PRICE DISCOVERY ON FOREIGN EXCHANGE MARKETS WITH DIFFERENTIALLY INFORMED TRADERS

Frank de Jong
University of Amsterdam and CEPR

Ronald Mahieu
Erasmus University Rotterdam

Peter Schotman
Limburg Institute of Financial Economics and CEPR

Irma van Leeuwen
Limburg Institute of Financial Economics

This version: August 2001

Abstract: This paper uses Reuters high-frequency exchange rate data to investigate the contributions to the price discovery process by individual banks in the foreign exchange market. We propose multivariate time series models in calendar time as well as in tick time to study the dynamic relations between the quotes of individual banks. We investigate the hypothesis that German banks are price leaders in the Deutschmark/dollar market. Our empirical results suggest an important but not exclusive role for German banks in the price discovery process. There is also a group of banks, German and non-German, that lags behind the market and does not contribute to the price discovery process. We do not find evidence for stronger price leadership of Deutsche Bank on days with suspected Bundesbank interventions in the foreign exchange market.

1 Introduction

It has often been denied that asymmetric information could be important in the foreign exchange market. Because all participants have access to the same public information, the existence of private information seems unlikely. However, Lyons (2001) discusses several channels through which private information could play an important role in the foreign exchange market. Because trading is structured as a decentralized market where traders advertise their quotes on trading screens and strike deals over the phone, the information on order flow and transaction prices remains largely private. Goodhart (1988) and Lyons (1997) argue that this market structure has important implications for the price discovery process on foreign exchange markets. The unavailability of the other traders' order flow induces differential information into the market: traders learn about the fundamental value of the foreign currencies from their own customers' orders only. In the words of Goodhart (1988), also stated in Peiers (1997), "A further source of informational advantage of traders is their access to, and trained interpretation of, the information contained in the order flow". Ito, Lyons and Melvin (1998) provide empirical evidence supporting this view by looking at the volatility of before and after trading restrictions in the Tokyo lunch hour were lifted.[1] Melvin and Covrig (1998) extend this analysis and find that Japanese banks have some form of price leadership in periods with potentially large differences in information about order flow. Lyons (1997) explains the information based trading in the foreign exchange market by the phenomenon of "hot potato trading". Dealers want to pass on inventory imbalances, and thereby have informational effects on the price.

Unfortunately, due to the lack of individual trading data, most of the empirical research does not analyze specific individual behaviour. Only a small number of papers utilize individual trade information. In the foreign exchange microstructure literature, the work of Lyons (1995) is the most prominent example. This lack of emphasis on individual behaviour is unfortunate since potentially much can be learned from individual trade patterns. A study of the behaviour of individual banks may also shed light on the as yet unresolved issue why the quotes are so noisy. We try to see if this is related to strategic interactions between market makers.

This paper therefore investigates the role of individual dealers in the price discovery process on foreign exchange markets. The central hypothesis in our paper is

---

[1]Their conclusions are disputed however, see Andersen, Bollerslev and Das (2001).

that large banks have an information advantage when trading in the foreign exchange market. As large banks will have a much larger customer base, their dealers will have a better view of demand and supply of a currency. We investigate whether the superior information on order flow has a systematic influence on the quotes originating from these banks. A related question is what one can learn about the underlying value from the quotes of large banks. This question is again motivated by the particular structure of the foreign exchange market. There are two indirect channels from which traders can learn about their competitor's orders: via the inter-dealer market and from their quotes.

A particular hypothesis we investigate is the "price leadership" hypothesis put forward by Goodhart (1988). Peiers (1997) recently investigated this hypothesis by studying the strategic interactions on the US dollar/Deutsche Mark market for periods around suspected interventions in the foreign exchange market (related to periods of tension in the EMS). The specific hypothesis tested in her paper was the price leadership of Deutsche Bank in such periods. We contrast the price discovery process around (suspected) interventions with the process on "ordinary" days.

As Peiers (1997) we address these issues using the Olsen HFDF93 dataset. This data set contains a continuous and complete record of buy and sell quotes from the Reuters trading screens. The records do not only give the price information but also reveal the identity of the quote issuer. Although these quotes are indicative and customer oriented, and are therefore probably less informative than firm quote or transaction data, the availability of the trader's identity is a substantial advantage over most micro data sets. An important difference with firm quote data is that the bid/ask spread in the HFDF93 is much wider, but according to Goodhart, Ito and Payne (1996) the midquotes seem to give a good indication of the actual market price.

The methodology of this paper is related to the work of Hasbrouck (1995), who analyzes the price discovery process on related financial markets. Hasbrouck proposes to fit a multivariate time series model to the vector of prices (the bank's quotes in our setup). The model allows for measurement of lead and lag relations between the quote revisions of individual banks, such as to identify price leaders in the market. The time series model must take into account the property that quotes share the same common (stochastic) trend in the long run. In the jargon of multivariate time series analysis, the quotes of different banks are cointegrated. The contributions to the price discovery process by individual market makers are measured by the "information share" of each individual trader, defined as the fraction of the total information in

the market that can be attributed to a particular bank.

Our methodology is similar in spirit to Hasbrouck's, but differs in the choice of the multivariate time series model. Whereas Hasbrouck works with a cointegrated vector autoregressive model, we propose a structural time series model of the type proposed by Harvey (1989). The structural model has some appealing properties. The model directly imposes cointegration between the price quotes of different banks. The model also allows for a parsimonious lag structure. In the structural model, the quotes are composed of two components, one common underlying random walk component, called the *efficient price*, and an idiosyncratic component specific for each bank. The common random walk component makes the quotes of individual banks cointegrated by construction. The random-walk-plus-noise structure is also well suited to describe the empirically observed strong negative first order serial correlation in the quote revisions.[2]

The cointegrated structural model allows for correlation between the innovations in the random walk component and the idiosyncratic component. We propose to use the contemporaneous covariance as a measure of the contribution to price discovery of each bank.

A final methodological point is the treatment of the irregular spacing of the quote data. We use two sampling models: tick time and calendar time. In tick time, each new quote is treated as one observation. The nature of the Olsen data is such that there is only one bank issuing a quote at a time. The other sampling assumption is calendar time, in which we sample quotes on fixed time intervals. Given the irregular quote pattern of banks there will be many intervals without quote updates. The usual procedure of dealing with these is to impute a zero return for such intervals. DeJong, Mahieu and Schotman (1998) demonstrate that this imputation procedure leads to a serious bias in the serial covariance estimates, which may bias the lead-lag patterns in the time series model and may also lead to an overstatement of the accuracy of the results. Therefore, we treat such events as missing observations and use estimation techniques specifically designed to deal with missing observations.

The remainder of the paper is organised as follows. Section 2 provides some key descriptive statistics of the data. Section 3 describes our price discovery model. Sections 4 and 5 report the empirical results for two different sampling assumptions, tick time and calendar time. Section 6 concludes.

---

[2] This point was first made by Zhou (1996), who fits a univariate random walk plus noise model to FX quote data.

2   Data

The data for this paper are taken from the HFDF93 dataset collected by Olsen and
Associates. This dataset contains all quotes on the Reuters screens in the period
October 1, 1992 trough September 30, 1993. A unique feature of this dataset is that
it contains the identity of the bank issuing the quotes. In this section we describe
the aspects of the data that are likely to be different across banks, when banks have
asymmetric information.

Following Peiers (1997) we identify a number of large banks with a potential
information advantage by their quote activity. Table 1 shows the number of quotes
from nine of the most actively quoting banks in the dollar/dmark market. Many of
them are also large in market capitalization. Although the foreign exchange market
is active around the clock, individual banks do not trade 24 hours a day and there are
fewer quotes outside the European business day hours (4-16GMT).[3] This relatively
low activity outside European hours is partly due to the limited use of Reuters screens
in Tokyo and the US (see Goodhart (1989)). Since most quotes are within European
business hours, it is not surprising that all of the most active banks on the Reuters
screen, except for Chemical Bank, are European. The market share of Deutsche Bank
is highest: when it is active, it accounts for almost 9% of all quotes. The selected
banks together make up half of all quotes during European business hours.[4]

The sequencing of quotes plays an important role in our empirical models. We
calculated the Markov transition matrix for the identity of the banks issuing quotes.
This matrix shows no particular pattern, as most conditional probabilities are close
to the marginals. The average duration between two quotes is also more or less
independent of the identity of the issuers. On average, the duration between two
quotes is around 9 seconds.

Huang and Masulis (1999) document that the quoted bid-ask spreads on the
Reuters screens are very stable over the trading day. A very remarkable feature
is the concentration of bid-ask spreads on round numbers, in particular 5 or 10 pips.
With quotes being only indicative in this market, banks do not use the spread be-
tween bid and ask as a signalling device, and spread data are unreliable. On the
other hand, Goodhart, Ito and Payne (1996) compare the Olsen quotes with data

---

[3] In the summer 4am GMT corresponds to 6am in continental Western Europe.

[4] Sapp (2000) reports a higher share in trading activity by US banks in the North American
hours, but Reuters has relatively few quote updates in that time segment. This may partly explain
why we find that European banks are the most active.

from the electronic Dealer 2000 trading system, and conclude that the mid-quote is a fairly good indication of the prices that the bank charges. Evans (1998) also provides evidence in the same direction. Therefore, in the remainder of this paper we will work with the bid-ask midpoint, and not use individual bid and ask quote data or data on spreads.

We now turn to an analysis of the time series properties of quote updates. From this point on we need to be specific on how quotes are sampled. The main choice is between sampling in tick time or in calendar time. Both have advantages and drawbacks. Data in tick time contain all information, automatically put more weight on the periods with most activity, and retain the exact sequencing of quotes. But tick by tick data are also very noisy, so that sampling at such a high frequency could induce erratic patterns in bank identity or quote levels. Therefore, we also take a look at aggregated returns over fixed calendar time intervals. The longer the time interval, the less we need to worry about noise, at the expense of losing information on microstructure detail. For the calendar time we aggregate returns over 30 seconds intervals.

Table 2 reports the variance and autocorrelations of percentage quote revisions in tick time, conditional on the identity of the bank issuing the last quote. There are some pronounced differences in the conditional variances. For example, the quote revisions by UBS are more volatile than average, whereas BHF, Rabobank and Den Norske Bank have less volatile quote updates. The first order autocorrelation is always negative and very close to minus one half. The second order covariance is significantly positive, but all higher order autocorrelations are virtually zero.

The strong negative first order autocorrelation was already documented by Dacorogna et al. (1996), and is typical for a random walk plus noise time series process. The term noise refers to the transitory component in the quotes, which could be the result of price discreteness or microstructure effects. Differences in quoting behavior of individual banks will show up as differences in the transitory components. The relative magnitude of the variance of the random walk component and the variance of the transitory "noise" component determines the strength of the first order autocorrelation. In a pure signal plus noise model, the first order autocorrelation converges to minus one half if the time between quotes shrinks to zero. In our estimates we see that for most banks, and on average, the autocorrelation is only slightly above this value.

For the calendar time analysis we sample the quotes of each bank on 30 second

intervals, thus creating a time series of quotes for each bank. It is important to note that some banks don't update their quote for long stretches of time so that their quotes are often out of line with the market. Because the quotes are indicative only, and carry no obligation to trade at these prices, we feel that such stale quotes are not very reliable indicators of the bank's information. We therefore define only quote updates as observations. If there is more than one quote update in a particular interval, the last quote is treated as the price observation for that interval. If there is no new quote, we treat this period as a missing observation. This approach differs from the usual procedure to substitute a zero return for intervals with no new quote. De Jong, Mahieu and Schotman (1998) show that substituting zero returns for missing observations leads to a serious bias in the estimates of the serial covariances. Instead, we use an estimator proposed by DeJong and Nijman (1997) that delivers consistent estimates of serial covariances of returns (quote revisions) in the presence of missing observations.

Table 3 reports the variance and autocorrelations of quote revisions on a 30 second interval. The table also reports the fraction of 30 second intervals with a quote revision by each bank. For most banks, this fraction is between 7% and 25%. Again, we see major differences in the variance and the first order serial correlation. Typically, the first order serial correlation is negative and stronger so for the banks with a higher variance. Again, this points at the presence of a substantial transitory ("noise") component in the quotes. The second order autocovariance is typically very small.

Finally, we look at the cross correlations between the quote revisions. In principle, we could correlate the returns of each pair of banks, but it turned out that due to the relatively large number of missing observations for some banks, and the distinct trading patterns, this didn't give very stable estimates. Instead, we calculate only the covariances of the returns of each bank with the "market", which is defined as all the other banks. Since the market is the aggregate of all banks but one, the time series of market quotes has far fewer missing observations: around 93% of the 30 second intervals has a quote update. Figure 1 depicts the cross correlations of the major banks with the market. For some of the banks the market is clearly leading (BHF, Rabobank). The peak in the cross correlation function is at negative lags, indicating that these banks lag behind the market. The cross covariances for positive lags (leads of the banks) are all virtually zero. For most other banks the contemporaneous cross correlation is sizeable. Three banks (UBS, Dresdner Bank, Societé Generale)

also have significant lead correlations with the market.[5] Notice that these lead and lag covariances are not an artifact of thin trading. Our estimation methodology is especially designed to estimate the lead-lag covariances of the "true" returns, even if there are intervals without a quote revision.

We now summarize the results in this section. There are substantial differences in the activity of banks, both in number of quotes and in the time of day when the banks are most active. Second, there is no clear pattern in the sequence of bank identities. The quote changes are small (in absolute value), but noisy. The variance of quote revisions differs substantially among banks, but for all banks there is a very pronounced first order serial correlation. There are significant covariances between the quote changes of the banks. Individual banks lag behind the market at most one minute, so convergence to the overall price level is relatively quick. A few banks actually lead the market.

3   Unobserved Components price discovery model

Although the previous results indicate important differences among major banks, we cannot say much about the relative contribution of each bank to the price discovery process yet. In this section we therefore propose a multivariate time series model for price discovery, that allows for an assessment of the individual bank's information.

The main idea of the model is that price quotes of all banks are derived from one common, but unobserved efficient price. We assume that the quotes equal the efficient price times an idiosyncratic component that can be either noise or reflect the strategic behavior of a bank. This component will be specific to the bank that sets the quote. The idea of separating the fundamental or efficient price from the idiosyncratic component was introduced in the market microstructure literature by Hasbrouck (1991, 1993). Zhou (1996) fits a univariate random walk plus noise model to the Olsen FX data. We extend these ideas to a multivariate model. To fix some notation, let P^* be the efficient price, P the vector of quoted bid-ask midpoints, and U the vector of idiosyncratic components, where the ith elements of P and U refer to bank i. We consider $n-1$ individual banks and a rest category that we will call the

---

[5] The figure does not show standard errors. Several estimators for the standard errors are discussed in DeJong and Nijman (1997). For the order of magnitude of the standard errors one could take the number of observations for which data in both series are available. In that case all correlations larger than 0.015 are significantly different from zero.

"market". Letting $p = \ln P$, $p^* = \ln P^*$, and $u = \ln U$, we have that the logarithm of the n-vector of quotes equals

$$p_t = \iota p_t^* + u_t, \tag{1}$$

where ι is an n-vector of ones. The aim of this section is to develop a multivariate time series model for this quote vector. The main assumption of the model is that the efficient price p^* is a random walk with serially uncorrelated increments. The noise term u is assumed to be transitory and takes account of all temporary deviations of the quotes from the efficient price. Notice that the quotes of all banks share the same random walk component. Therefore, by construction, the quote series are coin-tegrated. Economically this is a very intuitive restriction, since we expect the quotes of each bank to revert to the same efficient price in the long run. The random walk assumption for the efficient price excludes fundamental exchange rate predictability, but for intraday data this is a good approximation and it is standard in the price discovery literature (see e.g. Hasbrouck, 1995).[6]

We now state the assumptions of our model more formally. Suppose a specific sampling scheme for the data is given, and data are (potentially) recorded at times $t = 1, .., T$. Denote the change of the efficient price over the interval $(t-1, t)$ by

$$r_t = p_t^* - p_{t-1}^* \tag{2}$$

The maintained assumption of the price discovery model is that the unconditional serial covariances of r_t and u_t (the vector with elements u_{it}) are stable in the given sampling interval. We make the following assumptions

$$\mathsf{E}[r_t^2] = \sigma^2 \tag{3a}$$

$$\mathsf{E}[r_t u_t] = \psi \tag{3b}$$

$$\mathsf{E}[r_t u_{t+\ell}] = \gamma_\ell, \quad \ell \geq 1 \tag{3c}$$

$$\mathsf{E}[r_t u_{t-k}] = 0, \quad k > 0 \tag{3d}$$

$$\mathsf{E}[u_t u_t'] = \Omega \tag{3e}$$

$$\mathsf{E}[u_t u_{t-k}'] = 0, \quad k \neq 0, \tag{3f}$$

where ψ and γ_ℓ are $(n \times 1)$ and where Ω is an $(n \times n)$ matrix. These assumptions state that the fundamental news r_t is serially uncorrelated. The fundamental news r_t

---

[6] Harris, McInish and Wood (2000) propose an alternative permanent transitory decomposition, where the permanent component is a weighted average of the current price vector. However, in that decomposition the efficient price is not a random walk. This makes it hard to interpret the correlation structure between the permanent and transitory component.

and the idiosyncratic component u_t are uncorrelated at all lags, but may be correlated at leads of r_t to future u_t. The idiosyncratic component u_t is serially uncorrelated.[7] The model therefore has a specific unobserved components (UC) structure, namely a random walk plus noise (see Harvey, 1989, for an introduction to structural time series models). There is only contemporaneous correlation between the news and the noise; introducing serial correlation in r_t or cross correlation between r_t and u_t at leads and lags will lead to an underidentified model (again, see Harvey (1989) for a discussion of the structure and identification of the unobserved components model).

For the serial covariance properties of the model we consider the vector of quote changes Y_t with elements $y_{it} = p_{it} - p_{i,t-1}$,

$$Y_t \equiv p_t - p_{t-1} = \iota r_t + u_t - u_{t-1} \tag{4}$$

Given the assumptions in (3a)-(3f) the serial covariances of Y_t are

$$\mathsf{E}[Y_t Y_t'] = \sigma^2 \iota \iota' + \iota \psi' + \psi \iota' + 2\Omega \tag{5a}$$

$$\mathsf{E}[Y_t Y_{t-1}'] = -\psi \iota' - \Omega + \gamma \iota' \tag{5b}$$

$$\mathsf{E}[Y_t Y_{t-2}'] = -\gamma \iota' \tag{5c}$$

A very useful property of the 'random walk plus noise' structure of the model is its capability to describe the typical autocorrelation pattern of quote changes in a parsimonious way. As an illustration, consider the simplest model where $\gamma = 0$. In that case the non-zero moments are

$$\mathsf{E}[y_{it} y_{jt}] = \sigma^2 + 2\omega_{ij} + \psi_i + \psi_j \tag{6a}$$

$$\mathsf{E}[y_{it} y_{j,t-1}] = -\omega_{ij} - \psi_i \tag{6b}$$

The first order autocorrelation is

$$\rho_{1,ii} = \frac{-(\omega_{ii} + \psi_i)}{\sigma^2 + 2(\omega_{ii} + \psi_i)} \tag{7}$$

The lowest value this correlation can take is -0.5. We therefore expect $\omega_{ii} + \psi_i$ to be large relative to σ^2, given the empirically observed autocorrelations close to -0.5.

The most interesting parameter in the model is ψ, the covariance between the fundamental price change and the idiosyncratic shocks. This parameter determines

---

[7] Alternatively, one may assume that the lead and lag correlations between r_t and u_t are zero, and the autocovariance structure of u_t is unrestricted. This model is slightly more general than the model proposed, but the parameters are more difficult to interpret.

what we can learn from a quote adjustment of individual banks. To illustrate this point, consider the conditional covariance between quote changes and the fundamental price change,

$$\text{Cov}(p_{ti} - p_{t-1,j}, \, r_t) = \sigma^2 + \psi_i, \qquad (j = 1, \ldots, n), \tag{8}$$

given the quote by bank i. Inference on the fundamental news component r_t given observed quote changes is largely driven by this covariance. The higher ψ_i, the stronger the signal that a quote update by bank i sends and the more informative its quotes are. But on average, the information generated in each period or by each quote update should equal σ^2, the variance of r_t. Notice also that the n covariances between the quote updates and the news is determined by $n + 1$ parameters. In the model there is no way to identify these separately, so an identifying restriction is needed. The most natural restriction is found by considering the average covariance between the quote update of an arbitrarily selected bank and the news

$$\sum_{i=1}^{n} \pi_i \text{Cov}(p_{ti} - p_{t-1,j}, \, r_t) = \pi' \left(\sigma^2 \iota + \psi \right) = \sigma^2 + \pi' \psi \qquad (j = 1, \ldots, n) \tag{9}$$

where π is a vector of weights adding to one. In order to give the natural interpretation of σ^2 as the unconditional covariance of a quote update and the news, we impose the restriction $\pi' \psi = 0$. It turns out that this restriction is sufficient to identify the parameters σ^2 and ψ. This assumption also leads to a natural definition of the information share of a bank if we let π_i be the fraction of quotes issued by bank i in the sample. Consider the covariance between the fundamental price revision and the quote change of bank i in (8). Multiplying this covariance with the probability of the quote being originated by bank i gives a natural measure of how much information is generated by the quote updates of bank i. Dividing this through by the total information generated in the market, i.e. the variance of r_t, we obtain the information share of bank i

$$\beta_i = \frac{(\sigma^2 + \psi_i) \, \pi_i}{\sigma^2} = \pi_i \left(1 + \frac{\psi_i}{\sigma^2} \right) \tag{10}$$

Given the restriction $\pi' \psi = 0$, the information shares add up to 1. This definition of the information share has some appealing properties. First (ceteris paribus) a bank that issues more quotes has a larger information share. This definition also implies that the joint information share of two banks equals the sum of their individual information shares. A bank with a high contemporaneous covariance between its idiosyncratic term and the fundamental news (a high value of ψ_i) has a high information

share. A test of $\psi_i = 0$ amounts to a test of the hypothesis that a bank's information share equals its activity share (in terms of number of quote updates).

We now provide a comparison of the proposed information share measure to the more traditional one of Hasbrouck (1995) and the recently proposed measure based on the common factor components of Gonzalo and Granger (1995), which was applied in the market microstructure literature by Booth, So and Tse (1999), Chu, Hsieh and Tse (1999) and Harris, McInish and Wood (2000). Both the Hasbrouck and Granger-Gonzalo measures are based on the reduced form moving average representation

$$\Delta Y_t = \iota \theta' \epsilon_t + C \Delta \epsilon_t \tag{11}$$

Comparing this to the structural price discovery model

$$\Delta Y_t = \iota r_t + u_t - u_{t-1} \tag{12}$$

we can identify $r_t = \theta' \epsilon_t$ as the common permanent component. An interpretation of θ is as the regression coefficient in a regression of r_t on ϵ_t, and therefore can be written as $\theta = \mathrm{Var}(\epsilon_t)^{-1} \mathrm{Cov}(\epsilon_t, r_t)$. The Granger-Gonzalo information shares are given by θ itself. The information share in Hasbrouck's setup is given by the diagonal elements of the sum $\theta' \mathrm{Var}(\epsilon_t)\theta$. Our information share is defined as $\mathrm{Cov}(\Delta Y_t, r_t)$, appropriately normalized. Notice that because r_t is unpredictable from past information, $\mathrm{Cov}(\Delta Y_t, r_t) = \mathrm{Cov}(\epsilon_t, r_t) = \mathrm{Var}(\epsilon_t)\theta$. After normalization, our proposed vector of information shares is $\beta = \frac{1}{\sigma^2} \mathrm{Var}(\epsilon_t)\theta$ and can be interpreted as the coefficient vector in a regression of ϵ_t on r_t.

The advantage of our information share over Hasbrouck's definition is that it is uniquely defined. The advantage over the Granger-Gonzalo method is that our measure does take the variance of the innovations into account, so that a series with low innovation variance gets a low information share. The measure can be applied easily to both structural and the usual VECM models.

4 Price discovery in tick time

A typical feature of the Olsen dataset is that each quote is at least two seconds apart from the previous one; there are no simultaneous quote updates. Therefore, the observation scheme is somewhat different from the one assumed in the general discussion of the model. In this section we present the empirical model for the tick-by-tick data, and the empirical results of estimating the model in 'tick time'.

4.1 The model in tick time

In tick time the observed quote revision is the scalar $y_t = J'_t p_t - J'_{t-1} p_{t-1}$, where J_t is the vector with the bank's identity: $J_t = e_i$, the ith unit vector, if the quote update at time t is originated by bank i. Analogous to (4) the unobserved components model takes the form

$$y_t = r_t + J'_t u_t - J'_{t-1} u_{t-1}. \tag{13}$$

We first show that the parameters of the tick time price discovery model can be estimated by a regression analysis that exploits the variance and covariances of consecutive prices differences. For clarity, we discuss identification for the simplest model with serially uncorrelated idiosyncratic terms, but the results also hold in a model with a more general lag structure.[8]

The moments of the model, conditional on the sequence of bank identities, are given by

$$\mathsf{E}[y_t^2 | J_t, J_{t-1}] = \sigma^2 + (J_t + J_{t-1})' \omega^2 + 2J'_t \psi \tag{14a}$$

$$\mathsf{E}[y_t y_{t-1} | J_t, J_{t-1}] = -J'_{t-1} \omega^2 - J'_{t-1} \psi + J'_t \gamma \tag{14b}$$

$$\mathsf{E}[y_t y_{t-2} | J_t, J_{t-1}] = -J'_{t-1} \gamma \tag{14c}$$

where ω^2 is the vector with diagonal elements of Ω. Notice that the off-diagonal elements of Ω are absent from these equations. The reason for this is very simple: in tick time we never simultaneously observe quote revisions by two or more banks, we only have (by construction) sequential quote revisions. As a result, only the diagonal elements of the contemporaneous variance-covariance matrix of the vector of quote revisions can be identified. Moreover, due to the sampling in tick time, the bank identity vector J_t contains only one non-zero element. Hence, $\iota' J_t = 1$ for all t and therefore in the first equation one of the elements of J_t and J_{t-1} is redundant. As in the previous section we need the identifying restriction $\pi' \psi = 0$.

The moment conditions suggest a SUR regression model to estimate the parameter vector $\theta = (\sigma^2, \omega^2, \psi, \gamma)'$. Define the matrix of explanatory variables

$$X_t = \begin{pmatrix} 1 & 0 & 0 \\ J_t + J_{t-1} & -J_{t-1} & 0 \\ 2J_t & -J_{t-1} & 0 \\ 0 & J_t & -J_{t-1} \end{pmatrix} \tag{15}$$

---

[8] This is because the higher order serial covariances $\mathsf{E}[y_t y_{t-k} | J_t, .., J_{t-k}]$ for $k > 1$ are only determined by the higher order covariances of the noise term with the efficient price.

Let $z_t = (y_t^2, y_t y_{t-1}, y_t y_{t-2})'$, and let η_t be a (3×1) error vector. Consider the system of regression equations

$$z_t = X_t'\theta + \eta_t \tag{16}$$

Because the term y_t is common to y_t^2, $y_t y_{t-1}$ and $y_t y_{t-2}$ the elements of the error term η_t are likely to be mutually correlated. For the GLS estimator we assume that the covariance matrix of the errors is

$$\mathsf{E}[\eta_t \eta_t'] = \Sigma \tag{17}$$

In practice, we obtain initial estimates using OLS on the system and from there we apply two rounds of feasible GLS estimation.[9] Because of the underidentification of the model, ψ is estimated subject to the constraint $\pi'\psi = 0$.[10]

4.2 Empirical results for the full sample

Table 4 shows the two round SUR estimates of the tick time price discovery model over the full sample. For legibility, the returns have been multiplied by 10^4 before estimating the model, so the unit of the return is one basis point. The fundamental news innovation variance is estimated at 0.50, which (with an average duration between quotes of 9 seconds) corresponds to a standard deviations of returns over one trading day (4-16) of 0.5%. The estimates of the noise variance are much larger than the fundamental variance; for most banks the noise to signal ratio is around 6. Absent any other correlations, this would imply first order autocorrelations of around -0.43, which is close to the sample autocorrelations. For Societé Generale and ABN-AMRO the noise is smaller than for other banks, but still much larger than the fundamental news variance.

We now turn to the covariance between the news and noise term and the information shares, ψ_i, which forms the basis of our information share definition. There appears to be a dichotomy in the banks analysed. One group of banks has positive estimates for ψ_i and positive information shares. The banks that belong to this group are Deutsche Bank, Societé Generale, Dresdner Bank, ABN-AMRO and, perhaps,

---

[9] Standard errors are calculated from the usual SUR variance-covariance matrix $\mathrm{Var}(\hat{\theta}) = (X'(\hat{\Sigma}^{-1} \otimes I_n)X)^{-1}$ where X is the matrix that stacks $X_1', .., X_n'$.

[10] We replace the term $J_t'\psi$ in the model by $\bar{J}_t'\bar{\psi}$, where $\bar{J}_t = A'J_t$ and A is a $n \times (n-1)$ matrix with property $\pi'A = 0$. The moment equations of the SUR system are adjusted accordingly. The coefficients ψ are then found as $\psi = A\bar{\psi}$ and automatically satisfy $\pi'\psi = 0$.

UBS. Notice that these are large banks and two of these are the major German players in this market. The "rest" category of banks is relatively important as well, the information share of this category is higher than its activity share. The remaining banks have a negative estimate for ψ_i and very small or even negative implied information shares. As a formal test, a significantly positive ψ_i implies an information share that is significantly larger than the activity share of this bank. This holds for all banks in the first group, and for the "rest" category. The second group has significantly negative estimates of ψ_i, and hence information shares below the activity share. The estimates of γ_i for the second group of banks, i.e. the group with low information shares, is typically negative.

To summarize the results of this part, we draw the following conclusions. The quotes of the large German banks are informative, and these players appear to be price leaders. Price leadership is not exclusive for German banks, however. There are also non-German banks that provide relatively informative quotes and are price leaders. On the other hand, there is a second group of banks that lag behind the market and give quotes that are not informative.

4.3   Empirical results for intervention days

The sample period contains a number of days where the Deutsche Bundesbank intervened in the foreign exchange market. Peiers (1997) reports the suspected days and times of these interventions. The period around interventions is particularly interesting for the investigation of information effects in the foreign exchange markets because information differentials may be more pronounced in such periods. Peiers (1997) argues that Deutsche Bank is often used by the Bundesbank as their agent for interventions. This would give DB an informational advantage over other banks because they know an important part of the order flow, which is not observed by other participants. Indeed, Peiers (1997) concludes that DB is a price leader in the dmark-dollar market around interventions. This informational advantage could also hold for other German banks that are closer to rumours about interventions than non-German banks.[11]

In this part we will reconsider the evidence in Peiers using our price discovery model. We feel that our tick time model is more suitable for the question than the calendar time VAR model used by Peiers. The tick time model avoids the interpola-

---

[11] Melvin and Covrig (1998) find price leadership of Japanese banks in the yen/dollar market in periods with suspected strong information differentials.

tion of quotes over one-minute intervals that Peiers employs to deal with the irregular spacing of quotes. Instead, we look at the tick-by-tick price patterns. Another contribution of our analysis is the inclusion of all banks in the model.[12]

Table 5 reports the estimates of the tick time price discovery model for the sample period of one hour before to one hour after the intervention times as reported in Table I of Peiers (1997). In total, there are 10859 quote updates in this sample, which amounts to one quote per 9 seconds, which is very similar to the quote frequency in the full sample. Looking at the parameter estimates, we see that the fundamental variance more than doubles in the intervention periods. The estimates of the noise variance are not systematically bigger than the full sample estimates, however. Also, there are few changes in the estimates of the information shares. An exception is Dresdner Bank, which has a significantly negative estimated ψ_i and a very low information share. In contrast, in the full sample Dresdner Bank was one of the more informative players. We repeated this analysis for a smaller window of 25 minutes (both ways) around intervention time and found similar results.[13]

We conclude that there is more fundamental volatility during the intervention periods, but there are no clear differences in price leadership and information patterns. In particular, there seems to be little evidence for unusual price leadership of Deutsche Bank and no evidence for a special role of other German banks.

5 Price discovery in calendar time

The tick time assumption made in the previous section has some potential drawbacks. Because of the very short interval between quote updates, the sequencing of the quotes, and hence the bank identities, may contain some inaccuracies. Moreover, longer run price patterns may be obscured by the large "noise" component in the quotes. It would be interesting to see whether the same information shares and price leadership patterns are found when we use a different sampling assumption. In this section we therefore study the longer run interactions between banks in calendar time.

The calendar time analysis proceeds in two steps. First, we create a time series of quotes for each bank on 30 second intervals. This interval is short enough to show lead-lag relationships between the quote updates of individual banks, but it is longer than the average interval between two "ticks" (9 seconds), so that we smooth out some

---

[12] Peiers considers only the six largest banks.

[13] Details are available upon request from the authors of this paper.

potential erratic short term patterns.[14] In this sampling scheme there are sometimes multiple quote updates in a period, but sometimes there is no quote update at all. We deal with the former problem by taking the last price observed in an interval. The second problem (no quote updates) is dealt with using the DeJong and Nijman (1997) procedure, that allows for consistent estimates of serial covariances in time series with missing observations.

The second step in the calendar time model is to estimate the auto and cross serial covariances from the fixed interval time series. Due to the relatively large fraction of missing observations and the distinct trading pattern of the banks it was not feasible to estimate the full covariance matrices for a system of 10 banks. Instead, we estimated the serial covariances between the quote revisions of each bank and the "market", defined as the collection of all other banks.[15] So, only the "own" serial covariances $\mathsf{E}[y_{it}y_{i,t-k}]$ and the covariances with the market, $\mathsf{E}[y_{0t}y_{i,t-k}]$, were estimated. Following the moment conditions in equations (5a)–(5c), these covariances allow for identification of the parameters σ^2, ω_{ii}, ω_{i0}, ω_{0i}, for $i = 1, .., n-1$, and the vectors ψ and γ.

Table 6 reports the estimates of the price discovery model based on these moments. We let σ^2 be determined only by the summed autocovariances of the market. The reason for this is that the market series has the lowest fraction of missing observations. The variance of the fundamental news component is 1.81, which is about 3.6 times the estimate in the tick time model. This is what could be expected because on average there are around 3.4 quote updates in each 30 second interval. Just like in the tick time model we find that the variance of the noise component is large relative to the fundamental variance. As for the information shares, the lead of the market to the individual banks gives negative estimates of ψ_i and information shares that are lower than the activity share for the individual banks. Among banks we find the same pattern as in tick time, with Deutsche Bank and Societé Generale having the highest information share, and a group of uninformative banks. The market has the highest information share, also relative to its activity share.

---

[14] See DeJong, Mahieu and Schotman (1998) for a motivation of this interval length.

[15] For reasons of stability, we decided to estimate S and D from the sample covariances C_k over a relatively long lead and lag, up to $L = 10$.

6   Conclusion

In this paper we investigated the role of individual banks in the price discovery process on the foreign exchange market. Large banks are the main participants in this market and each bank has some private information concerning the order flow. We investigate whether this information structure leads to differences in the contribution to price discovery and to specific dynamic interaction between banks. In particular, we investigate the hypothesis of price leadership by German banks in the deutschmark/dollar market.

The empirical results indicate that there are clear differences in the contribution to price discovery among banks. There is one group of banks with a relatively high information share. We find an important, but not exclusive, role for large German banks in this group. For example, Deutsche bank is an important market leader, but there are other banks (also non-German) with comparable information shares. In contrast to this informative group of banks there is another group with very low information shares. The quote revisions of these banks typically lag behind the market quote revision.

To check the robustness of these results to the sampling assumption, we estimated the price discovery model both on a tick-by-tick basis and on fixed 30 second intervals. The findings of the tick time analysis are confirmed by the calendar time analysis, although we find slightly smaller information shares for the major dealers and somewhat higher share for the remaining banks.

The final topic we investigated is the price leadership pattern on days with suspected Bundesbank intervention in the foreign exchange market. In contrast to Peiers (1997) we do not find a special role for German banks in the hours around intervention times, and we do not find an unusual price leadership of Deutsche Bank in these periods either.

References

ANDERSEN, T.G., T. BOLLERSLEV, AND A. DAS (2001), Variance Ratio Statistics and High Frequency Data: Testing for Changes in Intraday Volatility Patterns, *Journal of Finance*, **56**, 305–327.

CHU, Q.C., W.G. HSIEH, AND Y. TSE (1999), Price Discovery on the S&P 500 Index Markets: An Analysis of Spot Index, Index Futures and SPDRs, **8**, 21–34.

DACOROGNA, M.M., C.L. GAUVREAU, U.A. MÜLLER, R.B. OLSEN, AND O.V.

PICTET (1996), Changing Time Scale for Short-term Forecasting in Financial Markets, *Journal of Forecasting*, **15**, 203–228.

DE JONG, F., R.J. MAHIEU, AND P.C. SCHOTMAN (1998), Price Discovery in the Foreign Exchange Market, *Journal of International Money and Finance*, **17**, 5–27.

DE JONG, F. AND T. NIJMAN (1997), High Frequency Analysis of Lead-Lag Relationships between Financial Markets, *Journal of Empirical Finance*, **4**, 259–277.

EVANS, M.D.D. (1998), The Microstructure of Foreign Exchange Dynamics, working paper, Georgetown University.

GONZALO, J. AND C.W.J. GRANGER (1995), Estimation of Common Long-Memory Components in Cointegrated systems, *Journal of Economics and Business Statistics*, **13**, 27–36.

GOODHART, C.A.E. (1988), The Foreign Exchange Market: A Random Walk with a Dragging Anchor, *Economica*, **55**, 437–460.

GOODHART, C.A.E. (1989), News and the Foreign Exchange Market, Working paper.

GOODHART, C.A.E., T. ITO, AND I. PAYNE (1996), One day in June 1993: A study of the working of Reuters D2000-2 electronic foreign exchange trading system, in G.Galli J. Frankel and A. Giovannini, (eds.), *The Microstructure of Foreign Exchange Markets*, University of Chicago Press.

HARRIS, R., T.H. MCINISH, AND R.A. WOOD (2000), Security Price Adjustment Across Exchanges: An Investigation of Common Factor Components for Dow Stocks, Working Paper.

HARVEY, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

HASBROUCK, J. (1991), The Summary Informativeness of Stock Trades: An Econometric Analysis, *Review of Financial Studies*, **4**, 571–595.

HASBROUCK, J. (1993), Assessing the Quality of a Security Market: A New Approach to Transaction-Cost Measurement, *Review of Financial Studies*, **6**, 191–212.

HASBROUCK, J. (1995), One Security, Many Markets: Determining the Contributions to Price Discovery, *Journal of Finance*, **50**, 1175–1199.

HUANG, R.D. AND R.W. MASULIS (1999), FX Spreads and Dealer Competition Across the 24 Hour Trading Day, *Review of Financial Studies*, **12**, 61–93.

ITO, T., R.K. LYONS, AND M.T. MELVIN (1998), Is there Private Information in the FX Market: The Tokyo Experiment, *Journal of Finance*, **53**, 1111–1130.

LYONS, R.K. (1995), Tests of Microstructural Hypotheses in the Foreign Exchange Market, *Journal of Financial Economics*, **39**, 321–351.

LYONS, R.K. (1997), A Simultaneous Trade Model of the Foreign Exchange Hot Potato, *Journal of International Economics*, **42**, 275–298.

LYONS, R.K. (2001), *The Microstructure Approach to Exchange Rates*, MIT Press, Forthcoming.

MELVIN, M. AND V. COVRIG (1998), Asymmetric Information and Price Discovery in the FX Market: Does Tokyo Know More about the Yen, working paper, Arizona State University.

PEIERS, B. (1997), Informed Traders, Intervention, and Price Leadership: A Deeper View of the Microstructure of the Foreign Exchange Market, *Journal of Finance*, **52**, 1589–1614.

SAPP, S. (2000), Price Leadership in the Spot Foreign Exchange Market, Working paper, University of Western Ontario.

ZHOU, B. (1996), High-Frequency Data and Volatility in Foreign-Exchange Rates, *Journal of Economics and Business Statistics*, **14**, 45–52.

Table 1: Most Active Banks

Bank	Acronym	Total quotes	quotes within trading hours
Deutsche Bank	DB	121,055	101,355
Chemical Bank	CHEM	67,554	54,372
BHF Bank	BHF	70,227	56,586
Rabobank	RABO	57,274	53,672
Societé Generale	SG	72,891	67,014
Der Norske Bank	NORS	38,958	38,902
Dresdner Bank	DRES	39,875	32,891
Union Bank of Switzerland	UBS	38,600	30,285
ABN-AMRO Bank	ABNA	35,971	25,595

Notes: Entries show the total number of quotes by a bank in the HFDF93 data. Trading hours are defined as the twelve hour period 4-16 GMT.

Table 2: Variance and autocorrelations of quote revisions in tick time

Bank	Var	ρ_1	ρ_2
DB	6.33	-0.38	-0.03
CHEM	5.93	-0.47	-0.05
BHF	4.37	-0.61	-0.10
RABO	4.97	-0.54	-0.01
SGEN	5.16	-0.48	0.01
NORS	4.74	-0.62	-0.04
DRES	6.92	-0.40	-0.01
UBS	8.10	-0.39	0.02
ABNA	5.24	-0.42	0.05
rest	6.37	-0.42	0.02
Overall	5.91	-0.46	0.03

Notes: Entries show the conditional variance and first and second order autocorrelations of mid-quote revisions $\Delta \ln P$ by bank k, irrespective of which bank issued the earlier quotes. Quote revisions have been multiplied by 10^4.

Table 3: Variance and autocorrelations of quote revisions in calendar time

	Var	ρ_1	ρ_2	obs
DB	8.33	-0.40	0.02	0.25
CHEM	6.48	-0.41	0.08	0.14
BHF	3.40	-0.04	-0.17	0.15
RABO	6.30	-0.36	0.05	0.14
SGEN	5.58	-0.25	0.02	0.16
NORS	5.18	-0.22	0.04	0.10
DRES	9.13	-0.29	-0.04	0.09
UBS	10.21	-0.29	-0.04	0.08
ABNA	6.61	-0.33	-0.04	0.07

Notes: Entries show the conditional variance and autocorrelations of quote revisions $\Delta \ln P$ in calendar time. Moments have been computed over 30 seconds intervals using the estimator of DeJong and Nijman (1997). "obs" is the fraction of 30 second intervals with a quote revision by that bank.

Table 4: Price discovery model in tick time

	σ^2	0.50	(0.02)					
	ω^2		ψ		γ		π	IS
DB	2.76	(0.05)	0.31	(0.03)	0.03	(0.02)	0.09	0.15
CHEM	3.03	(0.06)	-0.43	(0.04)	-0.23	(0.03)	0.05	0.01
BHF	2.59	(0.06)	-0.84	(0.04)	-0.29	(0.03)	0.05	-0.03
RABO	3.03	(0.06)	-0.74	(0.04)	-0.36	(0.03)	0.05	-0.02
SGEN	1.13	(0.05)	0.43	(0.04)	0.04	(0.02)	0.06	0.11
NORS	2.18	(0.07)	-0.70	(0.05)	-0.26	(0.03)	0.03	-0.01
DRES	2.63	(0.08)	0.35	(0.06)	-0.07	(0.03)	0.03	0.05
UBS	3.55	(0.08)	0.10	(0.06)	-0.30	(0.04)	0.03	0.03
ABNA	1.05	(0.09)	0.86	(0.06)	0.19	(0.04)	0.02	0.06
rest	2.48	(0.01)	0.06	(0.01)	-0.16	(0.01)	0.59	0.67

Notes: This table reports GLS estimates of the price discovery model in tick time with. Parameters: σ^2 is the variance of the fundamental news; ω_i is the variance of the idiosyncratic component; ψ_i is the covariance between news and the idiosyncratic component; γ_i is the covariance between the idiosyncratic term of bank i and the previous news term; π is the activity share and IS the information share of bank i defined in (10). Standard errors in parentheses.

Table 5: Price discovery model in tick time: Intervention days

	σ^2	1.11	(0.20)					
	ω^2		ψ		γ		π	IS
DB	3.71	(1.02)	1.61	(0.83)	0.78	(0.42)	0.09	0.32
CHEM	5.45	(1.23)	-1.82	(1.01)	-0.75	(0.51)	0.06	-0.10
BHF	1.98	(1.49)	-1.92	(1.23)	-1.02	(0.63)	0.05	-0.07
RABO	2.79	(1.39)	0.13	(1.14)	0.29	(0.58)	0.05	0.06
SGEN	2.20	(1.30)	1.07	(1.06)	0.68	(0.54)	0.06	0.14
NORS	1.29	(1.47)	-0.82	(1.21)	-0.66	(0.62)	0.04	-0.01
DRES	12.49	(1.40)	-3.75	(1.15)	1.25	(0.59)	0.05	-0.21
UBS	3.90	(1.98)	0.17	(1.63)	0.07	(0.84)	0.02	0.03
ABNA	-0.62	(1.81)	2.88	(1.49)	0.90	(0.77)	0.03	0.14
rest	3.91	(0.37)	0.19	(0.20)	0.15	(0.20)	0.55	0.70

Notes: This table reports GLS estimates of the price discovery model in tick time. Sample period is the period between one hour before and one hour after the suspected intervention times reported in Table I of Peiers (1997). Parameters: σ^2 is the variance of the fundamental news; ω_i is the variance of the idiosyncratic component; ψ_i is the covariance between news and the idiosyncratic component; γ_i is the covariance between the idiosyncratic term of bank i and the previous news term; π is the activity share and IS the information share of bank i. Standard errors in parentheses.

Table 6: Price discovery model in calendar time

σ^2	1.81				
	ψ_i	ω_{ii}	ω_{i0}	π	IS
DB	-0.53	3.42	-0.80	0.12	0.08
CHEM	-1.38	3.05	-0.77	0.07	0.02
BHF	-1.09	1.43	-0.81	0.07	0.03
RABO	-1.30	3.04	-0.83	0.07	0.02
SGEN	0.55	0.77	-1.28	0.08	0.10
NORS	-0.82	1.95	-0.84	0.05	0.03
DRES	0.34	3.03	-0.92	0.04	0.05
UBS	0.16	3.70	-0.82	0.04	0.04
ABNA	0.30	1.91	-0.81	0.03	0.04
market	0.65	2.72		0.44	0.60

Notes: This table reports the estimates of the price discovery model in calendar time. The individual banks are labelled $1, .., n$ and the market, defined as all quotes except those of bank i, is labelled 0. The parameters are: σ^2 is the variance of the fundamental news; ψ_i is the covariance between news and the idiosyncratic component of bank i; ω_{ii} is the variance of the idiosyncratic component of bank i; ω_{0i} is the covariance between the idiosyncratic term of bank i and the idiosyncratic term of the market; π is the activity share and IS the information share of bank i
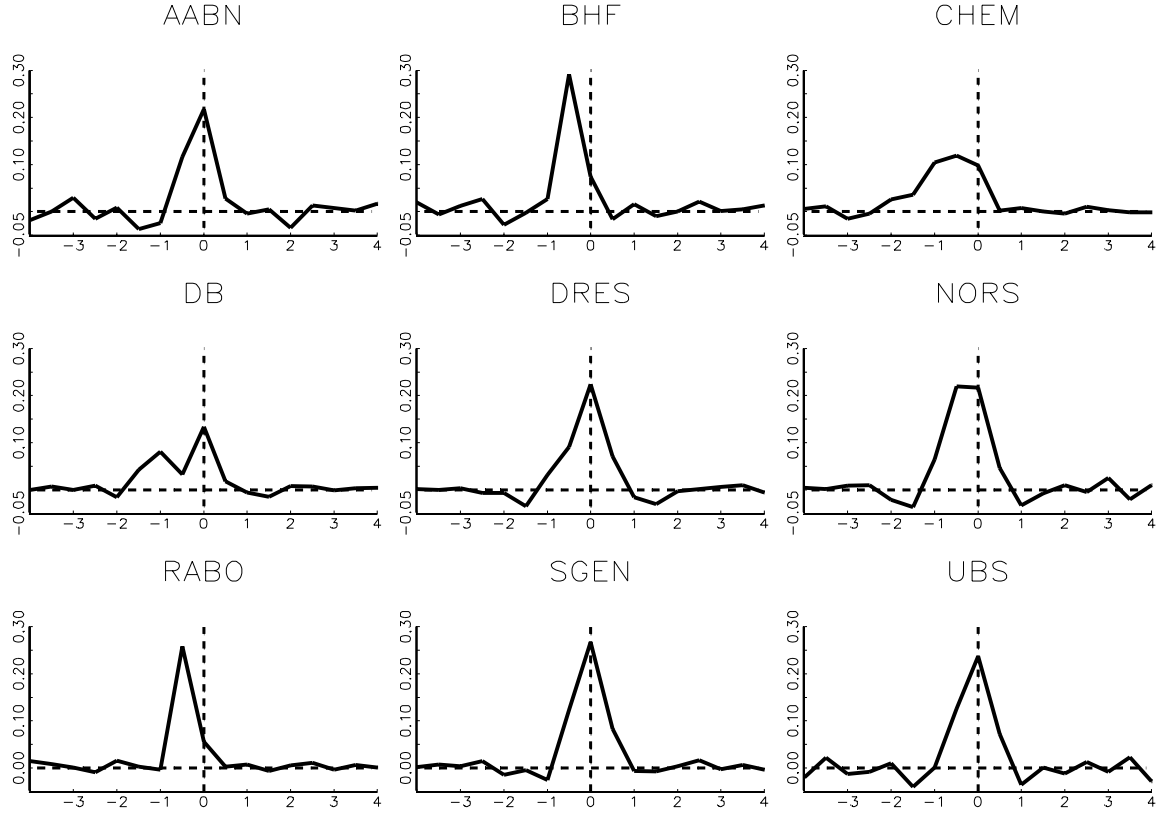
Figure 1: Cross correlation functions

Notes: This figure shows the estimated cross correlations between the quotes of an individual bank and the market (all other quotes) defined as $\mathrm{Cov}(p_{it} - p_{i,t-1}, p_{M,t} - p_{M,t-k-1})$, where i is the individual bank and M the market. Quotes are sampled every 30 seconds. Negative k thus indicate that the market is leading the individual bank. Units on the horizontal axis are minutes. Bank acronyms are explained in table 1.