Autoregressive Conditional Models for Interval-Valued Time Series Data

Ai Han
Chinese Academy of Sciences

Yongmiao Hong
Cornell University

Shouyang Wang
Chinese Academy of Sciences

This version, December 2013

We are benefited from the comments and suggestions from Hongzhi An, Donald Andrews, Gloria González-Rivera, Gil González-Rodríguez, Cheng Hsiao, James Hamilton, Jerry A. Hausman, Oliver Linton, Qiwei Yao, the seminar participants at Australian National University, Boston University, Cornell University, London School of Economics and Political Science, Yale University, and the conference participants at the Australian Econometric Society at Adelaide 2011, the conference in Honor of Halbert White, "Causality, Prediction, and Specification Analysis: Recent Advances and Future Directions" at San Diego 2011, the International Conference of the ERCIM Working Group on Computing & Statistics at London 2011, the International Conference on Computational Statistics at Limassol 2012. We gratefully acknowledge research support from the National Natural Science Foundation of China Grants No. 71201161.

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 top 10% income and bottom 10% income 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. More importantly, 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 outperforms 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, Level, Mean squared error, Minimum distance estimation, Range

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 $\mathbf{y} = [a, b] = \{y \in \mathbf{R}| \quad a \leq y \leq b$, 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 (1995, 2003, 2007, 2013), Manski and Tamer (2002), Andrews and Shi (2009), Andrews and Soares (2010), Beresteanu and Molinari (2008), Chernozhukov, Hong, and Tamer (2007), Chernozhukov, Rigobon and Stoker (2010), Bontemps, Magnac and Maurin (2012)). 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 captures both the 'range' (or 'variability') 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 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 bottom 10% and top 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 variability and the trend of economic processes. Russell and Engle (2009) argued that the high-frequency 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 alternative way to analyze intraday time series. Other 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). More 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 drawback of the CARR models is that using range as a volatility measure is unable to simultaneously capture the dual empirical features, i.e., 'variability' and 'level'. For example, the same range observations in different time periods may have the same range but distinct price levels.

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 and estimating the two univariate point-valued processes separately or joint modelling with vector autogression; see Maia, Carvalho and Ludermir (2008), Neto, Carvalho and Freire (2008), Neto and Carvalho (2010), Arroyo, Espínola and Maté (2011), Arroyo, González-Rivera, and Maté (2011), Lin and González-Rivera (2013), and the references therein. However, a bivariate point-valued sample may not efficiently make use of the information of the underlying interval process, and possible limitations often arise in handling separate classical studies; see Gil, González-Rodríguez, Colubi and Montenegro (2007), Blanco-Fernández, Corral and González-Rodríguez (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 model 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 interval 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 based on point-valued data. Consider the case of modelling the conditional range of the daily price of some asset where there are more variabilities in the level sample than in the range sample. Since 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 weighting 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 performs the best in finite sample, confirming our asymptotic analysis. 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 $Y : \Omega \rightarrow I_{\mathbf{R}}$ in Definition 2.1 is 'strongly measurable' with respect to 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 $X : \Omega_X \rightarrow \mathbf{R}^2$ 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 $Y : \Omega \rightarrow I_{\mathbf{R}}$ 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 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 conventional interval analysis (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 $\beta \cdot Y_t = -1 \times [1, 3] = [-1, -3] \in I_{\mathbf{R}}$.

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

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

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

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

The concept of extended interval together with Hukuhara's difference is suitable for econometric modelling 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 interval 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. Below our notation follows a convention throughout the paper: the scalar multiplication, e.g., $\beta \cdot A$, will be presented as βA, while the Hukuhara difference $A -_H B$ is simply represented as $A - B$.

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.

It may be noted that the concept of extended random interval differs from that of a confidence interval in statistical analysis, even if the restriction $Y_L \leq Y_R$ is imposed. 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 expectation of an interval' 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. Our main objective is to use the observed interval data to infer the dynamic structure of the interval time series $\{Y_t\}$ 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

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

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.

A 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), Li, Ogura and Kreinovich (2002), Molchanov (2005), Beresteanu and Molinari (2008), Beresteanu, Molchanov and Molinari (2011, 2012), Chandrasekhar, Chernozhukov, Molinari and Schrimpf (2012)). 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

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

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

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

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 a conditional expectation model of an interval process.

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

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

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 become extended random intervals, and the generalized $\mathcal{F}_c(\mathbf{R})$ space is $I_{\mathbf{R}}$. For any pair of extended intervals $A, B \in I_{\mathbf{R}}$,

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

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

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

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

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 $\lambda \geq 0$.

Our 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 $A_L \leq A_R$, 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 a Hukuhara difference between two extended intervals, is equal to the difference between the corresponding support functions of the two intervals. For more discussions on support functions, see Rockafellar (1970), Romanowska and Smith (1989), Choi and Smith (2003), Li, Ogura, and Kreinovich (2002), Molchanov (2005), Beresteanu and Molinari (2008), Beresteanu, Molchanov and Molinari (2011, 2012), Bontemps, Magnac and Maurin (2012), Chandrasekhar, Chernozhukov, Molinari and Schrimpf (2012).

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:

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

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

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

and similarly,

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

The D_K-metric has certain 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 the conditional mean of an interval process. This is stated in Lemma 2.1 below.

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

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

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).

Intuitively, 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:

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

Bertoluzza, Corral and Salas (1995) first 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

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

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 interesting to see that the D_K metric as a generalization of the d_W metric preserves this property (González-Rodríguez, Blanco-Fernández, Corral and Colubi (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,

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

Instead of considering 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 a certain 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 $K(u, v)$.

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

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

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 $\frac{3}{5} < \frac{b}{a} < 1$, $(A^r - B^r)^2$ 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 c and a respectively. If $0 < a < c$, $(A_L - B_L)^2$ 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)

$$\begin{aligned} & D_K^2(A, B) = a(A_R - B_R)^2 + c(A_L - B_L)^2 - 2b(A_R - B_R)(A_L - B_L) \\ = & \frac{a+2b+c}{4}(A^r - B^r)^2 + (a - 2b + c)(A^m - B^m)^2 + (a-c)(A^r - B^r)(A^m - B^m). \end{aligned}$$

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 equivalently, 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 autocovariance function of a stochastic interval time series process $\{Y_t\}$, *denoted* $\gamma_t(j)$, *is a scalar defined by*

$$\gamma_t(j) \equiv \text{cov}(Y_t, Y_{t-j}) = E\left\langle s_{Y_t} - s_{\mu_t}, s_{Y_{t-j}} - s_{\mu_{t-j}} \right\rangle_K, \; j = 0, \pm1, \pm2, ...,$$

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 the kernel* $K(u,v)$ *is symmetric.*

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

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

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* $\gamma_t(j)$, *for each* j, *of a stochastic interval time series process* $\{Y_t\}$ *depends on time* t, *then* $\{Y_t\}$ *is weakly stationary with respect to* D_K, *or covariance stationary with respect to* D_K.

Suppose $\{Y_t\}$ is a weakly stationary interval process with respect to D_K. Then an induced 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 $(\Omega, \mathcal{F}, P)$ 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 weakly stationary with respect to D_K.

If $\gamma(j) = 0$ for all $j \neq 0$, we say that the weakly stationary interval process $\{Y_t\}$ with respect to D_K is a *white noise* process with respect to D_K. 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 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), Molchanov (1993), Li, Ogura, and Kreinovich (2002). In partial identification analysis, related works applying random set theory include Molchanov (2005) who metrises the weak convergence of random closed sets; Beresteanu and Molinari (2008) who use limit theorems for *i.i.d.* random sets to establish consistency of their estimator for the sharp identification region of the parameter vector with respect to the Hausdorff metric; see also the references therein.

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 weakly stationary with respect to D_K interval process $\{Y_t\}$ with $E(Y_t) = \mu$ for all t, $E\langle s_{Y_t} - s_\mu, s_{Y_{t-j}} - s_\mu\rangle_K = \gamma(j)$ for *all t and j, and* $\sum_{j=-\infty}^{\infty} |\gamma(j)| < \infty$ *. Then* $\bar{Y}_T \xrightarrow{p} \mu$ *as* $T \to \infty$*, where* $\bar{Y}_T = T^{-1}\sum_{t=1}^{T} Y_t$ *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* $\lim_{T\to\infty} P\left[D_K(\bar{Y}_T, \mu) \geq \epsilon\right] = 0$*, 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 $\bar{Y}_T$ converges to the population mean μ of a weakly stationary interval process $\{Y_t\}$ with respect to D_K. In Theorem 2.1, the sample average $\bar{Y}_T$ 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

$$D_K(\bar{Y}_T, \mu) = d_H(\bar{Y}_T, \mu) = \left|\bar{Y}_T - \mu\right|,$$

subject to $\int_{(u,v)\in\mathbf{S}^0} dK(u,v) = 1$. Thus, Theorem 2.1 coincides with the familiar WLLN for a point-valued time series process, i.e., $\lim_{T\to\infty} P\left[\left|\bar{Y}_T - \mu\right| \geq \epsilon\right] = 0$ 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* $E\left\|Y_t\right\|_K^2 < \infty$ *for all t. Suppose the conditions of Theorem 2.1 hold. Then for each given* $j \in \{0, \pm 1, \pm 2, ...\}$,

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

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

Theorem 2.2 provides sufficient conditions that a weakly stationary interval process with respect to D_K is ergodic in second moments. Since the weighted inner product $\langle\cdot,\cdot\rangle_K$ is a scalar, the convergence in probability in Theorem 2.2 is with respect to either the D_K or d_H metric.

2.4 Autoregressive Conditional Interval Models

To capture the dynamics of a stochastic interval process $\{Y_t\}$, we first 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; $A(L)=1+\Sigma_{j=1}^{q}\gamma_j L^j$ 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) with respect to the information set I_{t-1}, that is, $E(u_t|I_{t-1})=[0,0]$ *a.s.* It is noted that the parameters in ACI models are scalar-valued rather than set-valued.

The ACI(p,q) model is an interval generalization of the well-known 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 $\{u_s, s\le t\}$, 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 $\{\alpha_j\}$ is given by $B(L)^{-1}A(L)=\Sigma_{j=0}^{\infty}\alpha_j L^j$. 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 $\{Y_s, s\le t\}$, 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 $B(L)^{-1}A(L)=\Sigma_{j=0}^{\infty}\lambda_j L^j$.

An ACI(p,q) model of an interval process can be extended to an 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 time series 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, which can capture range clustering.

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 a time series. The difference is that the derived range model in Eq.(2.18), with an ACIX(p, q, s) model as the data generating process (DGP), 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 the ACIX model more efficiently 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 $E(u_t|I_{t-1}) = 0$ $a.s.$ This can be used to forecast the trend of a time series process.

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

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

where $E(u_{L,t}|I_{t-1}) = E(u_{R,t}|I_{t-1}) = 0$ $a.s.$ given $E(u_t|I_{t-1}) = [0, 0]$ $a.s.$

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 $E\|Y_t\|_K^4 < \infty$, and it follows an ACIX(p, q, s) process in Eq.(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 $\mathbf{R}^{\mathbf{k}}$ where $\mathbf{k} = p+q+(s+1)J+2$. (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 Eq.(2.16).

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

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

3.1 Minimum D_K-Distance Estimation

Given that $E(Y_t|I_{t-1})$ is the optimal solution to minimize $E[D_K^2(Y_t, A(I_{t-1}))]$, as is established in Lemma 2.1, we will propose an estimation method that minimizes a sample analog of $E[D_K^2(Y_t, A(I_{t-1}))]$. 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:

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

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

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

$$q_t(\theta) = \|u_t(\theta)\|_K^2, \quad (3.2)$$

and

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

Since we only observe $\{Y_t, X_t'\}$ from time $t=1$ to time $t=T$, 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 interval error process $\{u_t(\theta)\}$.

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

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

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

Intuitively, the statistics $\hat{Q}_T(\theta)$ converges in probability to $E[D_K^2(Y_t, Z_t'(\theta)\theta)]$ uniformly in Θ as $T \to \infty$. Furthermore, the true model parameter θ^0 is the unique minimizer of $E[D_K^2(Y_t, Z_t'(\theta)\theta]$ given the IMDS condition on the interval innovation process $\{u_t\}$. It then follows from the extremum estimator theorem (e.g., Amemiya (1985)) that $\hat{\theta} \xrightarrow{p} \theta^0$ as $T \to \infty$.

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

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

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

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

The asymptotic variance of $\sqrt{T}(\hat{\theta}-\theta^0)$, 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 $T \to \infty$,

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

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

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

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

We note that the asymptotic variance of $\sqrt{T}\hat{\theta}$ 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 $\hat{\theta}$ has a convenient closed form that is similar 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* $\hat{\theta}$ *has the closed form*

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

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

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

Furthermore, as $T \to \infty$,

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

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

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

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

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

where $\hat{u}_{L,t}(\theta)$ and $\hat{u}_{R,t}(\theta)$ are the left and right bounds of $\hat{u}_t(\theta)$, $\hat{u}_t^r(\theta) = \hat{u}_{R,t}(\theta) - \hat{u}_{L,t}(\theta)$ and $\hat{u}_t^m(\theta) = \frac{1}{2}\hat{u}_{L,t}(\theta) + \frac{1}{2}\hat{u}_{R,t}(\theta)$ are the range and midpoint of $\hat{u}_t(\theta)$. Combining Eqs.(2.11) and (3.4), we obtain

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

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

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

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

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

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

Suppose we choose a kernel K with $(a, b, c) = (\frac{1}{4}, -\frac{1}{4}, \frac{1}{4})$. Then

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

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

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

The estimator $\hat{\theta}^m$ can consistently estimate the level parameter α_0, but it cannot identify the scale parameter β_0, because $\hat{\theta}^m$ 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}:

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

where $\hat{Y}_t^m$ and $\hat{Y}_t^r$ 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 $\hat{\theta}^r$ and $\hat{\theta}^m$ 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 use 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

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

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}:

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

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

Case 3: Constrained Conditional 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 constrained conditional quasi-maximum likelihood method (CCQML) based on the bivariate point-valued sample $\{Y_{L,t}, Y_{R,t}\}_{t=1}^T$. Assuming 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, we obtain the log-Gaussian likelihood function given the bivariate sample $\{Y_{L,t}, Y_{R,t}\}_{t=1}^T$ as follows:

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

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

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

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

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

where $\hat{\Sigma}_{ij}$ is the (i, j)-th component of the variance-covariance estimator $\hat{\Sigma}$. This first looks rather similar to the objective function $\hat{Q}_T(\theta)$ in Eq.(3.5) of the minimum D_K-distance estimator, with the choice of kernel K as $K(1,1) = \hat{\Sigma}_{11}$, $K(1,-1) = K(-1,1) = \hat{\Sigma}_{12} = \hat{\Sigma}_{21}$, $K(-1,-1) = \hat{\Sigma}_{22}$ (this correspondence between a kernel K and a matrix, e.g., $\hat{\Sigma}$, will be simply represented as $K = \hat{\Sigma}$, and our notation will follow this convention throughout this paper). However, we cannot interpret the CCQML estimator as 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 $\text{vech}(\Sigma)$ 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 in subsequent sections.

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 distribution of the interval population. Different choices of kernel K will deliver different minimum D_K-distance estimators for θ^0, and all of them are consistent for θ^0, provided the kernels satisfy 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 kernels that satisfy Eq.(2.6). We first impose a condition on the interval innovation process $\{u_t\}$.

Assumption 6. The interval innovation process $\{u_t\}$ satisfies $\text{var}(u_t|I_{t-1}) = \sigma_K^2 < \infty$, and the derived bivariate point process $\{u_{L,t}, u_{R,t}\}$ satisfies $\text{var}(u_{L,t}, u_{R,t}|I_{t-1}) = \Sigma^0$, where Σ^0 is a finite symmetric positive definite matrix.

This is a conditional homoskedasticity assumption on both $\{u_t\}$ and $\{u_{L,t}, u_{R,t}\}$. The *i.i.d.* condition for $\{u_t\}$ and $\{u_{L,t}, u_{R,t}\}$ 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

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

delivers a minimum D_K-distance estimator

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

which is asymptotically most efficient among all symmetric positive definite kernels K that satisfy Eq.(2.6), with the minimum asymptotic variance

$$\Omega^{opt} = \left\|K^{opt}\right\| \left[E^{-1} \left\langle s_{\frac{\partial u_t}{\partial \theta}}, s'_{\frac{\partial u_t}{\partial \theta}} \right\rangle_{K^{opt}}\right]$$

where $||K^{opt}|| \equiv K^{opt}(1,1)K^{opt}(-1,-1) - K^{opt}(-1,1)K^{opt}(1,-1)$.

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

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

Thus, K^{opt} downweights the sample squared distance components that have larger sampling variations. Specifically, it discounts the sum of squared residuals of the right bound if the right bound disturbance $u_{R,t}$ has a large variance, and discounts the sum of squared residuals of the left bound if 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 $c \neq 0$, the kernel cK^{opt} is also optimal.

The results in Theorem 4.1 do not apply if the conditional homoscedasticity condition in Assumption 6 is violated. We leave derivation of the optimal kernel under conditional heteroscedasticity for future study.

The optimal D_K-distance estimator is not feasible because the optimal kernel K^{opt}, which depends on the DGP, is infeasible. However, we can consider a two-stage minimum D_K-distance estimation method: In Step 1, we obtain a preliminary consistent estimator $\hat{\theta}$ 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}. Then, in Step 2, we obtain a minimum D_K-distance estimator with the choice of $K = \hat{K}^{opt}$:

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

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

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

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 Gaussian 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 $\{u_{L,t}, u_{R,t}\} \sim$i.i.d. $N(\mathbf{0}, \Sigma^0)$. Then as $T \to \infty$, the two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$ achieves the Cramer-Rao lower bound of the constrained maximum likelihood estimator for the bivariate ARMAX(p, q, s) model for the left and right bounds of the interval process $\{Y_t\}$.

Although they are asymptotically efficient, we note that the constrained maximum likelihood estimator 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 $\hat{\theta}^{opt}$.

When the bivariate process $(u_{L,t}, u_{R,t})'$ is not *i.i.d.* Gaussian, the CCQML estimator $\hat{\theta}_{QML}$ based on the Gaussian likelihood is consistent but not optimal for θ^0. It could be shown that the two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$ is asymptotically equivalent to $\hat{\theta}_{QML}$, but only in first order. Their efficiency differs in second order asympototic analysis, as is established in Theorem 4.4 below.

Assumption 7. (i) $\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}|E[\frac{\partial l_t(\varphi^0)}{\partial \theta}\frac{\partial l_{t-j}(\varphi^0)}{\partial h'}\frac{\partial^2 l_{t-l}(\varphi^0)}{\partial h \partial \theta'}]| < \infty$. The notation here indicates that each element in $E[\frac{\partial l_t(\varphi^0)}{\partial \theta}\frac{\partial l_{t-j}(\varphi^0)}{\partial h'}\frac{\partial^2 l_{t-l}(\varphi^0)}{\partial h \partial \theta'}]$ is absolute summable over all j and l. (ii) $\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}|E[\frac{\partial^2 l_t(\varphi^0)}{\partial \theta \partial h'}\frac{\partial l_{t-j}(\varphi^0)}{\partial h}\frac{\partial l_{t-l}(\varphi^0)}{\partial h'}\frac{\partial^2 l_{t-k}(\varphi^0)}{\partial h \partial \theta'}]| < \infty$. The notation indicates that each element in $E[\frac{\partial^2 l_t(\varphi^0)}{\partial \theta \partial h'}\frac{\partial l_{t-j}(\varphi^0)}{\partial h}\frac{\partial l_{t-l}(\varphi^0)}{\partial h'}\frac{\partial^2 l_{t-k}(\varphi^0)}{\partial h \partial \theta'}]$ is absolute summable over all j, k and l.

Theorem 4.4. Suppose Assumptions 1-5 and 7 hold. Then we have

$$\text{avar}(\sqrt{T}\hat{\theta}_{QML}) - \text{avar}(\sqrt{T}\hat{\theta}^{opt}) = T^{-1}\left(-H_{\theta\theta'}^{-1}\right)\Psi\left(-H_{\theta\theta'}^{-1}\right),$$

where

$$\Psi = -\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\left\{E\left[\frac{\partial l_t(\varphi^0)}{\partial \theta}\frac{\partial l_{t-j}(\varphi^0)}{\partial h'}H_{hh}^{-1}\frac{\partial^2 l_{t-l}(\varphi^0)}{\partial h \partial \theta'}\right] + E\left[\frac{\partial^2 l_t(\varphi^0)}{\partial \theta \partial h'}H_{hh}^{-1}\frac{\partial l_{t-j}(\varphi^0)}{\partial h}\frac{\partial l_{t-l}(\varphi^0)}{\partial \theta'}\right]\right\}$$

$H_{\theta\theta'} = E[\frac{\partial^2 l_t(\varphi^0)}{\partial \theta \partial \theta'}]$, $H_{hh} = E[\frac{\partial^2 l_t(\varphi^0)}{\partial h \partial h'}]$, *and* $\varphi^0 = (\theta^0, h^0)$ *with* $h^0 =$*vech*(Σ^0).

Theorem 4.4 suggests that the asymptotic variances of $\sqrt{T}\hat{\theta}_{QML}$ and $\sqrt{T}\hat{\theta}^{opt}$ are different in second order asymptotics, and the difference depends on the third order cumulants of the prespecified log-likelihood function, particularly on the interactions among $\frac{\partial l_t(\varphi^0)}{\partial \theta}$, $\frac{\partial l_t(\varphi^0)}{\partial h}$ and $\frac{\partial l_t^2(\varphi^0)}{\partial \theta \partial h'}$. The interaction terms are generally non-zero when $(u_{L,t}, u_{R,t})'$ is not Gaussian. Thus, we expect that their finite sample performances will differ. Since $\hat{\theta}_{QML}$ involves more parameters to estimate than $\hat{\theta}^{opt}$, it is expected that $\hat{\theta}^{opt}$ will be more efficient in small samples and finite samples, particularly when there exists conditional heteroscedasticity. This is confirmed in our simulation study.

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 $q \times \mathbf{k}$ nonstochastic matrix of full rank, $q \leq \mathbf{k}$, r is a $q \times 1$ nonstochastic vector, and $\mathbf{k}$ 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

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

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

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

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

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

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

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

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

where $\hat{M}_T(\tilde{\theta})$ and $\hat{V}_T(\tilde{\theta})$ are defined in the same way as $\hat{M}_T(\hat{\theta})$ and $\hat{V}_T(\hat{\theta})$ in Theorem 3.3 respectively, with the constrained minimum D_K-distance estimator $\tilde{\theta}$. Then $LM \xrightarrow{L} \chi_q^2$ 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 parameter restriction).

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

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

where $\hat{\theta}$, $\hat{M}_T(\hat{\theta})$ and $\hat{V}_T(\hat{\theta})$ are defined in the same way as $\hat{M}_T(\hat{\theta})$ and $\hat{V}_T(\hat{\theta})$ in Theorem 3.3 respectively. Then, $W \xrightarrow{L} \chi_q^2$ as $T \to \infty$.

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

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

6. Simulation Study

We now investigate the finite sample properties of conditional least squares (CLS), constrained conditional least squares (CCLS), CCQML, minimum D_K-distance (with a prespecified kernel K) 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 empirically relevant 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 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 with $(a, b, c) = (5, 3, 5)$. The minimum and maximum S&P 500 closing price values of day t form the raw interval-valued observations in this period, denoted as $\{P_1, ..., P_T\}$. Then we convert the raw interval price sample data to a weakly stationary interval sample, denoted $\{Y_1, ..., Y_T\}$, by taking the logarithm and Hukuhara difference as $Y_t = \ln(P_t) - \ln(P_{t-1})$. The initial values of Y_t and u_t for $t = 0$ are set to be $\bar{Y}_T$ and $[0, 0]$, respectively. We obtain the minimum D_K-distance parameter 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

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

based on the S&P 500 data. We then generate $\{u_t\}_{t=1}^T$ via the naive bootstrapping from $\{\hat{u}_t\}_{t=1}^T$, 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, CCLS, CCQML, minimum D_K-distance and two-stage minimum D_K-distance methods respectively. Two parameter estimates of CLS are obtained, i.e., $\hat{\theta}^r = (\hat{\beta}_0, \hat{\beta}_1, \hat{\gamma}_1)$ and $\hat{\theta}^m = (\hat{\alpha}_0, \hat{\beta}_1, \hat{\gamma}_1)$, based on range and midpoint data, respectively. We consider 4 kernels with $a = c$, one of which yields the CCLS estimator $\hat{\theta}_{CCLS}$ for the bivariate model of the left and right bounds of Y_t in Eq.(2.20). Another 6 kernels with the form of Case 5 in Section 2.1 are considered. The two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$ is obtained from a kernel K with $(a, b, c) = (10, 8, 16)$ in the first stage.

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

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

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

Tables 1-4 report Bias, SD, and RMSE of CLS, CCLS, CCQML, minimum D_K-distance (denoted as $\hat{\theta}$) and two-stage minimum D_K-distance estimators 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 $\hat{\theta}$ displays robust performance for various kernels. Second, both the interval-based minimum D_K-distance estimators and the bivariate-point based estimators outperform the estimators $\hat{\theta}^r$ and $\hat{\theta}^m$ in terms of RMSE. The two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$ dominates the minimum D_K-distance estimator $\hat{\theta}$ with most kernels, confirming the efficiency result in Theorems 4.1–4.2. The estimator $\hat{\theta}^{opt}$ outperforms $\hat{\theta}_{QML}$ for all parameters in θ^0 in terms of RMSE. Intuitively, CCQML has more unknown parameters to estimate than the two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$, thus $\hat{\theta}^{opt}$ has more desirable performance than $\hat{\theta}_{QML}$ in finite sample.

Lastly, comparing $\hat{\theta}$, $\hat{\theta}^{opt}$ and $\hat{\theta}_{QML}$ with $\hat{\theta}^m$ and $\hat{\theta}^r$, the efficiency gain over the CLS estimators based on either the level or range sample separately is enormous as T becomes large. This is apparently due to the fact that $\hat{\theta}$ and $\hat{\theta}^{opt}$ utilize the level, range and their correlation information in the interval data. On the other hand, while the estimators $\hat{\theta}^r$ and $\hat{\theta}^m$ can consistently estimate model parameters, $\hat{\theta}^m$ is better than $\hat{\theta}^r$. 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 with Conditional Homoscedasticity

This section investigates the finite sample properties of CCLS, CCQML, minimum D_K-distance and two-stage minimum D_K-distance estimators when the DGP of $(Y_{L,t}, Y_{R,t})'$ 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 and $\Sigma^0 = E[(u_{L,t}, u_{R,t})'(u_{L,t}, u_{R,t})]$.

We consider the following bivariate point process as the DGP:

$$\begin{cases} Y_{L,t} = \alpha_0 - \frac{1}{2}\beta_0 + \beta_1 Y_{L,t-1} + \gamma_1 u_{L,t-1} + u_{L,t}, \\ Y_{R,t} = \alpha_0 + \frac{1}{2}\beta_0 + \beta_1 Y_{R,t-1} + \gamma_1 u_{R,t-1} + u_{R,t}, \end{cases} \tag{6.2}$$

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 daily data. Bivariate point innnovation $\{u_{L,t}, u_{R,t}\}_{t=1}^T$ are generated with sample sizes of $T = 100$, 250, and 500 respectively, and three distributions are considered: bivariate Gaussian, bivariate Student-t_5, 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, $\text{corr}(u_{L,t}, u_{R,t}) = 0$ and -0.6 are considered. For each sample size T, we perform 1000 replications. For each replication, we compute CCQML estimator $\hat{\theta}_{QML}$, minimum D_K-distance estimators $\hat{\theta}$ from prespecified kernels and two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$. In particular, the prespecified kernels include the one that yields the CCLS estimator $\hat{\theta}_{CCLS}$ for Eq.(2.20), as well as a kernel that assigns the same weights to the midpoint and range (see K_{ab} in the tables below). $\hat{\theta}^{opt}$ is obtained from the kernel with $(a, b, c) = (10, 8, 16)$ in the first step. We also include the infeasible optimal kernel $K^{opt} = \Sigma^0$ to obtain the infeasible asymptotically most efficient minimum D_K-distance estimator $\hat{\theta}_{\Sigma^0}$; this allows us to study the impact of estimating the unknown K^{opt} in the two-stage minimum D_K-distance estimation.

We report Bias, SD, and RMSE of parameter estimates in Tables 5-1 to 8-1. All estimates converge to their true parameter values respectively in terms of RMSE as T increases. For a bivariate point $i.i.d.$ Gaussian innovation $(u_{L,t}, u_{R,t})'$, the two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$ is as efficient as the constrained maximum likelihood estimator for the bivariate model of the left and right bounds of Y_t, which is consistent with the result in Theorem 4.3. The estimator $\hat{\theta}^{opt}$ also significantly outperforms $\hat{\theta}$ with arbitrary choices of kernel K. It confirms the adaptive capability of our two-stage minimum D_K-distance estimator.

When the bivariate innovation $(u_{L,t}, u_{R,t})'$ follows a Student-t_5 or mixed distribution, $\hat{\theta}^{opt}$ is still the most efficient in the class of minimum D_K-distance estimators, which is consistent with Theorem 4.2. Moreover, $\hat{\theta}^{opt}$ generally outperforms $\hat{\theta}_{QML}$. Note that the efficiency gain of $\hat{\theta}^{opt}$ over the CCQML estimator is more substantial under asymmetric mixture distribution errors in finite samples. We also observe that $\hat{\theta}^{opt}$ outperforms $\hat{\theta}_{CCLS}$ when $\text{corr}(u_{L,t}, u_{R,t}) = -0.6$. This implies that since $\hat{\theta}_{CCLS}$ ignores the (negative) correlation between the left and right bounds, it is not efficient under the bivariate point-valued DGP. Finally, $\hat{\theta}^{opt}$ is almost the same efficient as the infeasible asymptotically efficient estimator $\hat{\theta}_{\Sigma^0}$ as T increases. This indicates that the first stage estimation has negligible impact on the efficiency of the two-stage minimum D_K-distance estimator.

6.3 Bivariate Point-Valued Data Generating Processes with Conditional Heteroscedasticity

To get an idea about the finite sample performances of different estimators under the neglected conditional heteroscedasticity in $(u_{L,t}, u_{R,t})'$, we consider a constant conditional correlation (CCC)-GARCH (1,1) model for $(u_{L,t}, u_{R,t})'$. Following DGP1 in McCloud and Hong (2011), we have $u_{L,t} = \sqrt{h_{L,t}} z_{L,t}$, $u_{R,t} = \sqrt{h_{R,t}} z_{R,t}$, and

$$\begin{cases} h_{L,t} = 0.4 + 0.15u^2_{L,t-1} + 0.8h_{L,t-1}, \\ h_{R,t} = 0.2 + 0.2u^2_{R,t-1} + 0.7h_{R,t-1}, \\ (z_{L,t}, z_{R,t})'|I_{t-1} \overset{i.i.d.}{\sim} N\left[\mathbf{0}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right] \end{cases} \tag{6.3}$$

where $\rho = 0$, and -0.6 respectively. We then generate the bivariate innovation $\{u_{L,t}, u_{R,t}\}_{t=1}^T$ from (6.3) with $T = 100$, 250, and 500 respectively. $\{Y_t\}_{t=1}^T$ is then generated from (6.2), where the true parameter values $\theta^0 = (\alpha_0, \beta_0, \beta_1, \gamma_1)'$ are obtained in the same way as previous experiments. For each sample size T, we perform 1000 replications. For each replication, we compute CCQML estimator $\hat{\theta}_{QML}$, minimum D_K-distance estimators $\hat{\theta}$ from prespecified kernels, and two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$. The prespecified kernels include both cases with $b > 0$ and $b < 0$. $\hat{\theta}^{opt}$ is obtained from a kernel with $(a, b, c) = (10, 8, 16)$ in the first step.

Several conclusions can be drawn from the results of parameter estimates reported in Tables 5-2 to 8-2. First, all minimum D_K-distance and CCQML estimators converge in terms of RMSE as T increases, under neglected conditional heteroskedasticity of $(u_{L,t}, u_{R,t})'$, although the bias and the variance of most estimates are larger than under conditional homoscedasticity. Second, $\hat{\theta}^{opt}$

clearly outperforms $\hat{\theta}_{QML}$ in finite samples. And compared to the results in Section 6.2, $\hat{\theta}^{opt}$ yields a larger gain over $\hat{\theta}_{QML}$ when there exists serial dependence in higher moments of $(u_{L,t}, u_{R,t})'$. In fact, the class of minimum D_K-distance estimators with arbitrary kernels with $b < 0$ also outperform $\hat{\theta}_{QML}$.

In addition to (6.3), we also examined DGP6 in McCloud and Hong (2011), i.e., DDC-GARCH (1,1) model, as our DGP for $(u_{L,t}, u_{R,t})'$. Due to the similar patterns of simulation results emerging from the DCC-GARCH(1,1) parameterization in terms of ranking different estimators, the experiment details are not reported here, yet are available from the authors on request.

Overall, the simulation results in Tables 1-8 generally reveal the desirable properties of the two-stage minimum D_K-distance estimator relative to many others.

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) find 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 an 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_t is the the difference between the return on the small portfolio and the return on the large portfolio, HML_t is the difference between the return on the high book-to-market portfolio and the return on the low book-to-market portfolio, and $TERM_t$ and DEF_t are defined as above.

Given the definition of variables in the Fama and French's (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 $i = 1, \ldots, 25$, $E(u_t|I_{t-1}) = [0, 0]$, $Y_t = [R_{ft}, R_t]$, $X_{1t} = [R_{ft}, R_{mt}]$,

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

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 Fama and French's (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 follows. We first estimate Fama and French's (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) are 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 Fama and French's (1993) model, and obtain estimate the interval version of Fama and French's (1993) model using the two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$. 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, the minimum D_K-distance estimators, the two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$, and the CCQML estimator $\hat{\theta}_{QML}$. The estimates for α_0 in (7.2) are not reported here, since Fama and French's (1993) model does not include this level parameter.

Table 9 shows some interesting findings. First, the minimum D_K-distance estimators, $\hat{\theta}_{QML}$ and $\hat{\theta}^{opt}$ for most of the 25 stock portfolios reveal strong evidence that the default risk factor

DEF_t is significant in capturing the variation of excess stock returns, compared to the critical value of 1.96 at the 5% significance level. Generally, $\hat{\theta}^{opt}$ yields larger t-statistics than $\hat{\theta}_{QML}$, and both of them have large t-statistics than OLS. On the other hand, there is not an overwhelming pattern for the effect of $TERM_t$ 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 result, the use of interval bond factor X_{4t} contains about the same information as the differenced $TERM_t$ 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' or 'difference' information.

8. Conclusion

Interval-valued data are not uncommon in economics. 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 or difference modelling. Interval forecasts are also often of direct interest in many applications in economics.

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, we propose a new 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, asymptotic 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. Finally, interval modelling can also be considered in cross-sectional econometrics. 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, Cambridge, MA.

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

Andrews, D. W. K. and Shi, X. (2009), "Inference Based on Conditional Moment Inequalities", mimeo.

Andrews, D. W. K. and Soares, G. (2010), "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection", Econometrica, 78, 119-157.

Arroyo, J., Espínola, R. and Maté, C. (2011), "Different Approaches to Forecast Interval Time Series: A Comparison in Finance", *Computational Economics*, 37 (2), 169-191.

Arroyo, J., González-Rivera, G. and Maté, C. (2011), "Forecasting with Interval and Histogram Data. Some Financial Applications", in *Handbook of Empirical Economics and Finance*, A. Ullah and D. Giles (eds.), Chapman and Hall, 247-280.

Artstein, Z. and Vitale, R. A. (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.

Beckers, S. (1983), "Variance of Security Price Return Based on High, Low and Closing Prices", *Journal of Business*, 56, 97-112.

Beresteanu, A. and Molinari, F. (2008), "Asymptotic Properties for a Class of Partially Identified Models", *Econometrica*, 76, 763–814.

Beresteanu, A., Molchanov, I. and Molinari, F. (2011), "Sharp Identification Regions in Models with Convex Moment Predictions", *Econometrica,* 79, 1785-1821.

Beresteanu, A., Molchanov, I. and Molinari, F. (2012), "Partial Identification Using Random Set Theory", *Journal of Econometrics,* 166, 17–32.

Bertoluzza, C., Corral, N. and Salas, A. (1995), "On a New Class of Distances between Fuzzy Numbers", *Mathware Soft Comput*, 2, 71–84.

Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.

Blanco-Fernández, A., Corral, N. and González-Rodríguez, G. (2011), "Estimation of a Flexible Simple Linear Model for Interval Data Based on Set Arithmetic", *Computational Statistics and Data Analysis*, 55, 2568-2578.

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, 5-59.

Bollerslev, T., Engle, R. and Nelson, B. (1994), "ARCH Models", Chapter 49, *Handbook of Econometrics*, North-Holland, Amsterdam.

Bontemps, C., Magnac, T. and Maurin, E. (2012), "Set Identified Linear Models", *Econometrica*, 80, 1129–1155.

Chandrasekhar, A., Chernozhukov, V., Molinari, F. and Schrimpf, P. (2012), "Inference for Best Linear Approximations to Set Identified Functions", CeMMAP Working Paper CWP 43/12.

Chernozhukov, V., Hong, H. and Tamer, E. (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models", *Econometrica*, 75, 1243-1284.

Chernozhukov, V., Rigobon, R. and Stoker, T. M. (2010), "Set Identification and Sensitivity Analysis with Tobin Regressors", *Quantitative Economics*, 1, 255–277.

Choi, D. H. and Smith, J. D. H. (2003), "Support Functions of General Convex Sets", *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.

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

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.

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.

Gil, M., González-Rodríguez, G., Colubi, A. and Montenegro, M. (2007), "Testing Linear Independence in Linear Models with Interval-Valued Data", *Computational Statistics and Data Analysis,* 51, 3002-3015.

Goldberger, A. (1964), *Econometric Theory*, John Wiley and Sons, New York.

González-Rodríguez, G., Blanco-Fernández, A., Corral, N. and Colubi, A. (2007), "Least Squares Estimation of Linear Regression Models for Convex Compact Random Sets", *Advance in Data Analysis and Classification* 1 (1), 6-81.

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

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

Körner, R. and 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*, 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.

Lin, W. and González-Rivera, G. (2013), "Constrained Regression for Interval-Valued Data", forthcoming in *Journal of Business & Economic Statistics*.

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*, 64 (3), 575-596.

Maia, A., Carvalho, F. and Ludermir T. (2008), "Forecasting Models for Interval-Valued Time Series", *Neurocomputing*, 71, 3344-3352.

Manski, C. (1995), *Identification Problems in the Social Sciences*, Harvard University Press, Cambridge, MA.

Manski, C. (2003), *Partial Identification of Probability Distributions*, Springer-Verlag, Heidelberg.

Manski, C. (2007), *Identification for Prediction and Decision*, Harvard University Press, Cambridge, MA.

Manski, C. (2013), *Public Policy in an Uncertain World: Analysis and Decisions*, Harvard University Press, Cambridge, MA.

Manski, C. and Tamer, E. (2002), "Inference on Regressions with Interval Data on a Regressor or Outcome", *Econometrica*, 70, 519-546.

McCloud, N. and Hong, Y. (2011), "Testing the Structure of Conditional Correlations in Multivariate Garch Models: A Generalized Cross-Spectrum Approach," *International Economic Review*, 52 (4), 991-1037.

Minkowsky, H. (1911), *Gesammelte Abhandlungen 2*, Band Teubner, Leipzig.

Molchanov, I. (1993), "Limit Theorems for Convex Hulls of Random Sets", *Advances in Applied Probability,* 25, 395-414.

Molchanov, I. (2005), *Theory of Random Sets*, Springer, London.

Moore, R. E., Kearfott, R. B., and Cloud, M. J. (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.

Neto, E. and Carvalho, F. (2010), "Constrained Linear Regression Models for Symbolic Interval-Valued Variables", *Computational Statistics and Data Analysis*, 54, 333-347.

Neto, E., Carvalho, F. and Freire, E. (2008), "Centre and Range Method for Fitting a Linear Regression Model to Symbolic Interval Data", *Computational Statistics and Data Analysis*, 52, 1500-1515.

Pagan, A. R. and Sabau, H. (1987), "On the Inconsistency of the MLE in Certain Heteroskedastic Regression Models," mimeo, University of Rochester Department of Economics.

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 Ralescu, D. A. (1983), "Strong Law of Large Numbers for Banach Space Valued Random Sets", *Annals of Probability*, 11, 222-224.

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

Rockafellar, R. (1970), *Convex Analysis*, Princeton University Press, New Jersey.

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

Russell, J. and Engle, R. (2009), "Analysis of High-Frequency Data", Chapter 7, *Handbook of Financial Econometrics, Volume 1: Tools and Techniques,* North-Holland, Amsterdam.

Sargent, T. J. (1987), *Macroeconomic Theory, 2nd ed*, Academic Press, Boston.

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

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

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

White, H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge University Press, Cambridge.

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)

$\hat{\alpha}_0(10^{-4})$												
	$T = 100$			$T = 250$			$T = 500$			$T = 1000$		
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
$CCQML$	−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: (1) ACI (1,1) Model: $Y_t = \alpha_0 + \beta_0 I_0 + \beta_1 Y_{t-1} + \gamma_1 u_{t-1} + u_t$.

(2) 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, CCQML,CCLS, and K^{opt} denote the estimates of $\hat{\theta}^m$, $\hat{\theta}^r$, $\hat{\theta}_{QML}$, $\hat{\theta}_{CCLS}$ and $\hat{\theta}^{opt}$ with special kernels, respectively.

(3) 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)

$\hat{\beta}_0(10^{-4})$												
	$T=100$			$T=250$			$T=500$			$T=1000$		
a/b/c	Bias	S.D	RMSE	Bias	S.D	RMSE	Bias	S.D	RMSE	Bias	S.D	RMSE
K_r	−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
$CCQML$	−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: (1) ACI (1,1) Model: $Y_t = \alpha_0 + \beta_0 I_0 + \beta_1 Y_{t-1} + \gamma_1 u_{t-1} + u_t$.

(2) 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, CCQML, CCLS, and K^{opt} denote the estimates of $\hat{\theta}^m$, $\hat{\theta}^r$, $\hat{\theta}_{QML}$, $\hat{\theta}_{CCLS}$ and $\hat{\theta}^{opt}$ with special kernels, respectively.

(3) 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)

$\hat{\beta}_1(10^{-2})$												
	$T=100$			$T=250$			$T=500$			$T=1000$		
$a/b/c$	$Bias$	$S.D$	$RMSE$	$Bias$	$S.D$	$RMSE$	$Bias$	$S.D$	$RMSE$	$Bias$	$S.D$	$RMSE$
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
$CCQML$	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: (1) ACI (1,1) Model: $Y_t = \alpha_0 + \beta_0 I_0 + \beta_1 Y_{t-1} + \gamma_1 u_{t-1} + u_t$.

(2) 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, CCQML,CCLS, and K^{opt} denote the estimates of $\hat{\theta}^m$, $\hat{\theta}^r$, $\hat{\theta}_{QML}$, $\hat{\theta}_{CCLS}$ and $\hat{\theta}^{opt}$ with special kernels, respectively.

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

(4) Bias is in -1.

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

$\hat{\gamma}_1(10^{-2})$												
	$T = 100$			$T = 250$			$T = 500$			$T = 1000$		
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
$CCQML$	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: (1) ACI (1,1) Model: $Y_t = \alpha_0 + \beta_0 I_0 + \beta_1 Y_{t-1} + \gamma_1 u_{t-1} + u_t$.

(2) 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, CCQML, CCLS and K^{opt} denote the estimates of $\hat{\theta}^m$, $\hat{\theta}^r$, $\hat{\theta}_{QML}$, $\hat{\theta}_{CCLS}$ and $\hat{\theta}^{opt}$ with special kernels, respectively.

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

TABLE 5-1. Bias, SD and RMSE of Estimates for α_0 in Bivariate Point Processes

$\hat{\alpha}_0$									
Gaussian	$T = 100$			$T = 250$			$T = 500$		
$\rho = 0$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
CML	0.8298	10.4484	10.4813	0.0875	4.5728	4.5736	0.1289	3.1884	3.1910
$CCLS/K_{\Sigma^0}$	0.8910	10.3927	10.4308	0.0874	4.5699	4.5708	0.1279	3.1871	3.1896
K_{ab}	1.6795	12.5311	12.6432	0.1126	4.6332	4.6345	0.1440	3.2005	3.2037
K_{abc}	1.9464	12.6272	12.7763	0.1127	4.6201	4.6214	0.1459	3.2097	3.2130
K^{opt}	0.9150	10.5101	10.5499	0.0888	4.5743	4.5751	0.1288	3.1883	3.1909
$\rho = -0.6$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
K_{Σ^0}	0.4817	4.9127	4.9363	0.2110	2.7353	2.7435	0.1624	1.8384	1.8456
CML	0.5179	4.9574	4.9844	0.2253	2.7462	2.7554	0.1721	1.8447	1.8527
$CCLS$	0.5455	5.1733	5.2020	0.2301	2.7679	2.7774	0.1811	1.8765	1.8852
K_{ab}	0.6453	5.4616	5.4996	0.2667	2.8377	2.8502	0.2037	1.9197	1.9305
K_{abc}	0.6742	5.5054	5.5465	0.2603	2.8170	2.8290	0.2093	1.9203	1.9316
K^{opt}	0.5043	4.9415	4.9672	0.2132	2.7366	2.7449	0.1640	1.8409	1.8482
Student-t_5	$T = 100$			$T = 250$			$T = 500$		
$\rho = 0$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$CCQML$	−0.0145	10.6060	10.6060	0.1766	6.5325	6.5349	0.1132	4.5059	4.5073
$CCLS/K_{\Sigma^0}$	−0.0037	10.5816	10.5816	0.1703	6.5337	6.5359	0.1089	4.5096	4.5110
K_{ab}	0.0553	10.6478	10.6479	0.1584	6.6086	6.6105	0.1037	4.4946	4.4958
K_{abc}	0.0737	10.6761	10.6763	0.1770	6.6169	6.6193	0.1172	4.5284	4.5299
K^{opt}	0.0086	10.5553	10.5554	0.1692	6.5298	6.5320	0.1067	4.5028	4.5040
$\rho = -0.6$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
K_{Σ^0}	0.4362	3.7112	3.7367	0.2854	2.1600	2.1787	0.1225	1.4629	1.4680
$CCQML$	0.6472	4.1205	4.1710	0.3542	2.2327	2.2606	0.1468	1.4707	1.4780
$CCLS$	0.6387	4.2634	4.3110	0.3469	2.3343	2.3600	0.1378	1.5118	1.5181
K_{ab}	0.9005	5.1942	5.2717	0.4344	2.5368	2.5737	0.1688	1.5575	1.5667
K_{abc}	0.9104	5.2289	5.3076	0.4208	2.5524	2.5869	0.1822	1.5898	1.6002
K^{opt}	0.5208	3.9127	3.9472	0.2948	2.1671	2.1871	0.1237	1.4593	1.4646
Mixture	$T = 100$			$T = 250$			$T = 500$		
$\rho = 0$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$CCQML$	0.0658	8.4540	8.4542	−0.0773	4.9371	4.9377	−0.1493	3.4264	3.4296
$CCLS/K_{\Sigma^0}$	0.0786	8.4136	8.4139	−0.0690	4.9420	4.9425	−0.1471	3.4214	3.4245
K_{ab}	0.0020	8.3888	8.3888	−0.0896	4.9567	4.9575	−0.1429	3.4239	3.4268
K_{abc}	0.0162	8.4071	8.4071	−0.0990	4.9636	4.9646	−0.1396	3.4177	3.4205
K^{opt}	0.0795	8.4058	8.4061	−0.0753	4.9324	4.9330	−0.1487	3.4213	3.4245
$\rho = -0.6$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
K_{Σ^0}	−0.0745	8.4408	8.4412	−0.0075	4.9528	4.9529	−0.1205	3.2849	3.2871
$CCQML$	−0.1694	8.7365	8.7381	−0.0158	4.9525	4.9525	−0.1258	3.2913	3.2937
$CCLS$	−0.0760	8.4720	8.4724	−0.0289	4.9360	4.9361	−0.1201	3.2890	3.2912
K_{ab}	−0.0592	8.6159	8.6161	−0.0395	4.9151	4.9153	−0.1169	3.2980	3.3001
K_{abc}	−0.0447	8.5978	8.5979	−0.0262	4.9635	4.9636	−0.1388	3.3032	3.3062
K^{opt}	−0.0922	8.4661	8.4666	−0.0102	4.9456	4.9456	−0.1203	3.2841	3.2863

Notes: (1) The first column with CML, CCQML, CCLS, K_{Σ^0} and K^{opt} denote the estimates of constrained maximum likelihood, $\hat{\theta}_{QML}$, $\hat{\theta}_{CCLS}$, $\hat{\theta}_{\Sigma^0}$ and $\hat{\theta}^{opt}$ respectively. K_{ab} and K_{abc} are with $(a, b, c) = (10, 6, 10)$ and $(a, b, c) = (10, 8, 19)$ respectively.

(2) Bivariate Gaussian, Student-t_5 and Mixture densities for $u_{L,t}$ and $u_{R,t}$ with $\rho = 0$ and $\rho = -0.6$ are considered respectively, where $\rho = \text{corr}(u_{L,t}, u_{R,t})$. $\hat{\theta}_{CCLS}$ coincides with $\hat{\theta}_{\Sigma^0}$ as $\rho = 0$.

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

TABLE 6-1. Bias, SD and RMSE of Estimates for β_0 in Bivariate Point Processes

$\hat{\beta}_0$									
Gaussian	$T=100$			$T=250$			$T=500$		
$\rho=0$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
CML	3.7740	27.4147	27.6733	0.5808	9.2854	9.3036	0.4462	6.3235	6.3393
$CCLS/K_{\Sigma^0}$	3.7685	26.8171	27.0805	0.5789	9.2782	9.2962	0.4447	6.3213	6.3370
K_{ab}	2.4366	26.6792	26.7903	0.5898	9.3354	9.3540	0.4737	6.3971	6.4146
K_{abc}	3.1966	27.6164	27.8008	0.5803	9.3357	9.3538	0.4819	6.4178	6.4359
K^{opt}	3.8291	27.4404	27.7062	0.5831	9.2838	9.3021	0.4456	6.3234	6.3390
$\rho=-0.6$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
K_{Σ^0}	−1.2941	18.3142	18.3599	0.0843	10.3981	10.3984	0.1530	7.4041	7.4057
CML	−1.3490	18.4992	18.5484	0.0659	10.4546	10.4548	0.1469	7.4240	7.4255
$CCLS$	−1.4203	18.9489	19.0020	0.0161	10.5036	10.5036	0.1333	7.4698	7.4710
K_{ab}	−1.5256	19.7844	19.8431	−0.0336	10.6743	10.6743	0.1183	7.5408	7.5418
K_{abc}	−1.5610	19.8056	19.8670	−0.0274	10.6681	10.6681	0.1143	7.5481	7.5490
K^{opt}	−1.3223	18.4163	18.4637	0.0749	10.4089	10.4092	0.1525	7.4032	7.4047
Student-t_5	$T=100$			$T=250$			$T=500$		
$\rho=0$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$CCQML$	0.3259	21.6623	21.6647	0.9509	13.0973	13.1318	0.3293	9.4160	9.4217
$CCLS/K_{\Sigma^0}$	0.1868	21.4437	21.4445	0.9124	13.0114	13.0433	0.3112	9.3943	9.3995
K_{ab}	0.0646	21.8261	21.8262	1.0303	13.2285	13.2685	0.3782	9.6245	9.6320
K_{abc}	0.0573	21.9014	21.9015	1.0184	13.2651	13.3041	0.3956	9.6312	9.6393
K^{opt}	0.1828	21.4460	21.4468	0.9210	13.0578	13.0902	0.3167	9.4036	9.4090
$\rho=-0.6$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
K_{Σ^0}	−0.2537	14.6822	14.6844	0.2862	7.7471	7.7524	0.0418	5.3364	5.3366
$CCQML$	−0.1864	15.5568	15.5579	0.3711	8.0034	8.0120	0.0776	5.3833	5.3839
$CCLS$	−0.1612	15.8848	15.8856	0.3681	7.9860	7.9945	0.0749	5.4009	5.4014
K_{ab}	−0.2414	17.2679	17.2695	0.4290	8.3241	8.3352	0.1053	5.4979	5.4989
K_{abc}	−0.0431	17.8152	17.8152	0.4169	8.3125	8.3230	0.0858	5.4759	5.4766
K^{opt}	−0.2248	14.9109	14.9126	0.2838	7.8137	7.8188	0.0419	5.3498	5.3500
Mixture	$T=100$			$T=250$			$T=500$		
$\rho=0$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$CCQML$	0.7685	17.0311	17.0484	0.8283	9.9686	10.0029	0.6215	6.9869	7.0144
$CCLS/K_{\Sigma^0}$	0.7274	16.8214	16.8372	0.8206	9.9568	9.9905	0.6232	6.9798	7.0076
K_{ab}	0.6616	17.2002	17.2129	0.8763	10.1292	10.1671	0.6565	7.1338	7.1639
K_{abc}	0.6536	17.1318	17.1443	0.8623	10.1256	10.1622	0.6373	7.1200	7.1485
K^{opt}	0.6925	16.8890	16.9032	0.8072	9.9304	9.9632	0.6110	6.9705	6.9972
$\rho=-0.6$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
K_{Σ^0}	0.7688	34.1301	34.1387	−0.1418	20.2516	20.2521	0.2496	13.7079	13.7102
$CCQML$	1.0859	36.2240	36.2403	−0.1072	20.4502	20.4505	0.4022	14.1563	14.1620
$CCLS$	0.7947	34.4177	34.4268	−0.1309	20.5699	20.5703	0.3367	13.9342	13.9383
K_{ab}	0.8134	35.0657	35.0751	−0.1126	20.8044	20.8047	0.3784	14.0688	14.0739
K_{abc}	0.8824	34.9502	34.9613	−0.1065	20.8457	20.8460	0.3369	14.0391	14.0432
K^{opt}	0.7295	34.0558	34.0636	−0.1729	20.3123	20.3130	0.2664	13.7573	13.7599

Notes: (1) The first column with CML, CCQML, CCLS, K_{Σ^0} and K^{opt} denote the estimates of constrained maximum likelihood, $\hat{\theta}_{QML}$, $\hat{\theta}_{CCLS}$, $\hat{\theta}_{\Sigma^0}$ and $\hat{\theta}^{opt}$ respectively. K_{ab} and K_{abc} are with $(a,b,c)=(10,6,10)$ and $(a,b,c)=(10,8,19)$ respectively.

(2) Bivariate Gaussian, Student-t_5 and Mixture densities for $u_{L,t}$ and $u_{R,t}$ with $\rho=0$ and $\rho=-0.6$ are considered respectively, where $\rho=\text{corr}(u_{L,t},u_{R,t})$. $\hat{\theta}_{CCLS}$ coincides with $\hat{\theta}_{\Sigma^0}$ as $\rho=0$.

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

TABLE 7-1. Bias, SD and RMSE of Estimates for β_1 in Bivariate Point Processes

$\hat{\beta}_1$									
Gaussian	$T = 100$			$T = 250$			$T = 500$		
$\rho = 0$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
CML	3.7740	27.4147	27.6733	0.5808	9.2854	9.3036	0.4462	6.3235	6.3393
$CCLS/K_{\Sigma^0}$	3.7685	26.8171	27.0805	0.5789	9.2782	9.2962	0.4447	6.3213	6.3370
K_{ab}	2.4366	26.6792	26.7903	0.5898	9.3354	9.3540	0.4737	6.3971	6.4146
K_{abc}	3.1966	27.6164	27.8008	0.5803	9.3357	9.3538	0.4819	6.4178	6.4359
K^{opt}	3.8291	27.4404	27.7062	0.5831	9.2838	9.3021	0.4456	6.3234	6.3390
$\rho = -0.6$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
K_{Σ^0}	−1.2941	18.3142	18.3599	0.0843	10.3981	10.3984	0.1530	7.4041	7.4057
CML	−1.3490	18.4992	18.5484	0.0659	10.4546	10.4548	0.1469	7.4240	7.4255
$CCLS$	−1.4203	18.9489	19.0020	0.0161	10.5036	10.5036	0.1333	7.4698	7.4710
K_{ab}	−1.5256	19.7844	19.8431	−0.0336	10.6743	10.6743	0.1183	7.5408	7.5418
K_{abc}	−1.5610	19.8056	19.8670	−0.0274	10.6681	10.6681	0.1143	7.5481	7.5490
K^{opt}	−1.3223	18.4163	18.4637	0.0749	10.4089	10.4092	0.1525	7.4032	7.4047
Student-t_5	$T = 100$			$T = 250$			$T = 500$		
$\rho = 0$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$CCQML$	0.3259	21.6623	21.6647	0.9509	13.0973	13.1318	0.3293	9.4160	9.4217
$CCLS/K_{\Sigma^0}$	0.1868	21.4437	21.4445	0.9124	13.0114	13.0433	0.3112	9.3943	9.3995
K_{ab}	0.0646	21.8261	21.8262	1.0303	13.2285	13.2685	0.3782	9.6245	9.6320
K_{abc}	0.0573	21.9014	21.9015	1.0184	13.2651	13.3041	0.3956	9.6312	9.6393
K^{opt}	0.1828	21.4460	21.4468	0.9210	13.0578	13.0902	0.3167	9.4036	9.4090
$\rho = -0.6$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
K_{Σ^0}	−0.2537	14.6822	14.6844	0.2862	7.7471	7.7524	0.0418	5.3364	5.3366
$CCQML$	−0.1864	15.5568	15.5579	0.3711	8.0034	8.0120	0.0776	5.3833	5.3839
$CCLS$	−0.1612	15.8848	15.8856	0.3681	7.9860	7.9945	0.0749	5.4009	5.4014
K_{ab}	−0.2414	17.2679	17.2695	0.4290	8.3241	8.3352	0.1053	5.4979	5.4989
K_{abc}	−0.0431	17.8152	17.8152	0.4169	8.3125	8.3230	0.0858	5.4759	5.4766
K^{opt}	−0.2248	14.9109	14.9126	0.2838	7.8137	7.8188	0.0419	5.3498	5.3500
Mixture	$T = 100$			$T = 250$			$T = 500$		
$\rho = 0$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$CCQML$	0.7685	17.0311	17.0484	0.8283	9.9686	10.0029	0.6215	6.9869	7.0144
$CCLS/K_{\Sigma^0}$	0.7274	16.8214	16.8372	0.8206	9.9568	9.9905	0.6232	6.9798	7.0076
K_{ab}	0.6616	17.2002	17.2129	0.8763	10.1292	10.1671	0.6565	7.1338	7.1639
K_{abc}	0.6536	17.1318	17.1443	0.8623	10.1256	10.1622	0.6373	7.1200	7.1485
K^{opt}	0.6925	16.8890	16.9032	0.8072	9.9304	9.9632	0.6110	6.9705	6.9972
$\rho = -0.6$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
K_{Σ^0}	0.7688	34.1301	34.1387	−0.1418	20.2516	20.2521	0.2496	13.7079	13.7102
$CCQML$	1.0859	36.2240	36.2403	−0.1072	20.4502	20.4505	0.4022	14.1563	14.1620
$CCLS$	0.7947	34.4177	34.4268	−0.1309	20.5699	20.5703	0.3367	13.9342	13.9383
K_{ab}	0.8134	35.0657	35.0751	−0.1126	20.8044	20.8047	0.3784	14.0688	14.0739
K_{abc}	0.8824	34.9502	34.9613	−0.1065	20.8457	20.8460	0.3369	14.0391	14.0432
K^{opt}	0.7295	34.0558	34.0636	−0.1729	20.3123	20.3130	0.2664	13.7573	13.7599

Notes: (1) The first column with CML, CCQML, CCLS, K_{Σ^0} and K^{opt} denote the estimates of constrained maximum likelihood, $\hat{\theta}_{QML}$, $\hat{\theta}_{CCLS}$, $\hat{\theta}_{\Sigma^0}$ and $\hat{\theta}^{opt}$ respectively. K_{ab} and K_{abc} are with $(a, b, c) = (10, 6, 10)$ and $(a, b, c) = (10, 8, 19)$ respectively.

(2) Bivariate Gaussian, Student-t_5 and Mixture densities for $u_{L,t}$ and $u_{R,t}$ with $\rho = 0$ and $\rho = -0.6$ are considered respectively, where $\rho = \text{corr}(u_{L,t}, u_{R,t})$. $\hat{\theta}_{CCLS}$ coincides with $\hat{\theta}_{\Sigma^0}$ as $\rho = 0$.

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

TABLE 8-1. Bias, SD and RMSE of Estimates γ_1 in Bivariate Point Processes

$\hat{\gamma}_1$									
Gaussian	$T = 100$			$T = 250$			$T = 500$		
$\rho = 0$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
CML	8.4453	9.4949	12.7073	0.8998	5.5124	5.5854	0.3648	3.9454	3.9622
$CCLS/K_{\Sigma^0}$	8.2679	8.8920	12.1419	0.8974	5.4600	5.5332	0.3725	3.9283	3.9460
K_{ab}	11.4850	7.7109	13.8334	0.9066	6.3690	6.4332	0.3545	4.5923	4.6059
K_{abc}	11.0291	7.0202	13.0738	0.9256	6.3887	6.4554	0.3883	4.6795	4.6955
K^{opt}	8.7537	9.0957	12.6237	0.9015	5.5045	5.5778	0.3650	3.9456	3.9625
$\rho = -0.6$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
K_{Σ^0}	1.9869	9.7370	9.9376	0.9657	5.7332	5.8140	0.3956	4.1090	4.1280
CML	2.2679	9.8667	10.1240	1.0748	5.7914	5.8903	0.4519	4.1387	4.1633
$CCLS$	2.3898	11.3345	11.5836	0.9910	6.5930	6.6671	0.4234	4.8319	4.8504
K_{ab}	2.8237	13.0718	13.3733	1.0923	7.5764	7.6548	0.4656	5.5542	5.5737
K_{abc}	2.9483	13.0215	13.3511	1.1055	7.5309	7.6117	0.4966	5.5726	5.5947
K^{opt}	2.0775	9.8796	10.0956	0.9675	5.7819	5.8623	0.3867	4.1259	4.1440
Student-t_5	$T = 100$			$T = 250$			$T = 500$		
$\rho = 0$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$CCQML$	3.1779	16.5411	16.8436	0.9417	10.0557	10.0997	0.2883	7.3200	7.3257
$CCLS/K_{\Sigma^0}$	2.9034	15.5630	15.8315	0.8499	9.7332	9.7702	0.2661	7.1633	7.1683
K_{ab}	2.7815	18.1555	18.3673	1.1267	11.0026	11.0602	0.2518	8.2875	8.2913
K_{abc}	2.5072	18.4031	18.5731	1.1503	11.0933	11.1527	0.2912	8.3810	8.3860
K^{opt}	2.9203	15.5951	15.8661	0.8445	9.7674	9.8039	0.2447	7.1731	7.1773
$\rho = -0.6$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
K_{Σ^0}	3.1537	15.9424	16.2514	0.8519	10.0569	10.0929	0.2965	7.3499	7.3559
$CCQML$	3.3597	17.6944	18.0106	1.0495	10.6465	10.6981	0.4882	8.2301	8.2445
$CCLS$	3.0205	21.7619	21.9705	1.2138	13.4193	13.4741	0.4726	9.9201	9.9314
K_{ab}	3.1531	22.9097	23.1257	1.2456	13.9726	14.0281	0.4933	10.3169	10.3287
K_{abc}	3.1721	22.6396	22.8608	1.2028	13.9364	13.9882	0.5022	10.3075	10.3197
K^{opt}	3.1857	16.0071	16.3210	0.8690	10.0662	10.1037	0.2759	7.3581	7.3633
Mixture	$T = 100$			$T = 250$			$T = 500$		
$\rho = 0$	$Bias$	SD	$RMSE$	$Bias$	SD	$RMSE$	$Bias$	SD	$RMSE$
$CCQML$	4.2318	17.2183	17.7307	1.5269	10.5734	10.6831	0.8042	7.3589	7.4027
$CCLS/K_{\Sigma^0}$	3.7932	16.2449	16.6819	1.4703	10.1369	10.2430	0.7806	7.1689	7.2112
K_{ab}	3.6155	19.0562	19.3961	1.6763	11.7443	11.8633	0.8165	8.5842	8.6230
K_{abc}	3.3665	19.3084	19.5997	1.7073	11.9231	12.0447	0.7111	8.5085	8.5382
K^{opt}	3.7985	16.2124	16.6515	1.4231	10.1247	10.2242	0.7491	7.1541	7.1932
$\rho = -0.6$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
K_{Σ^0}	4.2468	16.6675	17.2000	1.6779	9.7319	9.8755	0.8281	7.4169	7.4629
$CCQML$	5.5143	19.5446	20.3076	1.8861	10.7381	10.9025	1.1146	8.7967	8.8670
$CCLS$	4.6667	18.3164	18.9016	1.7353	11.4024	11.5337	0.9506	8.6531	8.7051
K_{ab}	4.9730	21.0036	21.5843	1.8271	13.1429	13.2693	1.0335	9.8146	9.8689
K_{abc}	4.6610	20.9544	21.4665	1.8625	12.9952	13.1280	0.9434	9.8626	9.9076
K^{opt}	4.1638	16.6509	17.1637	1.6462	9.7713	9.9090	0.8318	7.4221	7.4686

Notes. (1) The first column with CML, CCQML,CCLS, K_{Σ^0} and K^{opt} denote the estimates of constrained maximum likelihood, $\hat{\theta}_{QML}$, $\hat{\theta}_{CCLS}$, $\hat{\theta}_{\Sigma^0}$ and $\hat{\theta}^{opt}$ respectively. K_{ab} and K_{abc} are with $(a, b, c) = (10, 6, 10)$ and $(a, b, c) = (10, 8, 19)$ respectively.

(2) Bivariate Gaussian, Student-t_5 and Mixture densities for $u_{L,t}$ and $u_{R,t}$ with $\rho = 0$ and $\rho = -0.6$ are considered respectively, where $\rho = \text{corr}(u_{L,t}, u_{R,t})$. $\hat{\theta}_{CCLS}$ coincides with $\hat{\theta}_{\Sigma^0}$ as $\rho = 0$.

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

TABLE 5-2. Bias, SD and RMSE of Estimates for α_0 in CCC-GARCH (1,1)

$\hat{\alpha}_0$									
	$T=100$			$T=250$			$T=500$		
$\rho=0$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$\hat{\theta}_\rho$	0.3262	18.3641	18.3670	0.1281	11.0255	11.0262	0.1701	8.0849	8.0867
$CCQML$	0.3238	18.2934	18.2963	0.0844	11.0194	11.0197	0.1697	8.1314	8.1332
$CCLS$	0.2741	18.3073	18.3094	0.1177	10.9255	10.9261	0.1694	8.0684	8.0702
K_{ab}	0.2579	18.6976	18.6994	0.1007	10.9248	10.9253	0.1713	8.0712	8.0730
K_{abc1}	0.3400	18.7300	18.7331	0.1212	10.9768	10.9774	0.1820	8.0918	8.0939
K_{abc2}	0.3100	18.2593	18.2619	0.0892	10.8993	10.8996	0.1550	8.0230	8.0245
$K_{ab}^{(-)}$	0.3960	17.1907	17.1952	0.1161	10.9484	10.9490	0.1873	8.0161	8.0183
$K_{abc1}^{(-)}$	0.4379	17.3478	17.3533	0.0953	10.9437	10.9441	0.1777	8.0109	8.0128
$K_{abc2}^{(-)}$	0.4394	17.3298	17.3354	0.1033	10.9445	10.9450	0.1703	8.0031	8.0049
K^{opt}	0.4464	17.2188	17.2246	0.0831	10.8478	10.8482	0.1709	7.9976	7.9994
	$T=100$			$T=250$			$T=500$		
$\rho=-0.6$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$\hat{\theta}_\rho$	0.3730	12.1746	12.1803	0.2035	8.2096	8.2122	−0.1535	5.8634	5.8654
$CCQML$	0.6090	13.1052	13.1194	0.1867	8.2369	8.2390	0.1860	5.9337	5.9366
$CCLS$	0.4006	12.3911	12.3976	0.1885	8.1870	8.1892	−0.1420	5.8972	5.8989
K_{ab}	0.3177	12.6940	12.6980	0.1879	8.2176	8.2197	−0.1325	5.9305	5.9320
K_{abc1}	0.3294	12.7519	12.7562	0.1939	8.2223	8.2246	−0.1496	5.9215	5.9233
K_{abc2}	0.3253	12.5633	12.5675	0.1975	8.1671	8.1695	−0.1346	5.8999	5.9015
$K_{ab}^{(-)}$	0.4154	12.6040	12.6108	0.2015	8.2275	8.2300	0.2106	5.9590	5.9627
$K_{abc1}^{(-)}$	0.5273	12.3723	12.3836	0.2110	8.2136	8.2163	0.1867	5.9071	5.9101
$K_{abc2}^{(-)}$	0.5242	12.2964	12.3076	0.1951	8.2253	8.2276	0.1916	5.9155	5.9186
K^{opt}	0.5115	12.0509	12.0617	0.1662	8.1129	8.1146	−0.1502	5.8319	5.8338

Notes: (1) The first column with CCQML, CCLS and K^{opt} denote the estimates of $\hat{\theta}_{QML}$, $\hat{\theta}_{CCLS}$, and two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$ respectively. $K_{ab}/K_{ab}^{(-)}$ are with $(a,b,c)=(10,6,10)$ and $(10,-6,10)$ respectively. K_{abci} is with $(a,b,c)=(10,8,19)$ for $i=1$ and $(10,2,6)$ for $i=2$. $K_{abci}^{(-)}$ is with $(a,b,c)=(10,-3,2.5)$ for $i=1$ and $(10,-4,3)$ for $i=2$.

(2) Constant conditional correlation for $u_{L,t}$ and $u_{R,t}$ with $\rho=0$ and $\rho=-0.6$ are considered respectively. $\hat{\theta}_\rho$ is from the kernel with $(a,b,c)=(1,\rho,1)$.

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

TABLE 6-2. Bias, SD and RMSE of Estimates for β_0 in CCC-GARCH (1,1)

$\hat{\beta}_0$									
	$T=100$			$T=250$			$T=500$		
$\rho=0$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$\hat{\theta}_\rho$	1.5832	36.9517	36.9856	0.8096	22.2116	22.2264	0.5589	16.2638	16.2734
$CCQML$	1.7432	38.5877	38.6271	0.7815	22.1291	22.1429	0.7069	16.2564	16.2718
$CCLS$	1.6927	37.1014	37.1400	0.6939	21.8755	21.8865	0.5547	16.2169	16.2264
K_{ab}	1.9383	38.1556	38.2048	0.6352	21.8275	21.8368	0.5720	16.3784	16.3884
K_{abc1}	1.9302	37.8494	37.8986	0.6820	21.9591	21.9697	0.5775	16.4972	16.5073
K_{abc2}	1.5620	37.2145	37.2472	0.6712	21.7046	21.7149	0.5454	16.1487	16.1579
$K_{ab}^{(-)}$	-0.2536	34.7072	34.7081	0.6049	21.9392	21.9475	0.4842	15.9856	15.9929
$K_{abc1}^{(-)}$	-0.2246	34.9345	34.9352	0.5747	21.8082	21.8158	0.5303	15.9720	15.9808
$K_{abc2}^{(-)}$	-0.2628	34.7820	34.7830	0.5765	21.8167	21.8243	0.5503	15.9736	15.9831
K^{opt}	-0.1908	34.6822	34.6827	0.5420	21.7177	21.7245	0.5371	15.9058	15.9149
	$T=100$			$T=250$			$T=500$		
$\rho=-0.6$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$\hat{\theta}_\rho$	-0.1764	41.2049	41.2052	0.9344	26.2614	26.2781	0.6604	19.8438	19.8547
$CCQML$	-0.3959	41.1092	41.1111	1.1400	26.3765	26.4012	0.7314	19.6244	19.6380
$CCLS$	-0.3646	41.3780	41.3796	0.8330	26.1217	26.1350	0.7121	19.9028	19.9156
K_{ab}	-0.2591	41.7327	41.7335	0.8224	26.1503	26.1632	0.7377	19.9933	20.0069
K_{abc1}	-0.1719	41.6554	41.6558	0.8072	26.2531	26.2655	0.7449	20.0790	20.0929
K_{abc2}	-0.2549	41.4431	41.4439	0.8150	26.0364	26.0492	0.7160	19.8347	19.8476
$K_{ab}^{(-)}$	-0.2983	42.2806	42.2817	0.8938	26.4456	26.4607	0.6468	19.8235	19.8340
$K_{abc1}^{(-)}$	-0.9515	41.0943	41.1053	0.8497	26.0677	26.0816	0.6262	19.4292	19.4393
$K_{abc2}^{(-)}$	-0.8702	41.2807	41.2899	0.9330	26.1331	26.1497	0.6166	19.4309	19.4407
K^{opt}	-0.5937	40.1934	40.1977	0.9182	25.7977	25.8141	0.5763	19.3842	19.3928

Notes: (1) The first column with CCQML, CCLS and K^{opt} denote the estimates of $\hat{\theta}_{QML}$, $\hat{\theta}_{CCLS}$, and two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$ respectively. $K_{ab}/K_{ab}^{(-)}$ are with $(a,b,c)=(10,6,10)$ and $(10,-6,10)$ respectively. K_{abci} is with $(a,b,c)=(10,8,19)$ for $i=1$ and $(10,2,6)$ for $i=2$. $K_{abci}^{(-)}$ is with $(a,b,c)=(10,-3,2.5)$ for $i=1$ and $(10,-4,3)$ for $i=2$.

(2) Constant conditional correlation for $u_{L,t}$ and $u_{R,t}$ with $\rho=0$ and $\rho=-0.6$ are considered respectively. $\hat{\theta}_\rho$ is from the kernel with $(a,b,c)=(1,\rho,1)$.

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

TABLE 7-2. Bias, SD and RMSE of Estimates for β_1 in CCC-GARCH (1,1)

$\hat{\beta}_1$									
	$T=100$			$T=250$			$T=500$		
$\rho=0$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$\hat{\theta}_\rho$	−5.1700	23.3402	23.9059	−1.8568	15.8522	15.9606	−0.9422	11.7112	11.7491
$CCQML$	−6.8071	26.0490	26.9237	−1.8205	15.4964	15.6030	−1.1804	12.8178	12.8720
$CCLS$	−4.8740	22.0649	22.5968	−1.6444	14.9143	15.0046	−0.9085	11.3070	11.3434
K_{ab}	−5.1082	24.0989	24.6343	−1.5737	15.8136	15.8917	−0.8970	11.9478	11.9814
K_{abc1}	−5.2323	24.5890	25.1395	−1.5684	16.3361	16.4112	−0.9375	12.4469	12.4821
K_{abc2}	−4.5885	21.1834	21.6747	−1.5237	14.2489	14.3302	−0.7890	10.7557	10.7846
$K_{ab}^{(-)}$	−4.4438	20.6082	21.0819	−1.6554	14.2130	14.3091	−0.6782	10.3726	10.3948
$K_{abc1}^{(-)}$	−4.4396	20.2261	20.7076	−1.5819	13.6926	13.7837	−0.6446	10.0951	10.1156
$K_{abc2}^{(-)}$	−4.4845	20.1118	20.6057	−1.5886	13.6684	13.7604	−0.6692	10.0957	10.1178
K^{opt}	−3.9911	19.5291	19.9327	−1.4666	13.2740	13.3548	−0.6488	9.7382	9.7598
	$T=100$			$T=250$			$T=500$		
$\rho=-0.6$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$\hat{\theta}_\rho$	−5.1139	23.1467	23.7049	−1.8138	15.6059	15.7110	−1.3468	12.2553	12.3291
$CCQML$	−5.1057	22.4803	23.0528	−2.1822	15.1071	15.2639	−1.3685	11.2132	11.2964
$CCLS$	−5.3099	24.1066	24.6845	−1.7459	15.7767	15.8730	−1.4848	12.4950	12.5829
K_{ab}	−5.7424	25.9332	26.5613	−1.8016	16.7059	16.8027	−1.5657	13.1845	13.2771
K_{abc1}	−5.8670	25.7590	26.4187	−1.7448	16.8426	16.9327	−1.5415	13.2432	13.3327
K_{abc2}	−5.2287	24.1880	24.7466	−1.7667	15.7318	15.8307	−1.4833	12.4906	12.5784
$K_{ab}^{(-)}$	−5.2628	23.3868	23.9717	−1.7896	15.9658	16.0657	−1.0475	12.1638	12.2088
$K_{abc1}^{(-)}$	−3.7596	20.5259	20.8673	−1.8321	13.9451	14.0649	−0.8708	10.2275	10.2646
$K_{abc2}^{(-)}$	−3.7202	20.7477	21.0786	−1.8359	14.1548	14.2734	−0.9050	10.3940	10.4333
K^{opt}	−4.0383	19.7016	20.1112	−1.6706	13.4004	13.5041	−0.8012	9.9381	9.9703

Notes: (1) The first column with CCQML, CCLS, and K^{opt} denote the estimates of $\hat{\theta}_{QML}$, $\hat{\theta}_{CCLS}$, and two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$ respectively. $K_{ab}/K_{ab}^{(-)}$ are with $(a,b,c)=(10,6,10)$ and $(10,-6,10)$ respectively. K_{abci} is with $(a,b,c)=(10,8,19)$ for $i=1$ and $(10,2,6)$ for $i=2$. $K_{abci}^{(-)}$ is with $(a,b,c)=(10,-3,2.5)$ for $i=1$ and $(10,-4,3)$ for $i=2$.

(2) Constant conditional correlation for $u_{L,t}$ and $u_{R,t}$ with $\rho=0$ and $\rho=-0.6$ are considered respectively. $\hat{\theta}_\rho$ is from the kernel with $(a,b,c)=(1,\rho,1)$.

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

TABLE 8-2. Bias, SD and RMSE of Estimates for γ_1 in CCC-GARCH (1,1)

$\hat{\gamma}_1$									
$\rho = 0$	$T = 100$			$T = 250$			$T = 500$		
	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$\hat{\theta}_\rho$	3.4481	23.7807	24.0294	1.3455	16.3966	16.4517	0.5192	12.2486	12.2596
$CCQML$	5.1979	24.6054	25.1485	1.0617	15.7562	15.7919	0.6052	11.9287	11.9441
$CCLS$	3.1517	22.3783	22.5992	1.1039	15.3862	15.4257	0.4547	11.8826	11.8913
K_{ab}	3.3173	24.3267	24.5519	0.9726	16.2556	16.2846	0.3874	12.6170	12.6230
K_{abc1}	3.4418	24.8162	25.0538	1.0046	16.8454	16.8753	0.4035	13.1492	13.1554
K_{abc2}	2.8967	21.5026	21.6968	0.9317	14.6547	14.6843	0.3479	11.2991	11.3044
$K_{ab}^{(-)}$	3.3135	20.9494	21.2099	0.8890	14.7091	14.7360	0.2619	10.7972	10.8004
$K_{abc1}^{(-)}$	3.1719	20.6452	20.8874	0.7863	14.1628	14.1846	0.2258	10.5434	10.5458
$K_{abc2}^{(-)}$	3.2349	20.4939	20.7477	0.8111	14.1222	14.1455	0.2374	10.5563	10.5590
K^{opt}	3.3985	19.8049	20.0943	0.6973	13.6731	13.6908	0.2699	10.1557	10.1593
	$T = 100$			$T = 250$			$T = 500$		
$\rho = -0.6$	Bias	SD	RMSE	Bias	SD	RMSE	Bias	SD	RMSE
$\hat{\theta}_\rho$	3.9923	23.6568	23.9913	1.3351	16.0954	16.1507	0.7671	12.5065	12.5300
$CCQML$	3.8581	22.6455	22.9718	1.5001	15.0078	15.0826	0.8370	11.5454	11.5757
$CCLS$	4.0327	24.7236	25.0504	1.2187	16.3211	16.3665	0.9641	12.6448	12.6815
K_{ab}	4.4080	26.6215	26.9840	1.2331	17.2795	17.3234	1.0707	13.3057	13.3487
K_{abc1}	4.5906	26.3725	26.7691	1.2059	17.4410	17.4826	1.0302	13.3751	13.4148
K_{abc2}	3.8728	24.8842	25.1838	1.1989	16.2415	16.2857	0.9941	12.6078	12.6469
$K_{ab}^{(-)}$	4.1370	23.8643	24.2202	1.2463	16.5218	16.5687	0.5322	12.7415	12.7527
$K_{abc1}^{(-)}$	2.3757	21.3781	21.5097	1.1880	14.2453	14.2947	0.3812	10.7022	10.7089
$K_{abc2}^{(-)}$	2.4046	21.5325	21.6664	1.2171	14.4368	14.4880	0.4215	10.8099	10.8181
K^{opt}	2.7108	20.2394	20.4201	1.0267	13.7417	13.7800	0.2870	10.3725	10.3764

Notes: (1) The first column with CCQML, CCLS and K^{opt} denote the estimates of $\hat{\theta}_{QML}$, $\hat{\theta}_{CCLS}$, and two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$ respectively. $K_{ab}/K_{ab}^{(-)}$ are with $(a, b, c) = (10, 6, 10)$ and $(10, -6, 10)$ respectively. K_{abci} is with $(a, b, c) = (10, 8, 19)$ for $i = 1$ and $(10, 2, 6)$ for $i = 2$. $K_{abci}^{(-)}$ is with $(a, b, c) = (10, -3, 2.5)$ for $i = 1$ and $(10, -4, 3)$ for $i = 2$.

(2) Constant conditional correlation for $u_{L,t}$ and $u_{R,t}$ with $\rho = 0$ and $\rho = -0.6$ are considered respectively. $\hat{\theta}_\rho$ is from the kernel with $(a, b, c) = (1, \rho, 1)$.

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

TABLE 9. ***t-statistics*** **for the 5-factor CAPM**

		BE/ME Quantile Group Low						BE/ME Quantile Group 2					
Small		OLS	$CCLS$	K_{ab}	K_{abc}	$CCQML$	K^{opt}	OLS	$CCLS$	K_{ab}	K_{abc}	$CCQML$	K^{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	−1.64	−1.64	−1.61
	β_1	51.25	14.57	15.09	15.80	19.79	19.66	51.01	15.05	15.89	16.90	20.47	20.53
	β_2	−7.96	−0.14	−0.14	−0.16	−1.23	−1.21	−6.94	−1.58	−1.58	-1.91	-21.15	-20.81
	β_3	−15.67	−1.41	−1.31	−1.57	−15.16	−14.90	−0.32	1.21	1.14	1.40	15.76	15.60
	β_4	0.90	0.31	0.32	0.33	0.39	0.38	−0.50	0.30	0.31	0.32	0.34	0.33
	β_5	0.88	0.73	0.53	0.83	3.01	2.99	−1.39	0.19	0.14	0.23	−1.43	−1.45

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

(2) The first row with OLS, CCQML, CCLS and K^{opt} denote the estimates of OLS, $\hat{\theta}_{QML}$, $\hat{\theta}_{CCLS}$, and two-stage minimum D_K-distance estimator $\hat{\theta}^{opt}$ respectively. K_{ab} and K_{abc} are with $(a, b, c) = (10, 6, 10)$ and $(10, 8, 19)$ respectively.

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

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

		BE/ME Quantile Group 3						BE/ME Quantile Group 4					
Small		OLS	$CCLS$	K_{ab}	K_{abc}	$CCQML$	K^{opt}	OLS	$CCLS$	K_{ab}	K_{abc}	$CCQML$	K^{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	$CCLS$	K_{ab}	K_{abc}	$CCQML$	K^{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