

Bien, Katarzyna; Nolte, Ingmar; Pohlmeier, Winfried

## Working Paper

# An inflated Multivariate Integer Count Hurdle model: An application to bid and ask quote dynamics

CoFE Discussion Paper, No. 07/04

## Provided in Cooperation with:

University of Konstanz, Center of Finance and Econometrics (CoFE)

*Suggested Citation:* Bien, Katarzyna; Nolte, Ingmar; Pohlmeier, Winfried (2007) : An inflated Multivariate Integer Count Hurdle model: An application to bid and ask quote dynamics, CoFE Discussion Paper, No. 07/04, University of Konstanz, Center of Finance and Econometrics (CoFE), Konstanz,  
<https://nbn-resolving.de/urn:nbn:de:bsz:352-opus-32415>

This Version is available at:

<https://hdl.handle.net/10419/32170>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# An Inflated Multivariate Integer Count Hurdle Model:

An Application to Bid and Ask Quote Dynamics

Katarzyna Bien

University of Konstanz

Ingmar Nolte\*

University of Konstanz,  
CoFE

Winfried Pohlmeier

University of Konstanz,  
CoFE, ZEW

This Version: March 28, 2007

---

\*Department of Economics, Box D124, University of Konstanz, 78457 Konstanz, Germany. Phone +49-7531-88-3753, Fax -4450, email: Ingmar.Nolte@uni-konstanz.de. The work is supported in part by the European Community's Human Potential Program under contract HPRN-CT-2002-00232, Microstructure of Financial Markets in Europe; and by the Fritz Thyssen Foundation through the project 'Dealer-Behavior and Price-Dynamics on the Foreign Exchange Market'. For helpful comments we would like to thank Joachim Grammig and David Veredas.

## Abstract

In this paper we develop a model for the conditional inflated multivariate density of integer count variables with domain  $\mathbb{Z}^n$ . Our modelling framework is based on a copula approach and can be used for a broad set of applications where the primary characteristics of the data are: (i) discrete domain, (ii) the tendency to cluster at certain outcome values and (iii) contemporaneous dependence. These kind of properties can be found for high or ultra-high frequent data describing the trading process on financial markets. We present a straightforward method of sampling from such an inflated multivariate density through the application of an Independence Metropolis-Hastings sampling algorithm.

We demonstrate the power of our approach by modelling the conditional bivariate density of bid and ask quote changes in a high frequency setup. We show how to derive the implied conditional discrete density of the bid-ask spread, taking quote clusterings (at multiples of 5 ticks) into account.

*JEL classification:* G10, F30, C30

*Keywords:* Multivariate Discrete Distributions, Conditional Inflation, Copula Functions, Truncations, Metropolized-Independence Sampler

# 1 Introduction

In this paper a model for the conditional multivariate density of integer count variables with domain  $\mathbb{Z}^n$  is developed. The model specification is suited to cope with data obeying the following primary characteristics: first, their discrete domain, second, the tendency to cluster at certain outcome values and third, their contemporaneous dependence. We show that the model can be generalized by accounting for additional characteristics such as truncation regions and error correction mechanisms.

Although econometric modelling of univariate processes with discrete supports has been studied extensively, the multivariate counterpart is still underdeveloped. Most of the existing approaches (e.g. Kocherlakota & Kocherlakota (1992), Johnson, Kotz & Balakrishnan (1997)) concentrate on the parametric modelling of multivariate discrete distributions with a nonnegative domain and a nonnegative contemporaneous dependency only. Alternatively, Cameron, Li, Trivedi & Zimmer (2004) exploit the concept of copula functions to derive a more flexible form of the bivariate distribution for non-negative count variables that allows for both a positive or a negative dependence between the discrete random variables.

In the univariate framework, inflated models, and especially the zero-inflated ones (Mullahy (1986)) gained a large amount of attention in several academic disciplines (e.g. Lambert (1986), Cheung (2002), Jang (2005)). However, there are only a few contributions referring to their zero-inflated multivariate counterparts (Dixon & Coles (1997), Li, Lu, Park, Kim & Peterson (1999), Wahlin (2001), Wang, Lee, Yau & Carrivick (2003)). A more general treatment of multivariate inflation models has been proposed by Karlis & Ntzoufras (2003, 2005), who suggest an bivariate Poisson model where the diagonal points of the probability table are inflated.

The inflated multivariate integer count hurdle model (*MICH*) proposed here can be viewed as an inflated version of a model being a combination of the copula approach by Cameron et al. (2004) and the Integer Count Hurdle (ICH) model of Liesenfeld, Nolte & Pohlmeier (2006) which allows for the dynamic specification of univariate conditional distributions with discrete domain  $\mathbb{Z}$ . In the inflated MICH model specific outcomes on  $\mathbb{Z}^n$  are allowed to be inflated unconditionally as well as conditionally based on historical information.

Quite a number of applications of the inflated MICH model are conceivable. Most apparent are applications to high frequent financial data, which are characterized by a set of contemporaneously correlated trade marks, many of them are discrete in nature at high or ultra high frequencies and tend to cluster at certain round values. In empirical studies on financial market microstructure, characteristics of the multivariate time-varying conditional densities (moments, ranges, quantiles, etc.) are crucial. For instance, with our model we are able to derive multivariate conditional volatility or

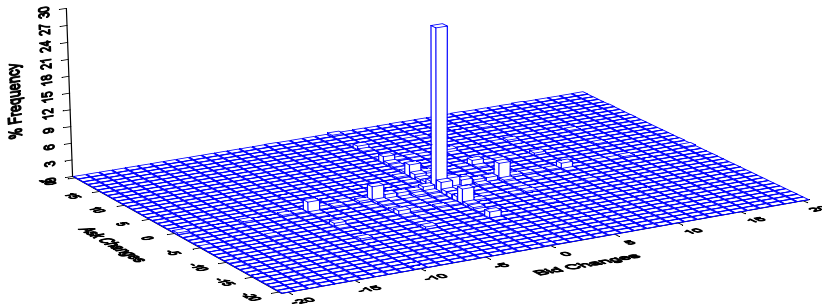
liquidity measures, such as the bid-ask spread and related figures.

The discreteness and the clustering of price and quote changes plays an important role for financial theory and applications. Harris (1991), Christie & Schultz (1994) and Christie, Harris & Schultz (1994) investigate clustering of bid and ask quote prices at round tick sizes, which can be attributed to investor and market maker trading preferences. Huang & Stoll (1994), Crack & Ledoit (1996) and Szpiro (1998) among others, show that discrete price changes and price clustering at round ticks imply a ray shaped pattern in the scatter plot of returns against one period lagged returns, which is referred to as the “compass rose”. The compass rose can be found for many financial instruments on different markets, such as futures (Lee, Gleason & Mathur (1999)), exchange rates (Gleason, Lee & Mathur (2000), Szpiro (1998) ) and stocks (Crack & Ledoit (1996), Antoniou & Vorlow (2005)).

It has several implications for the dynamics of the data generating process of asset returns which may render naively applied statistical tests such as the Brock, Dechert, Scheinkman & LeBaron (1996) test (Krämer & Runde (1997)), random walk tests or simple autocorrelation estimates (Fang (2002)), invalid. Moreover, GARCH models estimated for such data may be misspecified (Amilon (2003)) and the assumption of a geometric Brownian Motion as the true price process can at least be questioned, which has consequences, for instance, for option pricing (Ball (1988)) and the discrimination between the market microstructure noise and the underlying price process in the realized volatility literature (Andersen, Bollerslev, Diebold & Labys (1999), Oomen (2005), Hansen & Lunde (2006)). Furthermore, Vorlow (2004) analyzes to which extent such patterns can be exploited for forecasting issues. Our approach contributes to this literature since the inflated MICH is able to pick up complex nonlinear structure such as the compass rose in a multivariate setting.

In this paper, we demonstrate the power of our approach by modelling the bivariate time series process of bid and ask quote changes sampled a high frequency. We show how to derive the implied conditional discrete density of the bid-ask spread, which itself or whose characteristics (e.g. quantiles) can be considered as trivial proxies of market liquidity. The empirical application is based on bid and ask quote change data for the Citicorp stock, traded at the NYSE over the period from 20<sup>th</sup> February 2001 to 23<sup>rd</sup> February 2001. The bivariate histogram of these data is depicted in Figure 1. The distinct features of these times series are: Firstly, bid and ask quote changes are discrete because they are multiples of the tick size. Secondly, they tend to cluster at distinct points, since the bid-ask spread tends to cluster at multiples of 5 ticks. Thirdly, quotes changes reveal a strong instantaneous dependence, primarily because of the potential existence of a long run equilibrium relationship between the bid and ask quote processes. The fact that the bid-ask spread has to be positive additionally implies a certain co-movement between the quote changes, as well. The positivity constraint on the bid-ask spread, furthermore, implies a dynamic truncation region on

the conditional density of the bid and ask quote changes. In our model, we account for this truncation region by using a truncated copula so that we can avoid imposing restrictions on the marginal processes.



**Figure 1:** Histogram of bid and ask quote changes in number of ticks.

To evaluate the goodness-of-fit of our model, we propose to simulate the conditional bivariate density of the bid and ask quote process, with the help of a Metropolized Independence Sampling (MIS) algorithm of Hastings (1970). This approach allows us to inspect both the dynamics and the bivariate density specification in a straightforward way.

The paper is organized as follows. In Section 2 we describe the general modelling framework. Section 3 contains the descriptive analysis of the bid and ask change processes. In Section 4, we discuss specification issues and present the estimation results, whereas Section 5 contains the description of the MIS algorithm and the model evaluation. Section 6 concludes.

## 2 General Modelling Framework

Let  $Y_t = (Y_{1t}, \dots, Y_{nt})' \in \mathbb{Z}^n$ , with  $t = 1, \dots, T$ , denote the multivariate process of  $n$  integer count variables and let  $\mathcal{F}_{t-1}$  denote the associated filtration at time  $t - 1$ . Moreover, let  $F(y_{1t}, \dots, y_{nt} | \mathcal{F}_{t-1})$  denote the conditional cumulative density function of  $Y_t$  and  $f(y_{1t}, \dots, y_{nt} | \mathcal{F}_{t-1})$  its conditional density. Furthermore, let  $g_h \equiv g_h(\mathcal{F}_{t-1})$  denote mappings from  $\mathbb{Z}^n$  into a subset  $B_h \subset \mathbb{R}$ , where  $h = 1, \dots, H$ , i.e.  $g_h : \mathbb{Z}^n \rightarrow B_h$  and assume that  $B_h \cap B_{h'} = \emptyset$  for  $h \neq h'$ . Thus, the functions  $g_h(\mathcal{F}_{t-1})$  define distinct events at time  $t - 1$  on which we can condition at time  $t$ .

We assume, given condition  $g_h(\mathcal{F}_{t-1}) \in B_h$ , that certain outcomes ( $Y_t|g_h(\mathcal{F}_{t-1}) \in B_h$ ) are inflated. With  $A_h \equiv A_h|g_h(\mathcal{F}_{t-1}) \in B_h \subset \mathbb{Z}^n$  we denote the sets of these points, which should be inflated and we specify the conditional density of  $Y_t$  given the information up to  $\mathcal{F}_{t-1}$  as:

$$\begin{aligned}
f(y_{1t}, \dots, y_{nt}|\mathcal{F}_{t-1}) = & \\
& \sum_{h=1}^H \sum_{(i_1, \dots, i_n) \in A_h} \omega_{i_1, \dots, i_n|B_h} \cdot \mathbb{1}\{Y_{1t} = i_1, \dots, Y_{nt} = i_n\} \cdot \mathbb{1}\{g_h(\mathcal{F}_{t-1}) \in B_h\} \\
& + \left( 1 - \sum_{h=1}^H \sum_{(i_1, \dots, i_n) \in A_h} \omega_{i_1, \dots, i_n|B_h} \cdot \mathbb{1}\{g_h(\mathcal{F}_{t-1}) \in B_h\} \right) \cdot f^*(y_{1t}, \dots, y_{nt}|\mathcal{F}_{t-1})
\end{aligned} \tag{1}$$

where  $\omega_{i_1, \dots, i_n|B_h} \in (0, 1)$ , with  $h = 1, \dots, H$ , represents the inflation probability of a certain point  $(i_1, \dots, i_n)$  contained in  $A_h$ , given  $g(\mathcal{F}_{t-1}) \in B_h$ , and  $f^*(y_{1t}, \dots, y_{nt}|\mathcal{F}_{t-1})$  denotes a discrete multivariate density that is modelled with a multivariate integer count hurdle (MICH) specification as presented below. The multivariate density in equation (1) is generated by a mixture distribution consisting of a multivariate degenerate distribution and the multivariate density  $f^*(y_{1t}, \dots, y_{nt}|\mathcal{F}_{t-1})$ .

In order to illustrate the setup of the model, consider the following example which we will elaborate on in the empirical application below. Assume, that in a bivariate setting,  $Y_{1t}$  and  $Y_{2t}$  denote bid and ask quote changes of a given stock.  $g_h(\mathcal{F}_{t-1}) \in B_h$  could then denote the conditioning events of 5-tick-wide or 10-tick-wide bid-ask spreads at time  $t - 1$ . Conditional on these events, the outcomes  $A_h = \{(0, 0), (-5, 0), (0, 5)\}$  of the bivariate  $(Y_{1t}, Y_{2t})$  process appear relatively more frequent than other outcomes. Thus, the probabilities referring to the set of outcomes  $A_h$  should be appropriately inflated at time  $t$ .

As a benchmark specification for the joint density  $f^*(y_{1t}, \dots, y_{nt}|\mathcal{F}_{t-1})$  we propose the standard MICH model, where each of the marginal processes,  $Y_{kt}$ ,  $k = 1, \dots, n$  is assumed to follow a dynamic integer count hurdle (ICH) process as described in Liesenfeld et al. (2006) whereas the dependency between the marginals is modelled with a copula function, as in Cameron et al. (2004). Thus, in opposite to most of the existing approaches for discrete distributions modelling (e.g. Kocherlakota & Kocherlakota (1992), Johnson et al. (1997)), we neither restrict the domain of marginal densities to be nonnegative, nor we preclude negative contemporaneous dependence between them. In the following, we briefly describe the theoretical framework behind the standard MICH approach, where we first present the ICH model for the marginal densities. Second, we describe how to account for the contemporaneous dependence between the marginal processes with the help of a copula function.

## 2.1 Marginal Processes

The ICH model, is based on the decomposition of the process of discrete integer valued variables into two components, i.e., a process indicating whether the integer variable is negative, equal to zero or positive (the direction process) and a process for the absolute value of the discrete variable irrespective of its sign (the size process). We present here the simplest form of the ICH model and we refer to Liesenfeld et al. (2006) for a more elaborate presentation.

Let  $\pi_{jt}^k$ ,  $j \in \{-1, 0, 1\}$  denote the conditional probabilities of respectively: a negative  $P(Y_{kt} < 0 | \mathcal{F}_{t-1})$ , a zero  $P(Y_{kt} = 0 | \mathcal{F}_{t-1})$  or a positive  $P(Y_{kt} > 0 | \mathcal{F}_{t-1})$  value of the integer variable  $Y_{kt}$ ,  $k = 1, \dots, n$ , at time  $t$ . The conditional density of  $Y_{kt}$  is then specified as

$$f(y_{kt} | \mathcal{F}_{t-1}) = \pi_{-1t}^k \mathbb{1}_{\{y_{kt} < 0\}} \cdot \pi_{0t}^k \mathbb{1}_{\{y_{kt} = 0\}} \cdot \pi_{1t}^k \mathbb{1}_{\{y_{kt} > 0\}} \cdot f_{|Y_{kt}|}(|y_{kt}| | Y_{kt} \neq 0, \mathcal{F}_{t-1})^{(1 - \mathbb{1}_{\{y_{kt} = 0\}})},$$

where  $f_{|Y_{kt}|}(|y_{kt}| | Y_{kt} \neq 0, \mathcal{F}_{t-1})$  denotes the conditional density of the size process with support  $\mathbb{N} \setminus \{0\}$ . To get a parsimoniously specified model, we adopt the simplification of Liesenfeld et al. (2006), that the conditional density of an absolute value of a variable stems from the same distribution irrespective of whether the variable is positive or negative.

The conditional probabilities of the direction process are modelled with the autoregressive conditional multinomial model (ACM) of Russell & Engle (2002) using a logistic link function given by

$$\pi_{jt}^k = \frac{\exp(\Lambda_{jt}^k)}{\sum_{j=-1}^1 \exp(\Lambda_{jt}^k)} \quad (2)$$

where  $\Lambda_{0t}^k = 0$ ,  $\forall t$  is the normalizing constraint. The resulting vector of log-odds ratios  $\Lambda_t^k \equiv (\Lambda_{-1t}^k, \Lambda_{1t}^k)' = (\ln[\pi_{-1t}^k / \pi_{0t}^k], \ln[\pi_{1t}^k / \pi_{0t}^k])'$  is specified as a multivariate ARMA(1,1)-type model:

$$\Lambda_t^k = G_1 Z_{t-1}^k + \lambda_t^k \quad \text{with} \quad \Lambda_t^k = \mu + B_1 \Lambda_{t-1}^k + A_1 \xi_{t-1}^k. \quad (3)$$

The vector  $Z_{t-1}^k$  contains further explanatory variables and  $G_1$  denotes the corresponding coefficient matrix,  $\mu$  denotes the vector of constants, and  $B_1$  and  $A_1$  denote  $2 \times 2$  coefficient matrices. In the empirical application, we put the following symmetry and diagonality restrictions  $b_{11}^{(1)} = b_{22}^{(1)}$  and  $b_{12}^{(1)} = b_{21}^{(1)} = 0$  on the  $B_1$  matrix to obtain a parsimonious model specification. The innovation vector of the ARMA model is specified as a martingale difference sequence in the following way:

$$\xi_t^k \equiv (\xi_{-1t}^k, \xi_{1t}^k)', \quad \text{where} \quad \xi_{jt}^k \equiv \frac{x_{jt}^k - \pi_{jt}^k}{\sqrt{\pi_{jt}^k(1 - \pi_{jt}^k)}}, \quad j \in \{-1, 1\}, \quad (4)$$



and

$$x_t^k \equiv (x_{-1t}^k, x_{1t}^k)' = \begin{cases} (1, 0)' & \text{if } Y_{kt} < 0 \\ (0, 0)' & \text{if } Y_{kt} = 0 \\ (0, 1)' & \text{if } Y_{kt} > 0, \end{cases} \quad (5)$$

denotes the state vector, whether  $Y_{kt}$  decreases, stays equal or increases at time  $t$ . Thus,  $\xi_t^k$  represents the standardized state vector  $x_t^k$ .

The conditional density of the size process is modelled with an at-zero-truncated Negative Binomial (NegBin) distribution:

$$f_{|Y_{kt}|}(|y_{kt}| | Y_{kt} \neq 0, \mathcal{F}_{t-1}) \equiv \frac{\Gamma(\kappa + |y_{kt}|)}{\Gamma(\kappa)\Gamma(|y_{kt}| + 1)} \left( \left[ \frac{\kappa + \omega_t^k}{\kappa} \right]^\kappa - 1 \right)^{-1} \left( \frac{\omega_t^k}{\omega_t^k + \kappa} \right)^{|y_{kt}|}, \quad (6)$$

where  $|y_{kt}| \in \mathbb{N} \setminus \{0\}$ ,  $\kappa > 0$  denotes the dispersion parameter. The scaling parameter  $\omega_t^k$  is parameterized using the exponential link function with a generalized autoregressive moving average model (GLARMA(1,1)) of Shephard (1995) in the following way:

$$\ln \omega_t^k = \delta \tilde{D}_t + \tilde{\lambda}_t^k \quad \text{with} \quad \tilde{\lambda}_t^k = \tilde{\mu} + S^k(\nu, \tau, K) + \beta_1 \tilde{\lambda}_{t-1}^k + \alpha_1 \tilde{\xi}_{t-1}^k.$$

where  $\tilde{D}_t \in \{-1, 1\}$  indicates a negative or positive value of  $Y_{kt}$  at time  $t$  with the corresponding coefficient denoted by  $\delta$ .  $\tilde{\mu}$  denotes the constant term.  $\beta_1$  as well as  $\alpha_1$  denote coefficients and  $\tilde{\xi}_t^k$  being constructed as

$$\tilde{\xi}_t^k \equiv \frac{|Y_{kt}| - \mathbb{E}(|Y_{kt}| | Y_{kt} \neq 0, \mathcal{F}_{t-1})}{\text{V}(|Y_{kt}| | Y_{kt} \neq 0, \mathcal{F}_{t-1})^{1/2}},$$

is the innovation term that drives the GLARMA model in  $\lambda_t^k$ . The conditional moments of the at-zero-truncated NegBin distribution are given by

$$\begin{aligned} \mathbb{E}(|Y_{kt}| | Y_{kt} \neq 0, \mathfrak{F}_{t-1}) &= \frac{\omega_t^k}{1 - \vartheta_t^k}, \\ \text{V}(|Y_{kt}| | Y_{kt} \neq 0, \mathfrak{F}_{t-1}) &= \frac{\omega_t^k}{1 - \vartheta_t^k} - \left( \frac{\omega_t^k}{(1 - \vartheta_t^k)} \right)^2 \left( \vartheta_t^k - \frac{1 - \vartheta_t^k}{\kappa} \right), \end{aligned}$$

where  $\vartheta_t^k$  is given by  $\vartheta_t^k = [\kappa / (\kappa + \omega_t^k)]^\kappa$ . The Fourier flexible form

$$S^k(\nu, \tau, K) \equiv \nu_0 \tau + \sum_{l=1}^K \nu_{2l-1} \sin(2\pi(2l-1)\tau) + \nu_{2l} \cos(2\pi(2l)\tau) \quad (7)$$

captures potential diurnal seasonality evident in intraday high frequent financial data, where  $\tau$  is the intraday time standardized to  $[0, 1]$  and  $\nu$  is a  $2K + 1$  dimensional parameter vector.

## 2.2 Copula Function

The contemporaneous dependence between the  $n$  marginal processes is modelled through a copula function. Sklar (1959) introduced the copula concept and Patton (2001) extended it to conditional distributions. Within that framework the marginal distributions and/or the copula function can be specified conditional on  $\mathfrak{F}_{t-1}$ , so that the conditional multivariate distribution of  $Y_t$  can be modelled as:

$$F(y_{1t}, \dots, y_{nt} | \mathcal{F}_{t-1}) = \mathbf{C}(F(y_{1t} | \mathcal{F}_{t-1}), \dots, F(y_{nt} | \mathcal{F}_{t-1}) | \mathcal{F}_{t-1}), \quad (8)$$

where  $F(y_{kt} | \mathcal{F}_{t-1})$  denotes the conditional distribution function of the  $k^{\text{th}}$  component and  $\mathbf{C}(\cdot | \mathcal{F}_{t-1})$  the conditional copula function defined on the domain  $[0, 1]^n$ . This approach provides a flexible tool for modelling multivariate distributions as it allows for the decomposition of the multivariate distribution into the marginal distributions, which are interconnected by a copula function, being solely responsible for their contemporaneous dependence.

If the marginal distribution functions are continuous, the copula function  $\mathbf{C}$  is unique on its domain  $[0, 1]^n$ , because the random variables  $Y_{kt}$ ,  $k = 1, \dots, n$  are mapped through the strictly monotone increasing functions  $F(y_{kt} | \mathcal{F}_{t-1})$  onto the entire set  $[0, 1]^n$ . The joint density function can then be derived by differentiating  $\mathbf{C}$  with respect to the continuous random variables  $Y_{kt}$ , as:

$$f(y_{1t}, \dots, y_{nt} | \mathcal{F}_{t-1}) = \frac{\partial^n \mathbf{C}(F(y_{1t} | \mathcal{F}_{t-1}), \dots, F(y_{nt} | \mathcal{F}_{t-1}) | \mathcal{F}_{t-1})}{\partial y_{1t} \dots \partial y_{nt}}, \quad (9)$$

However, if the random variables  $Y_{kt}$  are discrete,  $F(y_{kt} | \mathcal{F}_{t-1})$  are step functions and the copula function is uniquely defined not on  $[0, 1]^n$ , but on the Cartesian product of the ranges of the  $n$  marginal distribution functions, i.e.,  $\bigotimes_{k=1}^n \text{Range}(F_{kt})$  so that it is impossible to derive the multivariate density function using equation (9). In order to overcome this problem, we apply a finite difference approximation to the derivatives of the copula function (e.g. Meester & J.MacKay (1994) and Cameron et al. (2004))

$$f(y_{1t}, \dots, y_{nt} | \mathcal{F}_{t-1}) = \Delta_n \dots \Delta_1 \mathbf{C}(F(y_{1t} | \mathcal{F}_{t-1}), \dots, F(y_{nt} | \mathcal{F}_{t-1}) | \mathcal{F}_{t-1}), \quad (10)$$

where  $\Delta_k$ , for  $k \in \{1, \dots, n\}$ , denotes the  $k^{\text{th}}$  component first order differencing operator being defined through

$$\begin{aligned} \Delta_k \mathbf{C}(F(y_{1t} | \mathcal{F}_{t-1}), \dots, F(y_{kt} | \mathcal{F}_{t-1}), \dots, F(y_{nt} | \mathcal{F}_{t-1}) | \mathcal{F}_{t-1}) = \\ \mathbf{C}(F(y_{1t} | \mathcal{F}_{t-1}), \dots, F(y_{kt} | \mathcal{F}_{t-1}), \dots, F(y_{nt} | \mathcal{F}_{t-1}) | \mathcal{F}_{t-1}) \\ - \mathbf{C}(F(y_{1t} | \mathcal{F}_{t-1}), \dots, F(y_{kt} - 1 | \mathcal{F}_{t-1}), \dots, F(y_{nt} | \mathcal{F}_{t-1}) | \mathcal{F}_{t-1}). \end{aligned}$$

The conditional multivariate density of  $Y_t$  can therefore be derived by specifying the cumulative distribution functions  $F(y_{1t} | \mathcal{F}_{t-1}), \dots, F(y_{nt} | \mathcal{F}_{t-1})$  and the copula function  $\mathbf{C}$  in equation (10).

### 3 Bivariate Density of Bid and Ask Quote Changes

The inflated MICH model is particularly suited for the analysis of market microstructure effects using high-frequency data. A particular feature of such data, especially with respect to the changes of transaction prices or quotes, is their discreteness which precludes modelling methods that rely on continuous distributions. On a high frequency level, prices change as multiples of a minimum tick size, which implies only a few stepwise downward or upward movements within a short time interval.

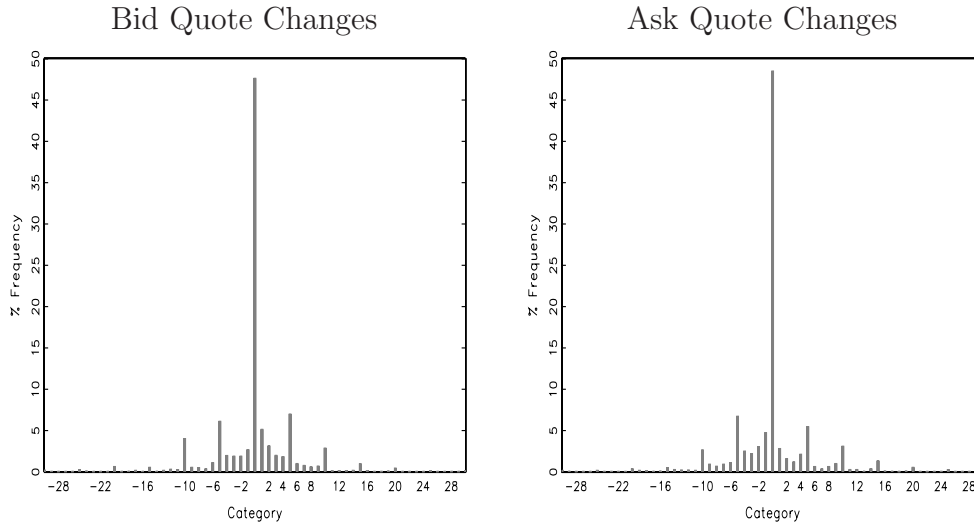
We apply our model to the bivariate process of bid and ask quote changes,  $(C_t^b, C_t^a)$ , where  $C_t^b = Q_t^b - Q_{t-1}^b$ ,  $C_t^a = Q_t^a - Q_{t-1}^a$  and  $Q_t^b$  and  $Q_t^a$  denote the bid and ask quotes at time  $t$ . In our analysis we use quote data of the Citigroup Inc. (C) stock, for the period from the 20<sup>th</sup> to the 23<sup>rd</sup> February 2001 traded at the New York Stock Exchange (NYSE). Citigroup is a very actively traded stock with a mean inter-quote duration of 8.2 seconds and a very high market capitalization of 231.14 bn. USD. The bid and ask quote series have been extracted from the Trades and Quotes (TAQ) Database of the NYSE. We sample the data at a frequency of 30 seconds, as on the one side – it allows to preserve the discrete nature of data, especially its clustering to round numbers, whereas and on the other side – the frequency is low enough to preserve a high correlation between the two marginal processes.

Figure 1, in the Introduction, presents the bivariate histogram of the bid and ask quote changes. We observe that the most frequent outcome is  $(0, 0)$  which corresponds to simultaneous zero movement on both market sides. Besides the discreteness of the data, its most striking feature is clustering to round values. In the histogram we can observe many picks corresponding to the outcomes:  $(-5, 0)$ ,  $(0, 5)$ ,  $(5, 0)$ ,  $(5, 5)$ ,  $(10, 0)$ ,  $(0, 10)$ , which means that bid and ask quotes evidence some tendency to gravitate or rather stick to 'round' states. After the decimalization at the NYSE on the 29<sup>th</sup> January 2001, those states are given by multiples of 5 ticks. Quote clustering, that takes place on a grid that is much rougher than mandated by the existing tick size, is a well-recognized topic in a market microstructure literature. Studies on that issue date back to times long before the decimalization, when the tick size at the NYSE accounted to one-eighth of a dollar. Harris (1991) states for example:

*“Stock prices cluster on round fractions. Integers are more common than halves; halves are more common than odd quarters; odd quarters are more common than odd eights; other fractions are rarely observed. The phenomenon is remarkably persistent through time and across stocks (...) Clustering on larger fractions can be explained if traders sometimes choose to restrict further their terms of trade to the sets of quarters, halves, or whole numbers. The use of these smaller sets may be result of explicit agreements among traders”.*

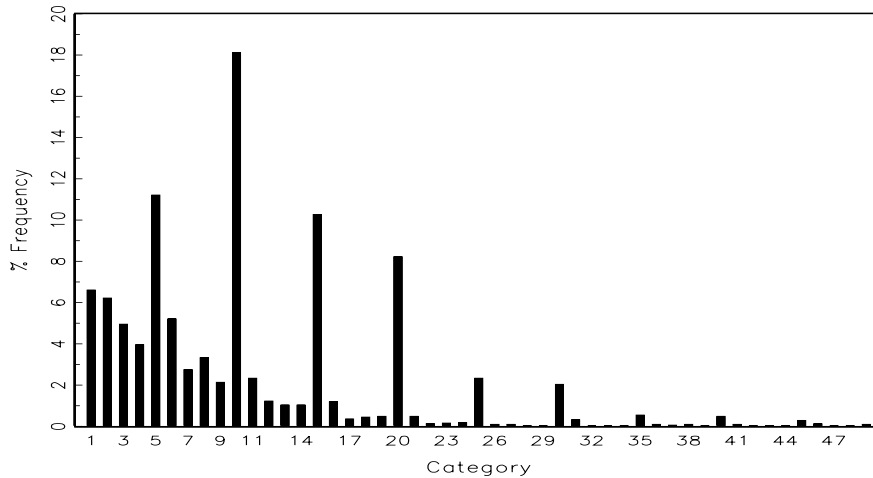
Because of trading convenience market participants prefer to trade at round prices, thus the quotes in the limit order book tend to cluster at round numbers. This phenomenon

can also clearly be observed in the histograms of the marginal processes in Figure 2. We see that both processes have a fairly large discrete support between between -20 and 20 ticks, but the frequencies of the 5, 10, 15 or 20-tick-large quote changes are far above the frequencies, which would be predicted from any smooth univariate discrete distribution.



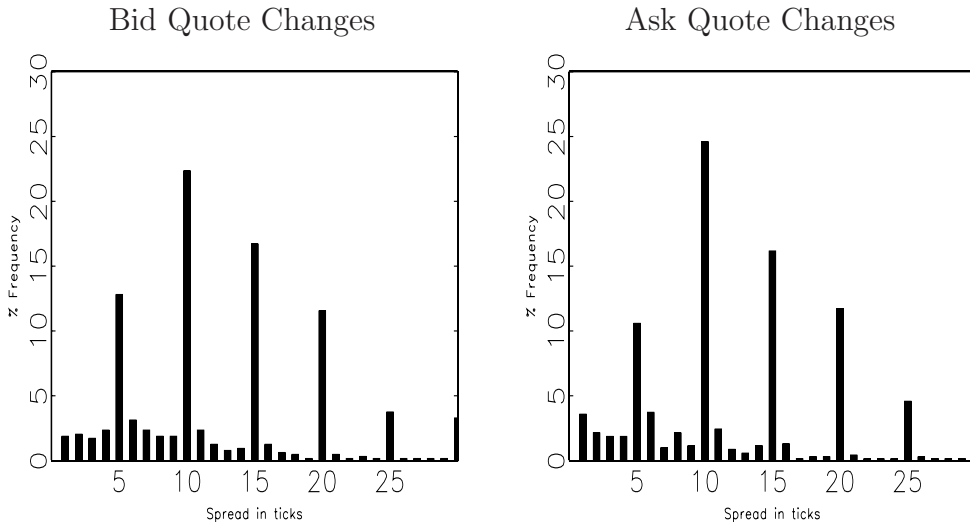
**Figure 2:** Histogram of bid and ask quote changes in number of ticks. Citigroup Inc. stock, traded at the NYSE, 20<sup>th</sup> to the 23<sup>rd</sup> February 2001, number of observations: 3104

Gravitating of quote changes to round values has a tremendous influence on the values of the bid-ask spread. Its distinctly multimodal histogram, presented in Figure 3, shows that nearly 12% of all observed spread values accounted to 5 ticks, about 18% to 10 ticks and about 10% and 8% percent to 15 and 20 ticks, respectively. Thus, even having in mind a stepwise bid-ask spread change induced by the tick size, the roughness of its distribution is much more severe. Non-round values of the variable occur very rarely in comparison to round ones.



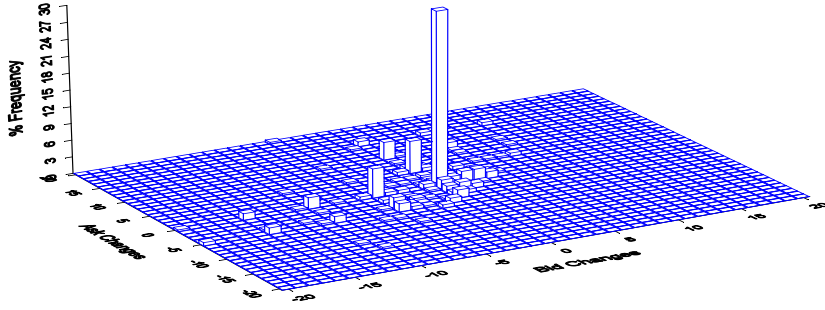
**Figure 3:** Histogram of the bid-ask spread in ticks.

Bid-ask spread clustering has also been analyzed in several market microstructure studies. Hasbrouck (1999) investigates spread clustering on the USD/DM exchange market and hypothesizes, that this phenomenon can reflect a multimodal cost scheme, as there are certain regimes in which cost of market making fixes at round values. He also notes that spread clustering stems from bid and ask quote clustering that is not independent, but takes place simultaneously on both market sides. Thus, if the bid quote lies on a 5 tick multiple, there is in comparison a relatively higher probability, that the ask quote also lies on a five-tick multiple. This observation can also be deduced from our data. In the Figure 4 we present the frequencies (y-axis) of the round ( $\pm 30, \pm 25, \dots, \pm 5$ ) realization of the bid (left panel) and ask (right panel) quote changes at time  $t$ , once the spread of a given size (x-axis) has been observed in time  $t - 1$ . We observe, that once a round value of the bid-ask spread stabilizes, the transmission to those states, where the bid-ask spread takes on round value again is more probable than the transmission to the remaining non-round states.

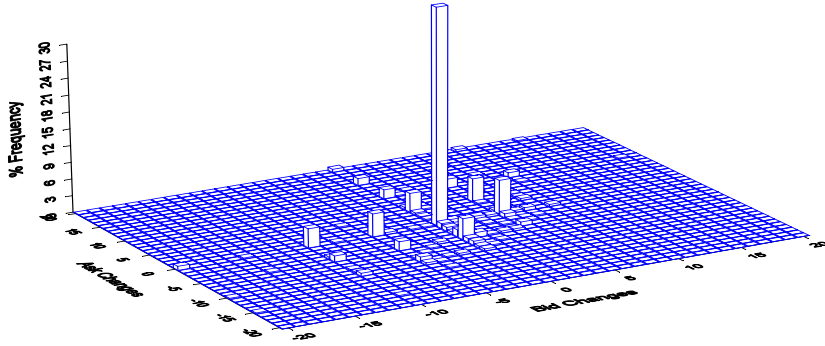


**Figure 4:** Histogram of round quotes changes conditional on a previous round bid-ask spread of the given size.

The simultaneous clustering on both market sides is also clearly visible in Figures 5 and 6. Here we plotted the bivariate histograms of bid and ask quote changes given a previous bid-ask spread of 5 and 10 ticks, respectively. Our modelling scheme, where particular states are conditionally inflated can be easily justified. For example, given a previous spread equal to 5 ticks, the bid and ask quote changes cluster mainly at two outcomes  $(-5, 0)$  and  $(0, 5)$ , which in fact results in a spread equal to 10 ticks (Figure 5). Given a spread equal to 10 ticks, the quote changes tend to cluster on a wider grid course, that is for example  $(-5, 0)$  and  $(0, 5)$  – which leads to a spread of 15,  $(5, 0)$  and  $(0, -5)$  – which leads to a spread of 5 ticks (Figure 6).

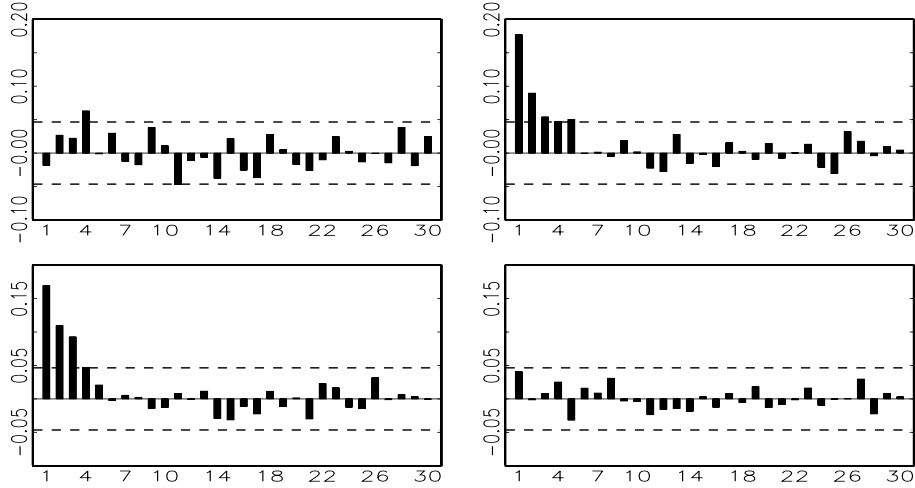


**Figure 5:** Bivariate histogram of bid and ask quote changes in number of ticks, conditional on a previous spread equal 5 ticks.



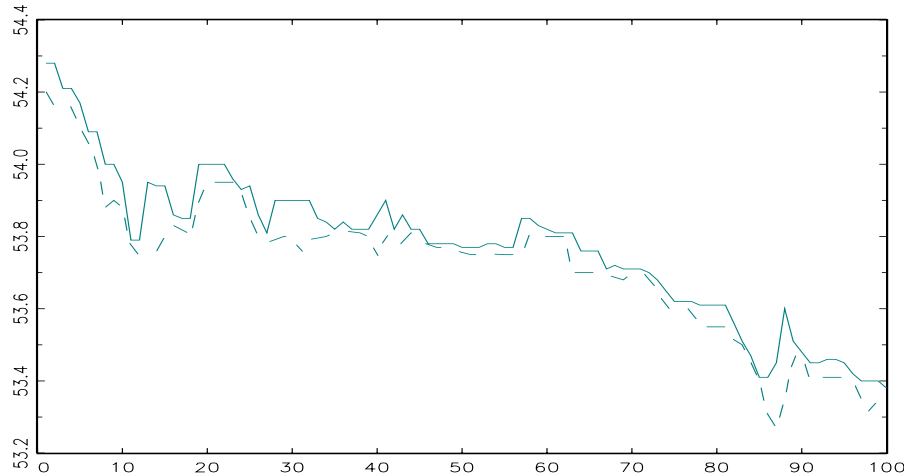
**Figure 6:** Bivariate histogram of bid and ask quote changes in number of ticks, conditional on a previous spread equal 10 ticks.

The dynamic properties of our process are reflected by the multivariate autocorrelation function between the bid and ask quote changes presented in Figure 7. We see that both marginal processes are positively cross-correlated and we observe in Figure 8 that the bid and ask quotes move together. The series  $Q_t^b$  and  $Q_t^a$  are non-stationary, but there must exist a long-run equilibrium between the two processes, as the size of the bid-ask spread, being the function of contemporaneous and lagged quotes changes, cannot increase or decrease infinitely over time.

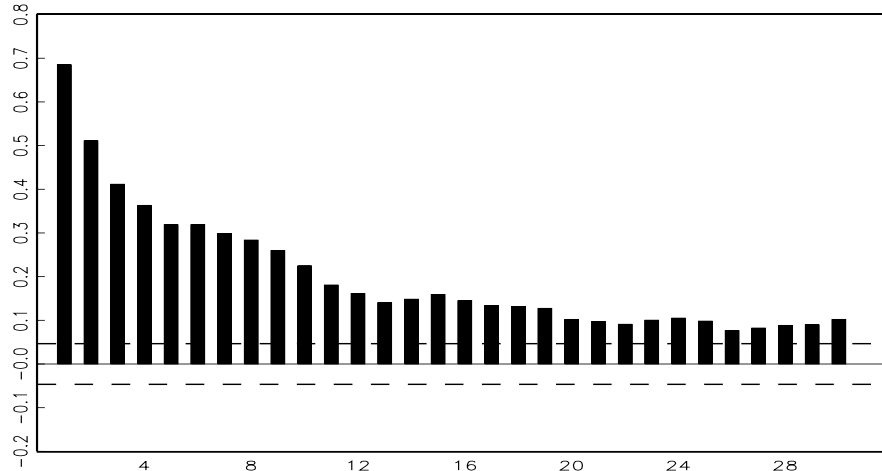


**Figure 7:** Multivariate autocorrelation function for bid and ask quote changes. Upper left panel:  $\text{corr}(c_t^a, c_{t-l}^a)$ ; upper right panel:  $\text{corr}(c_t^b, c_{t-l}^b)$ ; lower left panel:  $\text{corr}(c_{t-l}^a, c_t^b)$  and lower right panel:  $\text{corr}(c_t^b, c_{t-l}^b)$ . The dashed lines mark the approximate 99% confidence interval  $\pm 2.58/\sqrt{T}$ .

The particular behavior of the marginal processes leads to the strong positive autocorrelation of the bid-ask spread, which has been depicted in Figure 9. We therefore conclude, that the periods in which the observed difference between the quotes' levels is relatively wide (narrow) cluster together.



**Figure 8:** Bid and ask quote levels for the first 100 observations of the data sample.



**Figure 9:** Autocorrelation function of bid-ask spread. The dashed line marks the approximate 99% confidence interval  $\pm 2.58/\sqrt{T}$ .

## 4 Specification Issues

In the following we focus on two specifications that clarify the necessity of augmenting the bivariate density of the bid and ask quote changes by conditionally inflating certain quote change probabilities and by including an error correction mechanism:

1. Standard MICH model (without inflated outcomes and error correction mechanism.)
2. Inflated MICH model (augmented by inflated outcomes and error correction mechanism).

The first specification simply corresponds to the component  $f^*(y_{1t}, \dots, y_{nt} | \mathcal{F}_{t-1})$  of the equation (1) and can be perceived as a benchmark model, as it is a standard MICH model, in which we restrain ourselves from inflating any states. The inflated MICH model is an augmented specification as it relies on the whole parametrization stated in equation (1) and has been motivated by the time-varying modi of the bivariate density as well as the comovement between the levels of the bid and ask quotes. In order to reflect the clustering schemes in a possibly careful manner (see Figures 5 and 6), we adapt the following specification of equation (1).  $g_h(\mathcal{F}_{t-1}) = Q_{t-1}^b - Q_{t-1}^a$  denotes the function providing the size of the previous bid-ask spread and  $g_h(\mathcal{F}_{t-1}) \in B_h$  denotes the event that the bid-ask spread is equal to  $B_h$  ticks at time  $t-1$ , where  $h = 1, \dots, 4$ . Conditional on the value of the previous spread  $B_h$ , we inflate selected round outcomes  $A_h$  of the bid and ask quote change process. The detailed definition of the sets  $B_h$  (size of the bid-ask spread in  $t-1$ ) and  $A_h$  (set of inflated outcomes in  $t$ ) are presented in Table 1.



| $h$ | $B_h$          | $A_h$  |
|-----|----------------|--|
| 1   | $B_1 = \{5\}$  | $A_1 = \{(0, 0), (-5, 0), (0, 5), (-10, 0), (0, 10)\}$                 |
| 2   | $B_2 = \{10\}$ | $A_2 = \{(0, 0), (-10, 0), (-5, 0), (0, -5), (0, 5), (5, 0), (5, 5)\}$ |
| 3   | $B_3 = \{15\}$ | $A_3 = \{(0, 0), (0, -5), (5, 0)\}$                                    |
| 4   | $B_4 = \{20\}$ | $A_4 = \{(0, 0), (0, -10), (0, -5), (5, 0), (10, 0)\}$                 |

**Table 1:** Definition of the inflated outcomes.

We decide to inflate the most common outcomes of the bivariate quote processes conditionally on the value of the bid-ask spread. For example, we inflate the outcome  $C_t^b = -5$  and  $C_t^a = 0$ , given that the spread at time  $t - 1$  accounts to 5 ticks. In order to improve the goodness-of-fit of the inflated MICH model, we also decide to inflate certain univariate outcomes of the marginal distributions governed by the ICH specifications. In the light of the histograms presented in Figure 2, we also inflate the outcomes of the absolute quote changes from the set  $C = \{5, 10, 15, 20\}$ . The conditional probability of the absolute bid quote changes (ask quote changes analogically) then takes the following form:

$$\begin{aligned} \tilde{f}_{|C_t^b|}(|c_t| | C_t^b \neq 0, \mathcal{F}_{t-1}) = \\ \sum_{h=1}^4 \sum_{k \in C} \left( (\gamma_{0|k}^b + \gamma_{1|k}^b \cdot \mathbb{1}\{g_h \in B_h\}) \mathbb{1}\{|c_t| = k\} + (1 - \gamma_{0|k}^b - \gamma_{1|k}^b \cdot \mathbb{1}\{g_h \in B_h\}) \right. \\ \left. \cdot f_{|C_t^b|}(|c_t| | C_t^b \neq 0, \mathcal{F}_{t-1}) \right) \end{aligned} \quad (11)$$

where  $g_h(\mathcal{F}_{t-1}) \in B_h$ ,  $h = 1, \dots, 4$  (see Table 1),  $\gamma_{0|k}^b$  and  $\gamma_{1|k}^b$  denote additional inflation parameters and the density  $f_{|C_t^b|}(|c_t| | C_t^b \neq 0, \mathcal{F}_{t-1})$  has been already specified in equation (6). The round values of absolute bid quote changes can therefore be generated either by the truncated-at-zero Negbin distribution, which is specified by  $f_{|C_t^b|}(|c_t| | C_t^b \neq 0, \mathcal{F}_{t-1})$  or they can stem from degenerated distributions with a probability  $\gamma_{0|k}^b + \gamma_{1|k}^b \cdot \mathbb{1}\{g_h \in B_h\}$  for every  $k \in C$ . We decide to use the time-varying specification of the inflation component  $\gamma_{0|k}^b + \gamma_{1|k}^b \cdot \mathbb{1}\{g_h \in B_h\}$ , because after a round value of the spread at time  $t - 1$  the round quote changes at time  $t$  appear relatively more frequent (see Figure 4). Inflating the marginal processes as presented in equation (11) yields additional flexibility, because it allows us to inflate round outcomes of the two marginal processes separately. Moreover, inflating the marginal processes enables a more parsimonious specification by reducing the number of multivariate inflation parameters  $\omega_{i_1, i_2 | B_h}$ . However, inflating the marginal densities solely turns out to insufficient.

In order to provide a more parsimonious specification of the inflated MICH model, we impose in Table 2 the following parameter restrictions for the bivariate density  $\omega_{i_1, i_2 | B_h}$  and for the marginal densities  $\gamma_{0|k}^b$  and  $\gamma_{1|k}^b$ :

$$\begin{aligned}
\omega_{-5,0|5} &= \omega_{0,5|5}, \\
\omega_{-10,0|5} &= \omega_{0,10|5}, \\
\omega_{-10,0|10} &= \omega_{-5,0|10} = \omega_{0,-5|10} = \omega_{0,5|10}, \\
\omega_{5,0|10} &= \omega_{5,5|10}, \\
\omega_{0,-5|15} &= \omega_{5,0|15}, \\
\omega_{0,-10|20} &= \omega_{0,-5|20} = \omega_{5,0|20} = \omega_{10,0|20}, \\
\gamma_{0|k}^b &= \gamma_{0|k}^a, \\
\gamma_{1|k}^b &= \gamma_{1|k}^a,
\end{aligned}$$

**Table 2:** Parameter restrictions for  $k \in \{5, 10, 15, 20\}$ .

## 4.1 Long-Run Comovement

We account for the comovement of the bid and ask quotes, by including an error correction type mechanism into the inflated MICH specification. The intuition behind such a solution stems from the usual error correction reasoning for linear models, where

$$\hat{\epsilon}_t = Q_t^a - \hat{\gamma}_1 Q_t^b - \hat{\gamma}_0 \quad (12)$$

is the estimated deviation from the long-run relationship between the bid and ask quotes. A potential asymmetric adjustment to the equilibrium is captured by including the generated regressors:  $Z_{t-1,1} = |\hat{\epsilon}_{t-1}| \cdot \mathbf{1}\{\hat{\epsilon}_{t-1} > 0\}$  and  $Z_{t-1,2} = |\hat{\epsilon}_{t-1}| \cdot \mathbf{1}\{\hat{\epsilon}_{t-1} < 0\}$  into the ACM submodel for the direction of bid and ask quote changes (see equation (3)). For the ask quote process for example, we would await a negative (positive) impact of  $Z_{t-1,1}$  on the conditional probability of an upward (downward) price movement. A symmetrical influence of this regressor is expected for the bid quote. Thus, if during the last period the difference between the bid and ask quotes has been higher than the “equilibrium value”, the expected positive impact of the variable  $Z_{t-1,1}$  on the upward probability of the bid quote would narrow the current bid-ask spread. For  $Z_{t-1,2}$  we expect the opposite effect.

## 4.2 Copula Function

With the copula approach we can model the joint bid and ask quote change process in a flexible way without restricting the direction of the dependency. Moreover, truncations on the support of the joint density caused by the non-negativity of the bid-ask spread can easily be accounted for. More precisely, for the quote changes from  $t - 1$  to  $t$ , we need to ensure that:

$$C_t^a - C_t^b > Q_{t-1}^b - Q_{t-1}^a, \quad (13)$$

holds. We model the joint density  $f(c_t^b, c_t^a | \mathcal{F}_{t-1})$  with a truncated bivariate copula, based on the Joe-Clayton copula:

$$\tilde{C}(u, v; \tau_U, \tau_L) = 1 - (1 - \{[1 - (1 - u)^{\eta_U}]^{-\eta_L} + [1 - (1 - v)^{\eta_U}]^{-\eta_L} - 1\}^{\frac{-1}{\eta_L}})^{\frac{1}{\eta_U}} \quad (14)$$

with  $\eta_U = \frac{1}{\log_2(2-\tau_U)}$  and  $\eta_L = \frac{-1}{\log_2(\tau_L)}$ .  $u = F(c_t^b|\mathcal{F}_{t-1})$ ,  $v = F(c_t^a|\mathcal{F}_{t-1})$  and  $\tau_U, \tau_L \in (0, 1)$  denote the upper and lower tail dependency parameters of the Joe-Clayton copula. The Joe-Clayton copula is attractive for two reasons. First, contrary to elliptical copulas it has a simple form that does not involve double integrals and is therefore easy to estimate. Second, it allows for different upper and lower tail dependency. In our setup, upper (lower) tail dependency is a measure of a probability, that we observe an extreme large upward (downward) movement of the bid quote simultaneously with an the extreme large upward (downward) movement of the ask quote. The truncated Joe-Clayton copula accounting for the non-negativity restrictions takes the form

$$\mathbf{C}(u, v; \tau_U, \tau_L | C_t^a - C_t^b < Q_{t-1}^a - Q_{t-1}^b) = \frac{\tilde{\mathbf{C}}(u, v; \tau_U, \tau_L)}{\mathbb{P}(C_t^a - C_t^b < Q_{t-1}^a - Q_{t-1}^b)}. \quad (15)$$

Note, that as defined in equation (8),  $\mathbf{C}(\cdot)$  is by construction a conditional on  $\mathcal{F}_{t-1}$  copula, implying a time varying dependency.

Due to the specification of the models which involve inflation points and finite difference approximations of the copula derivatives, it is necessary to perform a one step Maximum Likelihood estimation. The copula function is given by equation (14) and the distribution functions  $F(c_t^b|\mathcal{F}_{t-1})$  (and analogously  $F(c_t^a|\mathcal{F}_{t-1})$ ) can be derived in the following way

$$F(c_t^b|\mathcal{F}_{t-1}) = \sum_{k=-50}^{c_t^b} \pi_{-1t}^b \mathbb{1}_{\{k<0\}} \cdot \pi_{0t}^b \mathbb{1}_{\{k=0\}} \cdot \pi_{1t}^b \mathbb{1}_{\{k>0\}} \cdot f_{|k|}(|k| | k \neq 0, \mathcal{F}_{t-1})^{(1-\mathbb{1}_{\{k=0\}})}$$

where we set the lower bound of the summation to  $-50$ . The probabilities of the downward, zero and upward movement of the bid or ask quote are specified with the logistic link function, as given in equation (2) while the density for the absolute value of the quote change is specified as a conditional NegBin distribution, as shown in equation (11).

### 4.3 Estimation and Simulation Results

The estimation results for the standard and the inflated MICH model are presented in Table 3. The Schwarz Information Criterion and the Wald tests reveal that the inflated MICH model captures the data generating process significantly better than the standard MICH model. Both the error correction parameters as well as the conditional inflation parameters turn out to be jointly highly significant.  $\omega_{0,0|5}$ , for example, shows that given a bid-ask spread equal to 5 ticks at time  $t-1$  the  $(0, 0)$  outcome is additionally, to the  $(0, 0)$  outcome being generated by the standard MICH model, inflated by 18%. Given a spread equal to 20 ticks the number of additional  $(0,0)$  outcomes rises even to more than 34 % (see  $\omega_{0,0|20}$ ).

Regarding the estimates for the ACM submodels of both specifications, we observe a significant persistency pattern ( $\hat{B}_1$  matrix) of the direction processes and we can conclude, that if the probability of an quote change has been high in the previous period, it is also expected to be high in the next period. Moreover, in the inflated MICH model we can easily confirm the existence of an error correction mechanisms. The parameters on the impact of  $Z_{t-1,1}$  and  $Z_{t-1,2}$  on the probability, that the quote moves down (up) are denoted by  $g_{1d}$  and  $g_{2d}$  ( $g_{1u}$  and  $g_{2u}$ ) and reveal an asymmetric error correction mechanism.

In the GLARMA parts of the MICH models, the values of the dispersion parameters  $\kappa^{-0.5}$  are significantly different from zero, allowing the rejection of the null hypothesis of an at-zero-truncated Poisson distribution in favor of at-zero-truncated Negbin one. The GLARMA models are characterized by high persistency, reflected by high values of  $\beta_1$ . Jointly significant coefficients of the seasonal component  $S(\nu, \tau, 2)$  indicate also, that there exists a diurnally seasonality pattern for the absolute bid and ask quote changes. We observe also the significant impact of the direction variable  $D_t$  on the size of the absolute price change. This impact is different for the absolute bid and ask quote changes and can be treated as a quasi additional correction mechanism. For the ask quote we observe that if the quotes move up, the absolute value of that movement is larger than if it moves down. For the bid quote change the opposite is true, as the downward movements are significantly bigger than the upward ones. This seems to be a natural mechanism to limit the eventual number of negative spread outcomes being predicted by the model. In the GLARMA part of the inflated MICH model the inflation parameters are also highly significant, which confirms the observations deduced from the descriptive analysis.

|   | Standard MICH Model             |        |          |        | Inflated MICH Model             |        |          |        |
|---|---------------------------------|--------|----------|--------|---------------------------------|--------|----------|--------|
| parameter   | ASK                             |        | BID      |        | ASK                             |        | BID      |        |
|   | estimate                        | std.   | estimate | std.   | estimate                        | std.   | estimate | std.   |
|   | ACM Model Parameters            |        |          |        | ACM Model Parameters            |        |          |        |
| $\mu_1$   | -0.1308                         | 0.0484 | -0.1339  | 0.0468 | 0.0267                          | 0.0226 | -0.0432  | 0.0177 |
| $\mu_2$   | -0.1865                         | 0.0680 | -0.1092  | 0.0385 | -0.0946                         | 0.0400 | 0.0205   | 0.0136 |
| $b_{11}$  | 0.7527                          | 0.0885 | 0.8032   | 0.0672 | 0.7759                          | 0.0661 | 0.8913   | 0.0290 |
| $a_{11}$  | 0.0189                          | 0.0304 | 0.0929   | 0.0328 | 0.0141                          | 0.0336 | 0.0659   | 0.0277 |
| $a_{12}$  | 0.1136                          | 0.0336 | 0.1406   | 0.0378 | 0.0910                          | 0.0343 | -0.0078  | 0.0284 |
| $a_{21}$  | 0.0894                          | 0.0353 | 0.0350   | 0.0288 | -0.0941                         | 0.0406 | -0.0211  | 0.0267 |
| $a_{22}$  | 0.1190                          | 0.0373 | 0.1066   | 0.0314 | 0.1543                          | 0.0420 | 0.1424   | 0.0304 |
|   | ACM Error Correction Parameters |        |          |        | ACM Error Correction Parameters |        |          |        |
| $g_{1d}$  |                                 |        |          |        | -1.3350                         | 0.7861 | -6.7183  | 1.2555 |
| $g_{2d}$  |                                 |        |          |        | -7.5018                         | 1.6373 | 6.2232   | 1.6305 |
| $g_{1u}$  |                                 |        |          |        | -8.2129                         | 1.1318 | 0.1917   | 0.9080 |
| $g_{2u}$  |                                 |        |          |        | 4.4870                          | 1.6312 | -13.2031 | 1.7727 |
|   | GLARMA Model Parameters         |        |          |        | GLARMA Model Parameters         |        |          |        |
| $\kappa^{0.5}$  | 0.8282                          | 0.0243 | 0.8480   | 0.0250 | 1.2385                          | 0.0694 | 1.4007   | 0.0928 |
| $\tilde{\mu}$   | 0.0312                          | 0.0067 | 0.0317   | 0.0061 | 0.0693                          | 0.0235 | 0.0293   | 0.0106 |
| $\beta_1$   | 0.9869                          | 0.0033 | 0.9897   | 0.0025 | 0.9679                          | 0.0116 | 0.9902   | 0.0040 |
| $\alpha_1$  | 0.0261                          | 0.0045 | 0.0211   | 0.0042 | 0.0905                          | 0.0203 | 0.0343   | 0.0095 |
| $\nu_0$   | 0.0155                          | 0.0097 | -0.0020  | 0.0077 | 0.0417                          | 0.0297 | -0.0024  | 0.0155 |
| $\nu_1$   | 0.0041                          | 0.0013 | 0.0036   | 0.0011 | 0.0148                          | 0.0061 | 0.0054   | 0.0023 |
| $\nu_2$   | -0.0006                         | 0.0010 | -0.0004  | 0.0008 | 0.0028                          | 0.0034 | -0.0003  | 0.0015 |
| $\nu_3$   | 0.0008                          | 0.0030 | -0.0046  | 0.0024 | 0.0024                          | 0.0091 | -0.0070  | 0.0048 |
| $\nu_4$   | 0.0007                          | 0.0017 | -0.0014  | 0.0015 | 0.0012                          | 0.0051 | -0.0021  | 0.0029 |
| $\delta$  | 0.0631                          | 0.0247 | -0.1261  | 0.0253 | 0.0471                          | 0.0475 | -0.2354  | 0.0494 |
|   | GLARMA Inflation Parameters     |        |          |        | GLARMA Inflation Parameters     |        |          |        |
| $\gamma_{1 5}$  |                                 |        |          |        | 0.2130                          | 0.0192 | 0.2130   | 0.0192 |
| $\gamma_{1 10}$   |                                 |        |          |        | 0.0850                          | 0.0091 | 0.0850   | 0.0091 |
| $\gamma_{1 15}$   |                                 |        |          |        | 0.0972                          | 0.0137 | 0.0972   | 0.0137 |
| $\gamma_{1 20}$   |                                 |        |          |        | 0.0571                          | 0.0066 | 0.0571   | 0.0066 |
| $\gamma_{0 5}$  |                                 |        |          |        | 0.0315                          | 0.0076 | 0.0315   | 0.0076 |
| $\gamma_{0 10}$   |                                 |        |          |        | 0.0128                          | 0.0038 | 0.0128   | 0.0038 |
| $\gamma_{0 15}$   |                                 |        |          |        | 0.0087                          | 0.0059 | 0.0087   | 0.0059 |
| $\gamma_{0 20}$   |                                 |        |          |        | 0.0124                          | 0.0035 | 0.0124   | 0.0035 |
|   | Joint Inflation Parameters      |        |          |        | Joint Inflation Parameters      |        |          |        |
|   |                                 |        |          |        | estimate                        | std.   | estimate | std.   |
| $\omega_{0,0 5}$  |                                 |        |          |        | 0.1808                          |        | 0.0329   |        |
| $\omega_{0,0 10}$   |                                 |        |          |        | 0.2892                          |        | 0.0278   |        |
| $\omega_{0,0 15}$   |                                 |        |          |        | 0.2140                          |        | 0.0350   |        |
| $\omega_{0,0 20}$   |                                 |        |          |        | 0.3441                          |        | 0.0404   |        |
| $\omega_{-5,0 5}$   |                                 |        |          |        | 0.0184                          |        | 0.0096   |        |
| $\omega_{-10,0 5}$  |                                 |        |          |        | 0.0120                          |        | 0.0062   |        |
| $\omega_{-10,0 10}$   |                                 |        |          |        | 0.0191                          |        | 0.0042   |        |
| $\omega_{5,0 10}$   |                                 |        |          |        | 0.0335                          |        | 0.0068   |        |
| $\omega_{0,-5 15}$  |                                 |        |          |        | 0.0343                          |        | 0.0117   |        |
| $\omega_{0,-10 20}$   |                                 |        |          |        | 0.0396                          |        | 0.0114   |        |
| $\omega_{5,0 20}$   |                                 |        |          |        | 0.0187                          |        | 0.0089   |        |
|   | Copula Parameters               |        |          |        | Copula Parameters               |        |          |        |
|   |                                 |        |          |        | estimate                        | std.   | estimate | std.   |
| $\tau_L$  |                                 |        |          |        | 0.0535                          | 0.0201 | 0.1762   | 0.0321 |
| $\tau_U$  |                                 |        |          |        | 0.2491                          | 0.0278 | 0.3504   | 0.0264 |
| # of obs.   | 3104                            |        |          |        | 3104                            |        |          |        |
| Mean log-likelihood   | -4.8402                         |        |          |        | -4.3827                         |        |          |        |
| SIC   | 9.7736                          |        |          |        | 8.9285                          |        |          |        |
| Wald Test (Significance Error Corr. Param., df= 8)              |                                 |        |          |        | 574.19 (0.0000)                 |        |          |        |
| Wald Test (Significance Inflation Param., df= 19)               |                                 |        |          |        | 1832.3 (0.0000)                 |        |          |        |
| Wald Test (Significance Error Corr. & Inflation Param., df= 27) |                                 |        |          |        | 2590.2 (0.0000)                 |        |          |        |

**Table 3:** Maximum Likelihood estimates of the standard and the inflated MICH model.

## 5 Simulation and Model Evaluation

### 5.1 Metropolis-Hastings Algorithm

The evaluation of the goodness-of-fit of the two models in terms of the conditional joint distribution relies on a simulation strategy. We simulate the two conditional density specifications ( $N$  draws from each) at every available data point  $t$  of our sample. Application of the Metropolis-Hastings drawing algorithm allows us to derive the truncated density, that accounts for the necessary time-varying restrictions to ensure a positive bid-ask spread. Based on the simulated conditional densities at every point  $t$ , we evaluate, whether the assumed parametrization of the inflated MICH model captures the dynamical behavior of the joint bivariate conditional density of the bid and ask quote change satisfactorily. This is achieved by applying a modified version of Diebold, Gunther & Tay (1998) density forecasting test suggested by Liesenfeld et al. (2006) to each component of the decomposition of the joint bivariate density into its marginal and conditional densities. Moreover, we are able to investigate the residual series for both specifications in a straightforward way. Having simulated the bivariate density for every data point, we are able to derive a conditional as well as unconditional density, not only for the marginal processes, but also for every function of them. A natural example of such a function is the bid-ask spread. Thus, we are able to evaluate, whether the proposed specification appropriately accounts for the bid-ask spread properties: its strong positive autocorrelation and the clustering schemes, leading to a sharply multimodal distribution.

Let  $f(c_t^b, c_t^a | \mathcal{F}_{t-1})$  denote the conditional target density function of  $(C_t^b, C_t^a)$  from which we want to sample using the Metropolized Independence Sampler (MIS) of Hastings (1970). The proposed density, denoted by  $\tilde{f}(c_t^b, c_t^a | \mathcal{F}_{t-1})$  is generated from an unrestricted MICH model with the same data generating processes for the marginal processes, but governed by a simple Gaussian copula with dependency parameter (correlation coefficient)  $\rho = 0.3$ . The algorithm of the MIS can then be summarized in the following steps, for  $i = 1, \dots, N$ :

Given the current state of vector  $(C_{i,t}^b, C_{i,t}^a)$  at time  $t$ :

1. Draw a candidate  $(C_t^b, C_t^a)$  from the proposed density  $\tilde{f}(c_t^b, c_t^a | \mathcal{F}_{t-1})$  in the following way:
  - compute Cholesky decomposition  $A$  ( $2 \times 2$ ) of the correlation matrix  $R$ , where  $R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ .
  - simulate  $x = (x_1, x_2)$  from the 2-dimensional standard normal distribution,
  - set  $y = Ax$ ,

- set  $u_1 = \Phi(y_1)$  and  $u_2 = \Phi(y_2)$  where  $\Phi$  denotes the univariate standard normal distribution function,
- set  $C_t^b = F_1^{-1}(u_1|\mathcal{F}_{t-1})$  and  $C_t^a = F_2^{-1}(u_2|\mathcal{F}_{t-1})$  where  $F_1$  and  $F_2$  denotes the marginal cumulative distribution functions of the bid and ask quote changes respectively.

2. Simulate  $u$  from the Uniform[0,1] and let

- set  $(C_{i+1,t}^b, C_{i+1,t}^a) = (C_t^b, C_t^a)$  if  $u \leq \min\left(1, \frac{w(C_t^b, C_t^a)}{w(C_{i,t}^b, C_{i,t}^a)}\right)$ ,
- set  $(C_{i+1,t}^b, C_{i+1,t}^a) = (C_{i,t}^b, C_{i,t}^a)$  otherwise

where  $w(C_t^b, C_t^a) = \frac{f(c_t^b, c_t^a|\mathcal{F}_{t-1})}{f(c_{i,t}^b, c_{i,t}^a|\mathcal{F}_{t-1})}$  is the usual importance sampling weight.

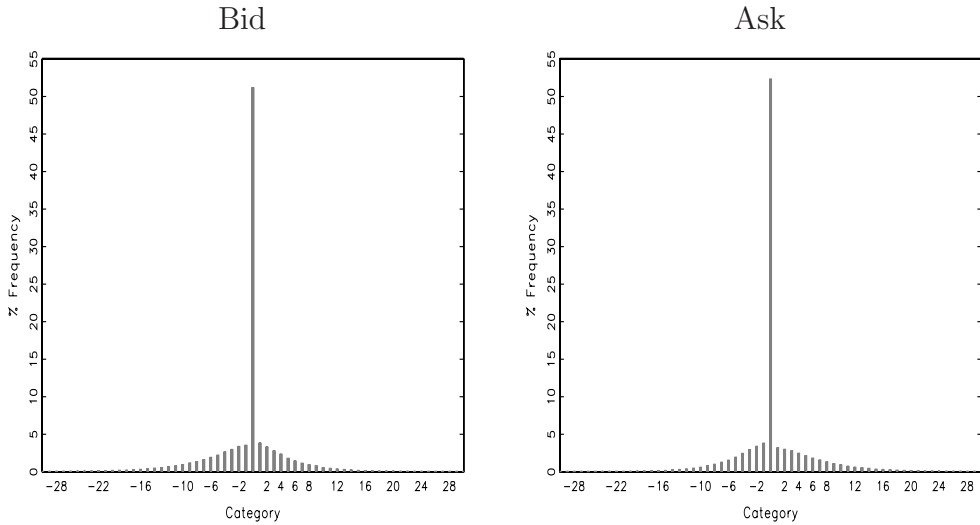
We simply conduct dependent draws from a candidate density which is non-truncated (and thus may generate negative bid-ask spreads), non-inflated, whose marginal distributions are given by  $F(c_t^b)$ ,  $F(c_t^a)$ , and a dependency being governed by the Gaussian copula function with correlation matrix  $R$ . The target density  $f(c_t^b, c_t^a|\mathcal{F}_{t-1})$ , however, accounts through the copula function (see equation (13)) for the truncations on the outcome space implied by the positive bid-ask spread. For every vector  $(C_{i,t}^b, C_{i,t}^a)$  that does not fulfil the restriction (13), the target density  $f(c_{i,t}^b, c_{i,t}^a|\mathcal{F}_{t-1})$  is equal to zero.

## 5.2 Model Evaluation

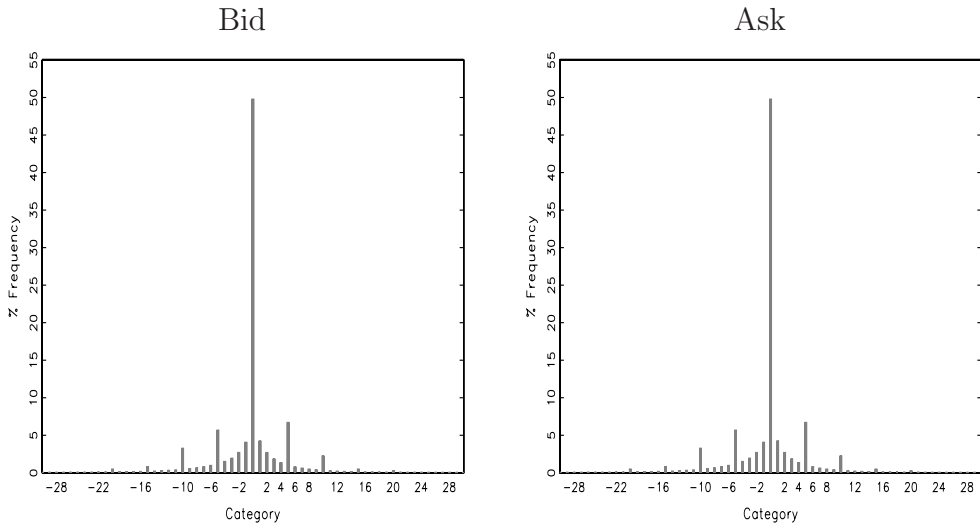
Based on the simulation results, we compare both specifications, assess the necessity of the conditional inflation, and evaluate to what extent the inflated MICH model is able to capture the discussed features of the data generating process. In Figure 10 we depict the histograms of the simulated marginal quote change processes for the standard MICH model and in Figure 11 for the inflated one. If we compare the obtained histograms with those of the raw series (see Figure 2), we easily observe that contrary to the standard MICH model, the latter specification successfully reflects the tendency to cluster at round bid and ask quotes changes.

The goodness-of-fit of the unconditional bivariate density can be studied in the light of the bivariate histograms depicted in Figures 12 and 13. There, we present the differences between the frequencies of the empirical and the simulated data points, in order to infer at which points  $(i, j)$  the estimated unconditional probabilities differ. To assess these differences graphically, we plotted absolute values of the negative differences in the left panel and positive differences in the right one. We see that the standard MICH model clearly underestimates the round outcomes of the quote changes, since in the right panel peaks of around 2 percent at multiples of 5 ticks are clearly visible. Moreover, since the MICH model tries to account for these inflated outcomes by adjusting and distorting the naturally smooth shape of the GLARMA density, it overestimates

the neighbor states of the inflated states clearly (for example  $(0,4)$ ,  $(-6,0)$ - see left panel of the Figure 12).



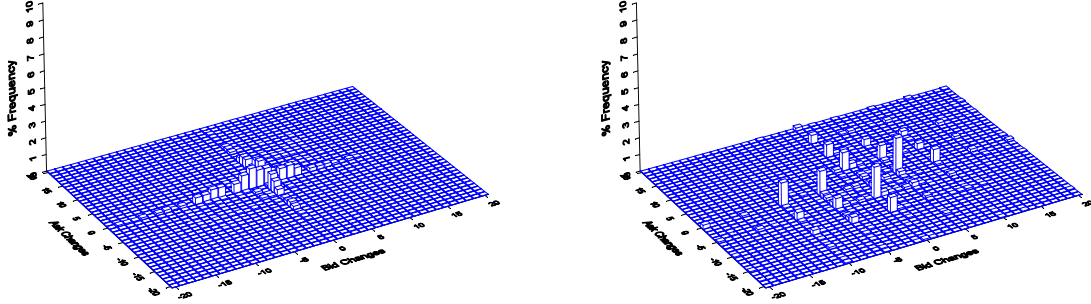
**Figure 10:** Histogram of simulated bid and ask quote changes in number of ticks for the standard MICH model.



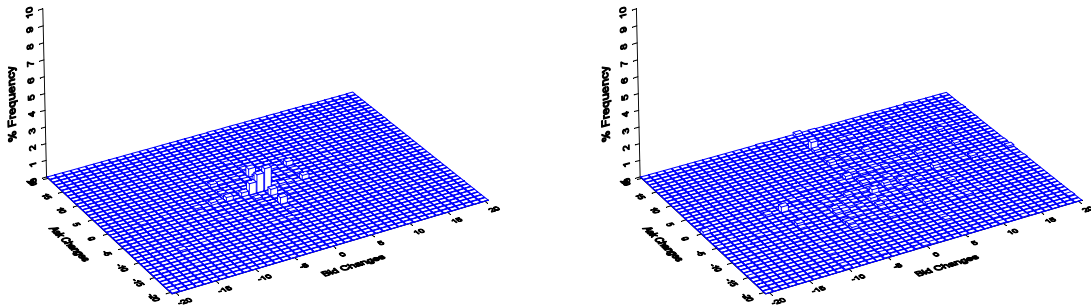
**Figure 11:** Histogram of simulated bid and ask quote changes in number of ticks for the inflated MICH model.

The application of the inflated MICH model improves the overall goodness-of-fit considerably. In the left panel of Figure 13, we see that we strongly limited or even entirely eliminated the picks induced by the simultaneous quote change clustering. The surface of the histogram for positive errors becomes much more plane, but the goodness-of-fit is still not perfect, since we overestimate the  $(0, 0)$  outcome of the unconditional density by about 1.3 percent.



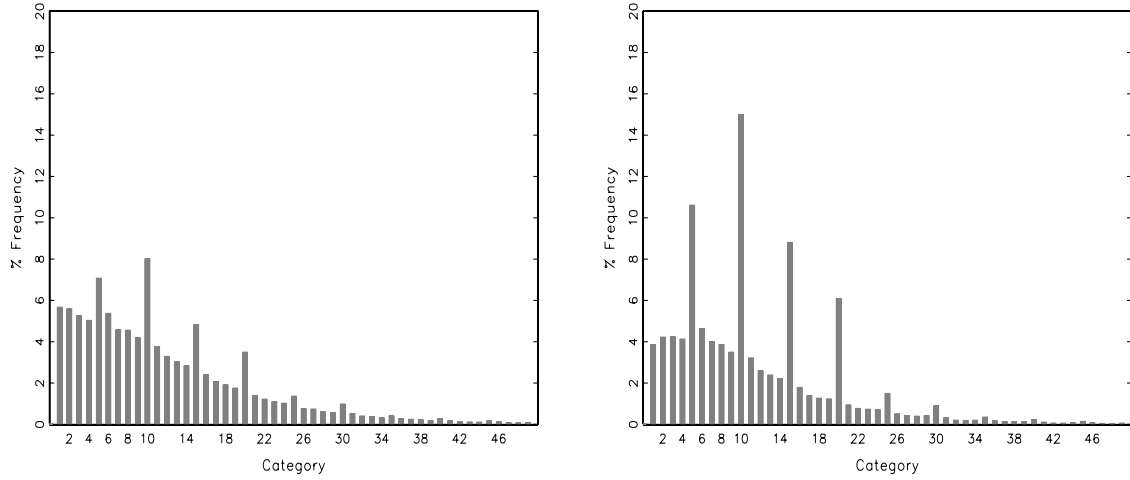


**Figure 12:** Bivariate histogram of the positive (left panel) and the negative (right panel) differences between the empirical and the simulated bivariate histogram of the quote changes for the standard MICH model.



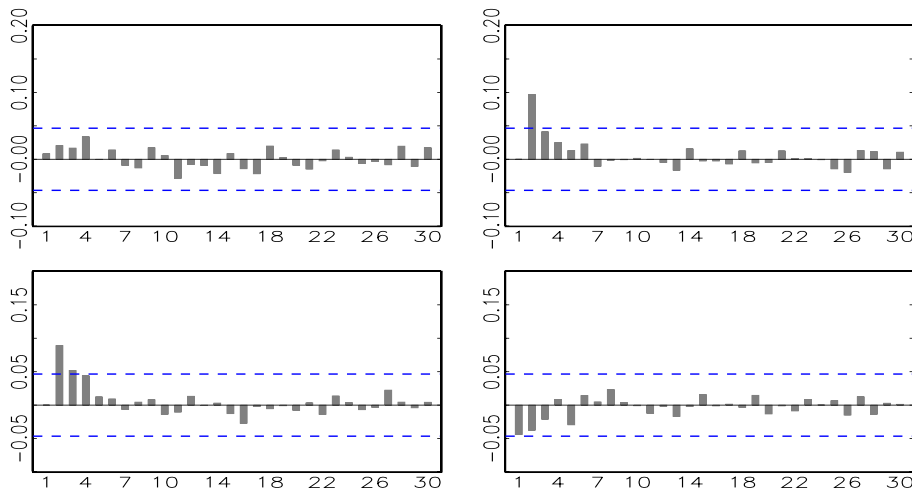
**Figure 13:** Bivariate histogram of the positive (left panel) and the negative (right panel) differences between the empirical and the simulated bivariate histogram of the quote changes for the inflated MICH model.

Given the simulated conditional densities for the bid and quote changes we are able to derive the conditional density of the bid-ask spread at every time point  $t$  as well. The histograms of the simulated bid-ask spread process are depicted in Figure 14. We see that contrary to the basic specification, the inflated MICH model mirrors the spread clustering scheme quite successfully. Bid-ask spread values stick to multiples of 5 ticks, and the shape of the frequency bar is very near to the histogram of the empirical bid-ask spread series (compare Figure 3). However, the unconditional probability for a very narrow spread values (1 and 2 ticks) are under-predicted by our model.

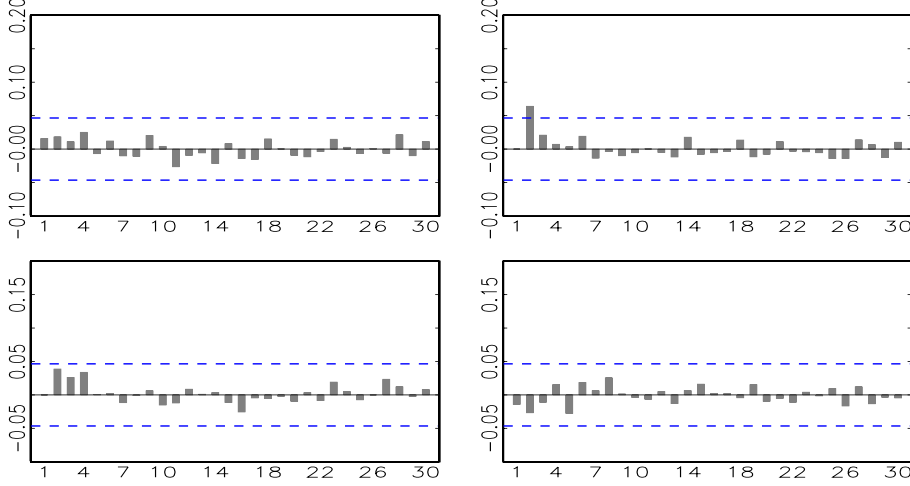


**Figure 14:** Histogram of the simulated bid-ask spread for the standard (left panel) and the inflated (right panel) MICH model specification.

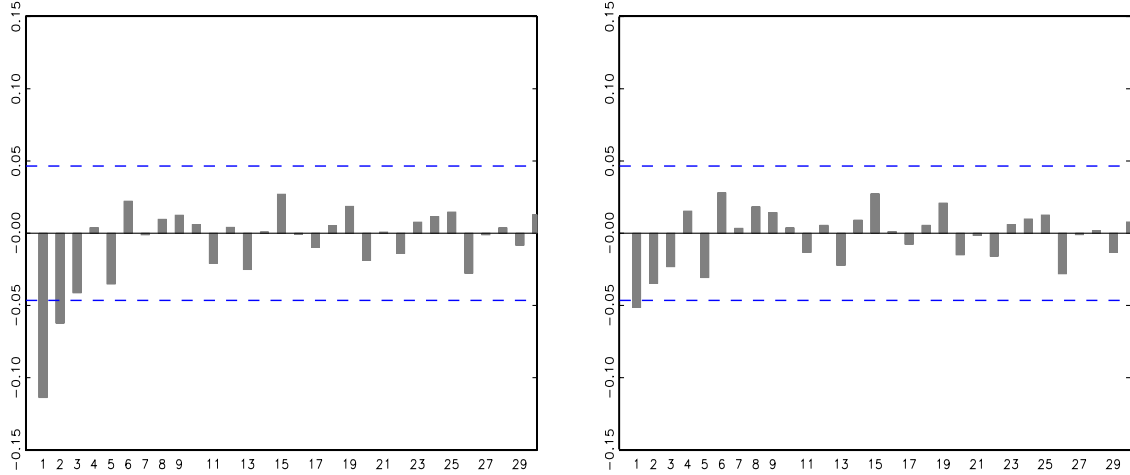
For finance applications the goodness-of-fit of the unconditional bivariate distribution is of minor importance. More interesting is the fit of the model with respect to the conditional density and its implied dynamics. First, we check the bivariate autocorrelation function of the residual series for the simulated marginal processes,  $\hat{\varepsilon}_t = (\hat{\varepsilon}_t^b, \hat{\varepsilon}_t^a)'$  depicted in Figures 15 and 16. We can see, that in contrary to the standard MICH model, the inflated specification almost entirely eliminates the strong cross correlation pattern that has been a leading feature of the bivariate quote change processes (see Figure 7). The adequacy of the dynamic specification can be checked by plotting the autocorrelation function of the bid-ask spread residuals. As shown in Figure 17, owing to the inflated model specification, we totally eliminate the significant positive autocorrelation that characterized the raw spread data series (compare Figure 9).



**Figure 15:** Multivariate autocorrelation function for residuals of bid and ask quote changes for the standard MICH model. Upper left panel:  $\text{corr}(\varepsilon_t^a, \varepsilon_{t-l}^a)$ ; upper right panel:  $\text{corr}(\varepsilon_t^a, \varepsilon_{t-l}^b)$ ; lower left panel:  $\text{corr}(\varepsilon_{t-l}^b, \varepsilon_t^a)$  and lower right panel:  $\text{corr}(\varepsilon_t^b, \varepsilon_{t-l}^b)$ . The dashed lines mark the approximate 99% confidence interval  $\pm 2.58/\sqrt{T}$ .



**Figure 16:** Multivariate autocorrelation function for residuals of bid and ask quote changes for the inflated MICH model. Upper left panel:  $\text{corr}(\varepsilon_t^a, \varepsilon_{t-l}^a)$ ; upper right panel:  $\text{corr}(\varepsilon_t^b, \varepsilon_{t-l}^b)$ ; lower left panel:  $\text{corr}(\varepsilon_{t-l}^b, \varepsilon_t^a)$  and lower right panel:  $\text{corr}(\varepsilon_t^b, \varepsilon_{t-l}^a)$ . The dashed lines mark the approximate 99% confidence interval  $\pm 2.58/\sqrt{T}$ .



**Figure 17:** Autocorrelation function for the residuals of the bid-ask spread for the standard (left panel) and the inflated (right panel) MICH model specification. The dashed lines mark the approximate 99% confidence interval  $\pm 2.58/\sqrt{T}$ .

We can also investigate the fit of the conditional densities through the application of integral density forecasting (IDF) testing procedures. IDF can be perceived as a way of deriving i.i.d  $U(0,1)$  distributed variables from any sample of variables with a known distribution. Since the joint density of  $(C_t^b, C_t^a)$  can be decomposed into a marginal and conditional component as:  $f(c_t^b, c_t^a) = f(c_t^b)f(c_t^a|c_t^b)$  or as  $f(c_t^b, c_t^a) = f(c_t^a)f(c_t^b|c_t^a)$ , the corresponding density forecast transformations of the vector  $(C_t^b, C_t^a)$  can be defined for the first decomposition as

$$\begin{aligned} u_1(c_t^b) &= P(C_t^b \leq c_t^b) = F(c_t^b), \\ u_2(c_t^a|c_t^b) &= P(C_t^a \leq c_t^a | C_t^b = c_t^b) = F(c_t^a|c_t^b), \end{aligned} \tag{16}$$

and analogously for the second decomposition. In a continuous variable setting, under the null hypothesis of a perfect fit (i.e. the correct bivariate density for the data generating process),  $u_1$  and  $u_2$  should be both uniformly and independently distributed on the set  $[0, 1]$ . As we operate in a discrete variable framework, the variables  $u_1$  and  $u_2$ , obtained by equation (16) are step functions, so they are obviously not continuously distributed on  $[0, 1]$ . We, therefore use the modification by Liesenfeld et al. (2006) for a discrete density setup.

Relying on  $N$  simulated vectors  $(C_{i,t}^b, C_{i,t}^a)$ ,  $i = 1, \dots, N$  which correspond to the simulated bid and ask quote changes at time point  $t$ , we can define the variables  $u_b^l$ ,  $u_b^u$ , and  $u_{a|b}^l$ ,  $u_{a|b}^u$  (analogously for the second decomposition) as:

$$u_b^l(C_t^b) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}\{C_{i,t}^b \leq c_t^b - 1\} \quad (17)$$

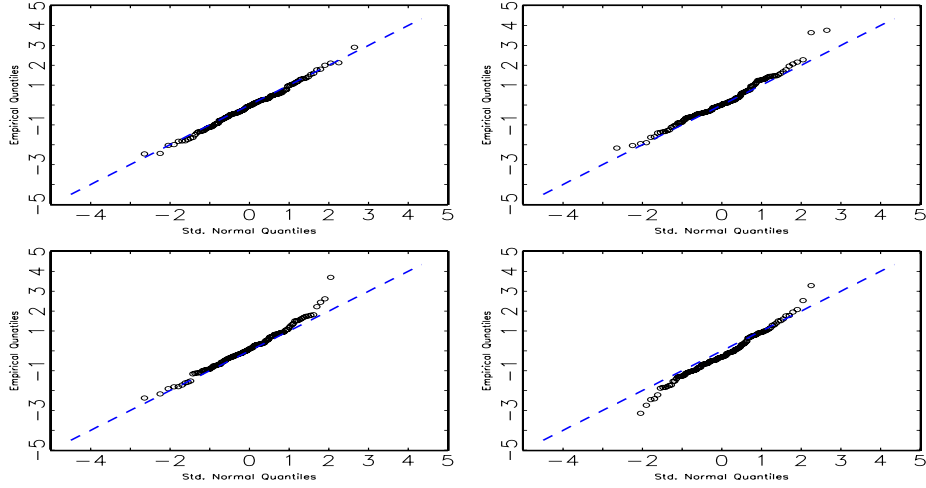
$$u_{a|b}^l(C_t^a) = \frac{1}{\sum_{i=1}^N \mathbb{1}\{C_{i,t}^b = c_t^b\}} \sum_{i=1}^N \mathbb{1}\{C_{i,t}^a \leq c_t^a - 1\} \cdot \mathbb{1}\{C_{i,t}^b = c_t^b\}$$

$$u_b^u(C_t^b) = \frac{1}{N} \sum_{i=1}^N \mathbb{1}\{C_{i,t}^b \leq c_t^b\} \quad (18)$$

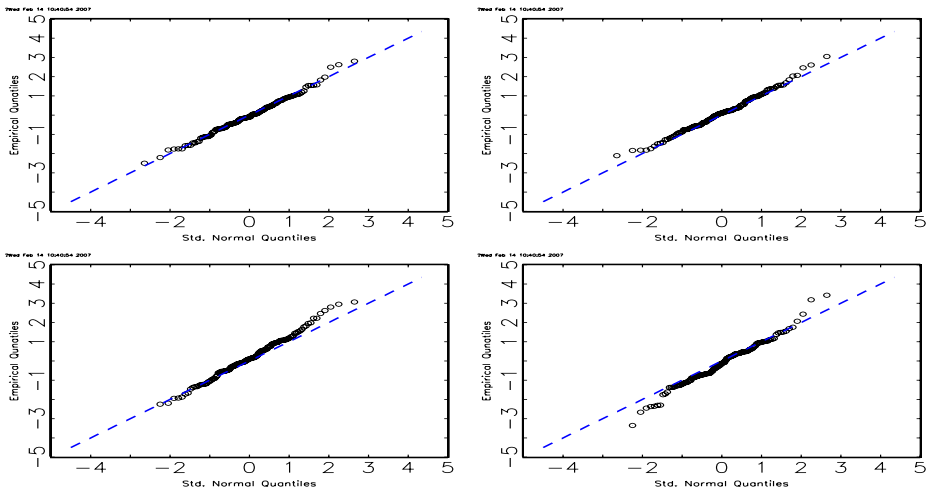
$$u_{a|b}^u(C_t^a) = \frac{1}{\sum_{i=1}^N \mathbb{1}\{C_{i,t}^b = c_t^b\}} \sum_{i=1}^N \mathbb{1}\{C_{i,t}^a \leq c_t^a\} \cdot \mathbb{1}\{C_{i,t}^b = c_t^b\}$$

Note, that the upper bound variables  $u_b^u$ ,  $u_{a|b}^u$  are derived according to equation (16), respectively as  $F(c_t^b)$ ,  $F(c_t^a|c_t^b)$ , whereas the lower bound variables  $u_b^l$ ,  $u_{a|b}^l$  as  $F(c_t^b - 1)$ ,  $F(c_t^a - 1|c_t^b)$ , respectively. If the model is correctly specified, the random variables  $u_t^b$  and  $u_t^{a|b}$ , drawn from the uniform distributions as  $u_t^b \sim U(u_b^l, u_b^u)$  and  $u_t^{a|b} \sim U(u_{a|b}^l, u_{a|b}^u)$  should be uniformly and independently distributed on  $[0, 1]$ . A similar statement can be made for the variables  $u_t^a$ ,  $u_t^{b|a}$  stemming from the second decomposition.

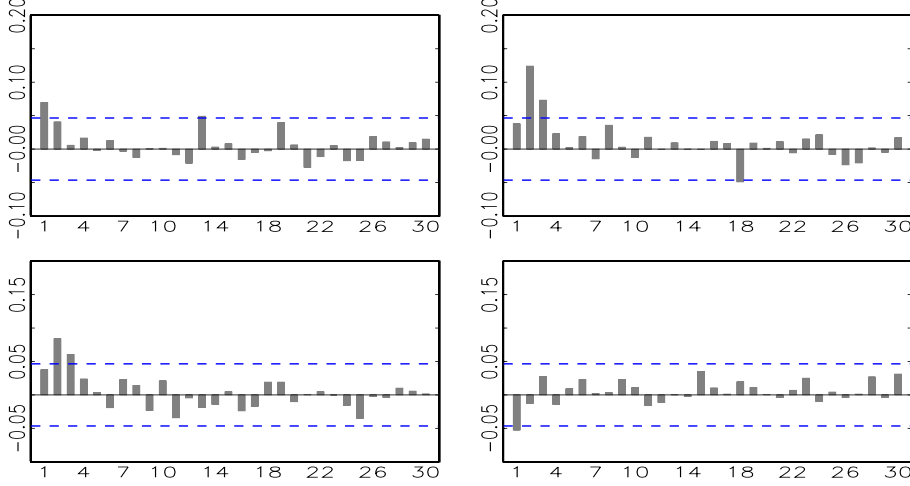
We have plotted the quantile–quantile (QQ) plots of the normalized variables  $u_t^b$ ,  $u_t^{a|b}$ ,  $u_t^a$  and  $u_t^{b|a}$  of the standard MICH model in Figure 18 and of the inflated MICH model in Figure 19. We have mapped these modified density forecast variables into a standard normal distribution so that under the correct model specification they should be i.i.d. standard normally distributed. We see that the deviations from normality are quite striking especially with respect to the upper and lower tails of the standard MICH specification. The inflated MICH model leads to some improvements especially with respect to the specifications of the marginal distributions, which can be seen in the upper panel of the Figure 19. If the joint density specification is correct, the normalized density forecast variables should be also independent. In order to test this aspect, we plot the multivariate cross correlation function for vectors  $(u_t^a, u_t^{b|a})$  (the figures for  $(u_t^b, u_t^{a|b})$  lead to the same conclusion and are therefore omitted) in Figures 20 to 21. Here, the improvements achieved by the inflated MICH model are also clearly visible.



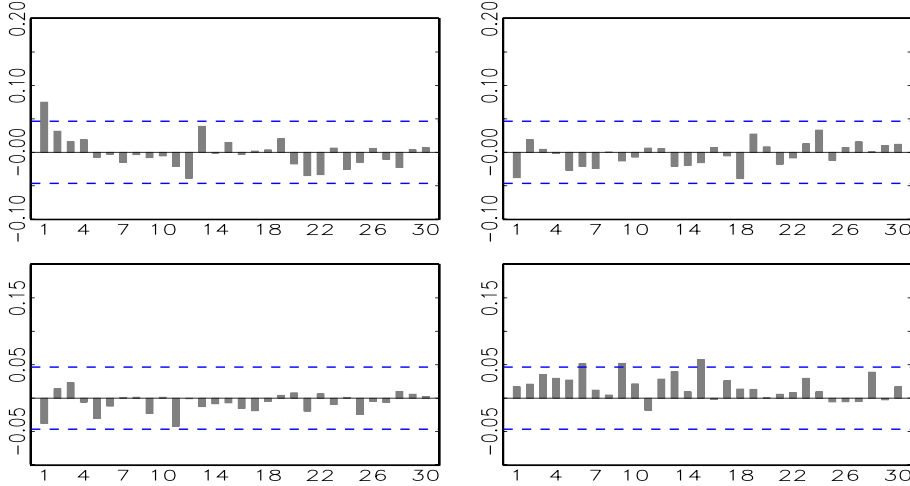
**Figure 18:** QQ plots of the normalized density forecast variables for the standard MICH model. Upper left panel:  $u_t^a$ ; upper right panel:  $u_t^b$ ; lower left panel:  $u_t^{a|b}$  and lower right panel:  $u_t^{b|a}$ .



**Figure 19:** QQ plots of the normalized density forecast variables for the inflated MICH model. Upper left panel:  $u_t^a$ ; upper right panel:  $u_t^b$ ; lower left panel:  $u_t^{a|b}$  and lower right panel:  $u_t^{b|a}$ .



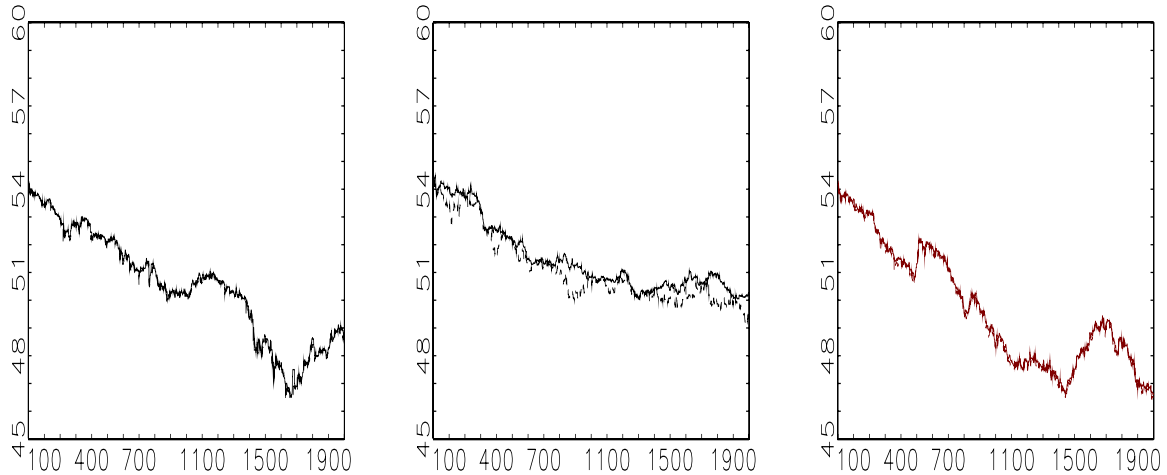
**Figure 20:** Multivariate autocorrelation function for normalized density forecast variables of the standard MICH model. Upper left panel:  $\text{corr}(u_t^b, u_{t-1}^b)$ ; upper right panel:  $\text{corr}(u_t^b, u_{t-1}^{a|b})$ ; lower left panel:  $\text{corr}(u_{t-1}^b, u_t^{a|b})$  and lower right panel:  $\text{corr}(u_t^{a|b}, u_{t-1}^{a|b})$ . The dashed lines mark the approximate 99% confidence interval  $\pm 2.58/\sqrt{T}$ .



**Figure 21:** Multivariate autocorrelation function for normalized density forecast variables of the inflated MICH model. Upper left panel:  $\text{corr}(u_t^b, u_{t-1}^b)$ ; upper right panel:  $\text{corr}(u_t^b, u_{t-1}^{a|b})$ ; lower left panel:  $\text{corr}(u_{t-1}^b, u_t^{a|b})$  and lower right panel:  $\text{corr}(u_t^{a|b}, u_{t-1}^{a|b})$ . The dashed lines mark the approximate 99% confidence interval  $\pm 2.58/\sqrt{T}$ .

Finally, we check the validity of the error correction mechanism. As before we start from simulating a sample of  $N$  possible realization of the bivariate quote processes for a given time point  $t$ . Then, we randomly choose one realization of the simulated vectors,  $(C_t^b, C_t^a)$  and given this outcome we recursively recalculate the conditional mean of both marginal processes and the bid and ask quote levels at time  $t$ . Relying on the implied value for the bid-ask spread we simulate the conditional density for time  $t + 1$ . We perform 2000 iterations of such a recursive simulation strategy. Possible trajectories of the bid and ask quote levels derived by such a simulation procedure are presented in Figure 22. In order to compare the achieved results with the real

data we also present the empirical time series of the bid and ask quote levels for the first 2000 observations of our sample in Figure 22. We observe, that in particular the omission of the error correction variables results in an unnaturally wide bid-ask spread induced by the standard MICH model (Figure 22, left panel). On the contrary, with the inflated MICH approach containing an error-correction mechanism, both time series move together and ensure a stationary bid-ask spread process.



**Figure 22:** True bid and ask quote levels (left panel) as well as simulated trajectory of the bid and ask quote levels for the standard (middle panel) and the inflated (right panel) MICH model.

## 6 Conclusion

In this paper we propose a modelling approach for complex multivariate processes of discrete random variables. Combining the approach by Cameron et al. (2004) with the ICH model by Liesenfeld et al. (2006) and allowing for inflated outcomes, truncation regions and error correction mechanisms, we model the joint process of discrete bid and ask quote changes in a high frequency setup.

The discrete bivariate quote change process is characterized by the following properties: i) clustering at round events (multiples of 5 ticks), ii) strong instantaneous dependence and iii) co-movement of both components as well as iv) truncations, since the implied bid-ask spread has to be positive.

We show that even without an extensive model selection procedure in comparison to the standard MICH model its inflated specification explains the underlying bivariate data generating process in terms of the dynamics and the density specification very well. We use the conditional bivariate density of bid and ask quote changes in order to derive the conditional density of the bid-ask spread, which can serve as a proxy for market liquidity. The data generating process of the bid-ask spread is also successfully captured by the proposed specification.

The potential merits of our approach should be checked in the light of real world applications such as the measurement of multivariate conditional volatilities and the

quantification of liquidity or value-at-risk applications. Obviously, our approach can easily be extended to the most general case of mixed multivariate distributions for continuous and discrete random variables. In the light of market microstructure studies, such an extension is particularly attractive, as it enables the joint analysis of several marks of the trading process (volumes, price and volume durations, discrete quote changes etc.).

## References

- AMILON, H. (2003): "GARCH estimation and discrete stock prices: an application to low-priced Australian stocks," *Economics Letters*, 81 (2), 215–222.
- ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, & P. LABYS (1999): "(Understanding, Optimizing, Using and Forecasting) Realized Volatility and Correlation," New York University, Leonard N. Stern School Finance Department Working Paper 99-061.
- ANTONIOU, A. & C. E. VORLOW (2005): "Price Clustering and Discreteness: Is there Chaos behind the Noise?" *Physica A*, 348.
- BALL, C. (1988): "Estimation Bias Induced by Discrete Security Prices," *Journal of Finance*, 43, 841–865.
- BROCK, W. A., W. D. DECHERT, J. A. SCHEINKMAN, & B. LEBARON (1996): "A Test for Independence Based on the Correlation Dimension," *Econometric Reviews*, 15 (3), 197–235.
- CAMERON, C., T. LI, P. TRIVEDI, & D. ZIMMER (2004): "Modelling the Differences in Counted Outcomes Using Bivariate Copula Models with Application to Mismeasured Counts," *Econometrics Journal*, 7, 566–584.
- CHEUNG, Y. (2002): "Zero-inflated models for regression analysis of count data: a study of growth and development." *Statistics in Medicine*, 21, 1461–1469.
- CHRISTIE, W. G. & P. SCHULTZ (1994): "Why Do NASDAQ Market Makers Avoid Odd-Eighth Quotes?" *Journal of Finance*, 49, 1813–1840.
- CHRISTIE, W. G., J. HARRIS, & P. SCHULTZ (1994): "Why Did NASDAQ Market Makers Stop Avoiding Odd-Eighth Quotes?" *Journal of Finance*, 49, 1841–1860.
- CRACK, T. F. & O. LEDOIT (1996): "Robust Structure without Predictability: The "Compass Rose" Pattern of the Stock Market," *Journal of Finance*, 51 (2), 751–62.
- DIEBOLD, F. X., T. A. GUNTHER, & A. S. TAY (1998): "Evaluating Density Forecasts, with Applications to Financial Risk Management," *International Economic Review*, 39, 863–883.



- DIXON, M. J. & S. COLES (1997): “Modelling association football scores and inefficiencies in the football betting market,” *Appl. Statist.*, 46, 265–280.
- FANG, Y. (2002): “The Compass Rose and Random Walk Tests,” *Computational Statistics & Data Analysis*, 39, 299–310.
- GLEASON, K. C., C. I. LEE, & I. MATHUR (2000): “An explanation for the compass rose pattern,” *Economics Letters*, 68 (2), 127–133.
- HANSEN, P. R. & A. LUNDE (2006): “Realized Variance and Market Microstructure Noise,” *Journal of Business and Economic Statistics*, 24, 127–218.
- HARRIS, L. E. (1991): “Stock Price Clustering and Discreteness,” *Review of Financial Studies*, 4, 389–415.
- HASBROUCK, J. (1999): “Security Bid/Ask Dynamics with Discreteness and Clustering: Simple Strategies for Modelling and Estimation,” *Journal of Financial Markets*, 2 (1), 1–28.
- HASTINGS, W. K. (1970): “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika*, 57, 97–109.
- HUANG, R. D. & H. R. STOLL (1994): “Market Microstructure and Stock Return Predictions,” *Review of Financial Studies*, 7 (1), 179–213.
- JANG, T. Y. (2005): “Count Data Models for Trip Generation,” *Journal of Transportation Engineering*, 131, 444–450.
- JOHNSON, N., S. KOTZ, & N. BALAKRISHNAN (1997): *Discrete Multivariate Distributions*, Wiley, New York.
- KARLIS, D. & I. NTZOUFRAS (2003): “Analysis of sports data by using bivariate Poisson models,” *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52, 381–393.
- (2005): “Bivariate Poisson and Diagonal Inflated Bivariate Poisson Regression Models in R,” *Journal of Statistical Software*, 14, 1–36.
- KOCHERLAKOTA & KOCHERLAKOTA (1992): *Bivariate Discrete Distributions*, Dekker, New York.
- KRÄMER, W. & R. RUNDE (1997): “Chaos and the compass rose,” *Economics Letters*, 54 (2), 113–118.
- LAMBERT, D. (1986): “Zero-Inflated Poisson Regression with an Application to Defects in Manufacturing,” *Technometrics*, 34, 1–14.

- LEE, C. I., K. C. GLEASON, & I. MATHUR (1999): "A comprehensive examination of the compass rose pattern in futures markets," *The Journal of Futures Markets*, 19 (5), 541–564.
- LI, C., J. LU, J. PARK, K. KIM, & J. PETERSON (1999): "Multivariate Zero-Inflated Poisson Models and Their Applications," *Technometrics*, 41, 29–38.
- LIESENFELD, R., I. NOLTE, & W. POHLMEIER (2006): "Modelling Financial Transaction Price Movements: A Dynamic Integer Count Data Model," *Empirical Economics*, 30, 795–825.
- MEESTER, S. & J. MACKAY (1994): "A parametric Model for Cluster Correlated Categorical Data," *Biometrics*, 50, 954–963.
- MULLAHY, J. (1986): "Specification and Testing of Some Modified Count Data Models," *Journal of Econometrics*, 33, 341–365.
- OOMEN, R. C. A. (2005): "Properties of Bias-Corrected Realized Variance under Alternative Sampling Schemes," *Journal of Financial Econometrics*, 3, 555–577.
- PATTON, A. (2001): "Modelling Time-Varying Exchange Rate Dependence Using the Conditional Copula," Discussion Paper, UCSD Dept. of Economics.
- RUSSELL, J. R. & R. F. ENGLE (2002): "Econometric Analysis of Discrete-Valued Irregularly-Spaced Financial Transactions Data," Revised Version of Discussion Paper 98-10, University of California, San Diego.
- SHEPHARD, N. (1995): "Generalized Linear Autoregressions," Working Paper, Nuffield College, Oxford.
- SKLAR, A. (1959): "Fonctions de répartition à n dimensions et leurs marges," *Public Institute of Statistics at the University of Paris*, 8, 229–231.
- SZPIRO, G. G. (1998): "Tick size, the compass rose and market nanostructure," *Journal of Banking & Finance*, 22 (12), 1559–1569.
- VORLOW, C. E. (2004): "Stock Price Clustering and Discreteness: The "Compass Rose" and Predictability," Working Paper, University of Durham.
- WAHLIN, J. (2001): "Bivariate ZIP Models," *Biometrical Journal*, 43, 147–160.
- WANG, K., A. LEE, K. YAU, & P. CARRIVICK (2003): "A bivariate zero-inflated Poisson regression model to analyze occupational injuries," *Accident Analysis and Prevention*, 35, 625–629.