The impact of order flow on the foreign exchange market: A copula approach

Yoshihiro Kitamura

Working Paper No.232
June, 2008

The impact of order flow on the foreign exchange market: A copula approach

Yoshihiro Kitamura*
Faculty of Economics, University of Toyama,
3190 Gofuku, Toyama 930-8555, JAPAN

Abstract

Using the copula function, I propose a new econometric method to measure the time varying impact of order flows on returns in foreign exchange markets and examine whether this impact is affected by market conditions such as the number of informed and liquidity traders. My results indicate that the impact of order flow decreases with both more informed trading and more liquidity trading. The former finding suggests an especially important theoretical implication that the effect of competition among informed traders tends to dominate that of the adverse selection problem faced by uniformed traders in the euro/dollar and yen/dollar markets.

JEL classification: F31; F33; G15
Keywords: Copula; Exchange rate; Order flow

---

* Tel.: +81 76 445 6475. *E-mail address*: kitamu@gmail.com
This research is supported by a Grant-in-Aid for Young Scientists (B: 19730216) from the Japan Society for the Promotion of Science.

1. Introduction

In foreign spot exchange (FX) rate economics, the role of order flow, which is defined as net buyer-initiated trade, has been stressed in recent years. Microstructure analysis, which stresses the role of order flow, explains the dynamics of FX rates more precisely than the macroeconomic approach. The theory of microstructure suggests that the better explanatory power of order flows is due to the fact that order flow signals information relevant to FX rates (Evans and Lyons, 2002a).

The relationship between order flows and FX rates is never negligible, and the relationship should be considered carefully: Berger et al. (2006) found that the variation of market sensitivity to arrival information plays an important role over time in explaining the persistence of FX return volatility. Berger et al. (2007) showed that the relationship between order flows and FX returns appears to be stronger when market liquidity is lower. These findings indicate that market sensitivity to order flow should be treated as time variant. Unless we consider the variation over time of market sensitivity, a misleading conclusion might be drawn regarding the relationship between order flows and returns. For example, assuming a time invariant relationship between a return and order flow, Boyer and Norden (2006) tested their stable long-term relationship using the data of Evans and Lyons (2002a, 2002b). They did not obtain supportive results for their null hypothesis on co-integration, and their findings do not seem to be consistent with Evans and Lyons (2002a, 2002b), who insist that order flow plays an important role in FX rate dynamics. However, as mentioned in their conclusion, this might change if we adopt a different approach, such as one that examines time *variant* relationships between order flows and returns.

Theoretically, Kyle (1985) and Admati and Pfleiderer (1988) insisted that the impact of order flows on corresponding asset prices is state dependent. Admati and Pfleiderer (1988) showed that the degree of that impact increases when there are more informed traders[1] and when the information gathered by informed traders is sufficiently imprecise. When the gathered information is relatively precise, the converse is expected (the impact is a decreasing function of informed trading). The former case is explained by the reaction of market makers[2], who are averse to trading disadvantageously with informed traders. Therefore, for uninformed (liquidity) traders, the existence of

[1] In my study, traders in the FX market are informed traders if they can access private information, which is not known by all people and produces a better forecast than public information alone (Lyons 2001, p.26).

[2] In my study, a word of "market maker" has an equivalent meaning to that of "uninformed (liquidity) trader".

informed traders amplifies an uncertainty in the market and heightens their cost of trading. Meanwhile, the latter case is the result of competition among informed traders, which, in turn, lowers the liquidity traders' cost of trading. Because competition among informed traders reveals private information rapidly, this competition sweeps away a market uncertainty. Admati and Pfleiderer (1988) also showed that the degree of this impact decreases with more liquidity traders. This is the result of competition among market makers, who undercut each other to avoid losing their trading opportunities.

The theory developed by Admati and Pfleiderer (1988) gives me testable implications: does the impact of order flow depend on market conditions, such as the number of informed and liquidity traders? These implications should be tested with a micro-data set that reflects market conditions in detail. To address this issue, I used 10-minute high frequency data sets and examined the theoretical implications proposed by Admati and Pfleiderer (1988).[3] My data set was purchased from Electronic Broking System Corporation (EBS), which is well known and represents most of the global interdealer trading activity in euro/US dollar and US dollar/yen. Moreover, my data contain transactable quotes, not indicative ones. Thus, I consider my data to be truly representative of the behavior of the global interdealer FX rate.

A notable feature of this study is to estimate the impact of order flow by using the copula function. The copula function enables me to obtain the impact of order flow, which is time *varying*. Using this time varying impact, I examined whether the sizes of informed and/or liquidity trading influence the impact of order flow on FX returns. As shown later, the selected Clayton=Gumbel copula may imply that a tail dependency exists between returns and order flows in the FX markets: the larger order flows can predict the movement in returns with more accuracy in the euro/dollar and yen/dollar markets. The one virtue of a copula is that it captures such tail dependency between random variables. Although I limited my study to the FX market, my method is easily extended to another asset markets. It is possible that the time varying impact of order flows gives us information on market "depth" (the ability of the market to absorb quantities (of order flows) without having a large effect on price; Kyle, 1985), and I believe that this information is useful for a policy-maker who wishes to understand market characteristics in detail and build adequate regularities for each market.

This paper is organized as follows. The next section summarizes my data set. In section 3, the concept of the copula is introduced, and I estimate the time varying impact of order flows. Section 4 provides the estimated result of the copula function and examines the theoretical arguments put forth by Admati and Pfleiderer (1988). The last section concludes my paper.

---

[3] The reason for the use of 10 minutes intervals is mentioned in the next section.

2. The data set

A survey released by the Bank for International Settlements (2007) reveals that euro /US dollar and yen /US dollar are heavily traded in the FX market. Therefore, I focused on these pairs. Hereafter, the euro/dollar and the yen/dollar refer to the euro and yen quoted against the United States dollar spot rate, respectively. The sample period ran from September 1, 2005, to August 31, 2006. I excluded all data from my sample that were collected from the Friday closing in the U.S.[4] to Sunday 23:30 Greenwich Mean Time (GMT) because trading activity during these hours is minimal. The data set was created in one-second time slices comprising a Price Record and a Deal Record. The Price Record lists the EBS best bid/ask prices at the end of a time slice. The Deal Record lists the highest paid and lowest given deal prices during the time slice. No information about trading volume or order type (market or limit order) is available. With a Deal Record, I constructed a trade indicator that takes the value 1 (-1) when a paid (given) rate shows up in the Deal Record. If there is no deal in the EBS system, then that indicator takes the value zero. I converted the original data set into 10-minute intervals. Bjønnes and Rime (2005), who investigated FX dealer behaviour, found that the half-life of inventory that an electronic broking dealer holds is about 7 minutes. Therefore, I am confident that my 10-minute intervals will capture an inventory effect in order flows, which enables me to split order flows into expected- and unexpected-components in the empirical part. I employed as a price S, the midpoint quote of the best bid and ask rates observed at the end of each interval. A return for each currency has been defined as $100\times\ln(S_t/S_{t-1})$. By summing up trade indicators in each interval, I calculated order flow (buyer-initiated trades minus seller-initiated trades).

Table 1 around here

Table 1 reports the summary statistics of the first and second moments of returns and order flows for each pair of currencies. For the second moment of the returns and order flows, the values of first order serial correlation show serial dependences for all the currencies, leading me to select the GARCH framework to capture the dynamics of return and order flow volatility.

It is well documented that the volatility of an FX rate return shows intraday seasonality (e.g., Ballie and Bollerslev, 1990; Ito and Hashimoto, 2006). This could have induced me to select a longer GARCH lag term because longer lags capture the spikes due to seasonality. Therefore, I removed the seasonality of volatility to reduce

[4]The timing of closing is defined in Table 5.

the number of estimated parameters. I also considered the timing of daylight saving time (DST) by America and Europe because this changes intraday seasonality in the FX market (Ito and Hashimoto, 2006). In my sample, I used four periods that correlated with the beginning or ending of DST. The first and fourth periods were 1) September 1, 2005, to October 30, 2005, and 4) April 2, 2006, to August 31, 2006, during which America and Europe adopted DST. The second one was 2) October 31, 2005, to March 25, 2006, where no DST applied, and the third one was 3) March 26, 2006, to April 1, 2006, in which only Europe adopted DST.

Figure 1 around here

The circles plotted in Figure 1 show averaged squared returns for each currency in each interval and period. We can observe the intraday U-shape pattern of averaged squared returns and the peaks corresponding to the timing of the opening and closing of the major markets in Tokyo, London and New York. To remove seasonality, I divided the sample returns by the root of five interval moving average of averaged squared returns. The bold line in Figure 1 is the moving average in each interval and period. Because I also observed seasonality in the squared order flows, which is quite similar to that of the squared returns, I removed the seasonality of the order flows using the same method adopted for squared returns.

3. Empirical methodology

3. 1 The underlying theory

Recent studies in FX rate economics stress the role of order flows, which are again defined as the net of buyer-initiated trades. Evans and Lyons (2002a) asserted that order flows convey information on aggregate portfolio shifts and, therefore, can substantially explain the dynamics of FX rates. Because order flows contain and/or signal information relevant to the FX rate, they explain the dynamics of FX rates well.

Figure 2 around here

It is reasonable to suppose that the sensitivity of the FX market to order flow is state dependent. Figure 2 shows the rolling regression parameter of order flows, which is obtained by regressing 10-minute returns onto order flows with 30-day windows. As regressors, I adopted order flows and constant term. Both returns and order flows are seasonally adjusted. The results of Fig.2 indicate that the impact of order flow is likely to be time varying in both the euro/dollar and yen/dollar cases.

Moreover, the theory developed by Admati and Pfleiderer (1988) asserts that the impact of order flow on the corresponding asset return is state dependent.

Lemma 3 (Admati and Pfleiderer, 1988): *Let* λ_t *be the impact of order flow on the corresponding asset price. Assume that* n_t *informed traders trade in period* t *and that each observes an independent signal* $\delta_t + \varepsilon_t^{\,i}$, *where* $\mathrm{var}(\delta_t) = 1$ *and* $\mathrm{var}(\varepsilon_t^i) = \phi_t$ *for all* i. *Let* Ψ_t *be the total variance of the liquidity trading in period* t. *Then*

$$\lambda_t = \frac{1}{1+n_t+2\phi_t}\sqrt{\frac{n_t(1+\phi_t)}{\Psi_t}}. \qquad (1)$$

This lemma has the two testable implications. The impact of order flow is 1) an increasing function of informed traders n_t if $\phi_t >> n_t$ or a decreasing one of n_t, and the impact is 2) decreasing with liquidity trading Ψ_t.

An increase in informed traders has two effects: The first effect is explained by the adverse selection problem faced by market makers. Because the probability that market makers will face disadvantageous deals increases with informed traders, the sensitivity of market makers to (unexpected) order flow increases with more informed traders in order to hedge against their losses, which are caused by deals with informed traders. The case $\phi_t >> n_t$ corresponds to the period when the information gathered by informed traders is sufficiently imprecise. The second effect is due to competition among informed traders, which in turn decreases a market uncertainty and lowers the cost of liquidity trading.

The effect of Ψ_t is due to competition among market makers, which forces them to set their prices to avoid losing business to competitors who undercut them.

Therefore, the lemma of Admati and Pfleiderer (1988) leads us to suppose that the relationship between order flow and an asset return is time varying. When there are more informed traders, order flow has more (less) impact on returns. Additionally, an increase in liquidity trading always decreases the impact of order flows.

Using a copula function, I estimated a time varying impact of order flow and examined whether the impact depends on market conditions. Let ρ_t be the correlation between return and order flow at t period. And let $\mathrm{var}(x_t)$ and $\mathrm{var}(y_t)$ be volatilities of return and order flow, respectively. These three variables are assumed to be time variant, and I define the impact of order flow on return as follows:

$$\Lambda_t = \text{cov}(x_t, y_t) / \text{var}(y_t)$$

$$= \rho_t \sqrt{\frac{\text{var}(x_t)}{\text{var}(y_t)}}, \qquad (2)$$

where $\rho_t = \text{cov}(x_t, y_t) / \sqrt{\text{var}(x_t)\text{var}(y_t)}$.

Equation (2) is drawn from the linear projection of x_t on y_t ($E[x_t|y_t]$=cov(x_t, y_t) var$(y_t)^{-1}y_t$; Hamilton 1994, p.94). As explained in the following subsection, the time varying parameters in eq. (2) are estimated using the copula function. Equation (2) is equivalent to the theoretical frameworks of Kyle (1985) and Admati and Pfleiderer (1988), in the sense that the pricing rule adopted by market makers in these studies is based on the linear projection of their price on observed order flows.

Some extant literature has indicated the time varying impact of order flow on FX returns. Payne (2003) divided his sample into some intra-day intervals and concluded that the impact of order flow decreases in proportion to the market activity in the Deutsche mark/dollar market. Evans (2002), who also examined the Deutsche mark/dollar market, assumed that the impact of order flow depends on trade intensity, which is measured by the number of price changes. He showed that unexpected order flows have comparatively small effects on prices when market conditions are characterized by low trade intensity.

Compared with the extant literature, the notable feature of this study is that the time varying impacts of order flow are measured directly without sample division and tested directly to determine whether these were state dependent. The copula method enabled me to implement a direct test of the lemma suggested by Admati and Pfleiderer (1988) by regressing the obtained time varying impact onto the proxy variables for informed and liquidity traders.

To obtain proxies for informed- and uninformed (liquidity)-trading for an empirical test, I adopted order flow. McGroarty et al (2007) split order flow into expected- and unexpected-components: Unexpected order flow is linked to private information. Expected order flow is not. This idea is based on the work of Easley and O'Hara (1992), who provided an important insight into the link between market activity and private information when they suggested that private information signals cause irregular (unexpected) trading to deviate from its normal level. Therefore, although it is well documented that order flows convey information relevant to FX returns (e.g., Evans and Lyons, 2002a), McGroarty et al (2007) supposed that the normal level of order flow should be attributed to *non*-informational factors. One factor related to non-information is possibly explained with an inventory control. Bjønnes and Rime

(2005) found intensive inventory control by FX dealers, which generates an auto-regressive (AR) process of trades: That is, a temporal inventory is unloaded over short periods of time in the FX market. Moreover, the size of inventory control seems to be time varying. It is likely that when trade intensity is higher, the size of inventory control becomes much larger in proportion to trade intensity. Because traders can easily find their counterparties to unload their inventory, the size of inventory on their hands tends to be larger in an active market. This causes fluctuation in the size of order flows, and this fluctuation generates the auto-regressive conditional heteroskedasticity (ARCH) pattern in the volatility of order flow. As explained later, I assumed that the uninformed (inventory) component of order flow is expected with the ARMA-GARCH process and took the absolute residual from the process as the informed component. Additionally, to check the robustness of my results, I tried alternative methods to obtain the proxies of informed and uninformed trading.

3. 2 The copula

The name "copula" was chosen to emphasize the manner in which a copula "couples" a joint distribution function to its univariate margins (Nelsen 1999, p.15). The merit of the copula is to decompose the density of a multivariate distribution function into the copula density function and the univariate marginal densities. This enables the analyst to construct a flexible model of a joint distribution function. For empirics, only a few multivariate distributions, such as the normal and *t*-distributions, are implemented in practice. Meanwhile, a copula enables us to flexibly generate various multivariate distributions without the *ex ante* specification of them. For example, it is well known that financial data are approximated with marginal distributions of fat tails (e.g., Bollerslev, 1987) and the copula permits us to approximate the univariate marginal densities of each series of data with a *t*-distribution. Given a copula and marginal densities, we can obtain the corresponding density of a joint distribution for the pair of data sets.

Throughout this paper, I denote the distribution of random variables ($c.d.f.$) using an uppercase letter and the corresponding density ($p.d.f.$) using a lowercase letter. All $c.d.f.$s are assumed to be sufficiently smooth for all required derivatives to exist. Let $D(x,y\,|\,\theta)$ and $C(F(x\,|\,\theta),G(y\,|\,\theta))$ be a conditional bivariate $c.d.f.$ and conditional copula *c.d.f.* of random variables x,y given θ, respectively. θ is a set of parameters. $F(x\,|\,\theta)$ and $G(y\,|\,\theta)$ are the conditional marginal $c.d.f.$s of random variables x and y, respectively. I assume that only the subset θ_c, θ_x and θ_y of θ affect $C(.\,|\,\theta), F(.\,|\,\theta)$ and $G(.\,|\,\theta)$, respectively. Therefore,

$C(.|\theta) = C(.|\theta_c)$, $F(.|\theta) = F(.|\theta_x)$ and $G(.|\theta) = G(.|\theta_y)$.

The copula theory (*see* Nelsen, 1999; Patton, 2006) tells us that the copula function of $F_t(.|\Phi_{t-1},\theta_x)$ and $G_t(.|\Phi_{t-1},\theta_y)$ generates the conditional bivariate $c.d.f.$ of x_t and y_t given Φ_{t-1} (available information at *t* period) and parameter set θ:

$$D_t(x_t, y_t | \Phi_{t-1}, \theta) = C_t(F_t(x_t | \Phi_{t-1}, \theta_x), G_t(y_t | \Phi_{t-1}, \theta_y) | \Phi_{t-1}, \theta_c), \tag{3}$$

where subscript t refers to a value evaluated or observed at period t. By differentiating the bivariate $c.d.f.$ with respect to x_t, y_t, I can express the conditional bivariate $p.d.f.$ as a product of the copula and the univariate conditional marginal densities:

$$d_t(x_t, y_t | \Phi_{t-1}, \theta) = c_t(u_t, v_t | \Phi_{t-1}, \theta_c) f_t(x_t | \Phi_{t-1}, \theta_x) g_t(y_t | \Phi_{t-1}, \theta_y), \tag{4}$$

where $u_t = F(x_t | \Phi_{t-1}, \theta_x)$ and $v_t = G(y_t | \Phi_{t-1}, \theta_y)$. Generally, it is known that u_t and v_t have a uniform distribution within the interval $[0,1]$, regardless of the original distribution (Patton, 2006).[5] Equation (4) indicates that, if we can obtain the conditional marginal $p.d.f.$s of x, y and the copula, then we also obtain the conditional bivariate $p.d.f.$ of x and y.

3. 3 Empirical model

Various copula functions have been suggested by researchers (*see* Nelsen, 1999), and this study adopts three copulas, a standard normal (Gaussian), the Clayton and the Gumbel copulas, which are customarily used in the extant financial economics literature (e.g., Patton, 2006; Bartram et al., 2007; Rodriguez, 2007).

Figure 3 around here

The one virtue of copula functions is that they enable the analyst to measure tail dependence. Intuitively, "dependence" corresponds to a correlation between random

[5] This is confirmed with the q-q plot in the following section.

variables.[6] The Gaussian copula has symmetric tail dependence. The Clayton and Gumbel copulas describe situations of asymmetric tail dependence; the former copula exhibits only lower tail dependence, and the later one exhibits higher tail dependence. Setting the measure of dependence to 0.5 in all the copulas, Figure 3 shows the scatter plots of simulated bivariate copulas. In all cases, 1000 observations were generated, and margins were selected as standard normal. Obviously, the Clayton (second panel) and the Gumbel (third panel) exhibit only lower and upper tail dependences, respectively. The Gaussian copula (first panel) shows symmetric dependence. These copulas enable me to model asymmetric (or symmetric) relationships between order flows and returns. For example, the existence of upper tail dependence suggested by the Gumbel might imply that the buying pressure for the U.S. dollar (positive order flows) expects a larger movement in the dollar than the selling pressure for it (negative order flows) does. The Clayton's copula implies the opposite. Although these scenarios have not been tested in literature as far as I know, I am confident that this copula method will produce new understanding of the FX market.

Hereafter, let the random values x_i and y_i be a return and order flow of currency i, respectively. Let $\Psi(.)$ be the Gaussian $c.d.f.$ and denote $a_{i,t} = \Psi(u_{i,t})^{-1}$, $b_{i,t} = \Psi(v_{i,t})^{-1}$, ($u_{i,t} = F(x_{i,t} \mid \Phi_{t-1}, \theta_x)$, $v_{i,t} = G(y_{i,t} \mid \Phi_{t-1}, \theta_y)$).

Let $\rho_{i,t}$ be the time varying correlation of $a_{i,t}, b_{i,t}$ (or equivalently $u_{i,t}, v_{i,t}$), and I then obtain the following Gaussian *p.d.f*:

$$c^N(u_{i,t}, v_{i,t} \mid \Phi_{t-1}, \theta_c^N = \rho^N_{i,t}) = \frac{1}{\sqrt{1-\rho_{i,t}^{N2}}} \times \exp\left\{ -\frac{1}{2(1-\rho_{i,t}^{N2})}[a_{i,t}^2 + b_{i,t}^2 - 2\rho_{i,t}^N a_{i,t} b_{i,t}] + \frac{1}{2}[a_{i,t}^2 + b_{i,t}^2] \right\} \tag{5}$$

$$\rho^N_{i,t} = \omega_i^N + \eta_i^N{}_i \rho_{i,t-1}^N + \gamma_i^N \frac{1}{w} \sum_{j=1}^{w} |u_{i,t-j} - v_{i,t-j}| \tag{6}$$

where $\rho^N_{i,t} \in [-1,1]$.

In the above equations, superscript "*N*" corresponds to the normal (Gaussian) copula. The AR term in eq. (6) captures any persistency of $\rho_{i,t}$. The third term in eq. (6)

---

[6] For a formal definition of dependence, see Patton (2006).

assumes that the mean difference of the last *w* observations of $u_{i,t}$ and $v_{i,t}$ captures any variation in dependence (Patton, 2006; Bartram et al., 2007). These reasonable assumptions enable me to estimate a time varying correlation coefficient between $x_{i,t}$ and $y_{i,t}$. I set the maximum lag length of *w* to 10 and select the value that maximizes the log-likelihood of eq. (13), which appears later. *w*=1 was selected for both the euro/dollar and the yen/dollar.

The Gumbel and Clayton copulas (*c.d.f.*s) are defined as follows：

$$C^G(u_{i,t}, v_{i,t} \mid \Phi_{t-1}, \theta_c^G = \rho_{i,t}^G) = \exp\left\{-[(-\ln u_{i,t})^{\delta} + (-\ln v_{i,t})^{\delta}]^{1/\delta}\right\} \tag{7}$$

$$\rho_{i,t}^G = \omega_i^G + \eta_i^G \rho_{i,t-1}^G + \gamma_i^G \frac{1}{w}\sum_{j=1}^{w} \left| u_{i,t-j} - v_{i,t-j} \right|, \tag{8}$$

where $\rho_{i,t}^G = 1 - \delta^{-1}$

and $\rho_{i,t}^G \in (0,1)$.

$$C^{Cl}(u_{i,t}, v_{i,t} \mid \Phi_{t-1}, \theta_c^{Cl} = \rho_{i,t}^{Cl}) = \max[(u_{i,t}^{-\kappa} + v_{i,t}^{-\kappa} - 1)^{-1/\kappa}, 0] \tag{9}$$

$$\rho_{i,t}^{Cl} = \omega_i^{Cl} + \eta_i^{Cl} \rho_{i,t-1}^{Cl} + \gamma_i^{Cl} \frac{1}{w}\sum_{j=1}^{w} \left| u_{i,t-j} - v_{i,t-j} \right|, \tag{10}$$

where $\rho_{i,t}^{Cl} = \kappa/(2+\kappa)$

and $\rho_{i,t}^{Cl} \in (0,1)$.

In eqs. (7)-(10), superscripts *G*, *Cl* represent Gumbel and Clayton copulas, respectively. Using the same method in the Gaussian, I set *w*=1 in all the cases. Unlike the Gaussian copula, the measures of dependence, $\rho_{i,t}^G$ and $\rho_{i,t}^{Cl}$, range in (0,1).

In order to build a more flexible model, a mixture of the Gaussian, Gumbel and Clayton copulas were also tested:
Gaussian=Gumbel copula (*N*=*G*)

$$C^{N=G} = (1-\pi^{N=G})C^{N} + \pi^{N=G}C^{G}, \tag{11}$$

Gaussian=Clayton copula (N=Cl)

$$C^{N=Cl} = (1-\pi^{N=Cl})C^{N} + \pi^{N=Cl}C^{Cl}, \tag{12}$$

Clayton=Gumbel copula (Cl=G)

$$C^{Cl=G} = (1-\pi^{Cl=G})C^{Cl} + \pi^{Cl=G}C^{G}, \tag{13}$$

where π is a weighting parameter, which is estimated.

As shown in Table 1, serial correlations of first and second moments are also statistically significant in returns and order flows.[7] Because serial correlations of the first and second moments indicate AR and ARCH effects, the result in Table 1 leads me to select an AR-ARCH model to depict the stochastic processes of these two variables. It is well known that the process of asset returns follows the AR-ARCH pattern (e.g., Engle, 1982). Additionally, as I explained, an inventory control by FX traders and intraday seasonality of the FX market possibly cause the AR-ARCH process of order flows.

As a stochastic process of these two variables, I selected the following ARMA(p,q)-GARCH(r,s) model of the t density distribution with the degree of freedom μ:

$$z_{i,t} = \text{const}_i + \sum\nolimits_{P=1}^{p} \varphi_{i,P} z_{i,t-P} + \sum\nolimits_{Q=1}^{q} \delta_{i,Q} \varepsilon_{i,t-Q} + \varepsilon_{i,t} \tag{14}$$

$$h_{i,t} = \kappa_i + \sum\nolimits_{R=1}^{r} \alpha_{i,R} \varepsilon_{i,t-R}^2 + \sum\nolimits_{S=1}^{s} \beta_{i,S} h_{i,t-S} \tag{15}$$

$$\varepsilon_{i,t} \sim t_{\mu_i}(0, h_{i,t}), \tag{16}$$

where $z_{i,t}$ is a return or order flow ($x_{i,t}/y_{i,t}$) of currency i at period t. Because the non-normality of the data is apparent from the coefficients of skewness and kurtosis in Table 1, I assume that returns and order flows of each currency have marginal t-distributions.[8] It is also well known that the t-distribution approaches a normal

[7] Table 1 shows the positive autocorrelation in order flow in both euro/dollar and yen/dollar. This is possibly explained by “hot potato” trading (Lyons, 1997) and/or order splitting, as a large order is split into a number of smaller orders of same sign that are executed sequentially (Berger et al., 2007).

[8] Strictly speaking, order flows (not seasonally adjusted) are integer-valued variables to be approximated by a non-continuous distribution. The q-q plot (Fig.5) does not reject

distribution with variance h_t for $1/\mu \to 0$, but the $t-$distribution has "fatter tails" than the corresponding normal distribution for $1/\mu > 0$ (Bollerslev, 1987). Therefore, I can say that my t-distribution assumption is more comprehensive than that of a normal distribution.

The processes of both returns and order flows are assumed to have ARMA-GARCH patterns, and this may prompt selection of a *multivariate* ARMA-GARCH framework in place of the copula approach. However, the former approach needs to be specified in the form of multi-distribution between returns and order flows, and this entails a more severe constraint. In practice, only bivariate normal or *t*-distributions are implemented for a bivariate GARCH and these symmetric distributions ignore a possible asymmetric dependence between order flows and returns. Therefore, I can consider that this copula approach is superior to a multivariate GARCH framework because of its flexibility.

Taking a logarithm of eq. (4) and summing it across a set of time, I obtain the following equation:

$$\begin{aligned}&\sum_{t=1}^{T} \ln d_t(x_{i,t}, y_{i,t} \mid \Phi_{t-1}, \theta) = \\ &\qquad \sum_{t=1}^{T} \ln c_t(u_{i,t}, v_{i,t} \mid \Phi_{t-1}, \theta_c) + \sum_{t=1}^{T} \ln f_t(x_{i,t} \mid \Phi_{t-1}, \theta_x) + \sum_{t=1}^{T} \ln g_t(y_{i,t} \mid \Phi_{t-1}, \theta_y),\end{aligned} \tag{17}$$

where T is the sample size of the data set. Because eq. (17) contains a large set of parameters, I have used a two-stage estimation procedure to obtain the set of estimated parameters, $\hat{\theta}$, a method that was also adopted by Bartram et al. (2007).

The two step-estimator for the time varying copula parameter is valid if the degree of dependence (θ_c) does not affect the conditional margins. Otherwise, the procedure will lead to misspecification. Although I should consider the possibility that the conditional margins will be affected by the degree of dependence and change over time, it is quite difficult to achieve the simultaneous optimization of eq. (17).[9] Therefore, I adopted the two-step estimator as a second best approach to this challenge.

As a first stage, I search θ_x and θ_y, which maximize the terms of the

---

the assumption of *t*-distribution for order flows strongly and, therefore, I do not consider my approximation to be perfectly invalid.

[9] I adopted the two-step estimator as initial values for the simultaneous optimization of eq. (17), but I could not achieve convergence.

univariate marginal densities in eq. (17), respectively.

$$\hat{\theta}_x \equiv \operatorname{argmax} \sum_{t=1}^{T} \ln f_t(x_{i,t} \mid \Phi_{t-1}, \theta_x) \tag{18}$$

$$\hat{\theta}_y \equiv \operatorname{argmax} \sum_{t=1}^{T} \ln g_t(y_{i,t} \mid \Phi_{t-1}, \theta_y), \tag{19}$$

where " $J = \operatorname{argmax} K$ " means that parameter J is to maximize function K. This process corresponds to the maximum likelihood estimation of eqs. (14)-(16).

Next, I estimate the copula parameter θ_c. Estimating eqs. (14)-(16) gives numerical values of $u_{i,t}$ and $v_{i,t}$ of eqs. (6)-(10). I also obtain those of $a_{i,t}$ and $b_{i,t}$ in the case of Gaussian copula (eq. (5)). Given $\hat{\theta}_x$, $\hat{\theta}_y$ and the numerical values of $u_{i,t}$, $v_{i,t}$, $a_{i,t}$ and $b_{i,t}$, I obtain the following $\hat{\theta}_c$ in each copula:

$$\hat{\theta}_c \equiv \operatorname{argmax} \sum_{t=1}^{T} \ln c_t(u_{i,t}, v_{i,t} \mid \Phi_{t-1} \theta_c, \hat{\theta}_x, \hat{\theta}_y). \tag{20}$$

4. Estimation result

4. 1 Estimation of conditional marginal distributions

In this study, the Akaike's information criterion (AIC) was used to select lag lengths of mean and volatility equations of the ARMA-GARCH model. I use the following definition as the AIC:

$$\text{AIC} \equiv -2 \ln L + 2n, \tag{21}$$

where ln L is the log-likelihood calculated in each step, and n is the number of estimated parameters. First, I considered GARCH(1,1) for the volatility equation in order to control the GARCH effect on the mean equation. Then, setting the maximum length of AR and MA lags to 2, I selected the pair of AR(p) and MA (q) that minimizes the AIC. With this specification of the mean equation, GARCH(1,1), GARCH(1,2) and GARCH(2,1) were tested by the AIC.

Table 2 around here

Table 2 (part A) reports the estimated parameter of eqs. (18)-(19). For returns and order flows of the euro/dollar, ARMA(2,2)-GARCH(2,1) and ARMA(1,2)-GARCH(2,1) were selected, respectively. ARMA(2,2)-GARCH(2,1) and

ARMA(2,1)-GARCH(2,1) were selected for returns and order flows of the yen/dollar, respectively. The number in parentheses is the standard error evaluated with the *sandwich estimator* (White, 1982) in each stage.

Upon first look at the result of the first stage estimation, the reciprocal of degree of freedom, $1/\mu$, is statistically significantly different from zero in each case. This result is consistent with the result of Table 1 and my modeling in eqs. (14)-(16), which assumes that the marginal $p.d.f.$s of return and order are approximated by a fatter tail distribution than a standard normal one. From the estimation results of eqs. (18) and (19), I obtain time varying variables $\widehat{\text{var}(x_t)}$ and $\widehat{\text{var}(y_t)}$, which are conditional volatilities of GARCH and are used to calculate the impact of order flow defined in eq. (2). I replace $\text{var}(x_t)$ and $\text{var}(y_t)$ in eq. (2) with the conditional ones, $\widehat{\text{var}(x_t)}$ and $\widehat{\text{var}(y_t)}$, respectively.

Figure 4 around here

In order to check the validity of the model specification, tests such as the Ljung-Box (Q statistics) and the Kolmogorov-Smirnov test are often implemented. My sample size is well over 30,000 and this leads the Ljung-Box test to detect a serial correlation between some moments of model's residuals even if the serial correlation is close to zero. Therefore, I checked the validity of the model specification with graphics. First, I checked the serial correlations (maximum lag is 10) of the first and second moments of standardized residuals of the specified ARMA-GARCH model. In Fig. 4, I confirm that the serial correlations of the first and second moments of standardized residuals are smaller than those of the original series in all cases. Also, because the transformed variables of the standardized residuals should have a uniform distribution within the interval $[0,1]$, I also checked this with the q-q plot in Fig. 5.

Figure 5 around here

The transformed variables of the standardized residuals of ARMA-GARCH are obtained with $H_{\hat{\mu}_i}(\hat{\varepsilon}_{i,t}/\sqrt{\hat{h}_t})$, where $H_{\hat{\mu}_i}$ is a *t*-distribution function with a degree of freedom $\hat{\mu}_i$, with the "hat" signifying estimated parameter and variables. If $H_{\hat{\mu}_i}(\hat{\varepsilon}_{i,t}/\sqrt{\hat{h}_t})$ has a perfect fit with uniform distribution within the interval $[0,1]$, the dashed line should correspond to a 45° solid line connecting (0,0) and (1,1) in Fig. 5. As a whole, the result of Fig. 5 does not strongly reject uniform distributions of the transformed variables. Therefore, with the results of Figs.4 and 5, I am confident that my margin specification is valid.

4. 2 Estimation of copulas

The AIC was also used to select copulas for the model. As mentioned before, six types of copula were tried: the Gaussian, Gumbel, Clayton, Gaussian=Gumbel, Clayton=Gumbel and Gaussian=Clayton copulas. The AIC selected the Clayton=Gumbel copula for both the euro/dollar and yen/dollar. Additionally, if $C(u,v)$ is a copula where $u=F(x)$ and $v=G(y)$ and if it is correctly specified, its first derivative with respect to each of its arguments, $\partial C / \partial u$, has a uniform distribution within the interval [0,1] (Klugman and Parsa, 1999).

Figure 6 around here

Figure 6 shows the q-q plots of the first derivative of the six copulas for both the euro/dollar and the yen/dollar. Excluding the Gaussian and the Gaussian=Clayton copulas in the euro/dollar and yen/dollar, the results of Fig. 6 indicate that estimated copulas seem to be relatively well specified. Therefore, the Clayton=Gumbel copula was selected for both the euro/dollar and yen/dollar FX markets as suggested by the AIC. This result possibly implies that a tail dependency exists between returns and order flows: the larger sizes of order flows can predict the movement in the returns with more accuracy in the euro/dollar and yen/dollar markets.

Table 3 around here

To address this issue, pairs in the standardized residuals of order flows and returns were sorted on the former and split into three ranks (top, middle and bottom). The standardized residuals were calculated from the specified GARCH models in Table 2(A). I then examined the coefficient of correlation between the standardized residuals of order flows and returns in each rank. Here, I adopted the standardized residuals in order to avoid a spurious dependency between random variables caused by heteroskedasticity (Rodriguez, 2007). Table 3 shows that the correlation of coefficients in the top and bottom ranks (except the 10 percent case) are higher than those in the middle. Therefore, I am confident that the results of Table 3 are consistent with my modelling of tail dependency.

Table 2 (part B) reports the estimated parameters of the second stage.[10] All estimated parameters are statistically significant; therefore, I can safely conclude that my specification of correlation coefficient is adequate. The result of the second stage yields the time varying correlation coefficient ρ_t of eq. (2). For the Clayton=Gumbel copula, the measure of dependence (correlation) $\rho_t^{Cl=G}$ is defined with $(1-\pi^{Cl=G})$

---

[10] w=1, which maximizes the log-likelihood, is selected for both the euro/dollar and the yen/dollar in eqs. (8) and (10).

$\rho_t^{Cl} + \pi^{Cl=G} \rho_t^{G}$.

4. 3 Empirical test of time varying impact of order flows

From the result of the two-stage estimation, I obtained the time varying impact of order flow on return (Λ_t) defined in eq. (2).

Figure 7 around here

Figure 7 shows the time varying and constant impacts of order flow.[11] The latter is obtained by regressing returns onto order flows of each currency. The (horizontal) bold and thin lines correspond to the constant and time varying impacts, respectively.

Table 4 around here

Table 4 reports the summary of statistics of the estimated time varying impact of order flows. The averages of time varying impact are 0.36 (euro/dollar) and 0.45 (yen/dollar). The values of constant impact are 0.41 (euro/dollar) and 0.49 (yen/dollar).The R-square values are 0.17 (euro/dollar) and 0.24 (yen/dollar) and calculated with the time varying impact of order flow in the regression of returns onto order flows. They are quite similar to the constant impacts of order flow, which are 0.16 (euro/dollar) and 0.24 (yen/dollar).

Payne (2003) found that the impact of order flow decreases with market activity. He showed that the impact of order flow on the Deutsche mark/dollar return tended to be lower in his subsamples of 8 to 10 GMT and 12 to 14 GMT, and the timings of these subsamples corresponded to the opening of the Europe and the N.Y. markets, respectively. His finding indicates that the impact of order flow has intra-day seasonality. Following Melvin and Melvin (2003), I account for such seasonality by identifying the following distinct regions: (1) Asia, (2) Asia-Europe trading overlap, (3) Europe, (4) Europe-America overlap, (5) America and (6) Pacific.

Table 5 around here

Table 5 considers the timing of daylight saving time (DST) by Europe and America, and I use dummy seasonality variables based on the definition in Table 5. If the impact of order flow has seasonality, the regression of the order flow impact onto these dummies should show the significant effects of these variables. I introduced these seasonal dummies in order to remove deterministic seasonality and capture the pure effects of informed and liquidity trading on the impact of order flow. These effects were predicted by Admati and Pfleiderer (1988), whose lemma was presented in the first subsection of last section.

[11] Because outliers are observed in the first periods and in order to keep figures easy to see, the first 100 variables are dropped in Fig. 7.

To examine this, there is a need for proxies of informed and liquidity traders. To address this, I obtained the expected and unexpected components of order flow by estimating eqs. (7)-(9). Assuming that these components reflect liquidity and informed trading, their absolute values are adopted as the proxies of sizes in liquidity and informed traders (e.g., Chung et al., 2005), respectively. That is, I assume that liquidity trading is observed systematically and, conversely, that the entry of informed traders follows an unexpected stochastic process.

$$\Lambda_{i,t} = \text{const}_i + \Pi_i \Lambda_{i,t-1} + \sum_{j=1}^{5} \Theta_{i,j} segD_j + \Xi_i \mid y^*_{i,t} \mid + \Omega_i \mid \hat{y}_{i,t} \mid + e_{i,t}, \tag{22}$$

where $segD_j$ is a dummy variable that takes unity and is otherwise zero when data timing corresponds to that of segment j $[j = 1(\text{Asia}),...,5(\text{America})]$. e_t is an error term. $\mid y^*_{i,t} \mid$ is the absolute value of residuals of order flows that is obtained by estimating eqs. (7)-(9) for currency i. $\mid \hat{y}_{i,t} \mid$ is an absolute value of predicted order flow estimated by eqs. (7)-(9). Inclusion of the AR term can capture any persistence of the impact of order flow. The theory of Admati and Pfleiderer (1988) predicts that the sign conditions are $\Xi_i > 0$ if the gathered private information by informed traders is sufficiently imprecise or $\Xi_i < 0$. The effect of liquidity traders is $\Omega_i < 0$.

Equation (1) also indicates that there may be a linear dependence between the logarithms of the impact of order flow, informed and liquidity traders. Specifically, the log of Λ_t linearly depends on the logarithms of informed and liquidity traders. Therefore, I also considered the logarithm version of eq. (22) as follows;

$$\ln \Lambda_{i,t} = \text{const}_i + \Pi_i \ln \Lambda_{i,t-1} + \sum_{j=1}^{5} \Theta_{i,j} segD_j + \Xi_i \ln \mid y^*_{i,t} \mid + \Omega_i \ln \mid \hat{y}_{i,t} \mid + e_{i,t}. \tag{23}$$

Moreover, to confirm the robustness of my result, I adopted the alternative proxies of uninformed- and informed-trades, which were proposed by McGroarty et al (2007).

$$EV_t = \sum_{d=1}^{D} V_{d,t} / D \tag{24}$$

$$UV_{d,t} = V_{d,t} - EV_t \tag{25}$$

$$EO_t = \sum_{d=1}^{D} | O_{d,t} | / D \tag{26}$$

$$UO_{d,t} = | O_{d,t} | - EO_t , \tag{27}$$

where d and t represent the day and the time of day (in 10-minute intervals), respectively. D represents the total number of days. $V_{d,t}$ and $O_{d,t}$ are the number of transactions and order flow at t period of day d, respectively. The number of transactions (hereafter, trading volume) is calculated by summing up absolute values of trade indicators in each 10-minute interval. EV_t and EO_t are the expected trading volume and order flow, respectively and they correspond to uninformed components. Considering the timing of adoption of DST, these two variables are calculated within each period in Table 5. $UV_{d,t}$ and $UO_{d,t}$ are unexpected trading volume and order flow, respectively, and they correspond to informed components. I also replaced $| y^*_{i,t} |$ and $| \hat{y}_{i,t} |$ of eq. (22) with either $UV_{d,t}$ and $EV_{,t}$ (eq. (28)) or $UO_{d,t}$ and EO_t (eq. (29)), as follows.

$$\Lambda_{i,t} = \text{const}_i + \Pi_i \Lambda_{i,t-1} + \sum_{j=1}^{5} \Theta_{i,j} segD_j + \Xi_i UV_{i,d,t} + \Omega_i EV_{i,t} + e_{i,t} \tag{28}$$

$$\Lambda_{i,t} = \text{const}_i + \Pi_i \Lambda_{i,t-1} + \sum_{j=1}^{5} \Theta_{i,j} segD_j + \Xi_i UO_{i,d,t} + \Omega_i EO_{i,t} + e_{i,t} \tag{29}$$

Table 6 around here

Table 6 shows the estimation results of eqs. (22), (23), (28) and (29). Numbers in parentheses are the standard errors evaluated with the Newey and West variance-covariance matrix. The lag of the Newey and West variance-covariance matrix is $T^{1/4}$, where T is the sample size.

First, the impact of order flows decreases with informed trading (Ξ_i<0), and this is observed significantly in all cases except eq. (22) for the yen/dollar. Therefore, I conclude that an increase in informed trading makes market makers less sensitive to observed order flow. This empirical evidence indicates that the effect of competition among informed traders dominates that of informational uncertainty caused by them.

Secondly, liquidity trading tends to decrease the impact of order flows (Ω_i<0), although, for the yen/dollar, this tendency is statistically significantly only in eq.(28) . In summary, it is safe to conclude that the results reported in Table 6 (partly) give empirical support to the theory of Admati and Pfleiderer (1988). In particular, my empirical finding stresses the role of informed trading in the impact of order flow. Accordingly, I propose that the time varying relationship between returns and order flow, which depends on market conditions, are an important consideration.

I find a negative effect of informed trading on the impact of order flow, and this finding is comparable to the result of Evans (2002), who found that the impact of (unexpected) order flow is an *increasing* function of trade intensity. Trade intensity reflects two factors; (1) informed and (2) liquidity components. The theory developed by Admati and Pfleiderer (1988) suggests that the entry of informed traders increases with more trade intensity because they prefer to trade in a period of high trade intensity in order to conceal their entry into a market. Therefore, according to Evans (2002), trade intensity is supposed to work as a proxy for informed trading, and informed traders bring an uncertainty into the market. In this case, the theory predicts that the impact of order flows is an increasing function of informed traders. The theory of Admati and Pfleiderer (1988) also suggests that liquidity trading concentrates on periods of high volume. This indicates that both liquidity and informed trading increase with more trade intensity, and trade intensity also works as a proxy for liquidity trading. However, the finding of Evans (2002) indicates that the effect of informed trading dominates the effect of liquidity and that trading intensity is likely to reflect the former effect. This is because the theory does not predict a positive effect of liquidity on the impact of order flows.

In Evans (2002), why do informed traders bring an uncertainty into the FX market rather than competitions among informed traders? The former uncertainty has a positive effect on the impact, and the latter has an opposite effect. The finding in Evans (2002) can possibly be attributed to the lack of transparency in the direct interdealer market. The data set used by Evans (2002) has only direct interdealer transactions, which are observed only by the counterparties. In contrast, the market of electronic broking is transparent because the transaction through the electronic broking system shows up on the screen, and information about transactions is available for all participants. Therefore, I can consider that the direct interdealer market is less transparent than the electronic broking one. This suggests that the information gathered by informed traders is sufficiently imprecise, and the effect of informational uncertainty caused by informed traders dominates that of the competition among them in the direct interdealer market. In this case, the theory of Admati and Pfleiderer (1988) predicts that an impact of order flows is an increasing function of informed traders. Conversely, my electronic broking transactions are transparent; in turn, the effect of competition among informed traders dominates that of informational uncertainty caused by them. In this case, the theory predicts that an impact of order flows is a decreasing function of informed traders. As shown in Table 6, my empirical results are consistent with the above scenario.[12]

The above discussion suggests that the time varying impact is useful in designing economic policy: its response to informed traders may be useful to gauge the level of transparency in each market, and it enables us to understand the market characteristics in details.

5. Conclusion

The results of this study support the theoretical argument that the impact of order flow on returns decreases with more informed traders and more liquidity trading (Admati and Pfleiderer, 1988). The former impact is due to competition among informed traders, and the latter is explained by competition among market makers. Moreover, I find that the effect of informed trading on the impact of order flow is stronger than that of liquidity trading. The selected Clayton=Gumbel copula implies that irregular sizes of order flows can predict the movement in the returns with more accuracy than the regular size in the euro/dollar and yen/dollar markets.

My findings suggest that we should treat the relationship between order flows and returns as time varying. The impact of order flows depends on market conditions, especially on the entry of informed traders. Therefore, the impact is time varying. This conclusion might challenge the results of extant literature, which points to the lack of a stable long-term relationship between order flows and returns (e.g., Froot and Ramadorai, 2005; Boyer and Morden, 2006; Berger et al., 2007).[13]

[12] The data set used by Evans (2002) is available at his web site. Unfortunately, the data timing is not available and this prevents me trying a comparative analysis with mine.

[13] I also use a co-integration test (ADF-type) for 10-minute returns and order flows. The test was performed for both time varying and constant co-integration parameters of order flows. I found co-integrational relationships in all cases (the four cases in this study).

References

Admati, A., Pfleiderer, P., 1988. A theory of intra-day patterns: volume and price variability. Review of Financial Studies 1, 3-40.

Baillie, R., Bollerslev, T., 1990. Intra-day and inter-market volatility in foreign exchange rates. Review of Economic Studies 58, 565-585.

Bartram, S.M., Taylor, S.J., Wang, Y., 2007. The euro and European financial market dependence. Journal of Banking & Finance 31, 1461-481.

Bank for International Settlements., 2007. Triennial central bank survey of foreign exchange and derivatives market activity in April 2007; preliminary global results.

Berger, D.W., Chaboud, A.P., Chernenko, S.V. Howorka, E., Wright J.H., 2007. Order flow and exchange rate dynamics in electronic brokerage system data. Journal of International Economics, forthcoming

Berger, D.W., Chaboud, A.P., Hjalmarsson, E., Howorka, E., 2006. What drives volatility persistence in the foreign exchange market? FRB International Finance Discussion Papers 862.

Bjønnes, Geir Hoidai. and Rime, Dagfinn., 2005. Dealer behavior and trading systems in foreign exchange markets. Journal of Financial Economics 75, 571-605.

Bollerslev, T., 1987. A conditional heteroskedastic time series model for speculative prices and rates of return. Review of Economics and Statistics 69(3), 542-547.

Boyer, M.M., Norden, S.V., 2006. Exchange rates and order flow in the long run. Finance Research Letters 3(4), 235-243.

Chung, K.H., Li, M., McInish, T.H., 2005. Information-based trading, price impact of trades, and trade autocorrelation. Journal of Banking & Finance 29, 1645-1669.

Easley, D., O'Hara, M., 1992. Time and the process of security price adjustment. Journal of Finance 47, 577-605.

Engle, R., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987-1007.

Evans, M.D.D., Lyons, R.K., 2002a. Order flow and exchange rate dynamics. Journal of Political Economy 110(1), 170-180.

Evans, M.D.D., Lyons, R.K., 2002b. Information integration and FX trading. Journal of International Money and Finance 21, 807-831.

Evans, M.D.D., 2002. FX trading and exchange rate dynamics. Journal of Finance. 57, 2405-2447.

Froot, K., Ramadorai, T., 2005. Currency returns, intrinsic value, and institutional-investor flows. Journal of Finance 60(3), 1535-1566.

Hamilton, J.D., 1994. *Time Series Analysis*. Princeton University Press, Princeton.

Ito, T., Hashimoto, Y., 2006. Intraday seasonality in activities of the foreign exchange markets: Evidence from the electronic broking system. Journal of the Japanese and International Economies 20, 637-664.

Klugman, S.A., Parsa, R., 1999. Fitting bivariate loss distribution with copulas. Insurance: Mathematics and Economics 24, 139-148.

Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 53(6), 1315-1335.

Lyons, R.K., 2001. *The Microstructure Approach to Exchange Rates*. The MIT Press, Cambridge,. MA.

Lyons, R.K., 1997. A simultaneous trade model of the foreign exchange hot potato. Journal of International Economics 42, 275-298.

McGroarty, F., Gwilym, O., Thomas, S., 2007. The role of private information in return volatility, bid-ask spreads and price levels in the foreign exchange market, Journal of International Financial Markets, Institutions & Money, doi:10.1016/j.intfin.2008.04.001

Melvin, M., Melvin, B.P., 2003. The global transmission of volatility in the foreign exchange market. Review of Economics and Statistics 85(3), 670-679.

Nelsen, R.B., 1999. *An Introduction to Copulas*. Springer, New York.

Patton, A.J., 2006. Modelling asymmetric exchange rate dependence. International Economic Review 47, 527- 556.

Payne, L., 2003. Informed trade in spot foreign exchange markets: an empirical investigation. Journal of International Economics. 61(2), 307-329.

Rodriguez, J.C., 2007. Measuring financial contagion: a copula approach. Journal of Empirical Finance 14, 401-423.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50(1), 1-25.

Table 1

Summary of statistics

Variable	Number of observation	Mean	SD	First order serial correlation	LB(6)		Skew	Kurtsis
R_{euro}	36623	-0.00010	0.045	-0.03	44.0	***	-0.206	14.294
R^2_{euro}	36623	0.00205	0.007	0.13	1518.9	***		
X_{euro}	36623	-0.14810	12.490	0.05	187.6	***	0.088	5.160
X^2_{euro}	36623	156.02993	318.057	0.13	3160.0	***		
R_{yen}	36623	0.00016	0.049	-0.03	35.8	***	-0.496	13.729
R^2_{yen}	36623	0.00239	0.009	0.14	1579.8	***		
X_{yen}	36623	0.96439	11.181	0.04	153.3	***	0.024	5.453
X^2_{yen}	36623	125.94886	264.951	0.14	2040.2	***		

Note: LB(6) is the Ljung-Box statistics with six lags. R, R^2, X and X^2 are the first and second moments of returns and order flows for each currency pair (euro/dollar and yen/dollar). *** indicates 1 percent level of significance. SD: standard deviation.

Table 2

Estimation results of the copula

Part A: ARMA-GARCH (first stage) estimation

i=euro

Dependent variable	Return			Order flow		
	Estimates			Estimates		
$const_i$	0.00030	(0.00040)		-0.00190	(0.00050)	***
$\varphi_{i,1}$	1.05720	(0.10990)	***	0.93190	(0.01200)	***
$\varphi_{i,2}$	-0.18960	(0.07400)	**			
$\delta_{i,1}$	-1.13780	(0.10960)	***	-0.89370	(0.01320)	***
$\delta_{i,2}$	0.23960	(0.08060)	***	-0.02030	(0.00560)	***
$\kappa_{i,1}$	0.03010	(0.00600)	***	0.01840	(0.00660)	***
$\alpha_{i,1}$	0.14870	(0.00990)	***	0.08270	(0.00730)	***
$\alpha_{i,2}$	-0.09310	(0.01150)	***	-0.04480	(0.00910)	***
$\beta_{i,1}$	0.91920	(0.01140)	***	0.94450	(0.01070)	***
$1/\mu_i$	0.22540	(0.00530)	***	0.10530	(0.00490)	***

i=yen

Dependent variable	Return			Order flow		
	Estimates			Estimates		
$const_i$	0.00040	(0.00010)	***	0.01080	(0.00360)	***
$\varphi_{i,1}$	1.40320	(0.04170)	***	0.88990	(0.04690)	***
$\varphi_{i,2}$	-0.42940	(0.04290)	***	-0.01840	(0.00690)	***
$\delta_{i,1}$	-1.46810	(0.03980)	***	-0.85000	(0.04660)	***
$\delta_{i,2}$	0.48740	(0.04100)	***			
$\kappa_{i,1}$	0.02050	(0.00370)	***	0.01460	(0.00580)	**
$\alpha_{i,1}$	0.14360	(0.00910)	***	0.11120	(0.00810)	***
$\alpha_{i,2}$	-0.08530	(0.01070)	***	-0.07050	(0.01030)	***
$\beta_{i,1}$	0.92480	(0.00870)	***	0.94590	(0.01080)	***
$1/\mu_i$	0.20280	(0.00520)	***	0.13140	(0.00490)	***

Part B: Clayton=Gumbel copula (second stage) estimation

	i=Euro			i=Yen		
	Estimates			Estimates		
ω^{Cl}_{i}	0.08140	(0.01490)	***	0.25670	(0.08220)	***
μ^{Cl}_{i}	0.89710	(0.02140)	***	0.52790	(0.17560)	***
γ^{Cl}_{i}	-0.15010	(0.03920)	***	-0.15760	(0.06460)	**
ω^{G}_{i}	0.05610	(0.00540)	***	0.05010	(0.00780)	***
μ^{G}_{i}	0.90740	(0.00900)	***	0.91930	(0.01240)	***
γ^{G}_{i}	-0.12690	(0.01800)	***	-0.09080	(0.01780)	***
$\pi^{Cl=G}$	0.70950	(0.01820)	***	0.71100	(0.01550)	***

Note: The numbers in parentheses are the robust standard errors suggested by White (1982). ***, ** and * indicate 1, 5 and 10 percent levels of significance, respectively.

Table 3

Coefficient of correlation between the standardized residuals of order flows and returns in each rank.

	#	Top # percent	Middle 100-2# percent	Bottom # percent
Euro	10	0.198	0.279	0.260
	20	0.208	0.207	0.249
	30	0.226	0.137	0.249
Yen	10	0.243	0.348	0.272
	20	0.272	0.262	0.303
	30	0.295	0.177	0.313

Note: Pairs in the standardized residuals of order flows and returns were sorted on the former and split into three ranks (top, middle and bottom). The standardized residuals were calculated from the specified GARCH models in Table 2(A).

Table 4

Summary of statistics for a time varying impact of order flows.

	Euro	Yen
Mean	0.361	0.453
Median	0.365	0.456
S.D.	0.074	0.035
Max	2.094	0.575
Min	0.003	0.157
R2	0.170	0.241

Note: S.D. is the standard errors of time varying impacts of order flows. R2 is the R-squared value when calculating the time varying impact of order flow in the regression of returns onto order flows.

Table 5

Regional time zone (GMT) for the six segments.

Sample Period	Asia	Asia-Europe Overlap	Europe	Europe-America Overlap	America	Pacific
1/9/05-30/10/05 (Eur. and Amer. DST)	23:30-5:30	5:30-8:00	8:00-11:30	11:30-15:30	15:30-20:00	20:00-23:30
31/10/05-25/3/06 (no DST)	23:30-6:30	6:30-8:00	8:00-12:30	12:30-16:30	16:30-21:00	21:00-23:30
26/3/06-1/4/06 (only Eur. DST)	23:30-5:30	5:30-8:00	8:00-12:30	12:30-15:30	15:30-21:00	21:00-23:30
2/4/06-31/8/06 (Eur. and Amer. DST)	23:30-5:30	5:30-8:00	8:00-11:30	11:30-15:30	15:30-20:00	20:00-23:30

Note. ‘DST’ is daylight saving time. ‘Eur’ and ‘Amer’ refer to Europe and America, respectively.

Table 6

Regression of the order flow impact

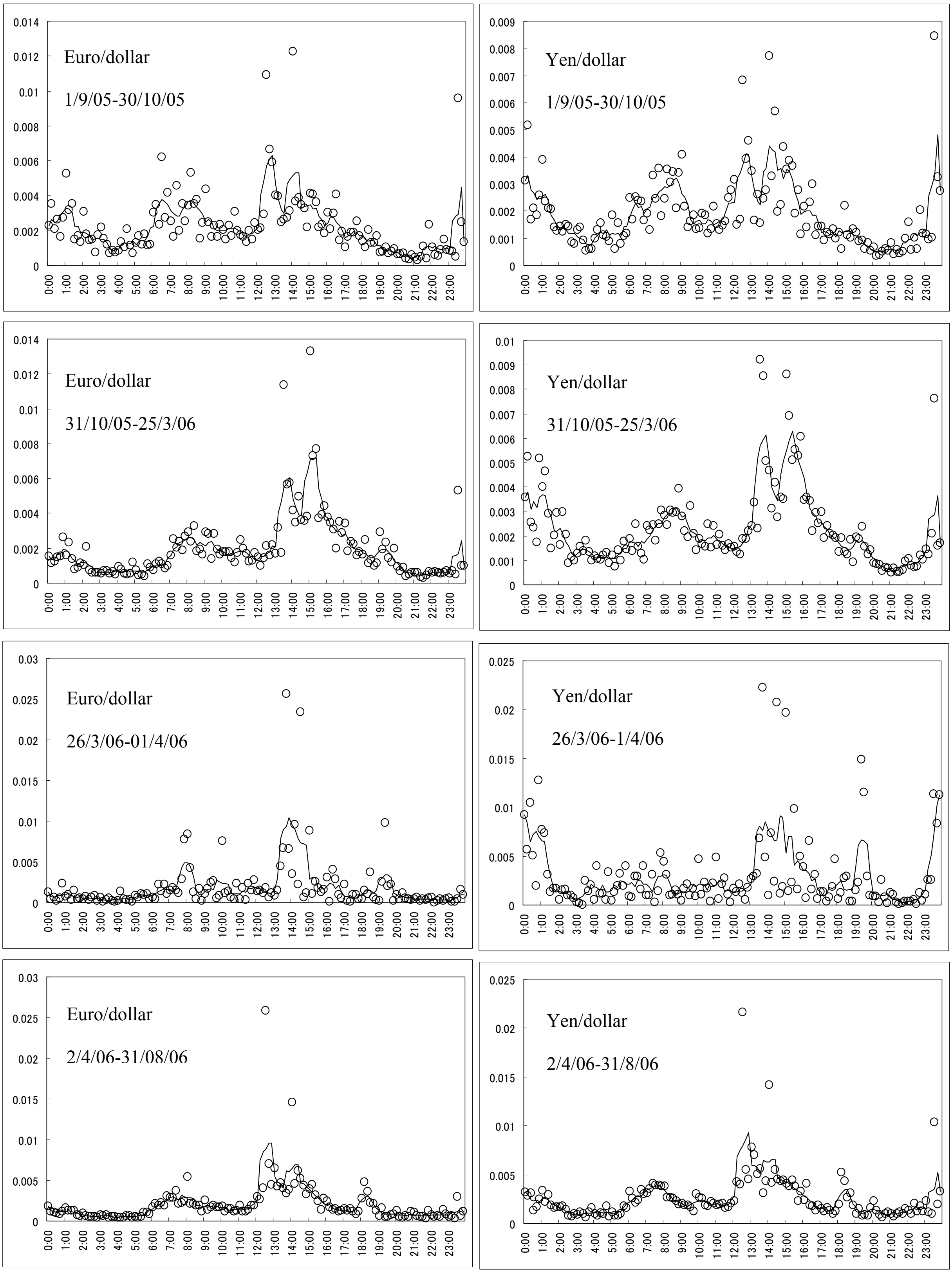
i=euro/dollar	Eq.(22) Estimates			Eq.(23) Estimates			Eq.(28) Estimates			Eq.(29) Estimates		
$const_i$	0.06995	(0.00443)	***	-0.15593	(0.00627)	***	0.06959	(0.00426)	***	0.07044	(0.00450)	***
Π_i	0.82123	(0.01134)	***	0.84959	(0.00578)	***	0.81925	(0.01128)	***	0.82183	(0.01128)	***
$\Theta_{i,1}$	0.00398	(0.00073)	***	0.00850	(0.00185)	***	0.00418	(0.00073)	***	0.00449	(0.00077)	***
$\Theta_{i,2}$	-0.00159	(0.00088)	**	-0.00568	(0.00226)	***	0.00041	(0.00102)		0.00096	(0.00116)	
$\Theta_{i,3}$	-0.00881	(0.00086)	***	-0.02417	(0.00231)	***	-0.00653	(0.00098)	***	-0.00575	(0.00107)	***
$\Theta_{i,4}$	-0.01183	(0.00097)	***	-0.03219	(0.00242)	***	-0.00740	(0.00126)	***	-0.00789	(0.00120)	***
$\Theta_{i,5}$	-0.00686	(0.00083)	***	-0.01695	(0.00205)	***	-0.00581	(0.00084)	***	-0.00540	(0.00085)	***
Ξ_i (informed component)	-0.00047	(0.00034)	*	-0.00193	(0.00049)	***	-0.00002	(3.05E-06)	***	-0.00008	(0.00003)	***
Ω_i (uninformed component)	-0.02551	(0.00766)	***	-0.00249	(0.00054)	***	-0.00002	(4.37E-06)	***	-0.00046	(0.00012)	***
R2	0.253			0.227			0.252			0.253		

i=yen/dollar	Eq.(22) Estimates			Eq.(23) Estimates			Eq.(28) Estimates			Eq.(29) Estimates		
$const_i$	0.11890	(0.00238)	***	-0.19870	(0.00501)	***	0.11957	(0.00234)	***	0.11914	(0.00241)	***
Π_i	0.74126	(0.00509)	***	0.74683	(0.00614)	***	0.74030	(0.00501)	***	0.74114	(0.00508)	***
$\Theta_{i,1}$	0.00019	(0.00044)		0.00038	(0.00097)		0.00080	(0.00048)	**	0.00037	(0.00049)	
$\Theta_{i,2}$	-0.00213	(0.00056)	***	-0.00475	(0.00125)	***	-0.00090	(0.00066)	*	-0.00186	(0.00065)	***
$\Theta_{i,3}$	-0.00350	(0.00050)	***	-0.00779	(0.00110)	***	-0.00251	(0.00059)	***	-0.00328	(0.00058)	***
$\Theta_{i,4}$	-0.00332	(0.00050)	***	-0.00747	(0.00113)	***	-0.00156	(0.00072)	**	-0.00298	(0.00064)	***
$\Theta_{i,5}$	-0.00198	(0.00047)	***	-0.00437	(0.00103)	***	-0.00159	(0.00048)	***	-0.00189	(0.00047)	***
Ξ_i (informed component)	-0.00015	(0.00020)		-0.00051	(0.00023)	**	-0.00001	(2.75E-06)	***	-0.00003	(0.00002)	**
Ω_i (uninformed component)	0.00011	(0.00330)		-0.00023	(0.00042)		-0.00001	(4.63E-06)	***	-0.00006	(0.00008)	
R2	0.437			0.437			0.437			0.427		

Note: The number in parentheses is the standard error calculated with the Newey and West variance-covariance matrix. ***, ** and * indicate 1, 5 and 10 percent levels of significance, respectively. R^2 is the coefficient of determination. The label "Eq. (#)" corresponds to the equation number in the text.

Figure 1

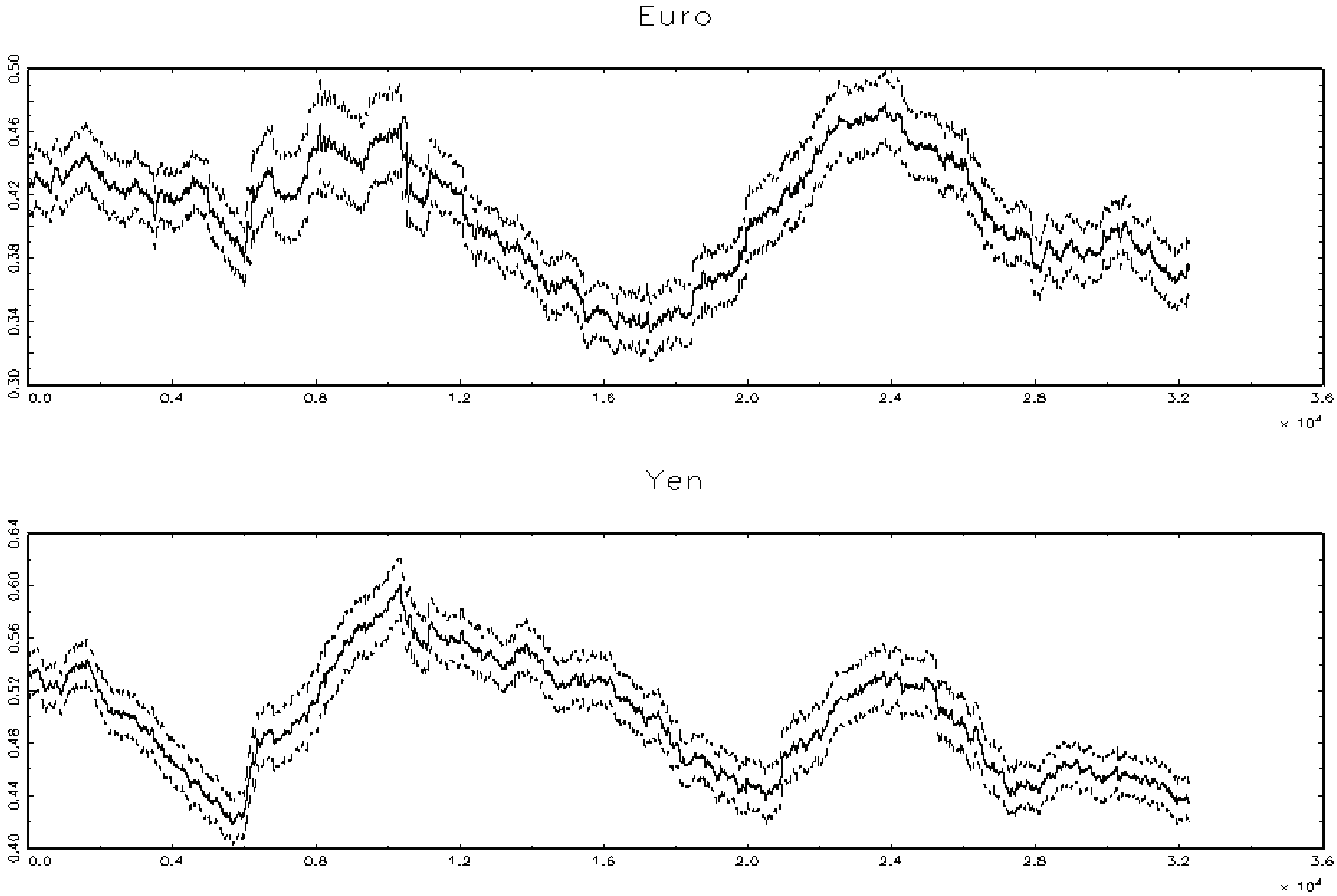
Seasonality of the *second* moments of the euro/dollar and the yen/dollar in each period



Note: The vertical axis shows the averaged second moments of the returns in each interval. The horizontal axis is Greenwich Mean Time (GMT). 1/9/05-30/10/05: America and Europe adopted the daylight saving time (DST). 31/10/05-25/3/06: No DST. 26/3/06-1/4/06: The DST only in Europe. 2/4/06-31/8/06: The DST in America and Europe.

Figure 2

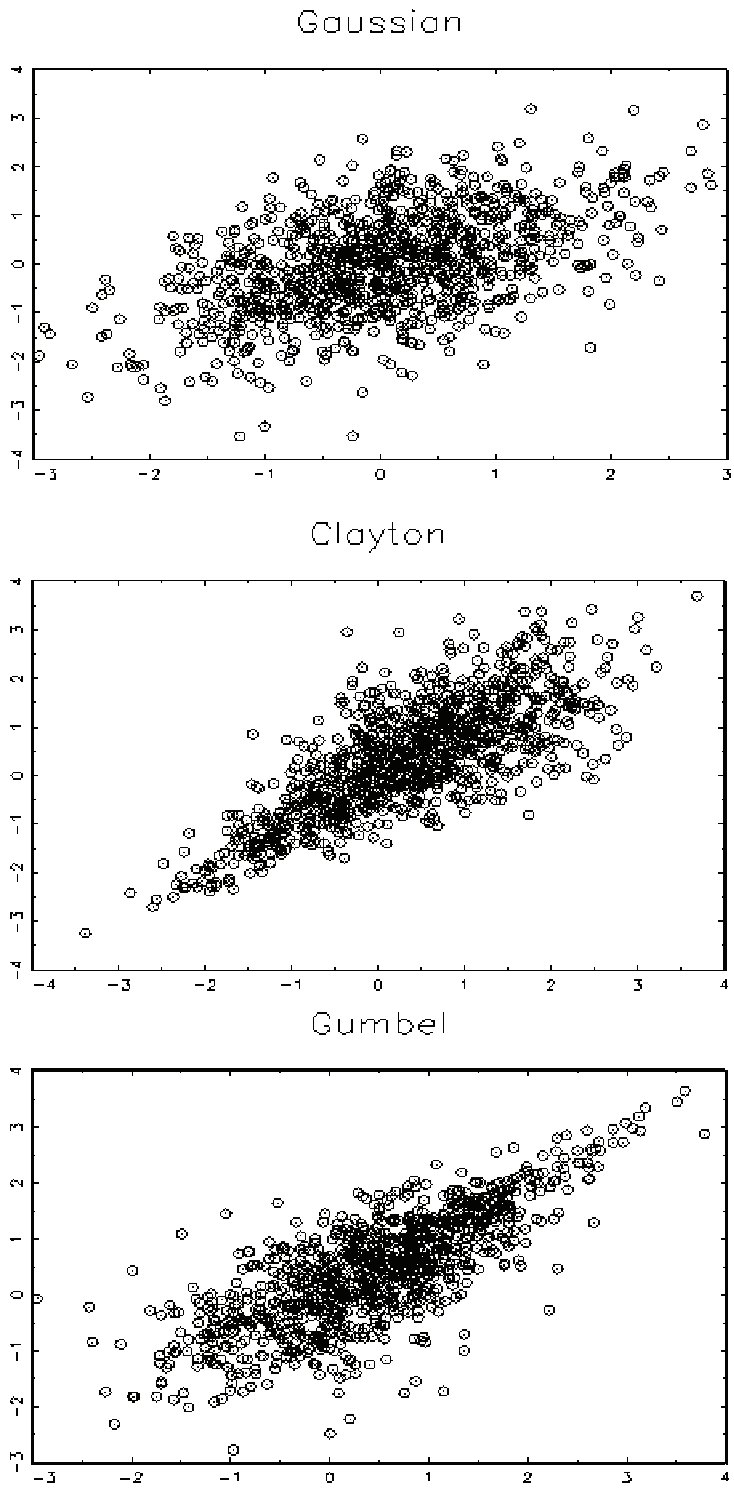
Rolling regressions of return regressed on order flow with 30-day windows



Note: Thin lines are the estimated two-standard error bound calculated with the Newey and West variance-covariance matrix. The lag of the Newey and West variance-covariance matrix is $T^{1/4}$, where T is the sample size. As regressors, I adopted order flows and constant term. Both returns and order flows are seasonally adjusted.

Figure 3

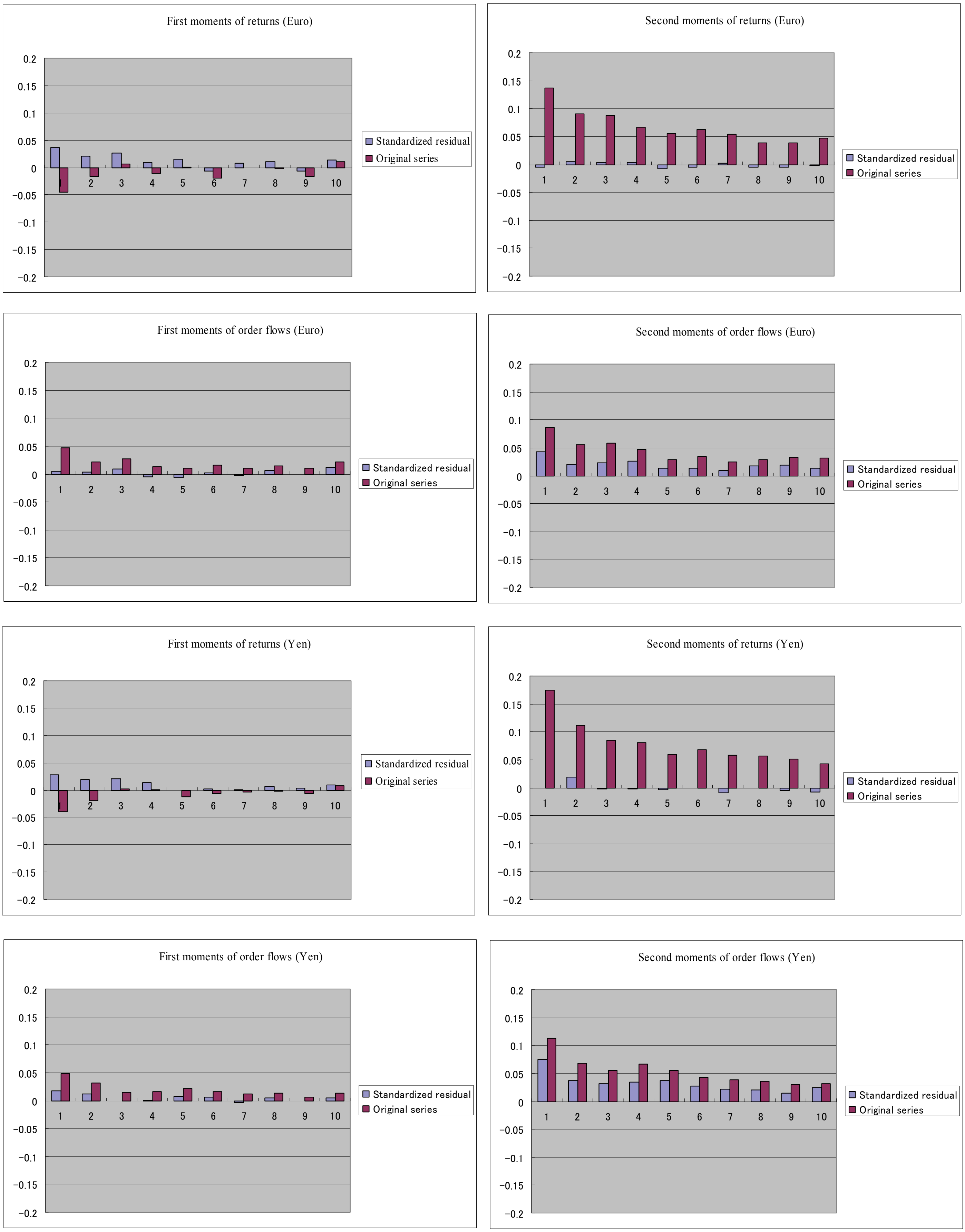
Simulated copulas



Note: In all cases, 1000 observations were generated, and margins were selected as standard normal. The measure of dependence is set to 0.5 in all copulas.

Figure 4

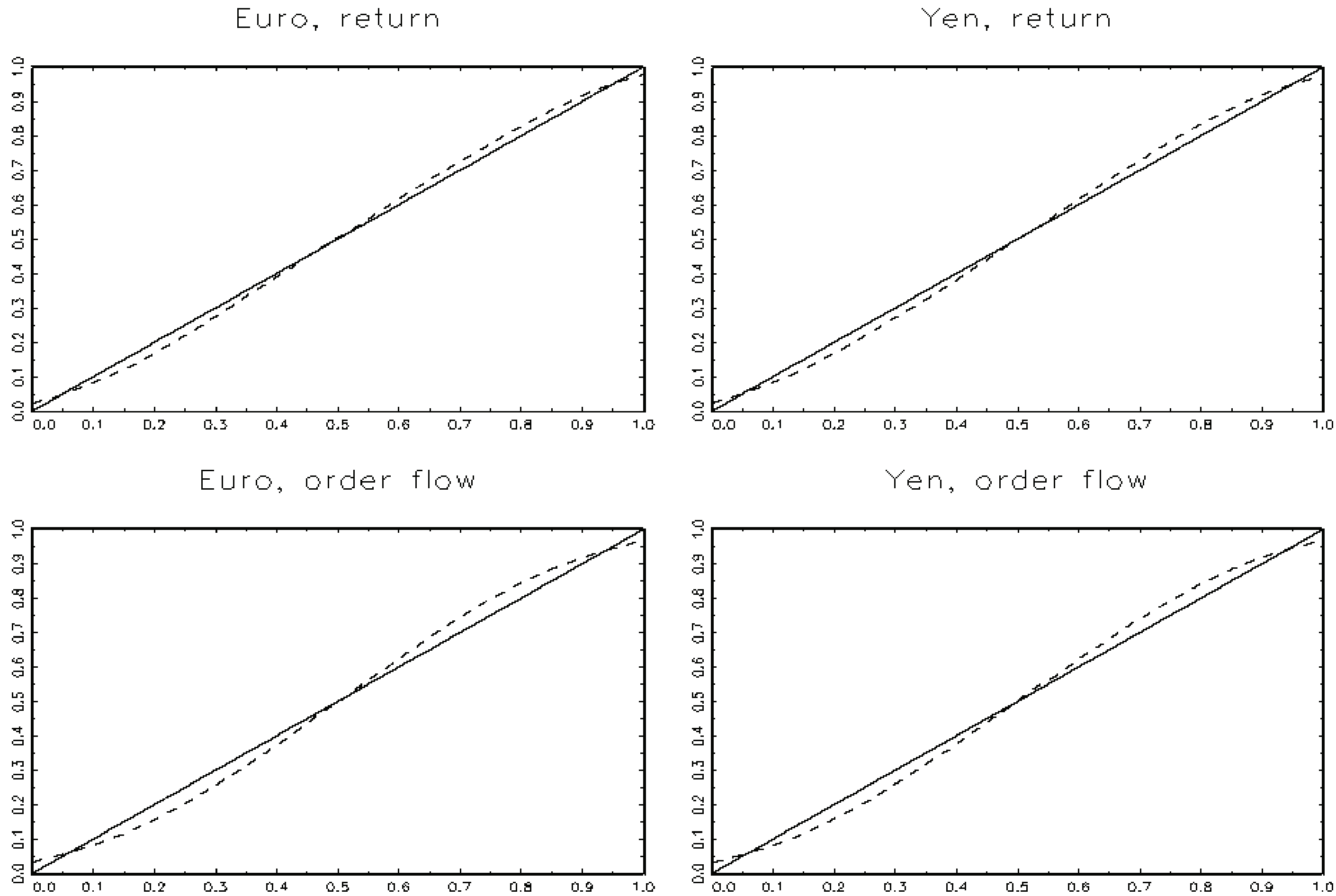
Serial correlations



Note: The original series are seasonally adjusted returns or order flows. Standardized residuals are obtained with estimated ARMA-GARCH models. Numbers of horizontal axis are the lag length of serial correlations.

Figure 5

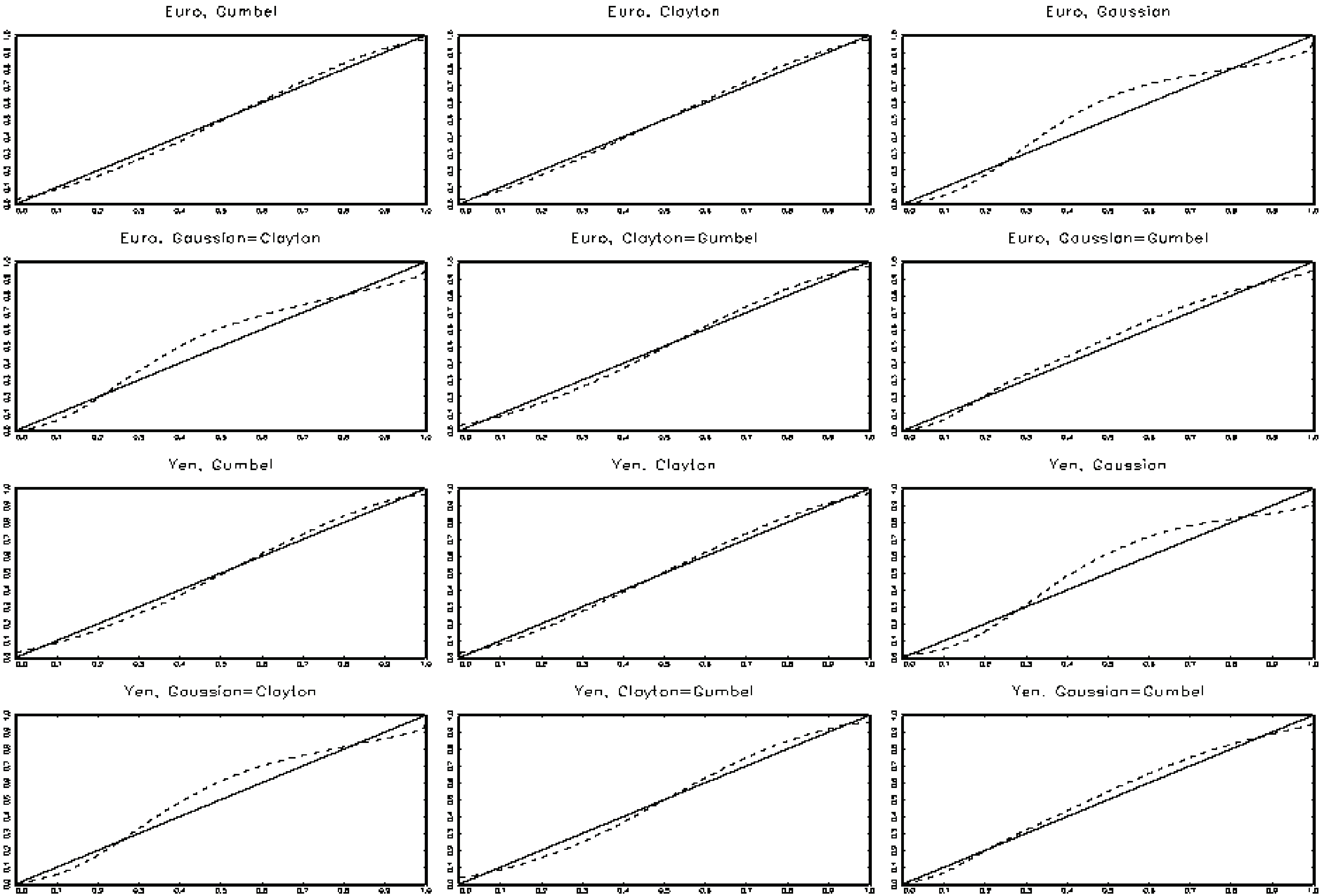
The q-q plots of specified ARMA-GARCH models



Note: The transformed variables of the standardized residuals of ARMA-GARCH is obtained with $H_{\hat{\mu}_i}(\hat{\varepsilon}_{i,t}/\sqrt{\hat{h}_t})$, where $H_{\hat{\mu}_i}$ is a *t*-distribution function with a degree of freedom $\hat{\mu}_i$ and "hat" indicates an estimated parameter or variable. If $H_{\hat{\mu}_i}(\hat{\varepsilon}_{i,t}/\sqrt{\hat{h}_t})$ has a perfect fit with uniform distribution within the interval $[0,1]$, the dashed line should correspond to a 45° solid line connecting (0,0) and (1,1).

Figure 6

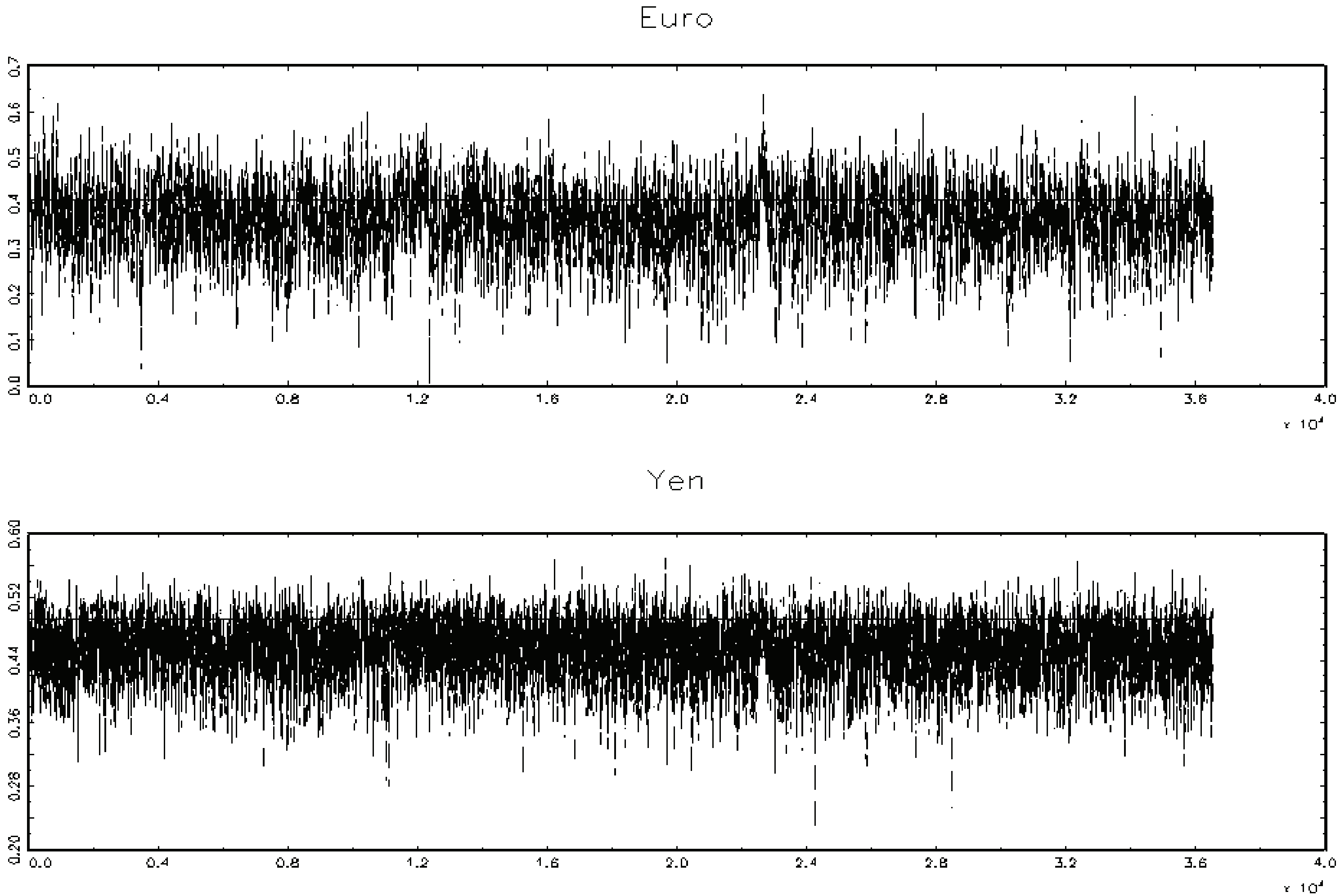
The q-q plots of the copulas



Note: If *C(u,v)* is a copula, where *u=F(x)* and *v=G(y)*, and it is correctly specified, its first derivative with respect to each of its arguments, $\partial C / \partial u$ (dashed line), has a uniform distribution within the interval [0,1] (solid line).

Figure 7

Time varying and constant impact of order flow



Note: The thin line is the time varying impact of order flow. The bold horizontal line is the constant impact. The horizontal axis shows each 10-minute interval (1 is one 10-minute interval) from 1 September, 2005 to 31 August, 2006.