premium data yields overwhelming evidence that the default risk factor is significant in explaining excess stock returns even when stock risk factors are controlled, a result that the previous literature and the OLS estimation fail to reveal (see Fama and French 1993). Section 8 concludes the paper. All mathematical proofs are collected in the Mathematical Appendix.

2. Interval Time Series and ACIX Model

In this section, we first introduce some basis concepts and analytic tools for stochastic interval time series. We then propose a parsimonious class of autoregressive conditional interval models with exogenous explanatory variables (ACIX) to capture the dynamics of interval time series processes. Both static and dynamic interval time series regression models are included as special cases.

2.1 Preliminary

To begin with, we first define an extended random interval.

Definition 2.1: An extended random interval Y on a probability space $(\Omega, \mathcal{F}, P)$ is a measurable mapping $Y : \Omega \to I_{\mathbf{R}}$, where $I_{\mathbf{R}}$ is the space of closed sets of ordered numbers in $\mathbf{R}$, *as* $Y(\omega) = [Y_L(\omega), Y_R(\omega)]$, where $Y_L(\omega), Y_R(\omega) \in \mathbf{R}$ for all $\omega \in \Omega$ *denote the left and right bounds of $Y(\omega)$ respectively, together with the following three compositions called addition, scalar multiplication and difference, respectively:*

(i) *Addition, symbolized by $+$, which is a binary composition in $I_{\mathbf{R}}$:*

$$A + B = [A_L + B_L, A_R + B_R];$$

(ii) *Scalar multiplication, symbolized by $\cdot$, which is a symmetric function from* $\mathbf{R} \times I_{\mathbf{R}}$ *to $I_{\mathbf{R}}$:*

$$\beta \cdot A = [\beta \cdot A_L, \beta \cdot A_R];$$

(iii) *Difference (Hukuhara (1967)), symbolized by $-_H$, which is a binary composition in $I_{\mathbf{R}}$:*

$$A -_H B = [A_L - B_L, A_R - B_R].$$

As a special case, a real-valued scalar $a \in \mathbf{R}$ can be presented by a 'degenerate interval', or a 'trivial interval' such that $a = [a, a]$. An example of degenerate intervals is the zero interval: $A = [0, 0]$. The mapping $Y : \Omega \to I_{\mathbf{R}}$ in Definition 2.1 is 'strongly measurable' with the σ-field generated by the topology induced by the Hausdorff metric d_H; see Li, Ogura, and Kreinovich

(2002, *Definition 1.2.1*). Specifically, for each interval X, we have $Y^{-1}(X) \in \mathcal{F}$, where $Y^{-1}(X) = \{\omega \in \Omega : Y(\omega) \cap X \neq \phi\}$ is the inverse image of Y.

For each $\omega \in \Omega$, $Y(\omega)$ is a set of ordered real-valued numbers, changing continuously from $Y_L(\omega)$ to $Y_R(\omega)$. To define the probability distribution of an extended random interval Y, we denote the Borel field of $I_{\mathbf{R}}$ as $\mathbf{B}(I_{\mathbf{R}})$. Given a $\mathbf{B}(I_{\mathbf{R}})$-measurable random interval Y, we define a sub-σ-field $\mathcal{F}_Y$ by

$$\mathcal{F}_Y = \sigma\left\{Y^{-1}(\upsilon), \upsilon \in \mathbf{B}(I_{\mathbf{R}})\right\},$$

where $Y^{-1}(\upsilon) = \{\omega \in \Omega : Y(\omega) \in \upsilon\}$. Then $\mathcal{F}_Y$ is a sub-σ-field of $\mathcal{F}$ with respect to which Y is measurable. The distribution of a random interval Y is a probability measure P on $\mathbf{B}(I_{\mathbf{R}})$ defined by

$$F_Y(\upsilon) = P\left[Y^{-1}(\upsilon)\right], \upsilon \in \mathbf{B}(I_{\mathbf{R}}).$$

Consider as an example the interval in which the S&P 500 stock index in day t fluctuates as an extended random interval Y_t defined on the probability space $(\Omega, \mathcal{F}, P)$, and the outcome of the experiment corresponds to a point $\omega \in \Omega$. Then the measuring process is carried out to obtain an interval in day t: $Y_t(\omega) = [Y_{L,t}(\omega), Y_{R,t}(\omega)]$. Unlike a bivariate random vector $X : \Omega_X \to \mathbf{R}^2$ of the left and right boundaries of Y where $X(\omega_X) = (Y_L(\omega_X), Y_R(\omega_X))'$ for $\omega_X \in \Omega_X$, the measurable mapping $Y : \Omega \to I_{\mathbf{R}}$ is a univariate random set of ordered numbers in the space of $I_{\mathbf{R}}$. Unless there exists a probability measure P_X on $\mathbf{B}(\mathbf{R}^2)$ such that

$$P_X\left[X^{-1}(\upsilon_X)\right] = P\left[Y^{-1}(\upsilon)\right],$$

for each $\upsilon_X \in \mathbf{B}(\mathbf{R}^2)$ and $\upsilon \in \mathbf{B}(I_{\mathbf{R}})$ such that $Y_L(\omega_X) = Y_L(\omega)$, $Y_R(\omega_X) = Y_R(\omega)$ and $X^{-1}(\upsilon_X) = \{\omega_X \in \Omega_X : X(\omega_X) \in \upsilon_X\}$, modelling an interval population Y cannot be simply equated to joint modelling a bivariate point-valued random vector for the left and right bounds of Y. The latter approach may lead to some information loss because it may not retain all information in a set of ordered numbers for each interval observation due to the fact that the two probability measures are not identical.

In Definition 2.1, we do not impose the conventional restriction of $Y_L \leq Y_R$ for regular intervals that has been imposed in the interval computing literature (see Moore, Kearfott, and Cloud (2009)). This is the reason we call Y as an extended interval. Our extension ensures the completeness of $I_{\mathbf{R}}$ and the consistency among the compositions introduced in Definition 2.1. Let $\beta = -1$ and $Y_t = [1, 3]$, for example. Then the extension ensures that $\beta Y_t = -1 \times [1, 3] = [-1, -3] \in I_{\mathbf{R}}$.[1]

[1] Our notation embodies a convention we follow throughout: the scalar multiplication, e.g., $\beta \cdot A$ will be presented as βA.

This is not a regular interval. Furthermore, $\forall \beta \in \mathbf{R}, Y_t \in I_{\mathbf{R}}$,

$$\beta Y_t + (-\beta) Y_t = [\beta Y_{L,t} - \beta Y_{L,t}, \beta Y_{R,t} - \beta Y_{R,t}] = [0, 0],$$

which implies that a symmetric element with respect to the addition exists. Conversely,

$$[0, 0] -_H (-\beta) Y_t = [0 + \beta Y_{L,t}, 0 + \beta Y_{R,t}] = \beta Y_t.$$

The concept of extended interval together with Hukuhara's difference is useful and suitable for econometric analysis of interval data. One example is the first difference of some interval process X_t:

$$Y_t = X_t -_H X_{t-1} = [X_{L,t} - X_{L,t-1}, X_{R,t} - X_{R,t-1}],$$

which becomes a stationary interval process although the original series X_t is not. Hukuhara introduced this difference operation to deal with the fact that the regular interval space, i.e., with the restriction $Y_{L,t} \leq Y_{R,t}$, is not a linear space due to the lack of a symmetric element with respect to the addition operation, which is addressed by our extension of the interval space.

Definition 2.1 also greatly extends the scope of applications of our methodology. For example, it covers the case of an extended interval with the risk-free rate as the left bound and the market portfolio return as the right bound, where the risk-free rate is not necessarily smaller than the market portfolio return. See Section 7 for applications to asset pricing modelling.

The concept of extended random interval differs from that of a confidence interval in statistical analysis, even if we impose the restriction $Y_L \leq Y_R$. The objective here is to learn about the probability distribution of an 'interval population' rather than a 'point population', and the forecast aims at the 'true interval' or the 'conditional interval expectation' of the underlying stochastic interval process. In contrast, the conventional confidence interval of a point-valued time series is to learn about the uncertainty or dispersion of a point population or its estimator given a prespecified confidence level.

Next, we define a stochastic interval time series process.

Definition 2.2: A stochastic interval time series process is a sequence of extended random intervals indexed by time $t \in \mathbf{Z} \equiv \{0, \pm 1, \pm 2, ...\}$, *denoted* $\{Y_t = [Y_{L,t}, Y_{R,t}]\}_{t=-\infty}^{\infty}$.

A segment $\{Y_1, Y_2, ..., Y_T\}$ from $t = 1$ to T of the interval time series $\{Y_t\}$ constitutes an interval time series random sample of size T. A realization of this random sample, denoted as $\{y_1, y_2, ..., y_T\}$, is called an interval time series data set with size T. The main objective here is to use the observed interval data to infer the dynamic structure of the interval time series process $\{Y_t\}$ and to use it for forecasts and other applications. For example, a leading object of interest is the conditional mean $E(Y_t|I_{t-1})$, where $I_{t-1} = \{Y_{t-1}, ..., Y_1\}$ is the information set available at time $t-1$.

Following Aumann' s (1965) definition of expectation of random sets, we now introduce the expectation of extended random intervals.

Definition 2.3: If Y_t is an extended random interval on $(\Omega, \mathcal{F}, P)$, then the expectation of Y_t is an extended interval defined by

$$\mu_t \equiv E(Y_t) = \left\{E\left(f\right)|f : \Omega \to \mathbf{R},\ f \in L^1,\ f \in Y_t \text{ a.s. } [P]\right\}$$

provided $E\left(|Y_t|\right) < \infty$ with $|Y_t| = \sup\{|y|,\ y \in Y_t(\omega)|$.

In order to quantify the variation of a random interval Y_t around its expectation μ_t, to define the autocovariance function of an interval time series process $\{Y_t\}$, and particularly to develop a minimum distance estimation method for an interval time series model, we need a suitable distance measure between intervals.

The basic idea of a distance measure between intervals is to consider the set of the absolute differences between all possible pairs of elements (points) of the intervals A and B, with respect to a suitable weighting function. The Hausdorff metric d_H (Munkres, 1999) has been widely used in measuring the distance between random sets (e.g., Artstein and Vitale (1975), Puri and Ralescu (1983, 1985), Cressie (1978), Hiai (1984), and Li, Ogura and Kreinovich (2002)). It is defined on a normed space Φ as follows:

$$d_H(A, B) = \max\left\{\sup_{a\in A}\inf_{b\in B} d(a, b), \sup_{b\in B}\inf_{a\in A} d(a, b)\right\},$$

where $d(a, b) = \|a - b\|_\Phi$ is the norm defined on Φ, and $A, B \in \varrho(\Phi)$ which is the family of all non-empty subsets of Φ. If Φ is a p-dimensional Euclidean space $\mathbf{R}^p$, $d_H(A, B)$ can be written as

$$d_H(A, B) = \max\left\{\sup_{a\in A} d(a, B), \sup_{b\in B} d(b, A)\right\} = \sup_{u\in\mathbf{S}^{p-1}} |s_A(u) - s_B(u)|, \quad (2.1)$$

where $\mathbf{S}^{p-1} = \{u \in \mathbf{R}^{p-1} : \|u\|_{\mathbf{R}^{p-1}} = 1\}$ is the unit sphere in $\mathbf{R}^p$, and $s_A(u)$ is called a *support function* of the set A defined as

$$s_A(u) = \sup_{a\in A} \langle u, a\rangle,\ u \in \mathbf{R}^{p-1}, \quad (2.2)$$

where $\langle\cdot,\cdot\rangle$ is an inner product. See Minkowsky (1911).

Eq.(2.1) indicates that d_H only considers the least upper bound of the set of absolute differences between all pairs of support functions in $p-1$ directions of tangent planes with weight 1. As shown in Näther (1997, 2000), the Fréchet expectation of a random set Y_t is not with respect to d_H. As a special case of random sets, the interval expectation $E(Y_t|I_{t-1})$ is not the optimal solution of the minimization problem, namely,

$$E(Y_t|I_{t-1}) \neq \arg\min_{A \in I_{\mathbf{R}}} E\left[d_H^2(Y_t, A(I_{t-1}))\right].$$

Thus, d_H is not a suitable metric to develop a minimum distance estimation method for time series models of conditional expectation of an interval process.

Körner and Näther (2002) developed a distance measure called D_K metric. For any pair of sets $A, B \in \mathcal{F}_c(\mathbf{R}^p)$,

$$D_K(A, B) = \sqrt{\int_{(u,v)\in \mathbf{S}^{p-1}} [s_A(u) - s_B(u)]\,[s_A(v) - s_B(v)]\, dK(u, v)},$$

where $\mathcal{F}_c(\mathbf{R}^p)$ is the space of convex compact sets, $\langle \cdot, \cdot \rangle_K$ denote the inner product in $\mathbf{S}^{p-1}$ with respect to kernel $K(u, v)$, and $K(u, v)$ is a symmetric positive definite weighting function on $\mathbf{S}^{p-1}$ which ensures that $D_K(A, B)$ is a metric for $\mathcal{F}_c(\mathbf{R}^p)$. When $p = 1$, the above random sets will be referred to the extended random intervals, and the generalized $\mathcal{F}_c(\mathbf{R})$ space is $I_{\mathbf{R}}$. For any pair of extended intervals $A, B \in I_{\mathbf{R}}$,

$$D_K(A, B) = \sqrt{\int_{(u,v)\in \mathbf{S}^{0}} [s_A(u) - s_B(u)]\,[s_A(v) - s_B(v)]\, dK(u, v)}, \tag{2.3}$$

where the unit space $\mathbf{S}^0 = \{u \in \mathbf{R}^1, |u| = 1\} = \{1, -1\}$ is a set consisting of 1 and -1. Here, the support function becomes

$$\begin{aligned} s_A(u) &= \begin{cases} \sup_{a\in A}\{u \cdot a | u \in \mathbf{S}^0\} & \text{if } A_L \le A_R, \\ \inf_{a\in A}\{u \cdot a | u \in \mathbf{S}^0\} & \text{if } A_R < A_L, \end{cases} \\ &= \begin{cases} A_R & u = 1, \\ -A_L & u = -1, \end{cases} \end{aligned} \tag{2.4}$$

and $s_A(u) = A$ if A is a degenerate interval as $A = A_L = A_R$.

The space of support functions $s_A(u)$ in Eq.(2.4) is linear, namely

$$\begin{aligned} s_{A+B} &= s_A + s_B, \\ s_{\lambda A} &= \lambda s_A, \text{ for all } \lambda \in \mathbf{R}, \\ s_{A-B} &= s_A - s_B. \end{aligned} \tag{2.5}$$

The usual support function in Eq.(2.2) is sublinear since that $s_{\lambda A} = \lambda s_A$ only holds for $\lambda \ge 0$. The extension of the regular interval space, which allows $A_L > A_R$ for $I_{\mathbf{R}}$, ensures that it holds for all $\lambda \in \mathbf{R}$. When $A_L \le A_R$, it is the usual support function as in Eq.(2.2). The result that $s_{A-B} = s_A - s_B$ shows that the support function of an extended interval produced from the Hukuhara difference between two intervals, is equal to the difference between the corresponding support functions of the two intervals. For more discussions on support functions, see Choi and Smith (2003), Romanowska and Smith (1989), and Li, Ogura, and Kreinovich (2002, Corollary 1.1.10).

The kernel $K(u, v)$ is a symmetric positive definite function such that for $u, v \in \mathbf{S}^0 = \{1, -1\}$,

$$\begin{cases} K(1,1) > 0, \\ K(1,1)K(-1,-1) > K(1,-1)^2, \\ K(1,-1) = K(-1,1). \end{cases} \tag{2.6}$$

For $A, B \in I_{\mathbf{R}}$, the mapping $\langle \cdot, \cdot \rangle_K : I_{\mathbf{R}} \to \mathbf{R}$ is a *linear* functional on $I_{\mathbf{R}.}$, with respect to any kernel K satisfying Eq.(2.6). This is because that the support functions form an inner product space (or unitary space), provided the inner product with respect to kernel K for each $A, B, C \in I_{\mathbf{R}}$ satisfies the following operation rules:

$$\begin{aligned} \langle s_A, s_B \rangle_K &= \langle s_B, s_A \rangle_K \,, \\ \langle s_{A+B}, s_C \rangle_K &= \langle s_A, s_C \rangle_K + \langle s_B, s_C \rangle_K \,, \\ \langle s_{\lambda A}, s_B \rangle_K &= \lambda \langle s_A, s_B \rangle_K \,, \text{ for all } \lambda \in \mathbf{R}, \\ \langle s_A, s_A \rangle_K &\geq 0, \\ \langle s_A, s_A \rangle_K &= 0 \text{ iff } A = [0, 0]. \end{aligned} \tag{2.7}$$

The norm for $A \in I_{\mathbf{R}}$ with respect to kernel K is defined as the nonnegative square root of $\langle s_A, s_A \rangle_K$,[2] i.e.,

$$\|A\|_K = D_K(A, [0,0]) = \langle s_A, s_A \rangle_K^{1/2} \,, \tag{2.8}$$

and similarly,

$$\|A - B\|_K = D_K(A, B) = \langle s_{A-B}, s_{A-B} \rangle_K^{1/2} \,. \tag{2.9}$$

The D_K-metric has some desirable properties. Most importantly, $s_A(u)$ is an isometry between $I_{\mathbf{R}}$ and a cone of the Hilbert subspace endowed with the generic L_2-type D_K distance respect to $K(u, v)$, which implies the suitability for the least squares estimation method of time series models for conditional mean of an interval process.

Lemma 2.1: Suppose $A(I_{t-1})$ is a measurable interval function of information set I_{t-1}. Then

$$E(Y_t | I_{t-1}) = \arg\min_{A \in I_{\mathbf{R}}} E\left[D_K^2(Y_t, A(I_{t-1})) \right]. \tag{2.10}$$

[2] Our notation embodies a convention we follow throughout: the support function $s_A(u)$ will be represented as s_A when u is not specified with particular values, while the difference $A -_H B$ will be represented as $A - B$.

See Näther (1997, 2000) for a generalized result of random sets, but not in a time series context.

Numerically the $D_K(A,B)$ in Eq.(2.3) has a simple quadratic form and is easy to compute. It follows from the definitions of $s_A(u)$ and $K(u,v)$ that

$$\begin{aligned} D_K^2(A,B) &= K(1,1)\,(A_R - B_R)^2 + K(-1,-1)(A_L - B_L)^2 - 2K(1,-1)(A_R - B_R)(A_L - B_L) \\ &= \begin{bmatrix} A_R - B_R \\ -(A_L - B_L) \end{bmatrix}' \begin{bmatrix} K(1,1) & K(1,-1) \\ K(-1,1) & K(-1,-1) \end{bmatrix} \begin{bmatrix} A_R - B_R \\ -(A_L - B_L) \end{bmatrix}. \end{aligned} \tag{2.11}$$

Recall that the crucial criterion of a distance between intervals A and B is to consider the set of the absolute differences between all possible pairs of elements (points) of A and B, with a proper weighting function to include the maximum amount of useful information contained in intervals. However, Eq.(2.11) might lead to a misunderstanding that $D_K^2(A,B)$ only considers a weighted average of distances between the two boundary points of intervals A and B, and ignores the distances between interior points. Below we elaborate $s_A(u)$ and $K(u,v)$ to gain insight into the numerical equality in Eq.(2.11).

The support function $s_A(u)$ is an alternate representation of $A \in I_{\mathbf{R}}$ in terms of the positions of two tangent planes, i.e., the left and right bounds, that enclose the interval A. Li, Ogura and Kreinovich (2002, Corollary 1.2.8) verify that $s_A(u)$ of the extended random interval A defined on $(\Omega, \mathcal{F}, P)$ is measurable, by which we can derive any point-valued random variable $A^{(\lambda)}(\omega) \in A(\omega)$:

$$A^{(\lambda)}(\omega) = \lambda s_{A(\omega)}(1) - (1-\lambda) s_{A(\omega)}(-1) = \lambda A_R + (1-\lambda) A_L \tag{2.12}$$

for $\lambda \in [0,1]$. For instance, for each $\omega \in \Omega$, $\lambda = 0$, 1 and 0.5 yield the left and right bounds, and the midpoint of $A(\omega)$ respectively:

$$\begin{aligned} A_L(\omega) &\equiv A^{(0)}(\omega) = -s_{A(\omega)}(-1), \\ A_R(\omega) &\equiv A^{(1)}(\omega) = s_{A(\omega)}(1), \\ A^m(\omega) &\equiv A^{(0.5)}(\omega) = \frac{1}{2} s_{A(\omega)}(1) - \frac{1}{2} s_{A(\omega)}(-1). \end{aligned} \tag{2.13}$$

Bertoluzza et al. (1995) introduced a d_W distance for intervals, which was later generalized to the D_K metric by Körner and Näther (2002). The d_W distance is defined as

$$d_W(A,B) = \sqrt{\int_{[0,1]} (A^{(\lambda)} - B^{(\lambda)})^2 dW(\lambda)} \quad \text{, for all } A, B \in I_{\mathbf{R}}$$

where $W(\lambda)$ is a probability measure on the real Borel space $([0,1])$, $\mathbf{B}([0,1])$. The $d_W(A,B)$ measure involves not only distances between extreme points with weights $W(0)$ and $W(1)$, but also distances between interior points in the intervals with weights $W(\lambda)$, $0 < \lambda < 1$.

It is appealing that the D_K metric as a generalization of the d_W metric preserves this property (González-Rodríguez et al. (2007)). The simpler expression of the D_K metric in Eq.(2.11) than $d_W(A,B)$ lies in the fact that it measures the distance between each pair of points in intervals A and B in terms of the support functions,

$$\begin{aligned}\left(A^{(\lambda)}-B^{(\lambda)}\right)^2 &= \left[\lambda A_R+(1-\lambda)A_L-\lambda B_R-(1-\lambda)B_L\right]^2\\ &= \lambda^2\left(A_R-B_R\right)^2+(1-\lambda)^2(A_L-B_L)^2+2\lambda(1-\lambda)\left(A_R-B_R\right)\left(A_L-B_L\right).\end{aligned} \tag{2.14}$$

Instead of an integral for $(A^{(\lambda)}-B^{(\lambda)})^2$ with respect to $W(\lambda)$, Eq.(2.14) suggests that the value of $K(u,v)$ for each pair of $(u,v)\in\mathbf{S}^0$ can be interpreted as

$$\begin{aligned}K(1,1) &= \int_0^1 \lambda^2 dW(\lambda),\\ K(1,-1) &= K(-1,1)=\int_0^1 \lambda(\lambda-1)dW(\lambda),\\ K(-1,-1) &= \int_0^1 (1-\lambda)^2 dW(\lambda).\end{aligned}$$

These identities suggest that the choice of kernel K is equivalent to the choice of some weighting function $W(\lambda)$. Thus, although $D_K^2(A,B)$ can be simply computed by the distances between extreme points with respect to kernel $K(u,v)$, it is in essence an integral over the distances between all pairs of points in intervals A and B with a weighting function $W(\lambda)$ implied by the choice of $K(u,v)$.

To gain further insight into the role of kernel K, we now explore some special choices of kernel $K(u,v)$ and discuss their implication on capturing the information contained in intervals. For notational convenience, we denote a generic choice of a symmetric kernel K as $K(1,1)=a$, $K(1,-1)=K(-1,1)=b$, $K(-1,-1)=c$, where a, b and c satisfy Eq.(2.6).

Case 1. $(a,b,c)=(\frac{1}{4},-\frac{1}{4},\frac{1}{4})$.

This kernel K corresponds to the choice of weighting function $W(\lambda)$ as a degenerate distribution: $W(\lambda)=1$ for $\lambda=\frac{1}{2}$ and 0 otherwise. The D_K metric becomes

$$D_K^2(A,B)=\left(A^m-B^m\right)^2,$$

which measures the distance between midpoints of A and B. Note that kernel K is not positive definite here.

Case 2. $(a,b,c)=(1,1,1)$.

In this case, we have

$$D_K^2(A,B) = (A^r - B^r)^2\,,$$

which measures the distance between ranges of A and B. Note that kernel K is not positive definite here.

Case 3. $a = c$, $|b| < a$. Then by Eq.(2.11),

$$D_K^2(A,B) = \frac{a+b}{2}\,(A^r - B^r)^2 + 2(a-b)\,(A^m - B^m)^2\,.$$

This measures the distance between the ranges A^r and B^r, and the distance between the midpoints A^m and B^m, with weights $\frac{a+b}{2}$ and $2(a-b)$ respectively. If $-1 < \frac{b}{a} < \frac{3}{5}$, $(A^m - B^m)^2$ receives a larger weight than $(A^r - B^r)^2$; if $\frac{3}{5} < \frac{3}{5} < 1$, $(A^r - B^r)^2$ receives a larger weight than $(A^m - B^m)^2$; and if $\frac{b}{a} = \frac{3}{5}$, the squared differences between ranges and between midpoints receive the same weight.

Case 4. $b = 0$. Then by Eq.(2.11),

$$D_K^2(A,B) = a\,(A_R - B_R)^2 + c\,(A_L - B_L)^2\,.$$

This measures the distance between the left bounds and the distance between the right bounds, with weights a and c respectively. If $0 < a < c$, $(A_L - A_L)^2$ receives a larger weight than $(A_R - B_R)^2$; if $0 < c < a$, $(A_R - B_R)^2$ receives a larger weight than $(A_L - B_L)^2$; and if $0 < a = c$, the squared differences between left bounds and right bounds receive the same weight. The choice of such a kernel K is equivalent to the choice of weighting function $W(\lambda)$ which follows a Bernoulli distribution with $W(0) = c, W(1) = a$, where $a + c = 1$.

Case 5. Suppose $a \neq c$, $b \neq 0$, where a, b and c satisfy Eq.(2.6). Then by Eq.(2.11)

$$\begin{aligned}
& D_K^2(A,B) \\
= \;& a\,(A_R - B_R)^2 + c(A_L - B_L)^2 - 2b\,(A_R - B_R)\,(A_L - B_L) \\
= \;& \frac{a+2b+c}{2}\,(A^r - B^r)^2 + (a-2b+c)\,(A^m - B^m)^2 + (a-c)\,(A^r - B^r)\,(A^m - B^m)\,.
\end{aligned}$$

Here, $D_K^2(A,B)$ can capture information in the left bound difference $A_L - B_L$, the right bound difference $A_R - B_R$, and their cross product $(A_R - B_R)\,(A_L - B_L)$, or the information in the range difference $A^r - B^r$, the level difference $A^m - B^m$, and their cross product $(A^r - B^r)\,(A^m - B^m)$. The utilization of the cross product information will enhance estimation efficiency, as will be seen below.

2.2 Stationarity of an Interval Time Series Process

To introduce the concept of weak stationarity for the interval time series process $\{Y_t\}$, we first define the autocovariance function of $\{Y_t\}$ based on support function s_A and kernel K.

Definition 2.4: The j^{th}-order autocovariance function of a stochastic interval time series process $\{Y_t\}$, denoted $\gamma_t(j)$, is a scalar defined by

$$\gamma_t(j) \equiv \mathrm{cov}(Y_t, Y_{t-j}) = E\left\langle s_{Y_t} - s_{\mu_t}, s_{Y_{t-j}} - s_{\mu_{t-j}} \right\rangle_K, \ j = 0, \pm 1, ...,$$

where $\mu_t = E(Y_t)$, and $\left\langle s_{Y_t} - s_{\mu_t}, s_{Y_{t-j}} - s_{\mu_{t-j}} \right\rangle_K$ is the inner product with respect to the kernel $K(u,v)$ on $\mathbf{S}^0 = \{-1, 1\}$. In particular, the variance of Y_t is

$$\gamma_t(0) = E\left\|Y_t - \mu_t\right\|_K^2 = E\left[D_K^2(Y_t, \mu_t)\right] = E\left\langle s_{Y_t} - s_{\mu_t}, s_{Y_t} - s_{\mu_t}\right\rangle_K,$$

and $\gamma_t(j) = \gamma_t(-j)$ for all integers j, provided kernel $K(u,v)$ is symmetric.

Note that $\gamma_t(j)$ has the form of covariance between two random intervals X and Z:

$$\mathrm{cov}(X, Z) = E\left\langle s_X - s_{\mu_X}, s_Z - s_{\mu_Z}\right\rangle_K.$$

Thus $\gamma_t(j)$ could be interpreted as the covariance of Y_t with its lagged value Y_{t-j}. When $\{Y_t\}$ is a stochastic point-valued process, we have

$$E\left\langle s_{Y_t} - s_{\mu_t}, s_{Y_{t-j}} - s_{\mu_{t-j}} \right\rangle_K = E\left[(Y_t - \mu_t)(Y_{t-j} - \mu_{t-j})\right],$$

subject to the restriction that $\int_{(u,v)\in\mathbf{S}^0} dK(u,v) = K(1,1) + K(-1,-1) + 2K(1,-1) = 1$, which is consistent with the definition of the autocovariance function of a point-valued time series.

We now define weak stationarity of a stochastic interval time series process.

Definition 2.5: If neither the mean μ_t nor the autocovariance function $\gamma_t(j)$, for each j, of a stochastic interval time series process $\{Y_t\}$ depends on time t, then $\{Y_t\}$ is D_K-weakly stationary, or D_K-covariance stationary.

Suppose $\{Y_t\}$ is a D_K-weakly stationary interval process. Then a derived stochastic point-valued process according to Eq. (2.12) is also weakly stationary. Given Eq.(2.13) and the interval process Y_t, we can obtain a bivariate point-valued process of the left and right bounds of Y_t :

$$\begin{cases} Y_t^{(0)} & = Y_{L,t}, \\ Y_t^{(1)} & = Y_{R,t}, \end{cases}$$

the range (or difference) of Y_t as a measure of 'volatility'

$$Y_t^r \equiv Y_t^1 - Y_t^0 = s_{Y_t}(1) + s_{Y_t}(-1) = Y_{R,t} - Y_{L,t},$$

and the midpoint of Y_t as a measure of 'level'

$$Y_t^m \equiv Y_t^{0.5} = s_{Y_t}\left(\frac{1}{2}\right) = \frac{Y_{L,t} + Y_{R,t}}{2}.$$

These point processes are in essence measurable linear transformations of Y_t based on its support function, and as a result, their probabilistic properties are determined by $(\Omega, \digamma, P)$ on which Y_t is defined. Thus $\{Y_t^r\}$, $\{Y_t^m\}$, and the bivariate point process $\{(Y_{L,t}, Y_{R,t})'\}$ are all weakly stationary processes if Y_t is D_K-weakly stationary.

If $\gamma(j) = 0$ for all $j \neq 0$, we say that the D_K-weakly stationary interval process $\{Y_t\}$ is a D_K-*weakly white noise* process. This arises when $\{Y_t\}$ is an independent and identically distributed (i.i.d.) sequence. Of course, zero autocorrelation of $\{Y_t\}$ across different lags does not necessarily imply serial independence of $\{Y_t\}$, as is the case with the conventional time series analysis.

Next we define strict stationarity of a stochastic interval time series process.

Definition 2.6: Let P_1 be the joint distribution function of the stochastic interval time series sequence $\{Y_1, Y_2, ...\}$, and let $P_{\tau+1}$ be the joint distribution function of the stochastic interval time series sequence $\{Y_{\tau+1}, Y_{\tau+2}, ...\}$. The stochastic interval time series process $\{Y_t\}$ is strictly stationary if $P_{\tau+1} = P_1$ for all $\tau \geq 1$.

In accordance with Definition 2.6, we could introduce the concept of *ergodicity* for a strictly stationary interval process, which is essentially the same as that for a point-valued process. For more discussion on ergodicity, see White (1999, Definition 3.33).

2.3 Law of Large Numbers for D_K-Weakly Stationary Interval Processes

The strong law of large numbers with the Hausdorff metric d_H of i.i.d. random compact subsets of finite-dimensional Euclidean space $\mathbf{R}^d$ was first proved by Artstein and Vitale (1975), and further studied by Cressie (1978), Hiai (1984), and Puri and Ralescu (1983, 1985). Li, Ogura, and Kreinovich (2002) proved a strong law of large numbers for i.i.d. compact convex subsets of a separable Banach space with the Hausdorff metric d_H.

However, these limit theories are not available for the D_K metric, particularly in a time series context. Below, we prove the weak law of large numbers (WLLN) for both the first and second moments of a stationary interval process.

Theorem 2.1. Let $\{Y_t\}_{t=1}^T$ be a random interval sample of size T from a D_K-weakly stationary interval process $\{Y_t\}$ with $E(Y_t) = \mu$ for all t, $E\left\langle s_{Y_t} - s_\mu, s_{Y_{t-j}} - s_\mu \right\rangle_K = \gamma(j)$ for all t and j, and $\sum_{j=-\infty}^{\infty} |\gamma(j)| < \infty$. Then $\overline{Y}_T \xrightarrow{p} \mu$ as $T \to \infty$, where $\overline{Y}_T = T^{-1}\sum_{t=1}^T Y_t$ is the sample mean of $\{Y_t\}_{t=1}^T$, and the convergence is with respect to the D_K metric in the sense that $\lim_{T\to\infty} P\left[D_K(\overline{Y}_T, \mu) \geq \epsilon\right] = 0$, for any given constant $\epsilon > 0$.

Theorem 2.1 provides the conditions of ergodicity in mean for a stochastic interval time series process, that is, when the autocovariance function $\gamma(j)$ is absolutely summable, the sample mean $\overline{Y}_T$ converges to the population mean μ of a D_K-weakly stationary interval process $\{Y_t\}$. In Theorem 2.1, the sample average $\overline{Y}_T$ and the population mean μ are both defined on $I_{\mathbf{R}}$, i.e., both are interval-valued. When they are point-valued, we have

$$D_K(\overline{Y}_T, \mu) = d_H(\overline{Y}_T, \mu),$$

subject to $\int_{(u,v)\in \mathbf{S}^0} dK(u,v) = 1$. Thus, Theorem 2.1 coincides with the familiar weak law of large numbers for a point-valued time series process, i.e., $\lim_{T\to\infty} P\left[\left|\overline{Y}_T - \mu\right| \geq \epsilon\right] = 0$ for each $\epsilon > 0$.

Next, we show that the sample autocovariance of a stationary interval process converges in probability to its autocovariance.

Theorem 2.2. Let $\{Y_t\}_{t=1}^T$ be a random sample of size T from a stationary ergodic stochastic interval time series process $\{Y_t\}$ such that $E\left\|Y_t\right\|_K^4 < \infty$ for all t and $j \geq 0$. Suppose the conditions of Theorem 2.1 hold. Then for each given j,

$$\widehat{\gamma}(j) \equiv T^{-1} \sum_{t=j+1}^{T} \left\langle s_{Y_t} - s_{\overline{Y}_T}, s_{Y_{t-j}} - s_{\overline{Y}_T} \right\rangle_K \xrightarrow{p} \gamma(j)$$

as $T \to \infty$, where $\overline{Y}_T = T^{-1}\sum_{t=1}^{T} Y_t$ is the sample mean of $\{Y_t\}_{t=1}^T$.

Theorem 2.2 provides sufficient conditions that a D_K-weakly stationary interval process is ergodic in second moments. Since the weighted inner product $\left\langle s_{Y_t} - s_{\overline{Y}_T}, s_{Y_{t-j}} - s_{\overline{Y}_T} \right\rangle_K$ is a scalar, the convergence in probability in Theorem 2.2 is with respect to either the d_H or D_K metric.

2.4 Autoregressive Conditional Interval Models

To capture the dynamics of a stochastic interval process $\{Y_t\}$, we propose a class of Autoregressive Conditional Interval (ACI) Models of order (p,q):

$$Y_t = \alpha_0 + \beta_0 I_0 + \sum_{j=1}^{p} \beta_j Y_{t-j} + \sum_{j=1}^{q} \gamma_j u_{t-j} + u_t, \tag{2.15}$$

or compactly,

$$B(L)Y_t = \alpha_0 + \beta_0 I_0 + A(L)u_t$$

where α_0, β_j $(j = 0, ..., p)$, γ_j $(j = 1, ..., q)$ are unknown scalar parameters, $I_0 = [-\frac{1}{2}, \frac{1}{2}]$ is a unit interval; $\alpha_0 + \beta_0 I_0 = [\alpha_0 - \frac{1}{2}\beta_0, \alpha_0 + \frac{1}{2}\beta_0]$ is a constant interval intercept; $A(L) = 1 + \sum_{j=1}^{q} \gamma_j L^j$ and $B(L) = 1 - \sum_{j=1}^{p} \beta_j L^j$, where L is the lag operator; u_t is an interval innovation. We assume that $\{u_t\}$ is a interval martingale difference sequence (IMDS thereof) with respect to the information set I_{t-1}, that is, $E(u_t|I_{t-1}) = [0, 0]$ $a.s.$

The ACI(p,q) model is an interval generalization of the popular ARMA (p,q) model for a point-valued time series process. It can be used to forecast intervals of economic processes, such as the GDP growth rate, the inflation rate, the stock price, the long-term and short-term interest rates, and the bid-ask spread. This is often of direct interest for policy makers and practitioners. When $q=0$, Eq.(2.15) becomes an ACI$(p,0)$ model, analogous to an AR(p) model for a point-valued time series:

$$Y_t = \alpha_0 + \beta_0 I_0 + \sum_{j=1}^{p} \beta_j Y_{t-j} + u_t.$$

When $p=0$, Eq.(2.15) becomes an ACI$(0,q)$ model, analogous to an MA(q) model for a point-valued time series:

$$Y_t = \alpha_0 + \beta_0 I_0 + \sum_{j=1}^{q} \gamma_j u_{t-j} + u_t.$$

If all the roots of $B(z)=0$ lie outside the unit circle, an ACI(p,q) process can be rewritten as a distributed lag of $\{u_s, s \leq t\}$, which is an ACI(0,∞) process,

$$\begin{aligned} Y_t &= B(L)^{-1}(\alpha_0 + \beta_0 I_0) + B(L)^{-1}A(L)u_t \\ &= B(1)^{-1}(\alpha_0 + \beta_0 I_0) + \sum_{j=0}^{\infty} \alpha_j u_{t-j}, \end{aligned}$$

where $\{a_j\}$ is given by $B(L)^{-1}A(L) = \Sigma_{j=0}^{\infty} a_j L^j$. On the other hand, if all the roots of $A(z)=0$ lie outside the unit circle, an ACI(p,q) model is an invertible process with u_t expressed as the linear summation of $\{Y_s, s \leq t\}$, which is an ACI$(\infty, 0)$ process,

$$\begin{aligned} u_t &= A(L)^{-1}B(L)Y_t - A(L)^{-1}(\alpha_0 + \beta_0 I_0) \\ &= -A(1)^{-1}(\alpha_0 + \beta_0 I_0) + \sum_{j=0}^{\infty} \lambda_j Y_{t-j}, \end{aligned}$$

where $\{\lambda_j\}$ is given by $B(L)^{-1}A(L) = \Sigma_{j=0}^{\infty} \lambda_j L^j$.

The ACI(p,q) model of an interval process can be extended to the ACIX(p,q,s) model by incorporating exogenous explanatory interval variables:

$$Y_t = \alpha_0 + \beta_0 I_0 + \sum_{j=1}^{p} \beta_j Y_{t-j} + \sum_{j=1}^{q} \gamma_j u_{t-j} + \sum_{j=0}^{s} \delta_j' X_{t-j} + u_t, \tag{2.16}$$

where $X_t = (X_{1t}, ..., X_{Jt})'$ is an exogenous stationary interval vector process, and $\delta_j = (\delta_{j,1}, ..., \delta_{j,J})'$ is the corresponding point-valued parameter vector. When $q=0$, i.e., when there is no MA component, the ACIX$(p,0,s)$ model is an interval time series regression model:

$$Y_t = \alpha_0 + \beta_0 I_0 + \sum_{j=1}^{p} \beta_j Y_{t-j} + \sum_{j=0}^{s} \delta_j' X_{t-j} + u_t, \tag{2.17}$$

where all explanatory interval variables are observable. This covers both static (with $p = 0$) or dynamic (with $p > 0$) interval regression models.

ACIX(p,q,s) models can be used to capture temporal dependence in an interval process. In particular, it can be used to capture some well-known empirical stylized facts in economics and finance, such as volatility (or range) clustering and level effect (i.e., correlation between volatility and level). For example, $\beta_1 > 0$ indicates that a wide interval at time t is likely to be followed by another wide interval in the next period.

Another advantage of modelling an ACIX(p,q,s) process is that one can derive some important univariate point-valued ARMAX(p,q,s) models as special cases, provided the derived point models are defined by the support function as in Eq.(2.12). For example, by Eq.(2.12) and taking the difference between $Y_t^{(1)}$ and $Y_t^{(0)}$, the left and right bounds of an ACIX(p,q,s) model, we obtain an ARMAX(p,q,s) type range model

$$Y_t^r = \beta_0 + \sum_{j=1}^{p} \beta_j Y_{t-j}^r + \sum_{j=1}^{q} \gamma_j u_{t-j}^r + \sum_{j=0}^{s} \delta_j' X_{t-j}^r + u_t^r, \tag{2.18}$$

where u_t^r is a MDS such that $E(u_t^r|I_{t-1}) = E(u_{R,t} - u_{L,t}|I_{t-1}) = 0$ $a.s$, given $E(u_t|I_{t-1}) = [0,0]$ $a.s.$ This delivers an alternative dynamic range model to Chou (2005) for modelling the range dynamics of an asset price. The difference is that the derived range model in Eq.(2.18), with an ACIX(p,q,s) model as the data generating process, has an additive innovation while Chou (2005) has a multiplicative innovation. Our approach has an advantage, that is, we can use an interval sample, rather than the range sample only, to estimate more efficiently the ACIX model even if the interest is in range modelling.

Similarly, we can obtain an ARMAX(p,q,s) level model with $\lambda = \frac{1}{2}$ in Eq. (2.12):

$$Y_t^m = \alpha_0 + \sum_{j=1}^{p} \beta_j Y_{t-j}^m + \sum_{j=1}^{q} \gamma_j u_{t-j}^m + \sum_{j=0}^{s} \delta_j' X_{t-j}^m + u_t^m, \tag{2.19}$$

where u_t^m is a MDS such that $E(u_t^m|I_{t-1}) = E(\frac{1}{2}u_{L,t} + \frac{1}{2}u_{R,t}|I_{t-1}) = 0$ $a.s$, given $E(u_t|I_{t-1}) = 0$ $a.s.$

Finally, we can obtain a bivariate ARMAX(p,q,s) model for the boundaries of Y_t :

$$\begin{cases} Y_{Lt} = \alpha_0 - \frac{1}{2}\beta_0 + \sum_{j=1}^{p} \beta_j Y_{L,t-j} + \sum_{j=1}^{q} \gamma_j u_{L,t-j} + \sum_{j=0}^{s} \delta_j' X_{L,t-j} + u_{L,t}, \\ Y_{Rt} = \alpha_0 + \frac{1}{2}\beta_0 + \sum_{j=1}^{p} \beta_j Y_{R,t-j} + \sum_{j=1}^{q} \gamma_j u_{R,t-j} + \sum_{j=0}^{s} \delta_j' X_{R,t-j} + u_{R,t}, \end{cases} \tag{2.20}$$

where $E\left(u_{L,t}|I_{t-1}\right) = E(u_{R,t}|I_{t-1}) = 0$ $a.s.$ given $E\left(u_t|I_{t-1}\right) = [0,0]$ $a.s.$ A similar result to Eq.(2.20) can be obtained by combining Eq. (2.18) and Eq. (2.19), as a bivariate ARMAX model for the midpoint and range processes.

3. Minimum Distance Estimation

We now propose a minimum distance estimation method for an ACIX(p,q,s) model. We first impose a set of regularity conditions:

Assumption 1. $\{Y_t\}$ is a strictly stationary and ergodic interval stochastic process with $E\left\|Y_t\right\|_K^4 < \infty$, and it follows an ACIX$(p,q,s)$ process in (2.16), where the interval innovation u_t is an IMDS with respect to the information set I_{t-1}, that is, $E(u_t|I_{t-1}) = [0,0]$ $a.s.$, and $X_t = (X_{1t},...,X_{Jt})'$ is an exogenous strictly stationary ergodic interval vector process.

Assumption 2. Put $A(z) = 1 + \sum_{j=1}^{q}\gamma_j z^j$ and $B(z) = 1 - \sum_{j=1}^{p}\beta_j z^j$. The roots of $A(z) = 0$ and $B(z) = 0$ lie outside the unit circle $|z| = 1$.

Assumption 3. (i) The parameter space Θ is a finite-dimensional compact space of $\mathbf{R}^{p+q+s+2}$. (ii) θ^0 is an interior point in Θ, where $\theta^0 = (\alpha_0, \beta_0, \beta_1, ..., \beta_p, \gamma_1, ..., \gamma_q, \delta_0', ..., \delta_s')'$ is the true parameter vector value given in (2.16).

Assumption 4. The assumed initial values are $Y_t = \widehat{Y}_0$ for $-p+1 \leq t \leq 0$, $u_t = \widehat{u}_0$ for $-q+1 \leq t \leq 0$ and $X_t = \widehat{X}_0$ for $-s \leq t \leq 0$, where there exists $0 < C < \infty$ such that $E\sup_{\theta\in\Theta}||\widehat{Y}_0||_K^2 < C$, $E\sup_{\theta\in\Theta}\|\widehat{u}_0\|_K^2 < C$, $E\sup_{\theta\in\Theta}||\widehat{X}_0||_K^2 < C$.

Assumption 5. The square matrices $E\left[\left\langle s_{\frac{\partial}{\partial\theta}u_t(\theta)}, s_{\frac{\partial}{\partial\theta'}u_t(\theta)}\right\rangle_K\right]$ and $E\left[\left\langle s_{\frac{\partial}{\partial\theta}u_t(\theta)}, s_{u_t(\theta)}\right\rangle_K \left\langle s_{u_t(\theta)}, s_{\frac{\partial}{\partial\theta}u_t(\theta)}\right\rangle_K\right]$ are positive definite for θ in a small neighborhood of θ^0.

3.1 Minimum D_K-Distance Estimation

We now propose a D_K-metric based estimation method for an ACIX(p,q,s) model. Given that $E(Y_t|I_{t-1})$ is the optimal solution to minimize $E\left[D_K^2(Y_t,A)|I_{t-1}\right]$, as is established in Lemma 2.1, we will propose an estimation method that minimizes a sample analog of $E\left[D_K^2(Y_t,A)|I_{t-1}\right]$. As an advantage, our method does not require specification of the distribution of the interval population. Also, the proposed method provides a unified framework that can generate various point-valued estimators (e.g., conditional least squares estimators based on the range and/or midpoint sample information) as special examples; see Section 3.2 below.

We define the minimum D_K-distance estimator as follows:

$$\widehat{\theta} = \arg\min_{\theta\in\Theta}\widehat{Q}_T(\theta),$$

where $T\widehat{Q}_T(\theta)$ is the sum of squared norm of residuals of the ACIX(p, q, s) model in (2.16), namely

$$\widehat{Q}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} q_t(\theta), \tag{3.1}$$

$$q_t(\theta) = \|u_t(\theta)\|_K^2 = D_K^2\left[u_t(\theta), 0\right] \tag{3.2}$$

and

$$u_t(\theta) = Y_t - \left[(\alpha_0 + \beta_0 I_0) - \sum_{j=1}^{p} \beta_j Y_{t-j} - \sum_{j=0}^{s} \delta_j' X_{t-j} - \sum_{j=1}^{q} \gamma_j u_{t-j}(\theta)\right]. \tag{3.3}$$

Since we only observe $\{Y_t\}$ from time $t = 1$ to time $t = T$. Therefore we have to assume some initial values for $\{Y_t\}_{t=-p+1}^{0}$, $\{X_t\}_{t=-s+1}^{0}$ and $\{u_t(\theta)\}_{t=-q+1}^{0}$ in computing the values for the unobservable interval error process $\{u_t(\theta)\}$:

We first establish consistency of $\widehat{\theta}$.

Theorem 3.1. Under Assumptions 1, 2, 3(i) and 4, as $T \to \infty$,

$$\widehat{\theta} \xrightarrow{p} \theta^0.$$

Intuitively, the statistic $\widehat{Q}_T(\theta)$ converges in probability to $E[D_K^2(Y_t, Z_t'\theta)]$ uniformly in Θ as $T \to \infty$. Furthermore, the true model parameter θ^0 is the unique minimizer of $E[D_K^2(Y_t, Z_t'\theta)]$ given the IMDS condition on the interval innovation process $\{u_t\}$. It then follows from the extreme estimator theorem (e.g., White (1994)) that $\widehat{\theta} \xrightarrow{p} \theta^0$ as $T \to \infty$.

Next, we derive the asymptotic normality of $\widehat{\theta}$.

Theorem 3.2. Under Assumptions 1-5, as $T \to \infty$,

$$\sqrt{T}(\widehat{\theta} - \theta^0) \xrightarrow{L} N(0, M^{-1}(\theta^0)V(\theta^0)M^{-1}(\theta^0)),$$

where $V(\theta^0) = E\left[\frac{\partial q_t(\theta^0)}{\partial \theta}\frac{\partial q_t(\theta^0)}{\partial \theta'}\right]$, $M(\theta^0) = -E\left[\frac{\partial^2}{\partial\theta\partial\theta'} q_t(\theta^0)\right]$, $q_t(\theta)$ *is defined as in Eq.(3.2) and all the derivatives are evaluated at* θ^0.

The asymptotic variance of $\sqrt{T}(\widehat{\theta} - \theta^0)$, i.e., $M^{-1}(\theta^0)V(\theta^0)M^{-1}(\theta^0)$, can be consistently estimated, as shown below.

Theorem 3.3. Under Assumptions 1-5, as $T \to \infty$,

$$\widehat{M}_T(\widehat{\theta}) = -\frac{1}{T}\sum_{t=1}^{T} \frac{\partial^2 q_t(\widehat{\theta})}{\partial\theta\partial\theta'} \xrightarrow{p} M(\theta^0),$$

$$\widehat{V}_T(\widehat{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial q_t(\widehat{\theta})}{\partial\theta}\frac{\partial q_t(\widehat{\theta})}{\partial\theta'} \xrightarrow{p} V(\theta^0),$$

where $q_t(\theta)$ is defined as in Eq.(3.2) and all the derivatives are evaluated at the estimator $\widehat{\theta}$ and the assumed initial values for $Y_t, X_t, u_t(\theta)$ with $t \leq 0$. Hence, as $T \to \infty$,

$$\widehat{M}_T^{-1}(\widehat{\theta})\widehat{V}_T(\widehat{\theta})\widehat{M}_T^{-1}(\widehat{\theta}) - M^{-1}(\theta^0)V(\theta^0)M^{-1}(\theta^0) \xrightarrow{p} 0.$$

We note that the asymptotic variance of $\sqrt{T}\widehat{\theta}$ cannot be simplified even under conditional homoskedasticity that $var(u_t|I_{t-1}) = \sigma_K^2$ for an arbitrary kernel K.

When the ACIX(p, q, s) model becomes an ACIX$(p, 0, s)$ model as in Eq.(2.17), namely, when there is no MA component in the ACIX(p, q, s) model, the minimum D_K-distance estimator $\widehat{\theta}$ has a closed form that is in a similar spirit to the conventional OLS estimator. This is stated below.

Corollary 3.1. Suppose Assumptions 1-5 hold, and $\{Y_t\}$ follows the ACIX$(p, 0, s)$ process in Eq.(2.17). Then the minimum D_K-distance estimator $\widehat{\theta}$ has the closed form

$$\widehat{\theta} = \left[\sum_{t=1+\max(p,s)}^{T} \left\langle s_{Z_t}, s'_{Z_t} \right\rangle_K \right]^{-1} \sum_{t=1+\max(p,s)}^{T} \left\langle s_{Z_t}, s_{Y_t} \right\rangle_K,$$

where $Z_t = ([1,1], I_0, Y_{t-1}, ..., Y_{t-p}, X'_t, X'_{t-1}, ..., X'_{t-s})'$. When $T \to \infty$, $\widehat{\theta} \xrightarrow{p} \theta^0$, and

$$\sqrt{T}(\widehat{\theta} - \theta^0) \xrightarrow{L} N(0, E^{-1}\left[\left\langle s_{Z_t}, s'_{Z_t} \right\rangle_K\right] E\left[\left\langle s_{Z_t}, s_{u_t} \right\rangle_K \left\langle s_{u_t}, s'_{Z_t} \right\rangle_K\right] E^{-1}\left[\left\langle s_{Z_t}, s'_{Z_t} \right\rangle_K\right]).$$

Furthermore, as $T \to \infty$,

$$T^{-1} \sum_{t=1+\max(p,s)}^{T} \left\langle s_{Z_t}, s'_{Z_t} \right\rangle_K \xrightarrow{p} E\left[\left\langle s_{Z_t}, s'_{Z_t} \right\rangle_K\right],$$
$$T^{-1} \sum_{t=1+\max(p,s)}^{T} \left\langle s_{Z_t}, s_{\widehat{u}_t} \right\rangle_K \left\langle s_{\widehat{u}_t}, s'_{Z_t} \right\rangle_K \xrightarrow{p} E\left[\left\langle s_{Z_t}, s_{u_t} \right\rangle_K \left\langle s_{u_t}, s'_{Z_t} \right\rangle_K\right],$$

where $\widehat{u}_t = Y_t - Z'_t\widehat{\theta}$.

3.2 Examples of Minimum $\mathbf{D_K}$-Distance Estimators

This section explores how the results in Theorems 3.1–3.3 can be used to derive various estimators as special cases. Based on the estimated interval residuals $\{\hat{u}_t(\theta)\}_{t=1}^T$, define

$$\begin{cases} \widehat{Q}_T^L(\theta) = \sum_{t=1}^T \widehat{u}_{L,t}^2(\theta), & \widehat{Q}_T^R(\theta) = \sum_{t=1}^T \widehat{u}_{R,t}^2(\theta), \quad \widehat{Q}_T^{LR}(\theta) = \sum_{t=1}^T \widehat{u}_{L,t}(\theta)\widehat{u}_{R,t}(\theta) \\ \widehat{Q}_T^r(\theta) = \sum_{t=1}^T \left[\widehat{u}_t^r(\theta)\right]^2, & \widehat{Q}_T^m(\theta) = \sum_{t=1}^T \left[\widehat{u}_t^m(\theta)\right]^2, \quad \widehat{Q}_T^{mr}(\theta) = \sum_{t=1}^T \widehat{u}_t^r(\theta)\widehat{u}_t^m(\theta), \end{cases} \tag{3.4}$$

where $\widehat{u}_{L,t}(\theta)$ and $\widehat{u}_{R,t}(\theta)$ are the left and right bounds of $\widehat{u}_t(\theta)$, $\widehat{u}_t^r(\theta) = \widehat{u}_{R,t}(\theta) - \widehat{u}_{L,t}(\theta)$ and $\widehat{u}_t^m(\theta) = \frac{1}{2}\widehat{u}_{L,t}(\theta) + \frac{1}{2}\widehat{u}_{R,t}(\theta)$ are the range and midpoint of $\widehat{u}_t(\theta)$. Combining Eqs.(2.11) and (3.6), we obtain

$$\begin{aligned}
\widehat{Q}_T(\theta) &= a\widehat{Q}_T^R(\theta) + c\widehat{Q}_T^L(\theta) - 2b\widehat{Q}_T^{LR}(\theta) \\
&= \frac{a+2b+c}{4}\widehat{Q}_T^r(\theta) + (a-2b+c)\,\widehat{Q}_T^m(\theta) + (a-c)\,\widehat{Q}_T^{mr}(\theta).
\end{aligned} \tag{3.5}$$

Case 1: Conditional Least Squares Estimators Based on Univariate Point Data

Suppose we choose a kernel K with $(a,b,c) = (1,1,1)$. Then

$$\widehat{Q}_T(\theta) = \widehat{Q}_T^r(\theta^r),$$

which is the sum of squared residuals of the conditional dynamic range model in Eq.(2.18). In this case, the D_K-minimum estimator solves

$$\widehat{\theta}^r = \arg\min_{\theta\in\Theta} \widehat{Q}_T^r(\theta).$$

The estimator $\widehat{\theta}^r$ cannot identify the level parameter α_0, because $\widehat{\theta}^r$ is based on the range sample $\{Y_t^r, X_t^r\}_{t=1}^T$, which contains no level information of the interval process $\{Y_t\}$.

To estimate α_0, we can use a kernel K with $(a,b,c) = (\frac{1}{4}, -\frac{1}{4}, \frac{1}{4})$. Then

$$\widehat{Q}_T(\theta) = \widehat{Q}_T^m(\theta),$$

which is the sum of squared residuals of the conditional dynamic level (i.e., midpoint) model in Eq.(2.19). In this case, the D_K-minimum estimator solves

$$\widehat{\theta}^m = \arg\min_{\theta\in\Theta} \widehat{Q}_T^m(\theta).$$

The estimator $\widehat{\theta}^m$ can consistently estimate the level parameter α_0, but it cannot identify the scale parameter β_0, because $\widehat{\theta}^m$ is based on the midpoint sample $\{Y_t^m, X_t^m\}_{t=1}^T$, which contains no range information of the interval process $\{Y_t\}$.

Given the fitted values for both range and mid-point processes, we can construct a one-step-ahead predictor for interval variable Y_t using information I_{t-1}:

$$\widehat{E}(Y_t|I_{t-1}) = \left[\widehat{Y}_t^m - \frac{1}{2}\widehat{Y}_t^r, \widehat{Y}_t^m + \frac{1}{2}\widehat{Y}_t^r\right],$$

where $\widehat{Y}_t^m$ and $\widehat{Y}_t^r$ are one-step-ahead point predictors for Y_t^m and Y_t^r based on Eqs.(2.19) and (2.18) respectively.

Both estimators $\widehat{\theta}^r$ and $\widehat{\theta}^m$ are convenient and they can consistently estimate partial parameters in the ACIX(p,q,s) model. However, besides the failure in identifying level parameter α_0 or scale parameter β_0, these estimators are not expected to be most efficient because they utilize the range and level sample information separately.

Case 2: Constrained Conditional Least Squares Estimators Based on Bivariate Point Samples

Now we consider the choice of kernel K with $a = c > 0$ and $b = 0$. Then

$$\frac{1}{a}\widehat{Q}_T(\theta) = \widehat{Q}_T^L(\theta) + \widehat{Q}_T^R(\theta) = \sum_{t=1}^{T}\left[\hat{u}_{L,t}^2(\theta) + \widehat{u}_{R,t}^2(\theta)\right].$$

This is the sum of squared residuals of the bivariate ARMAX model in Eq. (2.20) for the left bound $Y_{L,t}$ and right bound $Y_{R,t}$ of the interval process $\{Y_t\}$. Thus, the minimum D_K-distance estimator $\hat{\theta}$ becomes the constrained conditional least squares estimator for the bivariate ARMAX(p,q,s) model for the left and right bounds of Y_t; it is consistent for all parameters θ^0 in the ACIX model.

Given the fitted values for the bivariate ARMAX(p,q,s) model for $Y_{L,t}$ and $Y_{R,t}$, we can also construct a one-step-ahead predictor for interval variable Y_t using information I_{t-1}:

$$\widehat{E}(Y_t|I_{t-1}) = \left[\widehat{Y}_{L,t}, \widehat{Y}_{R,t}\right],$$

where $\widehat{Y}_{L,t}$ and $\widehat{Y}_{R,t}$ are one-step-ahead point predictors for $Y_{L,t}$ and $Y_{R,t}$ based on Eq.(2.20).

Case 3: Constrained Quasi-Maximum Likelihood Estimators

The bivariate ARMAX(p,q,s) model for the $(Y_{L,t}, Y_{R,t})'$ can also be consistently estimated by the conditional constrained quasi-maximum likelihood method (CCQML) based on the bivariate point-valued sample $\{Y_{L,t}, Y_{R,t}\}_{t=1}^T$. Assume that the bivariate innovation $\{u_{L,t}, u_{R,t}\}'$ follows i.i.d.$N(\mathbf{0}, \Sigma^0)$, where Σ^0 is a 2×2 unknown variance-covariance matrix.

The log-Gaussian likelihood function given the bivariate sample $\{Y_{L,t}, Y_{R,t}\}_{t=1}^T$ is given by

$$\hat{L}(\theta, \Sigma) = \frac{T}{2}\ln|\Sigma^{-1}| - \frac{1}{2}\sum_{t=1}^{T}(u_{L,t}(\theta), u_{R,t}(\theta))\,\Sigma^{-1}(u_{L,t}(\theta), u_{R,t}(\theta))',$$

where $u_{Lt}(\theta)$ and $u_{R,t}(\theta)$ are the left and right bounds of $u_t(\theta)$ defined in Eq. (3.3). Then the CCQML estimator,

$$\left(\hat{\theta}, \hat{\Sigma}\right) = \arg\max_{(\theta,\Sigma)\in\Theta\times\mathbf{R}^{2\times 2}} \hat{L}(\theta, \Sigma),$$

consistently estimate the unknown parameter θ^0 given the IMDS condition that $E(u_t|I_{t-1}) = 0$. We note that

$$-\hat{L}(\hat{\theta}, \hat{\Sigma}) = \hat{\Sigma}_{11}\hat{Q}_T^R(\hat{\theta}) + \hat{\Sigma}_{22}\hat{Q}_T^L(\hat{\theta}) - 2\hat{\Sigma}_{12}\hat{Q}_T^{LR}(\hat{\theta}),$$

where $\hat{\Sigma}_{ij}$ is the (i,j)-th component of the variance-covariance estimator $\hat{\Sigma}$. This looks rather similar to the objective function $\hat{Q}_T(\theta)$ in Eq. (3.5) of the minimum D_K-distance estimator, with the kernel $K = \hat{\Sigma}$. However, we cannot interpret the CCQML is a special case of the minimum D_K-distance estimator because for the minimum D_K-distance estimation, the kernel K is prespecified, whereas for the CCQML, both θ and Σ are unknown parameters and have to be estimated simultaneously. We will examine the relative efficiency between the minimum D_K-distance estimator and various alternative estimators for θ^0.

4. Efficiency and Two-Stage Minimum Distance Estimation

The minimum D_K-distance method provides consistent estimation for an ACIX model without having to specify the full density of the interval population. Different choice of kernel K will deliver different minimum D_K-distance estimators for θ^0, and all of them are consistent for θ^0, provided kernel satisfies Eq. (2.6). As discussed earlier, different choices of K imply different ways of utilizing the sample information of the interval process. Now, a question arises naturally: What is the optimal choice of kernel K, if any? Below, we derive an optimal kernel that yields a minimum D_K-distance estimator with the minimum asymptotic variance among a large class of kernel functions that satisfy Eq. (2.6). We first impose a condition on the interval innovation process $\{u_t\}$.

Assumption 6. The interval innovation process $u_t = [u_{L,t}, u_{R,t}]$ satisfies $\mathrm{var}(u_t|I_{t-1}) = \sigma_K^2 < \infty$.

This is a conditional homoskedasticity assumption on $\{u_t\}$. The i.i.d. condition for $\{u_t\}$ is a sufficient but not necessary condition for Assumption 6.

Theorem 4.1: Under Assumptions 1-6, the choice of kernel $K^{opt}(u, v)$ with

$$\begin{aligned} K^{opt}(1,1) &= \mathrm{var}(u_{L,t}), \\ K^{opt}(-1,1) &= K^{opt}(1,-1) = \mathrm{cov}(u_{L,t}, u_{R,t}), \\ K^{opt}(-1,-1) &= \mathrm{var}(u_{R,t}) \end{aligned}$$

delivers a minimum D_K-distance estimator

$$\widehat{\theta} = \arg\min_{\theta \in \mathbf{\Theta}} \frac{1}{T} \sum_{t=1}^{T} D^2_{K^{opt}} \left[Y_t, Z_t'(\theta)\theta \right],$$

which is asymptotically most efficient among all symmetric positive definite kernels K that satisfy

Eq. (2.6), *with the minimum asymptotic variance*

$$\begin{aligned}\Omega^{opt} &= -E\left[\frac{\partial^2}{\partial\theta\partial\theta'}q_t(\theta^0)|K=K^{opt}\right] \\ &= 2E\left\langle s_{\frac{\partial}{\partial\theta}u_t(\theta^0)}, s_{\frac{\partial}{\partial\theta'}u_t(\theta^0)}\right\rangle_{K^{opt}}.\end{aligned}$$

To explore the intuition behind Theorem 4.1, we observe that when kernel K^{opt} is used, the objective function of the minimum D_K-distance estimator becomes

$$\widehat{Q}_T(\theta) = var(u_{L,t})\widehat{Q}_T^R(\theta) + \mathrm{var}(u_{R,t})\widehat{Q}_T^L(\theta) - 2\mathrm{cov}(u_{L,t}, u_{R,t})\widehat{Q}_T^{LR}(\theta).$$

Thus, K^{opt} downweighs the sample squared distance components that have larger sampling variations. Specifically, it discounts the sum of squared residuals of the right bound when the right round disturbance $u_{R,t}$ has a large variance, and discounts the sum of squared residuals of the left bound when the left bound disturbance $u_{L,t}$ has a large variance. The use of K^{opt} also corrects correlations between the left and right bound disturbances. Such weighting and correlation correction are similar in spirit to the optimal weighting matrix in GLS. We note that the optimal choice of kernel K^{opt} is not unique. For any constant $c > 0$, the kernel cK^{opt} is also optimal.

The results of Theorem 4.1 do not apply when there exists conditional heteroskedasticity in the sense that $\mathrm{var}(u_t|I_{t-1}) = \sigma_t^2(K)$ is a time-varying function. In this case, one could first obtain a consistent estimator for the conditional variance $var(u_t|I_{t-1})$, and then construct a feasible adaptive D_K-minimum distance estimator. We leave it for future study.

The optimal D_K-distance estimator is not feasible because the optimal kernel K^{opt}, which depends on the data generating process, is infeasible. However, we can consider a two-stage minimum D_K-distance estimation method: In Step 1, obtain a preliminary consistent estimator $\widehat{\theta}$ of θ^0. For example, it can be a minimum D_K-distance estimator with an arbitrary prespecified kernel K satisfying Eq.(2.6). We then compute the estimated residuals $\{\hat{u}_t(\hat{\theta})\}$ and construct an estimator for the optimal kernel K^{opt} :

$$\hat{K}^{opt} = T^{-1}\sum_{t=1}^{T}\begin{bmatrix}\hat{u}_{L,t}^2(\hat{\theta}), & \hat{u}_{L,t}(\hat{\theta})\hat{u}_{R,t}(\hat{\theta}) \\ \hat{u}_{R,t}(\hat{\theta})\hat{u}_{L,t}(\hat{\theta}), & \hat{u}_{R,t}^2(\hat{\theta})\end{bmatrix}.$$

This is consistent for K^{opt}. In Step 2, we obtain a minimum D_K-distance estimator with the choice of $K = \hat{K}^{opt}$:

$$\widehat{\theta}^{opt} = \arg\min_{\theta\in\Theta}\frac{1}{T}\sum_{t=1}^{T}D_{\hat{K}^{opt}}^2\left[Y_t, Z_t'(\theta)\theta\right].$$

This two-stage minimum D_K-distance estimator is asymptotically most efficient among the class of kernels satisfying Eq. (2.6), as is shown in Theorem 4.2 below.

Theorem 4.2. Under Assumptions 1-6, as $T \to \infty$, the two-stage minimum D_K-distance estimator

$$\sqrt{T}(\widehat{\theta}_K^{opt} - \theta^0) \xrightarrow{p} N(0, \Omega^{opt}),$$

where Ω^{opt} is the minimum asymptotic variance as given in Theorem 4.1.

Interestingly, when the left and right bounds $u_{L,t}$ and $u_{R,t}$ of the interval innovation u_t follow an i.i.d. bivariate normal distribution, the two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$ achieves the Cramer-Rao lower bound. This is stated in Theorem 4.3.

Theorem 4.3: Suppose Assumptions 1-6 hold and $\{u_{L,t}, u_{R,t}\}$ follows a bivariate Gaussian distribution. Then as $T \to \infty$, the two-stage minimum D_K-distance estimator $\widehat{\theta}_K^{opt}$ achieves the Cramer-Rao lower bound of the constrained MLE for the bivariate ARMAX(p, q, s) model for the left and right bounds of the interval process $\{Y_t\}$.

We note that the constrained MLE for the bivariate ARMAX(p, q, s) model for the left and right bounds of the interval process $\{Y_t\}$ is not numerically identical to the two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$, although they are equally asymptotically efficient.

When the bivariate process $\{u_{L,t}, u_{R,t}\}$ does not follow a joint Gaussian distribution, the constrained conditional quasi-maximum likelihood estimator (CCQML) is generally not asymptotically most efficient, unless the dynamic information matrix equality holds. Furthermore, the asymptotic variance of the CCQML estimator generally differs from the asymptotic variance Ω^{opt} of the two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$. Note that the CCQML cannot be viewed as a special case of the minimum D_K-distance estimator with the choice of kernel $\hat{\Sigma}$ because Σ^0 and θ^0 are estimated simultaneously. In contrast, for two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$, the kernel $\hat{\Sigma}$ is given, and θ^0 is the only unknown parameter. We will investigate the relative efficiency among $\hat{\theta}^{opt}$ and other estimators via simulation.

5. Hypothesis Testing

In this section, we are interested in testing the hypothesis of interest:

$$H_0 : R\theta^0 = r,$$

where R is a $q \times k$ nonstochastic matrix of full rank, $q \leq k$, r is a $q \times 1$ nonstochastic vector, and k is the dimension of parameter θ in the ACIX(p, q, s) model of Eq.(2.16).

We will propose a Lagrange Multiplier (LM) test and a Wald test based on the minimum D_K-distance estimation. We first consider the LM test. Consider the following constrained D_K-distance minimization problem

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} \widehat{Q}_T(\theta),$$

subject to $R\theta = r$. Define the Lagrange function

$$L_T(\theta, \lambda) = \widehat{Q}_T(\theta) + \lambda'(r - RQ),$$

where λ is the multiplier. Let $\widetilde{\theta}$ and $\widetilde{\lambda}$ denote the solutions that maximize $L_T(\theta, \lambda)$, that is,

$$(\widetilde{\theta}, \widetilde{\lambda}) = \arg\min_{\theta\in\Theta} L_T(\theta, \lambda).$$

Then we can construct a LM test for H_0 based on $\widetilde{\lambda}$.

Theorem 5.1: Suppose Assumptions 1-5 and H_0 hold. Define

$$LM = \left[T\widetilde{\lambda}' R'\widehat{M_T}(\widetilde{\theta})R\right]\left[R'\widehat{M}_T^{-1}(\widetilde{\theta})\widehat{V}_T(\widetilde{\theta})\widehat{M}_T^{-1}(\widetilde{\theta})R\right]^{-1}\left[R'\widehat{M_T}(\widetilde{\theta})R\widetilde{\lambda}\right]$$

where $\widehat{M_T}(\widetilde{\theta})$ and $\widehat{V}_T(\widetilde{\theta})$ are defined in the same way as $\widehat{M_T}(\widehat{\theta})$ and $\widehat{V}_T(\widehat{\theta})$ in Theorem 3.3 respectively, with the constrained minimum D_K-distance estimator $\widetilde{\theta}$. Then $LM \xrightarrow{d} \chi^2_q$ as $T \to \infty$.

We note that the LM test only requires the minimum D_K-distance estimation under H_0.

Alternatively, we can construct a Wald test statistic that only involves the minimum D_K-distance estimation under the alternative hypothesis to H_0 (i.e., without restriction).

Theorem 5.2: Suppose Assumptions 1-5 and H_0 hold. Define a Wald test statistic

$$W = \left[T(R\widehat{\theta} - r)'\right]\left[R\widehat{M}_T^{-1}(\widehat{\theta})\widehat{V}_T(\widehat{\theta})\widehat{M}_T^{-1}(\widehat{\theta})R'\right]^{-1}\left[(R\widehat{\theta} - r)\right]$$

where $\widehat{\theta}$, $\widehat{M_T}(\widehat{\theta})$ and $\widehat{V}_T(\widehat{\theta})$ are defined in the same way as $\widehat{M_T}(\widehat{\theta})$ and $\widehat{V}_T(\widehat{\theta})$ in Theorem 3.4. Then, $W \xrightarrow{d} \chi^2_q$ as $T \to \infty$.

The Wald test W is essentially based on the comparison between the unrestricted and restricted minimum D_K-distance estimators $\widehat{\theta}$ and $\widetilde{\theta}$, but the test statistic W only involves the unrestricted parameter estimator $\widehat{\theta}$.

Because we do not assume a probability distribution for the interval process $\{Y_t\}$, we cannot construct a likelihood ratio test for the hypothesis H_0 of interest here.

6. Simulation Study

We now investigate the finite sample properties of CCLS, CCQML, minimum D_K-distance and two-stage minimum D_K-distance estimators via a Monte Carlo study. We will consider two sets of experiments. In the first experiment, the interval data are generated from an ACI process. In the second set of experiments, the interval data are constructed from a bivariate ARMA process.

6.1 ACI-Based Data Generating Processes

We first consider an ACI$(1,1)$ model as the data generating process (DGP):

$$Y_t = \alpha_0 + \beta_0 I_0 + \beta_1 Y_{t-1} + \gamma_1 u_{t-1} + u_t, \tag{6.1}$$

where parameter values $\theta^0 = (\alpha_0, \beta_0, \beta_1, \gamma_1)'$ are obtained from the minimum D_K-distance estimates of the ACI$(1,1)$ model based on the real interval data of the S&P 500 daily index from January 3, 1988 to September 18, 2009, and the kernel K used is $K(1,1) = K(-1,-1) = a$, $K(1,-1) = K(-1,1) = b$, with $\frac{b}{a} = \frac{3}{5}$. The minimum and maximum S&P 500 closing price values of day t form the raw interval-valued observations in this period, denoted $\{P_1, ..., P_T\}$. Then we convert the raw interval price sample data to be a D_K-weakly stationary interval sample, denoted $\{Y_1, ..., Y_T\}$, by taking the logarithm and Hukuhara difference as $Y_t = \ln(P_t) -_H \ln(P_{t-1})$. The initial values of Y_t and u_t for $t = 0$ are set to be $\overline{Y}_T$ and $[0,0]$, respectively. We obtain the minimum D_K-distance estimates and use them as the true parameter values in DGP (6.1). To simulate the interval innovations $\{u_t\}$ in (6.1), we first compute the estimated model residuals

$$\widehat{u}_t = Y_t -_H (\widehat{\alpha}_0 + I_0 \widehat{\beta}_0 + \widehat{\beta}_1 Y_{t-1} + \widehat{\gamma}_1 \widehat{u}_{t-1})$$

based on the S&P 500 data. We then generate $\{u_t\}_{t=1}^T$ via the naive bootstrapping from $\{\widehat{u}_t\}_{t=1}^T$, with $T = 100, 250, 500$, and 1000, respectively. For each sample size T, we perform 1000 replications. For each replication, we estimate model parameters of an ACI(1,1) model using CLS, CCQML, minimum D_K-distance and two-stage minimum D_K-distance methods. Two parameter estimates of CLS are produced, i.e., $\widehat{\theta}^r = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\gamma}_1)$ and $\widehat{\theta}^m = (\widehat{\alpha}_0, \widehat{\beta}_1, \widehat{\gamma}_1)$, based on range and midpoint data, respectively. We consider 4 kernels with the form of K_{ab} with $a = c$, which yields the constrained conditional least squares (CCLS) estimator $\widehat{\theta}_{CCLS}$ for the bivariate model of the left and right bounds of Y_t in Eq.(2.20). We also consider 6 kernels with the form K_{abc}. The two-stage minimum D_K-distance estimator is obtained from a kernel K_{abc} with $(a,b,c) = (10, 8, 16)$ in the first stage. For the values of a, b and c of each kernel, see the tables below.

We compute the bias, standard deviation (SD), and root mean square error (RMSE) for each estimator:

$$
\begin{aligned}
Bias(\widehat{\theta}_i) &= \frac{1}{1000}\sum_{m=1}^{1000}(\widehat{\theta}_i^{(m)} - \widehat{\theta}_i^{0}),\\
SD(\widehat{\theta}_i) &= \left[\frac{1}{1000}\sum_{m=1}^{1000}(\widehat{\theta}_i^{(m)} - \overline{\theta}_i)^2\right]^{1/2},\\
RMSE\left(\widehat{\theta}_i\right) &= \left[Bias^2(\widehat{\theta}_i) + SD^2(\widehat{\theta}_i)\right]^{1/2},
\end{aligned}
$$

where $\overline{\theta_i} = \frac{1}{1000}\sum_{m=1}^{1000}\widehat{\theta}_i^{(m)}$, and $\widehat{\theta}_i = \widehat{\alpha}_0, \widehat{\beta}_0, \widehat{\beta}_1, \widehat{\gamma}_1$, respectively.

Tables 1-4 report Bias, SD, and RMSE of $\widehat{\theta}^m$, $\widehat{\theta}^r$, $\widehat{\theta}_{QML}$, $\widehat{\theta}$ and $\widehat{\theta}^{opt}$ respectively. Several observations emerge. First, for all estimators, the RMSE converges to zero as the sample size T increases. In particular, the minimum D_K-distance estimator $\widehat{\theta}$ displays robust performance for various kernels. Second, both the interval-based minimum D_K-distance estimators, i.e., $\widehat{\theta}$ and $\widehat{\theta}^{opt}$, and the bivariate-point based CCQML estimator $\widehat{\theta}_{QML}$ outperform $\widehat{\theta}^r$ and $\widehat{\theta}^m$ in terms of RMSE. The two-stage minimum D_K-distance estimator $\widehat{\theta}^{opt}$ dominates the minimum D_K-distance estimator $\widehat{\theta}$ with most kernels, confirming the efficiency result in Theorems 4.1–4.2. The estimator $\widehat{\theta}^{opt}$ outperforms $\widehat{\theta}_{QML}$ for all parameters in θ^0 in terms of $RMSE$. Among other things, CCQML has more unknown parameters in the parameter space than that of two-stage minimum D_K-distance estimation.

Lastly, comparing $\widehat{\theta}$, $\widehat{\theta}^{opt}$ and $\widehat{\theta}_{QML}$ with $\widehat{\theta}^m$ and $\widehat{\theta}^r$, the efficiency gain over the CLS estimators based on level or range samples separately is enormous as T becomes large. This is apparently due to the fact that $\widehat{\theta}$ and $\widehat{\theta}^{opt}$ utilize the level, range and their correlation information contained in the interval data. On the other hand, while the estimators $\widehat{\theta}^r$ and $\widehat{\theta}^m$ can consistently estimate model parameters, $\widehat{\theta}^m$ is better than $\widehat{\theta}^r$. Data examination shows that this is due to more variations in level of Y_t rather than in range over time. This highlights the importance of utilizing level information of asset prices even when interest is in modelling the range (or volatility) dynamics .

6.2 Bivariate Point-valued Data Generating Processes

We now investigate the relative performances of the minimum D_K-distance estimators and the CCQML estimator when the data generating processes are various bivariate point processes with innovations $(u_{L,t}, u_{R,t})' \sim$ i.i.d. $f(0, \Sigma^0)$, where $f(0, \Sigma^0)$ is a bivariate density function. We then form interval sample data $\{Y_t\}_{t=1}^T$. Three densities for $f(\cdot,\cdot)$ are considered— bivariate normal, bivariate student-t with 5 degrees of freedom, and bivariate mixture with $u_{L,t} = a_1\varepsilon_{0t}+\varepsilon_{1t}$, $u_{R,t} = a_2\varepsilon_{0t}+\varepsilon_{2t}$ where ε_{it} follows i.i.d. $EXP(1)-1$ for $i = 0, 1, 2$, and they are jointly independent. Different values of constants a_1, a_2 result in different Σ^0 for the mixed distribution. For each distribution, $corr(u_{L,t}, u_{R,t}) = 0$ and -0.6 are considered with $T = 100$, 250 and 500 replications.

Specifically, we consider the following bivariate point process as the DGP:

$$\begin{cases} Y_{L,t} = \alpha_0 - \frac{1}{2}\beta_0 + \beta_1 Y_{L,t-1} + \gamma_1 u_{L,t-1} + u_{L,t}, \\ Y_{R,t} = \alpha_0 + \frac{1}{2}\beta_0 + \beta_1 Y_{R,t-1} + \gamma_1 u_{R,t-1} + u_{R,t}, \end{cases}$$

where parameter values $\theta^0 = (\alpha_0, \beta_0, \beta_1, \gamma_1)'$ are obtained in the same way as in Section 6.1 based on the actual S&P 500 data.

For each replication, we compute CCQML estimator $\widehat{\theta}_{QML}$, minimum D_K-distance estimator $\widehat{\theta}$ and two-stage minimum D_K-distance estimator $\widehat{\theta}^{opt}$, in addition to the same kernels used in Section 6.1. We include the infeasible optimal kernel $K^{opt} = \Sigma^0$ to obtain the infeasible asymptotically most efficient minimum D_K-distance estimator $\widehat{\theta}_{\Sigma^0}$; this allows us to study the finite sample behaviors between $\widehat{\theta}^{opt}$ and $\widehat{\theta}_{K_{\Sigma^0}}$.

In assessing the closeness of $\widehat{\theta}$ and $\widehat{\theta}^{opt}$ to $\widehat{\theta}_{QML}$ (which becomes MLE when $(u_{L,t}, u_{R,t})'$ follows a bivariate Gaussian distribution), we compute the following RMSE ratios:

$$R(i) = \frac{RMSE(\widehat{\theta}_i)}{RMSE(\widehat{\theta}_{QML,i})} \text{ and } R^{opt}(i) = \frac{RMSE(\widehat{\theta}_i^{opt})}{RMSE(\widehat{\theta}_{QML,i})}.$$

We first report the RMSE ratios of minimum D_K-distance estimator $\widehat{\theta}$ with different choices of kernels, including $\widehat{\theta}_{CCLS}$ that is obtained from a kernel K giving the same weight to the left and right bounds of Y_t, $\widehat{\theta}_{ab}$ from a kernel with $\frac{b}{a} = \frac{3}{5}$ assigning the same weights to the midpoint and range, $\widehat{\theta}_{abc}$ from a kernel K with $(a, b, c) = (10, 8, 19)$, two-stage minimum D_K-distance estimator $\widehat{\theta}^{opt}$ and infeasible asymptotically efficient minimum D_K-distance estimator $\widehat{\theta}_{\Sigma^0}$ with $K = \Sigma^0$. Note that when $corr(u_{L,t}, u_{R,t}) = 0$, $\widehat{\theta}_{CCLS}$ coincides with $\widehat{\theta}_{\Sigma^0}$ since we specify Σ^0 as an identity matrix here.

Tables 5-8 report the values of $R(i)$ and $R^{opt}(i)$ of various minimum D_K-distance estimators $\widehat{\theta}$ and $\widehat{\theta}^{opt}$. For a bivariate point i.i.d. Gaussian $(u_{L,t}, u_{R,t})$, $\widehat{\theta}^{opt}$ is asymptotically as efficient as the MLE $\widehat{\theta}_{ML}$ for the bivariate model of the left and right bounds of Y_t; indeed, their RMSE rations are very close to 1. The two-stage minimum D_K-distance estimator $\widehat{\theta}^{opt}$ also significantly improves the efficiency of $\widehat{\theta}$ with arbitrary choices of kernel. It confirms the adaptive capability of K^{opt} in correcting the inefficient kernel in the first stage for minimum D_K-distance estimation.

When $(u_{L,t}, u_{R,t})'$ follows a Student-t or mixed distribution, $\widehat{\theta}^{opt}$ is still the most efficient in the class of minimum D_K-distance estimators, which is consistent with Theorem 4.2. Like in Section 6.1, $\widehat{\theta}^{opt}$ has a smaller RMSE than $\widehat{\theta}_{QML}$,and the gain is more significant with $corr(u_{L,t}, u_{R,t}) = -0.6$ than with $corr(u_{L,t}, u_{R,t}) = 0$. In particular, the gain of $\widehat{\theta}^{opt}$ over $\widehat{\theta}_{QML}$ under asymmetric bivariate mixed distributions are more significant than under symmetric bivariate student$-t$ distributions. This is because the departure of the student$-t$ distribution from a bivariate normal relative to that of the mixed distribution is not very large, which leads to the smaller difference in RMSE. We also observe that $\widehat{\theta}^{opt}$ outperforms $\widehat{\theta}_{CCLS}$ when $corr(u_{L,t}, u_{R,t}) = -0.6$. This implies that the conditional least squares estimator $\widehat{\theta}_{CCLS}$ is not efficient even under the bivariate point valued date generating process. This is apparently due to the fact that $\widehat{\theta}_{CCLS}$ ignores the (negative correlation) between the left and right bounds. Finally, $\widehat{\theta}^{opt}$ is almost the same efficient as the infeasible asymptotically efficient estimator $\widehat{\theta}_{\Sigma^0}$ as T increases. This implies that the first stage estimation has negligible impact on the efficiency of $\hat{\theta}^{opt}$.

7. Empirical Application

In this section, we examine the explanatory power of bond market factors for excess stock returns when stock market factors are present. Fama and French (1993) consider two bond market factors, $TERM_t$ and DEF_t, where $TERM_t$ is the difference between the monthly long-term government bond return LG_t and the risk-free interest rate R_{ft}, and DEF_t is the difference between the return on a market portfolio of long-term corporate bonds LC_t and LG_t. Fama and French (1993) finds that these two bond market factors alone are significant in explaining excess stock returns. However, they find that the inclusion of three stock-factors (i.e., $R_{mt} - R_{ft}$, SMB_t, HML_t) in regressions for stocks kill the significance of $TERM_t$ and DEF_t. There are at least two possibilities for insignificance of $TERM_t$ and DEF_t. The first is that the three stock market factors contain all information in $TERM_t$ and DEF_t, and thus the bond market factors become insignificant when the stock market factors are included. The second possibility is that the OLS estimator used in Fama and French (1993) is not efficient because it does not exploit the level information of asset returns and interest rates. In this case, it may become significant if we use the more efficient two-stage minimum D_K-distance estimator. Our aim here is to explore whether the significance of bond market factors will be wiped out by the stock-market factors by using the interval CAPM model when a more efficient estimation method is used.

Fama and French's (1993) five-factor Capital Asset Pricing Model (CAMP) is

$$R_{it} - R_{ft} = \beta_0 + \beta_1(R_{mt} - R_{ft}) + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 TERM + \beta_5 DEF + \varepsilon_t, \qquad (7.1)$$

where R_t is a portfolio return, R_{ft} is the risk-free interest rate, R_{mt} is the market portfolio return, SMB is the the difference between the return on the small portfolio and the return on the large portfolio, HML is the difference between the return on the high book-to-market portfolio and the return on the low book-to-market portfolio, $TERM_t$ and DEF_t are defined as above.

Given the definition of variables in the Fama and French (1993) model, (7.1) can be viewed as a 'range' or 'difference' model of the following interval CAPM:

$$Y_{it} = \alpha_0 + \beta_0 I_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + u_t, \tag{7.2}$$

where $i = 1,..., 25$, $E(u_t|I_{t-1}) = [0, 0]$, $Y_t = [R_{ft}, R_t]$, $X_{1t} = [R_{ft}, R_{mt}]$,

$$\begin{aligned} X_{2t} &= \left[\frac{1}{3}(B/L_t + B/M_t + B/H_t), \frac{1}{3}(S/L_t + S/M_t + S/H_t)\right], \\ X_{3t} &= \left[\frac{1}{2}(S/L_t + B/L_t), \frac{1}{2}(S/H_t + B/H_t)\right], \end{aligned}$$

and $X_{4t} = [R_{ft}, LG_t]$, $X_{5t} = [LG_t, LC_t]$.

Using the monthly data from French's website, we estimate model parameters β_1, β_2, β_3, β_4, β_5 by OLS based on the Fama and French (1993) model (7.1) and by the two-stage minimum D_K-distance estimator based on the interval CAPM model (7.2) for each portfolio. To obtain a reliable standard error for each parameter estimator, we use the bootstrap method as the follow. We first estimate the Fama and French (1993) model in (7.1) with OLS and the interval CAPM in (7.1) with the minimum D_K-distance method for each of the 25 portfolios, and use the obtained parameter estimates as the true parameter values in the corresponding model. The estimation is based on the monthly data with the same sample period as in Fama and French (1993). The generations of the point innovations $\{\varepsilon_t\}_{t=1}^T$ for (7.1) and the interval innovation $\{u_t\}_{t=1}^T$ for (7.2) is the same as described in Section 6.1. We generate 500 bootstrap samples and obtain 500 bootstrap estimates for each parameter, which are then used to compute the estimated standard error of each parameter estimate and the associated t-test statistic. For each bootstrap sample, we estimate model parameters using the OLS estimator for the Fama and French (1993) model, and obtain estimate the interval version of the Fama and French (1993) model using the two-stage minimum D_K-distance estimator $\widehat{\theta}^{opt}$. For comparison, we also include minimum D_K-distance estimators with various choices of kernel K, and CCQML.

Table 9 reports the t-statistics for 5 groups of stock returns in terms of the book-to-market quantiles, each of which includes 5 groups in terms of the size quantiles. For each combination of two kinds of quantiles, we report the t-statistics of the OLS, and minimum D_K-distance estimates $\widehat{\theta}_{ab}$, $\widehat{\theta}_{abc}$ and $\widehat{\theta}^{opt}$, as well as the CCQML $\hat{\theta}_{QML}$. The estimates for α_0 in (7.2) are not reported here, since the FF model does not include this level parameter.

Table 9 shows some interesting findings. First, $\widehat{\theta}_{ab}$, $\widehat{\theta}_{abc}$, $\widehat{\theta}_{QML}$ and $\widehat{\theta}^{opt}$ for most of the 25 stock portfolios reveal strong evidence that the default risk factor DEF is significant in capturing the variation of excess stock returns, compared to the critical value of 1.96 at the 5% significance level. Generally, $\widehat{\theta}^{opt}$ yields larger t-statistics than $\widehat{\theta}_{QML}$, and both of them have large t-statistics than

OLS. On the other hand, there is not an overwhelming pattern for the effect of $TERM$ on excess stock returns for 25 portfolios. Data inspection shows the risk-free rate R_{ft} does not variate much over time relative to the long-term government bond return LG_t. As a consequence, the use of interval bond factor X_{4t} with about the same information as the differenced $TERM$ factor. In contrast, the significance of the two bond-market factors is still wiped out in the OLS regression on stock returns, as has been documented in Fama and French (1993). Thus, our evidence confirms the invaluable 'level' information contained in interval data compared to the point-valued data used in Fama and French (1993) which only contains the 'range' information only.

8. Conclusion

Interval-valued data are not uncommon in economics and econometrics. Compared to the point-valued data, interval-valued data contains more information including both level and range characteristics of the underlying stochastic process. This informational advantage can be exploited for more efficient estimation and inference, even if the interest is in range modelling. Interval forecasts are also often of direct interest in many applications.

This paper is perhaps the first attempt to model interval-valued time series data. We introduce an analytical framework for stationary interval-valued time series processes. To capture the dynamics of a stationary interval time series process, we propose a class of autoregressive conditional interval (ACIX) models with exogenous variables and develop a class of minimum D_K-distance estimators. We establish the asymptotic theory for consistency, normality and efficiency of the proposed estimators and exploit the relationships among various estimators that utilizes the interval sample information in different ways. In particular, we derive the optimal kernel function that yields an asymptotically most efficient estimator for an ACIX model among the class of symmetric positive definite kernels, and propose an asymptotically efficient two-stage minimum D_K-distance estimator. Simulation studies show that the two-stage minimum D_K-distance estimator outperform various estimators such as the conditional least squares estimators that are based on the range information and/or midpoint information of the interval sample, and the conditional qausi-maximum likelihood estimator based on the bivariate model for the left and right bounds of the interval process. In an empirical study on asset pricing, we document that unlike the conclusion of Fama and French (1993), some bond market factors, particularly the default risk factor, are significant in explaining the variation of excess stock returns even after the stock market factors are controlled. This highlights the gain of utilizing the level information of risk premium even when the interest is in range or difference modelling (i.e., excess risk premium).

The proposed ACIX models are the interval version of the ARMAX models for point-valued time series data. More flexible nonlinear models for interval time series, such as Markov-Chain regime switching models, autoregressive threshold models, and smooth transition models, can also be considered to capture nonlinear (e.g., asymmetric) features in the dynamic structure of stationary interval time series. On the other hand, the interval version of vector autoregression (VAR) or VARMA models can be considered to explore cross-dependence between different time series processes. Furthermore, one can consider nonstationary interval time series and the cointegrating relationships between nonstationary interval time series. Also, the optimal kernel specification needs further study for model application. All of these will be explored in future research.

References

Alizadeh, S., Brandt, M. and Diebold, F.X. (2002), "Range-Based Estimation of Stochastic Volatility Models," *Journal of Finance*, 57, 1047-1092.

Amemiya, T. (1985). Advanced Econometrics. Harvard University Press.

Andrews, D.W.K.. (1992), "Generic Uniform Convergence", *Econometric Theory*, 8, 241-257.

Artstein, Z. and R.A. Vitale (1975), "A Strong Law of Large Numbers for Random Compact Sets", Annals of Probability, 3: 879-882.

Aumann, R.J. (1965), "Integrals of Set-valued Functions", *Journal of Mathematical Analysis and Applications*, 12, 1-12.

Bates, C. and White, H. (1985), "A Unified Theory of Consistent Estimation for Parametric Models", *Econometric Theory,* 1, 151-178.

Bollerslev, T., Engle, R., and Nelson, B. (1994), "ARCH Models", Chapter 49 of the Handbook of Econometrics.

Beckers, S. (1983), "Variance of Security Price Return Based on High, Low and Closing Prices", *Journal of Business*, 56, 97-112.Choi1, Billingsley, P. (1968), "Convergence of Probability Measures. Wiley: New York.

Blanco, A., A. Colubi, N. Corral, and G. González-Rodríguez (2008), "On a Linear Independence Test for Interval-Valued Random Sets", D. Dubois *et al.* (eds.): *Soft Methods for Hand. Var. and Imprecision, ASC* 48, pp. 111–117.

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroscedasticity", *Journal of Econometrics* 31, 307-327.

Bollerslev, T., Chou, R., and Kroner, K. F. (1992), "ARCH Modeling in Finance: A Review of The Theory and Empirical Evidence", *Journal of Econometrics,* 52(1-2), 5-59.

Choi, D. H., and Smith, J. D. H., "Support functions of general convex sets" (2003), *Algebra Universalies*, 49(3), 305-319.

Chou, R. (2005), "Forecasting Financial Volatilities with Extreme Values: The Conditional Autoregressive Range (CARR) Model", *Journal of Money, Credit, and Banking*, 37 (3), 561-582.

C. Manski (2007). Identification for Prediction and Decision. Harvard University Press: Cambridge .

Cressie, N. (1978), "A Strong Limit Theorem for Random Sets", *Supplement to Advances in Applied Probability*, 10, 36-46.

Daffer, P.Z. and R. L. Taylor (1982), "Tightness and Strong Law of Large Numbers in Banach Spaces", *Bull. Inst. Math. Acad. Sinica*, 10(3), 251-263.

Diebold, F.X. and Yilmaz, K. (2009), "Measuring Financial Asset Return and Volatility Spillovers, With

Application to Global Equity Markets," *Economic Journal*, 119, 158-171.

Engle, R. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation", *Econometrica* 50, 987-1008.

Engle, R. (2002), "New Frontiers for ARCH Models," *Journal of Applied Econometrics*, 17, 425-446.

Fama, E. F. and French, K. R. (1993), " Common Risk Factors in The Returns on Stocks and Bonds", *Journal of Financial Economics,* 13 (33), 3-56.

González-Rodríguez, G., Blanco, A., Corral, N., Colubi, A. (2007), "Least Squares Estimation of Linear Regression Models for Convex Compact Random Sets", *Advance in Data Analysis and Classification* 1(1), 67–81.

González-Rodríguez, G., Blanco, A., Colubi, A., Lubiano M. A. (2009), "Estimation of a Simple Linear Regression Model for Fuzzy Random Variables", *Fuzzy Sets and Systems*, 160: 357–370.

H. Minkowsky, Gesammelte Abhandlungen 2. Band Teubner, Leipzig 1911.

Hiai, F. (1984), "Strong Laws of Large Numbers for Multivalued Random Variables', *Multifunctions and Integrands* (G. Salinetti, ed.), *Lecture Notes in Math.*, 1091, Springer-Verlag, Berlin, 160-172.

Hukuhara, M. (1967), "Intégration des Applications Measurable Dont la Valeur Est un Compact Convexe", *Funkcial Ekvac* 10, 205–223.

Körner, R. (1997), "On the Variance of Fuzzy Random Variables", *Fuzzy Sets and Systems*, 92, 83-93.

Körner, R., Näther, W. (2002), "On the Variance of Random Fuzzy Variables", In: *Bertoluzza*, C., Gil, M., Ralescu, D. (eds.) *Statistical Modeling, Analysis and Management of Fuzzy Data*, pp. 22–39. Physica-Verlag, Heildelberg.

Lee, S. W. and Hansen, B. E. (1994), "Asymptotic Theory for The GARCH(1,1) Quasi-Maximum Likelihood Estimator", *Econometric Theory*, 10, 29-52.

Li, S.M., Ogura, Y and Kreinovich, V. (2002). Limit Theorems and Applications of Set-Valued and Fuzzy Set-Valued Random Variables. Kluwer Academic Publishers: Dordrecht, Netherland.

Lumsdaine, R. . (1996), "Consistency and Asymptotic Normality of The Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models", *Econometrica*, Vol. 64 (3), 575-596.

Moore, R. E., Kearfott, R. B., and Michael J. C., (2009) *Introduction to Interval Analysis*, SIAM, Philadelphia, PA.

Munkres, J. (1999), *Topology*, 2nd edition, Prentice Hall: London.

Näther, W. (1997), "Linear Statistical Inference for Random Fuzzy Data", *Statistics*, 29: 221-240.

Näther, W. (2000), "On Random Fuzzy Variables of Second Order and Their Application to Linear Statistical Inference with Fuzzy Data," *Metrika* 51(3), 201–221.

Parkinson, M. (1980), "The Extreme Value Method for Estimating the Variance of the Rate of Return", *Journal of Business*, 53, 61-65.

Puri, M.L. and D. A. Ralescu (1983), “Strong Law of Large Numbers for Banach Space Valued Random Sets”, *Annals of Probability*, 11, 222-224.

Puri, M.L. and D. A. Ralescu (1985), “The Concept of Normality for Fuzzy Random Variables’, *Annals of Probability*, 13, 1373-1379.

Romanowska, A. B., and Smith, J. D. H. (1989), “Support Functions and Ordinal Products”, *Geometriae Dedicata,* 30(3), 281-296.

Stout, W. F. (1974), Almost Sure Convergence, Academic Press: New York.

Serfling, R. J. (1968), “Contributions to Central Limit Theory for Dependent Variables”, *The Annals of Mathematical Statistics*, Vol. 39 (4), 1158-1175.

Weiss, A. A. (1986), “Asymptotic Theory for ARCH Models: Estimation and Testing”, *Econometric Theory,* 2(1), 107-131.

White, H. (1999), Asymptotic Theory for Econometricians, Academic Press: New York.

TABLE 1. Bias, SD and RMSE of Estimates for Parameter α_0 in ACI (1,1)

$\widehat{\alpha}_0(10^{-4})$												
	$T=100$			$T=250$			$T=500$			$T=1000$		
a/b/c	Bias	S.D	RMSE	Bias	S.D	RMSE	Bias	S.D	RMSE	Bias	S.D	RMSE
K_r	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
K_m	−0.2444	2.6801	2.6912	0.4071	1.3415	1.4019	1.1184	0.9330	1.4565	0.4688	0.6446	0.7970
$CQML$	−0.2657	2.3823	2.3971	0.3592	1.2203	1.2721	1.0741	0.7732	1.3235	0.4505	0.5451	0.7071
$CCLS$	−0.2512	2.5193	2.5318	0.3867	1.3134	1.3691	1.0977	0.8682	1.3995	0.4487	0.5478	0.7081
10/2/10	−0.2347	2.5253	2.5362	0.3712	1.2564	1.3101	1.0924	0.8737	1.3989	0.4484	0.5668	0.7227
10/6/10	−0.2395	2.4510	2.4627	0.3680	1.2540	1.3069	1.0820	0.8274	1.3621	0.4605	0.5750	0.7367
10/8/10	−0.2344	2.5636	2.5743	0.3694	1.2334	1.2876	1.0951	0.8402	1.3803	0.4697	0.5795	0.7460
10/8/16	−0.2794	2.4169	2.4330	0.3576	1.2124	1.2640	1.0679	0.7677	1.3152	0.4508	0.5409	0.7041
10/8/17.5	−0.2985	2.5048	2.5225	0.3602	1.2129	1.2653	1.0783	0.8376	1.3654	0.4506	0.5392	0.7027
10/8/19	−0.2796	2.4242	2.4403	0.3588	1.2284	1.2797	1.0641	0.7643	1.3101	0.4523	0.5421	0.7060
10/6/6	−0.2438	2.4409	2.4531	0.3611	1.2251	1.2772	1.0766	0.7798	1.3293	0.4542	0.5481	0.7119
10/4/6	−0.2591	2.3690	2.3831	0.3516	1.2017	1.2521	1.0708	0.7713	1.3197	0.4479	0.5354	0.6981
10/2/6	−0.2494	2.3760	2.3891	0.3555	1.2028	1.2542	1.0688	0.7685	1.3164	0.4495	0.5361	0.6996
K^{opt}	−0.2817	2.3445	2.3613	0.3404	1.2074	1.2545	1.0541	0.7661	1.3031	0.4471	0.5390	0.7003

Notes: (a) ACI (1,1) Model: $Y_t = \alpha_0+\beta_0 I_0+\beta_1 Y_{t-1}+\gamma_1 u_{t-1}+u_t$.

(b) The kernel K used is of the form $K(1,1)=a$, $K(1,-1)=K(-1,1)=b$,and $K(-1,-1)=c$, and the values of $a/b/c$ are listed in the first column of the table. K_m, $K_r, CQML, CCLS$, and K^{opt} denote the estimates of $\widehat{\theta}^m, \widehat{\theta}^r, \widehat{\theta}_{CQML}, \widehat{\theta}_{CCLS}$ and $\widehat{\theta}^{opt}$ with special kernels, respectively.

(c) Bias, SD and the standard error of each parameter are computed based on 1000 bootstrap replications.

TABLE 2. Bias, SD and RMSE of Estimates for β_0 in ACI (1,1)

$\widehat{\beta}_0(10^{-4})$												
	$T = 100$			$T = 250$			$T = 250$			$T = 1000$		
a/b/c	Bias	S.D	RMSE	Bias	S.D	RMSE	Bias	S.D	RMSE	Bias	S.D	RMSE
K_r	−0.9955	3.0874	3.2439	−0.4164	1.4739	1.5316	−0.1285	0.9769	0.9853	−0.0436	0.6091	0.6102
K_m	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
$CQML$	−0.8701	2.6894	2.8267	−0.4099	1.3943	1.4533	−0.1311	0.8321	0.8424	−0.0344	0.5611	0.5622
$CCLS$	−0.9051	2.7350	2.8809	−0.4257	1.4680	1.5285	−0.1262	0.8351	0.8446	−0.0310	0.5660	0.5669
10/2/10	−0.9247	2.8410	2.9877	−0.4149	1.4129	1.4726	−0.1258	0.8431	0.8524	−0.0303	0.5662	0.5670
10/6/10	−0.9024	2.8262	2.9668	−0.4096	1.4204	1.4783	−0.1328	0.8514	0.8617	−0.0315	0.5817	0.5825
10/8/10	−0.9095	2.9480	3.0851	−0.3979	1.4372	1.4912	−0.1307	0.9091	0.9185	−0.0364	0.5838	0.5849
10/8/16	−0.8614	2.7421	2.8743	−0.3985	1.3815	1.4378	−0.1290	0.8311	0.8411	−0.0340	0.5617	0.5627
10/8/17.5	−0.8656	2.7661	2.8983	−0.4011	1.3816	1.4386	−0.1282	0.8289	0.8387	−0.0331	0.5633	0.5643
10/8/19	−0.8615	2.7690	2.8999	−0.4045	1.4038	1.4609	−0.1291	0.8267	0.8367	−0.0337	0.5647	0.5657
10/6/6	−0.8810	2.6996	2.8397	−0.4035	1.3844	1.4420	−0.1316	0.8297	0.8401	−0.0354	0.5644	0.5655
10/4/6	−0.8805	2.6806	2.8216	−0.4019	1.3806	1.4379	−0.1340	0.8278	0.8386	−0.0344	0.5612	0.5622
10/2/6	−0.9015	2.7442	2.8884	−0.4110	1.4060	1.4649	−0.1347	0.8357	0.8465	−0.0337	0.5644	0.5654
K^{opt}	−0.8521	2.6601	2.7933	−0.3998	1.3662	1.4235	−0.1373	0.8267	0.8380	−0.0393	0.5618	0.5632

Notes: (a) ACI (1,1) Model: $Y_t = \alpha_0 + \beta_0 I_0 + \beta_1 Y_{t-1} + \gamma_1 u_{t-1} + u_t$.

(b) The kernel K used is of the form $K(1,1) = a$, $K(1,-1) = K(-1,1) = b$,and $K(-1,-1) = c$, and the values of $a/b/c$ are listed in the first column of the table. K_m, $K_r, CQML, CCLS$, and K^{opt} denote the estimates of $\widehat{\theta}^m, \widehat{\theta}^r, \widehat{\theta}_{CQML}, \widehat{\theta}_{CCLS}$ and $\widehat{\theta}^{opt}$ with special kernels, respectively.

(c) Bias, SD and the standard error of each parameter are computed based on 1000 bootstrap replications..

TABLE 3. Bias, SD and RMSE of Estimates for β_1 in ACI (1,1)

$\widehat{\beta}_1(10^{-2})$												
	$T=100$			$T=250$			$T=500$			$T=1000$		
$a/b/c$	$Bias$	$S.D$	$RMSE$	$Bias$	$S.D$	$RMSE$	$Bias$	$S.D$	$RMSE$	$Bias$	$S.D$	$RMSE$
K_r	3.3167	12.6959	13.1219	1.8339	9.2751	9.4547	1.7655	11.8061	11.9374	1.5155	7.6049	7.7544
K_m	2.7914	9.6545	10.0499	2.0051	8.2841	8.5233	1.1157	5.7954	5.9018	1.0054	7.0858	7.1567
$CQML$	1.4442	4.8959	5.1045	1.1878	3.3584	3.5623	0.6951	2.5448	2.6380	0.4484	1.8921	1.9445
$CCLS$	2.1260	7.3187	7.6213	1.5549	6.3869	6.5735	0.9379	4.0115	4.1197	0.5439	3.2607	3.3057
10/2/10	2.1730	8.7214	8.9880	1.3087	4.7696	4.9459	0.8060	4.1959	4.2726	0.5650	4.7373	4.7709
10/6/10	1.7548	6.1015	6.3489	1.3739	6.4817	6.6257	0.7111	4.2881	4.3467	0.7729	5.1305	5.1884
10/8/10	2.6150	10.6366	10.9533	1.4524	5.7874	5.9669	1.1199	6.2604	6.3598	1.0061	4.2246	4.3427
10/8/16	1.6027	5.4063	5.6388	1.0325	3.1755	3.3391	0.5714	2.1095	2.1855	0.4718	1.6899	1.7545
10/8/17.5	1.8857	8.3394	8.5499	1.0520	2.9983	3.1775	0.6714	3.5870	3.6493	0.4593	1.5850	1.6502
10/8/19	1.6029	5.8121	6.0291	1.2322	3.8259	4.0195	0.5014	1.9390	2.0027	0.5116	1.9221	1.9890
10/6/6	1.8598	5.8567	6.1449	1.1452	3.9028	4.0673	0.6328	2.0452	2.1409	0.6074	2.1654	2.2490
10/4/6	1.3525	4.6540	4.8465	0.9408	3.1199	3.2587	0.5693	2.0440	2.1218	0.4444	1.9204	1.9711
10/2/6	1.6464	5.6329	5.8686	1.1112	3.7431	3.9046	0.6017	2.3274	2.4039	0.4506	1.9171	1.9693
K^{opt}	1.4759	3.8888	4.1594	1.0640	2.7109	2.9123	0.5954	1.7252	1.8251	0.4791	1.4757	1.5516

Notes: (a) Notes: (a) ACI (1,1) Model: $Y_t = \alpha_0 + \beta_0 I_0 + \beta_1 Y_{t-1} + \gamma_1 u_{t-1} + u_t$.

(b) The kernel K used is of the form $K(1,1) = a$, $K(1,-1) = K(-1,1) = b$,and $K(-1,-1) = c$, and the values of $a/b/c$ are listed in the first column of the table. K_m, $K_r, CQML, CCLS$, and K^{opt} denote the estimates of $\widehat{\theta}^m, \widehat{\theta}^r, \widehat{\theta}_{CQML}, \widehat{\theta}_{CCLS}$ and $\widehat{\theta}^{opt}$ with special kernels, respectively.

(c) Bias, SD and the standard error of each parameter are computed based on 1000 bootstrap replications

(d) Bias is in -1.

TABLE 4. Bias, SD and RMSE of Estimates for γ_1 in ACI (1,1)

$\widehat{\gamma}_1(10^{-2})$												
	$T = 100$			$T = 250$			$T = 500$			$T = 1000$		
a/b/c	Bias	S.D	RMSE	Bias	S.D	RMSE	Bias	S.D	RMSE	Bias	S.D	RMSE
K_r	1.3155	11.1962	11.2732	0.9540	8.7449	8.7968	1.3237	11.7092	11.7837	1.3032	7.3769	7.4911
K_m	0.9474	7.5063	7.5659	0.8086	7.1238	7.1695	0.6311	5.5730	5.6086	0.8032	7.0894	7.1347
$CQML$	0.0119	3.8593	3.8594	0.3591	2.5143	2.5398	0.2542	2.1876	2.2023	0.2390	1.7857	1.8016
$CCLS$	0.6861	5.3655	5.4092	0.6021	5.6838	5.7156	0.4957	3.6479	3.6814	0.3393	3.1898	3.2078
10/2/10	0.8254	7.2288	7.2758	0.4247	4.0132	4.0356	0.3595	3.9539	3.9702	0.3710	4.7699	4.7843
10/6/10	0.3976	4.6541	4.6710	0.5765	6.1216	6.1487	0.3275	4.1325	4.1454	0.5852	5.0710	5.1046
10/8/10	1.1149	9.4532	9.5187	0.6830	5.2189	5.2634	0.7946	6.0790	6.1307	0.8215	4.0581	4.1404
10/8/16	0.0414	4.0590	4.0592	0.1274	2.3714	2.3748	0.1520	1.9601	1.9660	0.2547	1.5249	1.5460
10/8/17.5	0.3295	7.5442	7.5514	0.1354	2.2202	2.2243	0.2504	3.5176	3.5265	0.2404	1.4375	1.4574
10/8/19	0.0339	4.3048	4.3049	0.2961	3.2550	3.2684	0.1059	1.9063	1.9283	0.2901	1.7531	1.7563
10/6/6	0.3309	4.3804	4.3929	0.4110	2.9218	2.9505	0.2118	1.8934	1.9052	0.4031	1.8694	1.9124
10/4/6	−0.0512	3.6302	3.6305	0.1862	2.5041	2.5110	0.1328	1.9850	1.9894	0.2586	1.7289	1.7482
10/2/6	0.2150	4.3349	4.3403	0.2654	2.9104	2.9224	0.1566	2.4103	2.4154	0.2663	1.7962	1.8158
K^{opt}	0.1942	2.2471	2.2555	0.2623	1.6412	1.6621	0.1756	1.4768	1.4872	0.2766	1.3554	1.3833

Notes: (a) ACI (1,1) Model: $Y_t = \alpha_0 + \beta_0 I_0 + \beta_1 Y_{t-1} + \gamma_1 u_{t-1} + u_t$.

(b) The kernel K used is of the form $K(1,1) = a$, $K(1,-1) = K(-1,1) = b$,and $K(-1,-1) = c$, and the values of $a/b/c$ are listed in the first column of the table. K_m, K_r, $CQML$, $CCLS$, and K^{opt} denote the estimates of $\widehat{\theta}^m, \widehat{\theta}^r, \widehat{\theta}_{CQML}, \widehat{\theta}_{CCLS}$ and $\widehat{\theta}^{opt}$ with special kernels, respectively.

(c) Bias, SD and the standard error of each parameter are computed based on 1000 bootstrap replications.

TABLE 5. Ratios of RMSE of Estimates for α_0 in Bivariate Point Processes

$\widehat{\alpha}_0$						
Normal	$corr(u_{Lt}, u_{Rt}) = 0$			$corr(u_{Lt}, u_{Rt}) = -0.6$		
	$T = 100$	$T = 250$	$T = 500$	$T = 100$	$T = 250$	$T = 500$
$\widehat{\theta}_{CCLS}()$				1.0437	1.0080	1.0175
$\widehat{\theta}_{ab}$	1.0427	1.0133	1.0040	1.1034	1.0344	1.0419
$\widehat{\theta}_{abc}$	1.0592	1.0132	1.0063	1.1166	1.0298	1.0432
$\widehat{\theta}^{opt}$	1.0018	1.0003	1.0000	0.9966	0.9962	0.9975
$\widehat{\theta}_{K_{\Sigma 0}}$	0.9972	0.9994	0.9996	0.9904	0.9957	0.9961
StudentT	$corr(u_{Lt}, u_{Rt}) = 0$			$corr(u_{Lt}, u_{Rt}) = -0.6$		
	$T = 100$	$T = 250$	$T = 500$	$T = 100$	$T = 250$	$T = 500$
$\widehat{\theta}_{CCLS}$				1.0336	1.0440	1.0271
$\widehat{\theta}_{ab}$	1.0040	1.0116	0.9975	1.2639	1.1385	1.0600
$\widehat{\theta}_{abc}$	1.0031	1.0121	1.0033	1.2746	1.1497	1.0705
$\widehat{\theta}^{opt}$	0.9952	0.9996	0.9993	0.9463	0.9675	0.9909
$\widehat{\theta}_{K_{\Sigma 0}}$	0.9977	1.0002	1.0008	0.8959	0.9638	0.9932
Mixture	$corr(u_{Lt}, u_{Rt}) = 0$			$corr(u_{Lt}, u_{Rt}) = -0.6$		
	$T = 100$	$T = 250$	$T = 500$	$T = 100$	$T = 250$	$T = 500$
$\widehat{\theta}_{CCLS}$				0.9696	0.9967	0.9992
$\widehat{\theta}_{ab}$	0.9923	1.0040	0.9992	0.9860	0.9925	1.0019
$\widehat{\theta}_{abc}$	0.9951	1.0042	0.9981	0.9804	1.0041	1.0053
$\widehat{\theta}^{opt}$	0.9943	0.9990	0.9985	0.9689	0.9986	0.9977
$\widehat{\theta}_{K_{\Sigma 0}}$	0.9952	1.0010	0.9985	0.9660	1.0001	0.9980

Notes: (a) $\widehat{\theta}_{ab}$ and $\widehat{\theta}_{CCLS}$ are from the kernel with the same weights for midpoint and range, and left and right bounds, respectively. $\widehat{\theta}_{abc}$ is from the kernel with $a/b/c = 10/8/19$, $\widehat{\theta}^{opt}$ is the two-stage minimum D_K-distance estimator, and $\widehat{\theta}_{K_{\Sigma 0}}$is the minimum D_K-distance estimator produced from Σ^0 .

(b) Bivariate Normal, Student T (df=5) and Mixture densities for u_{Lt} and u_{Rt} with $Corr(u_{Lt}, u_{Rt}) = 0$ and -0.6 are considered respectively. $\widehat{\theta}_{CCLS}$ is $\widehat{\theta}_{K_{\Sigma 0}}$ as $corr(u_{Lt}, u_{Rt}) = 0$

(c) Ratio of each parameter are computed based on 1000 bootstrap replications.

TABLE 6. Ratios of RMSE of Estimates for β_0 in Bivariate Point Processes

$\widehat{\beta}_0$						
Normal	$corr(u_{Lt}, u_{Rt}) = 0$			$corr(u_{Lt}, u_{Rt}) = -0.6$		
	$T = 100$	$T = 250$	$T = 500$	$T = 100$	$T = 250$	$T = 500$
$\widehat{\theta}_{CCLS}()$				1.0245	1.0047	1.0061
$\widehat{\theta}_{ab}$	1.0346	1.0054	1.0119	1.0698	1.0210	1.0157
$\widehat{\theta}_{abc}$	1.0419	1.0064	1.0159	1.0738	1.0215	1.0167
$\widehat{\theta}^{opt}$	1.0008	0.9998	1.0000	0.9954	0.9956	0.9972
$\widehat{\theta}_{K_{\Sigma 0}}$	0.9929	0.9992	0.9996	0.9898	0.9946	0.9973
StudentT	$corr(u_{Lt}, u_{Rt}) = 0$			$corr(u_{Lt}, u_{Rt}) = -0.6$		
	$T = 100$	$T = 250$	$T = 500$	$T = 100$	$T = 250$	$T = 500$
$\widehat{\theta}_{CCLS}$				1.0211	0.9978	1.0033
$\widehat{\theta}_{ab}$	1.0075	1.0104	1.0223	1.1100	1.0403	1.0214
$\widehat{\theta}_{abc}$	1.0132	1.0150	1.0258	1.1116	1.0398	1.0190
$\widehat{\theta}^{opt}$	0.9899	0.9968	0.9986	0.9585	0.9759	0.9937
$\widehat{\theta}_{K_{\Sigma 0}}$	0.9898	0.9933	0.9976	0.9722	0.9757	0.9990
Mixture	$corr(u_{Lt}, u_{Rt}) = 0$			$corr(u_{Lt}, u_{Rt}) = -0.6$		
	$T = 100$	$T = 250$	$T = 500$	$T = 100$	$T = 250$	$T = 500$
$\widehat{\theta}_{CCLS}$				0.9500	1.0059	0.9842
$\widehat{\theta}_{ab}$	1.0096	1.0164	1.0213	0.9679	1.0173	0.9938
$\widehat{\theta}_{abc}$	1.0085	1.0181	1.0220	0.9651	1.0178	0.9931
$\widehat{\theta}^{opt}$	0.9915	0.9960	0.9975	0.9399	0.9933	0.9716
$\widehat{\theta}_{K_{\Sigma 0}}$	0.9876	0.9988	0.9990	0.9363	0.9916	0.9703

Notes: (a) $\widehat{\theta}_{ab}$ and $\widehat{\theta}_{CCLS}$ are from the kernel with the same weights for midpoint and range, and left and right bounds, respectively. $\widehat{\theta}_{abc}$ is from the kernel with $a/b/c = 10/8/19$, $\widehat{\theta}^{opt}$ is the two-stage minimum D_K-distance estimator, and $\widehat{\theta}_{K_{\Sigma 0}}$is the minimum D_K-distance estimator produced from Σ^0 .

(b) Bivariate Normal, Student T (df=5) and Mixture densities for u_{Lt} and u_{Rt} with $Corr(u_{Lt}, u_{Rt}) = 0$ and -0.6 are considered respectively. $\widehat{\theta}_{CCLS}$ is $\widehat{\theta}_{K_{\Sigma 0}}$ as $corr(u_{Lt}, u_{Rt}) = 0$

(c) Ratio of each parameter are computed based on 1000 bootstrap replications.

TABLE 7. Ratios of RMSE of Estimates for β_1 in Bivariate Point Processes

$\widehat{\beta}_1$						
$Normal$	$corr(u_{Lt}, u_{Rt}) = 0$			$corr(u_{Lt}, u_{Rt}) = -0.6$		
	$T = 100$	$T = 250$	$T = 500$	$T = 100$	$T = 250$	$T = 500$
$\widehat{\theta}_{CCLS}()$				1.1576	1.0963	1.1890
$\widehat{\theta}_{ab}$	1.1582	1.1238	1.1846	1.3805	1.2677	1.3872
$\widehat{\theta}_{abc}$	1.2024	1.1395	1.2256	1.3945	1.2723	1.4012
$\widehat{\theta}^{opt}$	1.0022	0.9990	1.0000	1.0044	0.9987	0.9916
$\widehat{\theta}_{K_{\Sigma^0}}$	0.9843	0.9925	0.9980	0.9844	0.9946	0.9957
$StudentT$	$corr(u_{Lt}, u_{Rt}) = 0$			$corr(u_{Lt}, u_{Rt}) = -0.6$		
	$T = 100$	$T = 250$	$T = 500$	$T = 100$	$T = 250$	$T = 500$
$\widehat{\theta}_{CCLS}$				1.0405	1.1057	1.1498
$\widehat{\theta}_{ab}$	1.0814	1.1040	1.1337	1.2811	1.3202	1.3416
$\widehat{\theta}_{abc}$	1.0994	1.1307	1.1618	1.3064	1.3234	1.3401
$\widehat{\theta}^{opt}$	0.9493	0.9697	0.9807	0.8845	0.9468	0.9781
$\widehat{\theta}_{K_{\Sigma^0}}$	0.9539	0.9621	0.9803	0.8399	0.9345	0.9785
$Mixture$	$corr(u_{Lt}, u_{Rt}) = 0$			$corr(u_{Lt}, u_{Rt}) = -0.6$		
	$T = 100$	$T = 250$	$T = 500$	$T = 100$	$T = 250$	$T = 500$
$\widehat{\theta}_{CCLS}$				0.8817	0.8656	1.0118
$\widehat{\theta}_{ab}$	1.0736	1.1181	1.1582	1.0034	1.1528	0.9855
$\widehat{\theta}_{abc}$	1.0970	1.1440	1.1775	1.0016	1.1557	0.9923
$\widehat{\theta}^{opt}$	0.9442	0.9649	0.9746	0.7991	0.8774	0.7446
$\widehat{\theta}_{K_{\Sigma^0}}$	0.9403	0.9643	0.9768	0.8001	0.8790	0.7418

Notes: (a) $\widehat{\theta}_{ab}$ and $\widehat{\theta}_{CCLS}$ are from the kernel with the same weights for midpoint and range, and left and right bounds, respectively. $\widehat{\theta}_{abc}$ is from the kernel with $a/b/c = 10/8/19$, $\widehat{\theta}^{opt}$ is the two-stage minimum D_K-distance estimator, and $\widehat{\theta}_{K_{\Sigma^0}}$ is the minimum D_K-distance estimator produced from Σ^0 .

(b) Bivariate Normal, Student T (df=5) and Mixture densities for u_{Lt} and u_{Rt} with $Corr(u_{Lt}, u_{Rt}) = 0$ and -0.6 are considered respectively. $\widehat{\theta}_{CCLS}$ is $\widehat{\theta}_{K_{\Sigma^0}}$ as $corr(u_{Lt}, u_{Rt}) = 0$

(c) Ratio of each parameter are computed based on 1000 bootstrap replications.

TABLE 8. Ratios of RMSE of Estimates for γ_1 in Bivariate Point Processes

$\widehat{\gamma}_1$						
Normal	$corr(u_{Lt}, u_{Rt}) = 0$			$corr(u_{Lt}, u_{Rt}) = -0.6$		
	$T = 100$	$T = 250$	$T = 500$	$T = 100$	$T = 250$	$T = 500$
$\widehat{\theta}_{CCLS}$				1.1442	1.1319	1.1650
$\widehat{\theta}_{ab}$	1.1568	1.1518	1.1625	1.3324	1.3044	1.3445
$\widehat{\theta}_{abc}$	1.1869	1.1734	1.1961	1.3438	1.3078	1.3567
$\widehat{\theta}^{opt}$	1.0004	0.9986	1.0000	1.0002	0.9990	0.9953
$\widehat{\theta}_{K_{\Sigma^0}}$	0.9905	0.9907	0.9959	0.9539	0.9593	0.9809
Student T	$corr(u_{Lt}, u_{Rt}) = 0$			$corr(u_{Lt}, u_{Rt}) = -0.6$		
	$T = 100$	$T = 250$	$T = 500$	$T = 100$	$T = 250$	$T = 500$
$\widehat{\theta}_{CCLS}$				1.0713	1.1626	1.1874
$\widehat{\theta}_{ab}$	1.0905	1.0951	1.1318	1.2935	1.3548	1.3696
$\widehat{\theta}_{abc}$	1.1155	1.1169	1.1578	1.3176	1.3529	1.3740
$\widehat{\theta}^{opt}$	0.9420	0.9707	0.9797	0.9173	0.9889	0.9635
$\widehat{\theta}_{K_{\Sigma^0}}$	0.9399	0.9674	0.9785	0.8827	0.9824	0.9613
Mixture	$corr(u_{Lt}, u_{Rt}) = 0$			$corr(u_{Lt}, u_{Rt}) = -0.6$		
	$T = 100$	$T = 250$	$T = 500$	$T = 100$	$T = 250$	$T = 500$
$\widehat{\theta}_{CCLS}$				0.9308	1.0579	0.9817
$\widehat{\theta}_{ab}$	1.0939	1.1105	1.1648	1.0629	1.2171	1.1130
$\widehat{\theta}_{abc}$	1.1229	1.1402	1.1744	1.0639	1.2141	1.1222
$\widehat{\theta}^{opt}$	0.9391	0.9570	0.9717	0.8452	0.9089	0.8423
$\widehat{\theta}_{K_{\Sigma^0}}$	0.9408	0.9588	0.9741	0.8470	0.9058	0.8417

Notes. (a) $\widehat{\theta}_{ab}$ and $\widehat{\theta}_{CCLS}$ are from the kernel with the same weights for midpoint and range, and left and right bounds, respectively.

$\widehat{\theta}_{abc}$ is from the kernel with $a/b/c = 10/8/19$, $\widehat{\theta}^{opt}$ is the two-stage minimum D_K-distance estimator, and $\widehat{\theta}_{K_{\Sigma^0}}$is the minimum D_K-distance estimator produced from Σ^0 .

(b) Bivariate Normal, Student T (df=5) and Mixture densities for u_{Lt} and u_{Rt} with $Corr(u_{Lt}, u_{Rt}) = 0$ and -0.6 are considered respectively. $\widehat{\theta}_{CCLS}$ is $\widehat{\theta}_{K_{\Sigma^0}}$ as $corr(u_{Lt}, u_{Rt}) = 0$

(c) Ratio of each parameter are computed based on 1000 bootstrap replications.

TABLE 9. *t-statistics* **for 5-Factor CAPM**

		BE/ME Quantile Group Low						BE/ME Quantile Group 2					
Small		OLS	$\widehat{\theta}_{CCLS}$	$\widehat{\theta}_{ab}$	$\widehat{\theta}_{abc}$	$\widehat{\theta}_{QML}$	$\widehat{\theta}^{opt}$	OLS	$\widehat{\theta}_{CCLS}$	$\widehat{\theta}_{ab}$	$\widehat{\theta}_{abc}$	$\widehat{\theta}_{QML}$	$\widehat{\theta}^{opt}$
	β_0	−3.69	−1.03	−1.04	−1.04	−1.04	−1.04	−1.06	−1.17	−1.18	−1.18	−1.18	−1.17
	β_1	39.79	8.39	9.03	9.68	10.21	10.68	46.57	8.60	9.43	10.25	11.16	11.41
	β_2	35.19	8.22	8.31	10.27	14.62	19.79	45.16	7.89	8.04	10.27	17.37	20.98
	β_3	−5.76	−6.99	−6.56	−8.32	−12.23	−16.05	3.28	−5.93	−5.55	−7.30	−12.55	−14.83
	β_4	−2.21	−0.61	−0.64	−0.65	−0.65	−0.67	−2.50	−0.43	−0.45	−0.46	−0.47	−0.48
	β_5	−1.54	−3.35	−2.74	−4.05	−5.37	−5.81	−2.86	−3.44	−2.92	−4.31	−5.84	−6.08
2	β_0	−1.45	−1.10	−1.10	−1.10	−1.10	−1.10	−0.27	−1.28	−1.29	−1.29	−1.29	−1.27
	β_1	50.19	9.63	10.26	10.93	11.88	12.17	56.87	9.97	10.88	12.05	13.51	13.82
	β_2	32.39	6.79	6.83	8.32	14.83	18.71	36.38	6.23	6.32	8.81	19.38	21.10
	β_3	−13.10	−6.30	−5.90	−7.38	−13.57	−16.71	0.89	−4.82	−4.51	−6.72	−14.80	−16.04
	β_4	−1.05	−0.04	−0.04	−0.04	−0.05	−0.06	−0.62	−0.11	−0.12	−0.24	−0.24	−0.25
	β_5	−2.33	−3.01	−2.37	−3.64	−5.78	−6.07	−1.08	−2.52	−2.07	−3.12	−5.43	−5.57
3	β_0	−0.37	−1.21	−1.21	−1.21	−1.21	−1.20	1.45	−1.40	−1.41	−1.41	−1.41	−1.37
	β_1	53.22	10.75	11.44	12.20	13.62	13.72	48.27	11.22	12.21	13.25	15.15	15.15
	β_2	23.27	5.42	5.44	6.63	15.93	18.26	21.97	4.51	4.57	5.81	19.88	20.84
	β_3	−12.87	−5.22	−4.88	−6.12	−15.00	−16.96	1.35	−3.43	−3.20	−4.20	−14.43	−15.08
	β_4	−0.36	0.00	0.00	0.01	0.00	−0.02	0.76	0.25	0.27	0.27	0.28	0.27
	β_5	−0.93	−2.17	−1.68	−2.66	−5.22	−5.32	−0.36	−1.83	−1.47	−2.38	−5.19	−5.32
4	β_0	1.70	−1.29	−1.29	−1.30	−1.30	−1.27	−2.09	−1.51	−1.52	−1.52	−1.52	−1.49
	β_1	50.28	12.14	12.81	13.58	15.46	15.42	45.92	12.83	13.83	14.92	17.47	17.52
	β_2	10.32	3.53	3.53	4.24	15.45	16.21	8.33	2.14	2.16	2.70	20.14	20.30
	β_3	−14.44	−3.96	−3.69	−4.56	−17.00	−17.77	0.55	−1.61	−1.49	−1.94	−14.69	−14.80
	β_4	0.32	0.37	0.39	0.40	0.41	0.41	0.86	0.47	0.49	0.50	0.58	0.58
	β_5	−1.54	−1.61	−1.20	−1.95	−5.23	−5.22	−0.72	−1.14	−0.88	−1.48	−6.89	−7.03
Big	β_0	3.32	−1.42	−1.43	−1.43	−1.43	−1.38	−0.31	−1.63	−1.64	−1.89	−1.89	−1.84
	β_1	51.25	14.57	15.09	15.80	19.79	19.66	51.01	15.05	15.89	16.63	20.16	20.49
	β_2	−7.96	−0.14	−0.14	−0.16	−1.23	−1.21	−6.94	−1.58	−1.58	0.48	2.76	2.79
	β_3	−15.67	−1.41	−1.31	−1.57	−15.16	−14.90	−0.32	1.21	1.14	3.34	21.49	21.39
	β_4	0.90	0.31	0.32	0.33	0.39	0.38	−0.50	0.30	0.31	0.61	0.70	0.71
	β_5	0.88	0.73	0.53	0.83	3.01	2.99	−1.39	0.19	0.14	−1.25	−2.73	−2.92

Notes: (a) FF's 5-Factor CAPM: $ER_{it}= \beta_0+\beta_1 EM_t+\beta_2 SMB_t+\beta_3 HML_t+\beta_4 TERM_t+\beta_5 DEF_t+\varepsilon_t$, with OLS. Interval CAPM: $Y_t= \alpha_0+\beta_0 I_0+\beta_1 X_{1t}+\beta_2 X_{2t}+\beta_3 X_{3t}+\beta_4 X_{4t}+\beta_5 X_{5t}+u_t$, QML and minimum Dk-distance estimation, where $i = 1,.., 25$.

(b) $\widehat{\theta}_{CCLS}$ is from the CCLS estimator for the left and right bounds , $\widehat{\theta}_{ab}$ is from the kernel with the same weights for midpoint and range, $\widehat{\theta}_{abc}$ is from the kernel with $a/b/c = 10/8/19$, $\widehat{\theta}_{QML}$ and $\widehat{\theta}^{opt}$ are the QML and the two-stage minimum D_K-distance estimator. OLS denote the estimators of the F-F 5-factor model.

(c) The standard error of each parameter estimate is compared based on 500 bootstrap replications.

TABLE 9. [Continued] *t-statistics* **for 5-Factor CAPM**

		BE/ME Quantile Group 3						BE/ME Quantile Group 4					
Small		OLS	$\widehat{\theta}_{CCLS}$	$\widehat{\theta}_{ab}$	$\widehat{\theta}_{abc}$	$\widehat{\theta}_{QML}$	$\widehat{\theta}^{opt}$	OLS	$\widehat{\theta}_{CCLS}$	$\widehat{\theta}_{ab}$	$\widehat{\theta}_{abc}$	$\widehat{\theta}_{QML}$	$\widehat{\theta}^{opt}$
	β_0	−1.23	−1.28	−1.28	−1.28	−1.28	−1.27	1.26	−1.36	−1.37	−1.37	−1.37	−1.35
	β_1	53.51	8.93	9.90	10.85	12.04	12.15	52.48	9.20	10.30	11.38	12.76	12.80
	β_2	46.59	7.39	7.56	9.87	19.26	21.75	47.03	7.06	7.25	9.65	20.70	22.50
	β_3	9.72	−5.09	−4.77	−6.42	−12.57	−13.98	14.39	−4.40	−4.12	−5.67	−12.09	−13.01
	β_4	−0.15	−0.20	−0.21	−0.22	−0.22	−0.24	−1.38	−0.64	−0.68	−0.70	−0.72	−0.74
	β_5	−0.29	−2.69	−2.32	−3.43	−4.73	−4.87	0.82	−2.34	−2.04	−3.01	−4.12	−4.23
2	β_0	2.58	−1.39	−1.40	−1.40	−1.40	−1.37	2.86	−1.58	−1.59	−1.59	−1.59	−1.56
	β_1	54.15	10.14	11.22	12.30	13.95	13.95	56.81	10.98	12.31	13.66	15.81	15.79
	β_2	34.86	5.77	5.89	7.69	20.67	21.98	30.51	4.37	4.49	6.05	23.78	24.25
	β_3	8.85	−3.81	−3.56	−4.81	−12.96	−13.70	17.00	−1.92	−1.78	−2.50	−9.97	−10.16
	β_4	1.61	0.41	0.44	0.46	0.47	0.46	3.57	0.58	0.62	0.64	0.68	0.69
	β_5	−0.67	−2.35	−1.98	−3.07	−5.01	−5.15	2.11	−1.27	−1.09	−1.70	−2.98	−3.13
3	β_0	0.02	−1.56	−1.56	−1.56	−1.56	−1.53	2.39	−1.72	−1.73	−1.73	−1.73	−1.65
	β_1	46.61	11.56	12.81	14.08	16.37	16.37	51.57	12.36	13.84	15.35	19.24	19.37
	β_2	18.83	3.69	3.77	4.96	23.23	23.55	17.29	2.15	2.20	2.96	25.16	25.19
	β_3	9.77	−1.86	−1.73	−2.36	−11.34	−11.50	17.20	0.21	0.21	0.28	2.58	2.56
	β_4	2.00	0.54	0.58	0.60	0.65	0.66	3.05	0.64	0.69	0.71	0.97	0.97
	β_5	0.37	−1.44	−1.20	−1.93	−4.18	−4.36	1.66	−0.69	−0.59	−0.94	−2.74	−2.78
4	β_0	0.49	−1.67	−1.68	−1.68	−1.68	−1.57	1.07	−1.80	−1.81	−1.81	−1.81	−1.75
	β_1	45.41	13.21	14.47	15.80	19.67	19.66	46.39	13.25	14.74	16.27	19.76	19.87
	β_2	7.80	1.10	1.11	1.44	19.29	19.20	8.31	0.11	0.10	0.14	0.92	0.91
	β_3	8.61	0.25	0.24	0.31	4.74	4.65	16.21	2.16	2.03	2.78	20.99	20.87
	β_4	1.30	0.63	0.67	0.69	0.95	0.95	4.53	1.54	1.64	1.68	2.11	2.14
	β_5	−0.50	−0.87	−0.70	−1.16	−5.59	−5.57	0.66	−0.37	−0.31	−0.51	−1.42	−1.47
Big	β_0	−0.44	−1.74	−1.75	−1.75	−1.75	−1.72	−0.65	−1.96	−1.97	−1.97	−1.97	−1.95
	β_1	38.59	15.56	16.49	17.57	20.64	20.73	52.41	15.87	17.13	18.47	20.98	21.09
	β_2	−7.28	−3.24	−3.24	−3.95	−17.92	−17.83	−7.83	−4.82	−4.87	−6.10	−14.98	−14.89
	β_3	5.52	3.28	3.06	3.81	17.46	17.40	17.53	6.38	5.94	7.64	18.91	18.81
	β_4	0.47	0.60	0.62	0.63	0.67	0.66	−0.62	−0.03	−0.03	−0.03	−0.05	−0.06
	β_5	−0.76	0.80	0.61	0.98	3.17	3.16	−0.14	1.11	0.89	1.38	2.64	2.66

TABLE 9. [Continued] *t-statistics* **for 5-Factor CAPM**

		BE/ME Quantile Group High					
Small		OLS	$\widehat{\theta}_{CCLS}$	$\widehat{\theta}_{ab}$	$\widehat{\theta}_{abc}$	$\widehat{\theta}_{QML}$	$\widehat{\theta}^{opt}$
	β_0	1.02	−1.48	−1.48	−1.48	−1.48	−1.46
	β_1	51.01	9.28	10.54	11.79	13.31	13.34
	β_2	46.60	6.86	7.11	9.75	21.73	23.58
	β_3	20.76	−3.51	−3.29	−4.68	−10.32	−11.08
	β_4	−1.92	−0.80	−0.86	−0.89	−0.92	−0.96
	β_5	0.36	−2.45	−2.20	−3.17	−4.14	−4.28
2	β_0	1.25	−1.70	−1.70	−1.71	−1.71	−1.67
	β_1	57.83	11.12	12.66	14.23	16.67	16.74
	β_2	32.13	4.21	4.36	6.05	24.54	25.18
	β_3	23.54	−0.96	−0.89	−1.30	−5.53	−5.65
	β_4	−0.95	−0.22	−0.24	−0.24	−0.30	−0.35
	β_5	−1.07	−2.09	−1.85	−2.80	−4.47	−4.79
3	β_0	0.53	−1.76	−1.77	−1.77	−1.77	−1.73
	β_1	45.46	11.66	13.27	14.90	17.75	17.91
	β_2	21.51	3.01	3.11	4.31	20.72	21.02
	β_3	21.28	0.26	0.26	0.35	1.75	1.76
	β_4	1.42	0.88	0.95	0.98	1.14	1.18
	β_5	−2.48	−2.30	−2.03	−3.10	−5.50	−5.87
4	β_0	0.64	−1.88	−1.89	−1.89	−1.89	−1.84
	β_1	44.92	13.27	14.93	16.63	20.16	20.49
	β_2	9.74	0.35	0.35	0.48	2.76	2.79
	β_3	17.23	2.53	2.37	3.34	21.49	21.39
	β_4	1.22	0.55	0.59	0.61	0.70	0.71
	β_5	−0.47	−0.92	−0.79	−1.25	−2.73	−2.92
Big	β_0	−1.90	−2.08	−2.08	−2.09	−2.09	−2.07
	β_1	36.22	15.38	17.00	18.64	21.13	21.28
	β_2	−1.37	−4.09	−4.19	−5.45	−11.94	−11.87
	β_3	17.03	7.07	6.59	8.83	19.51	19.39
	β_4	−2.18	−0.28	−0.29	−0.29	−0.32	−0.33
	β_5	−2.62	−0.80	−0.68	−1.03	−1.62	−1.64