Economic Modelling xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.elsevier.com/locate/econmod

A mixed data sampling copula model for the return-liquidity dependence in stock index futures markets[☆]

Yuting Gong[a,*], Qiang Chen[b], Jufang Liang[c]

[a] *Department of Economics, SHU-UTS SILC Business School, Shanghai University, China*
[b] *School of Economics, Shanghai University of Finance and Economics, China*
[c] *School of Finance and Statistics, Hunan University, China*

A R T I C L E I N F O

Keywords:
Mixed data sampling
Copula
CSI 300 index futures
Liquidity
Return

A B S T R A C T

Understanding and quantifying the dependence of returns and liquidity is critical for liquidity risk management. In this paper the idea of mixed data sampling (MIDAS) is extended from linear correlation in Colacito et al. (2011) to the more general dependence measure: copula, and a copula-MIDAS model is proposed to describe the asymmetric return-liquidity dependence of CSI 300 index futures with short-run and long-run components. Based on the skewed t copula-MIDAS model, it is found that extreme decreases in returns tend to be accompanied by extreme increases in bid-ask spreads, but extreme increases in returns may not coincide with extreme reductions in bid-ask spreads. Furthermore, the return-spread dependence consists of both short-run and long-run components, and the long-run component will influence the return-spread dependence in the next two weeks. Last, the out-of-sample forecast of liquidity risk stresses the importance of considering asymmetry and long-run trend in return-spread dependence as it enables investors to well predict liquidity risk in times of market crashes. The results imply that high frequency trading investors of CSI 300 index futures should pay more attentions to prevent the potential liquidity risk when the bid-ask spreads are widened. And investors are suggested to use the past two-week high frequency data to forecast the current return-spread dependence in liquidity risk management.

1. Introduction

Understanding and quantifying the dependence of returns and liquidity is critical for liquidity risk management. Liquidity risk has been of major concern to both Chinese regulators and investors, since the stock market collapses in 2015. In the first half year of 2015, the prices of stock indices and futures are rising rapidly with excessive liquidity, which is fueled by the introduction of stock with-funding. However, from June the stock market starts to crash and liquidity is dried up due to panic investor sentiment and irrational selling. The unprecedented turmoil brings great challenges to regulators and incurs unexpected losses to most investors. Hence, it is necessary to take the return-liquidity dependence into account when investors measure and forecast the liquidity-adjusted risk by means of a joint distributional model.

This paper analyzes the dependence of high-frequency returns and bid-ask spreads in the Chinese CSI 300 (China Securities Index) futures market. The CSI 300 index futures, with CSI 300 index as the underlying asset, are traded in China Financial Futures Exchange and regarded as one of the world's most traded equity futures before the stock market crash in 2015. The topic is interesting because how to model their dependence structure affects the judgment and forecast of liquidity risk in the stock index futures market. The return-liquidity dependence is shown to exhibit two features: (1) asymmetric and (2) having short-run and long-run components. First, returns and liquidity measures have asymmetric dependence patterns in normal or extreme periods. Take the CSI 300 index futures as an example. In 2014 when the market is relatively tranquil, the 5-minute returns and bid-ask spreads (a proxy for liquidity) are weakly correlated with average correlation −0.02. However, in June 2015 with extreme stock crash and liquidity dry-ups, the return-spread correlation drops sharply to −0.23, ten times stronger than in 2014. The asymmetric behaviors of return-spread tail dependence have also been discussed in Weiß and Supper (2013) and Hameed et al. (2010). Second, the high frequency

☆ We would like to thank the editors and three anonymous referees for many helpful comments that helped in improving the quality of the paper. However, any remaining errors are solely ours. This work was supported by the Chinese National Science Foundation through grant number 71601108.

* Correspondence to: Department of Economics, SHU-UTS SILC Business School, Shanghai University, Room 511, Wenhui Building, 20 Chengzhong Road, Jiading District, Shanghai 201800, China.
E-mail address: yutinggong1@i.shu.edu.cn (Y. Gong).

http://dx.doi.org/10.1016/j.econmod.2017.03.023
Received 31 July 2016; Received in revised form 29 December 2016; Accepted 28 March 2017
0264-9993/

return-liquidity dependence display a long-run trend as it is likely to be affected by the market-wide factors sampled at low frequency, such as market turmoil and funding tightness. For example, in December 2014 during the times of tightness in funding market, the returns and bid-ask spreads have pronounced negative correlation at −0.27. While in August 2015 when the central bank reduces interest rates and the funding condition is improved, the return-spread correlation goes up to −0.03.

Current econometric models are able to model either of the two features individually, but it is still difficult to capture both of them simultaneously. First, the asymmetric return-liquidity dependence can be well characterized by copula models, which are believed to be the more general dependence measure than linear correlations. Copula models are widely employed to examine the asymmetric dependence of financial markets, such as Okimoto (2008) in equity markets and Wang and Xie (2016) in foreign exchange markets. But existing copulas are unable to distinguish the short-run and long-run components in the evolution of nonlinear dependence. Second, the return-liquidity dependence can be decomposed into short-run and long-run components via the mixed data sampling (MIDAS) scheme in Colacito et al. (2011). But their component model is built upon Engle (2002)'s dynamic conditional correlation (DCC) and hence is not flexible enough to specify the short-run and long-run components in nonlinear dependence patterns.

To overcome the problem, this paper proposes the copula-MIDAS model by extending the idea of mixed data sampling from correlation to copula. The newly proposed model enables us not only to describe the nonlinear dependence structures unaffected by the marginal distributions, but also to extract the short-run fluctuations and the long-run trend from the dynamics of nonlinear dependence.

In line with earlier MIDAS models, the copula-MIDAS model extends the idea of mixed data sampling from modelling linear correlation to modelling the more general dependence structure. So far the development of MIDAS models has experienced three periods: MIDAS on mean, MIDAS on volatility and MIDAS on dependence. (1) MIDAS on mean. Ghysels et al. (2004) propose the first basic MIDAS model, which accommodates the low frequency explained variable and the high frequency explanatory variables using highly parsimonious distributed lag polynomials. Ghysels et al. (2007) then include the autoregressive explained variable into the regression (AR-MIDAS). These models are widely used in nowcasting and short-term forecasting macroeconomic indicators, such as Clements and Galvão (2008), Andreou et al. (2011), Ferrara and Marsilli (2013), Monteforte and Moretti (2013) and Foroni et al. (2015). (2) MIDAS on volatility. Ghysels et al. (2005) incorporate the mixed data sampling scheme into modelling volatility dynamics and investigate the risk-return trade-off based on monthly and daily stock data. Engle et al. (2013) develop the GARCH-MIDAS model to revisit the relation between US stock market volatility and macroeconomic activities (inflation and industrial product growth). Kong et al. (2008) and Girardin and Joyeux (2013) use the above models to study the Chinese A-share and B-share stock markets. (3) MIDAS on dependence. Colacito et al. (2011) propose the DCC-MIDAS model of dynamic correlations with short-run and long-run component specifications and explore the correlation dynamics of industry portfolios and the10-year bond. Conrad et al. (2014) endogenize the variation of key macroeconomic figures into DCC-MIDAS and examine the effect of US macroeconomic environment on the long-term stock-oil correlation. Gong et al. (2016) also propose a copula model with the MIDAS mechanism. The copula-MIDAS model in this paper differs from it in two aspects. First, the dynamic mechanism proposed here has no exogenous explanatory variables, while Gong et al. (2016) utilize exogenous macroeconomic variables to explain the stock-bond dependence based on the t copula. Second, this paper analyzes the high-frequency return-spread dependence, while Gong et al. (2016) investigate the factors affecting stock-bond dependence based on daily and monthly variables. Overall, most of the MIDAS models only specify the short-run and long-run components in correlation, which is inadequate to capture the dynamics of nonlinear dependence, in particular, the dynamics of tail dependence. To this end, the copula-MIDAS model fulfills this gap by extending the MIDAS method into dynamic copula models.

Also, the empirical findings about the return-spread dependence complement the literature on the contemporaneous dependence of return and liquidity. The dependence of return and bid-ask spread is shown to be negative, consistent with Amihud (2002), Bekaert et al. (2007), Hameed et al. (2010) and Bali et al. (2014). Unlike them, the analysis using copula-MIDAS model makes one-step further by stating that the dependence structure of returns and bid-ask spreads are not only asymmetric but also having short-run and long-run components. The negative dependence is more pronounced in the case of extreme returns decrease and large bid-ask spreads than the case of extreme returns increase and small bid-ask spreads. Furthermore, the return-spread dependence exhibits a long-run trend that is less fluctuated within two weeks. It implies to investors to use high frequency returns and bid-ask spreads data in the past two weeks to forecast the current return-liquidity dependence in liquidity risk management.

The paper is organized as follows. Section 2 presents the copula-MIDAS model and its estimation. Section 3 provides a preliminary analysis of the 5-minute returns and bid-ask spreads data. Section 4 investigates the short-run and long-run components in the tail dependence of returns and bid-ask spreads and then forecasts the liquidity risk of the CSI 300 futures market. Section 5 concludes.

2. Econometric methodology

2.1. Copula-MIDAS model

The copula-MIDAS model is a natural extension of the DCC-MIDAS model by Colacito et al. (2011) to the copula framework. Consider the bivariate time series process (y_{1t}, y_{2t}), $t = 1, \cdots, T$. Let F_1 and F_2 represent the marginal distribution functions (CDF) of y_{1t} and y_{2t} at time t. By the probability integral transformation, $u_{it} = F_i(y_{it})$ for $i = 1, 2$. The copula-MIDAS model is written as follows:

$$(u_{1t}, u_{2t}) \sim C(u_{1t}, u_{2t}; \theta_t, \eta), \qquad \theta_t = \Lambda(\lambda_t), \tag{1}$$

$$\lambda_t = \lambda_\tau + \alpha g(u_{1t-1}, u_{2t-1}) + \beta\lambda_{t-1}, \tag{2}$$

$$\lambda_\tau = \sum_{k=1}^{K} \varphi_k(\omega)\Lambda^{-1}(\gamma_{\tau-k}). \tag{3}$$

Eq. (1) shows that the copula-MIDAS model belongs to the time-varying copula family. θ_t is the time-varying parameter and $\eta = (\eta^1, \cdots, \eta^Q)'$ is the parsimonious parameter vector. θ_t is assumed to be driven by an unobserved stochastic process λ_t such that $\theta_t = \Lambda(\lambda_t)$, where $\Lambda(\cdot)$ is a non-decreasing transformation to ensure that θ_t remains in its domain as in Patton (2006).

Eq. (2) differentiates the copula-MIDAS model from the extant dynamic copula models. The latent variable has two sampling schemes: one is the high frequency λ_t with subscript t, $t = 1, \cdots, T$, the same sampling frequency as (y_{1t}, y_{2t}); the other is the low frequency λ_τ with subscript τ, $\tau = 1, \cdots, [T/N_c]$, λ_τ changes its value after N_c periods of t. Then the dynamics of dependence structure can be decomposed into two components: a short-run component and a long-run component (discussed in (3)). The specification of the short-run component follows Patton (2006) and the short-lived effects are captured by an autoregressive lag λ_{t-1} and a data-driven term $g(u_{1t-1}, u_{2t-1})$ whose functional form depends on the copula types. α and β are the coefficients and $\beta \in (-1, 1)$.

Eq. (3) gives the long-run component of dependence structure λ_τ that reflects the fundamental or secular causes of the time variation in dependence. γ_τ is the transformed realized correlations of the high frequency observations for $t = (\tau - 1)N_c + 1, \cdots, \tau N_c$, and λ_τ is a