



ELSEVIER

Journal of International Money and Finance
17 (1998) 5–27

Journal of
International
Money
and Finance



Provided by Research Papers in Economics

Price discovery in the foreign exchange market: an empirical analysis of the yen/dmark rate^{*,**}

Frank De Jong^a, Ronald Mahieu^b, Peter Schotman^{c,†}

^aTilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

^bErasmus University, Postbus 1738, Rotterdam, 3000 DR Rotterdam, The Netherlands

^cLIFE, Maastricht University, PO Box 616, 6200 Maastricht, The Netherlands

Abstract

Using Reuters exchange rate data we investigate the dynamic relations between the direct quotes of the yen/dmark rate and the rate implied by yen/dollar and dmark/dollar rates. Since these high frequency data are observed at irregular intervals, technical problems arise in calculating auto-correlations and cross-correlations. We propose a covariance estimator for irregularly spaced data. The empirical results show lagged adjustment of the direct yen/dmark cross-rate to changes in the dollar implied rate. However, since the dollar implied rate is extremely noisy, substantial price discovery takes place through the direct yen/dmark market, especially during the most busy parts of the day. © 1998 Elsevier Science Ltd. All rights reserved.

JEL classifications: F31; C21

* We would like to thank participants at the HFDF-I conference in Zurich, EC2 meeting in Aarhus, ESEM in Istanbul, JIMF-LIFE workshop in Maastricht, conference on High Frequency Financial Data in San Diego and at seminars at University of Bern, Freie Universität Berlin, London School of Economics, Tinbergen Institute Rotterdam, GREMAQ and GREQAM, for constructive comments. In particular we would like to thank Philipp Hartmann, Richard Lyons, Franz Palm and Neil Shephard and two referees. Of course, all errors are our own.

** Note this article was presented at the JIMF-LIFE conference and it was submitted by the authors to the JIMF. It was considered by referees and accepted for publication by Michael Melvin, Co-Editor.

† Corresponding author. Tel.: +31 43 3883862; fax: +31 43 3258530; e-mail: p.schotman@berfin.unimaa.nl

This article investigates the price discovery process in the foreign exchange markets for the US dollar, the yen and the deutschmark currencies. In foreign exchange the dollar often acts as a vehicle currency (see Krugman, 1980, 1984) and hence functions as the international ‘money’. When exchanging one currency into another, a currency is first converted into dollars, which are then traded for the other currency. When observing exchange rate data on a daily or lower frequency the triangular relations among exchange rates are virtually an identity, but for high frequency data, such as the quotes on the Reuters screens, the ratio of two dollar exchange rates need not always be equal to the cross-rate. However, since foreign exchange markets are among the most liquid financial markets in the world, one expects price adjustment in these markets to occur very quickly.

Our aim in this article is to investigate the price adjustment and discovery in the yen/dmark cross-rate. More specifically the observations on the liquidity in the dollar markets lead to the following questions. Do changes in the yen/dmark rate first show up in either of the two dollar rates, or is news immediately reflected in the yen/dmark rate itself? Does the yen/dmark rate simply follow any change in the more active dollar markets, i.e. is it an informationally redundant market? Which market has the biggest impact on the *efficient* price, which is defined by Hasbrouck (1995) as the value to which both the cross and the dollar-implied exchange rate converge in the long run, absent new shocks. A related question is how much of the observed volatility in the actual and implied yen/dmark rate is due to changes in the efficient price and how much due to microstructure noise.

To answer such questions, we estimate the dynamics of the actual yen/dmark rate and the implied exchange rate calculated from the yen/dollar and dmark/dollar rates. The data are from the HFDF-93 dataset made available by Olsen and Associates, which has separate quotes for the yen/dmark rate as well as the yen/dollar and dmark/dollar rates. We identify a bivariate time series process for the actual and implied yen/dmark rate from the various auto- and cross-covariance functions. This time series model is then used to identify the relative contributions of actual and implied rates to the price discovery process.

An important assumption of our methodology is the stability of the covariances in real time, although possibly different for different segments of the day. In microstructure research one often uses the interval between two transactions or quotes as the basic time unit; this is the so called ‘transaction time’ or ‘tick time’ assumption. Under this assumption data are sampled more frequently during the most active times of the market. Although for certain problems, such as estimating bid-ask spreads, this sampling scheme is convenient and appropriate, we cannot rely on tick time for our analysis for several reasons.

First and foremost, we are interested in the real time lags between different markets. On foreign exchange markets quotes are indicative, so it is not obvious that there will be new quotes immediately after new information arrives. The traditional argument that information arrival and trading are almost perfectly correlated need not hold for quote data. A closely related reason for using calendar time is related to the specific properties of the HFDF-93 dataset. The data only contain quotes that appeared on the Reuters screens, which is a subset of

all information about the exchange rates, since, for example, competing data providers exist like Telerate and Knight Ridder. If during some minute the Reuters screen does not show a new quote, this does not necessarily imply that there have not been any new quotes on any other system.

A second, equally important, reason has a more technical nature. In our empirical analysis we will combine data from three markets, which have very different trading (or quote placement) intensity and also a very different timing of new quotes.¹ The three exchange rates have unequal numbers of observations and observations are irregularly spaced, hence we are confronted with a non-synchronous trading problem. Since the activity pattern in the markets is so different, there is no natural way to define ‘tick time’ while preserving the real time covariance structure of the data. Harris et al. (1995) succeed in sampling in ‘tick time’, which is determined however by the pace of the slowest market.

For these reasons, we choose real time as the relevant clock. This is a strong assumption, but we feel it is the most appropriate to make in the context of this article. To deal with the irregular spacing of the data, we use the covariance estimator developed by De Jong and Nijman (1997).

We will treat intra-day seasonality by examining the covariance structure for different segments of the day. That way we can test for different dynamics during different phases of a trading day. During Japanese business hours the microstructure dynamics for the yen/dmark exchange rate are likely to be different than they are during European or US business hours. Hartmann (1996) conjectures that the importance of the dollar depends on the time zone. Yen and dmark could function as a vehicle during Japanese business hours and in the European morning, when American banks are still closed. If the dollar is the only vehicle currency we would expect that price discovery would take place in the highly liquid dollar markets and that the cross-rates would simply follow. If the dmark and the yen also function as vehicle currencies we would expect that price discovery takes place in the cross-markets as well.

The article is organized as follows. In Sec. 1 we describe a statistical method for the estimation of the moments with irregularly spaced observations. Section 2 considers the data and auto- and cross-correlation functions. Section 3 develops a stylized microstructure model for the interpretation of the lead–lag relations in the covariance functions. Section 4 concludes.

1. Estimation of calendar time correlations with irregularly spaced observations

The basis of our approach is an estimator for the auto- and cross-covariances between two time series of returns, taking into account the irregular spacing of the

¹In previous empirical work with these exchange rate quotes, Müller et al. (1990), Dacorogna et al. (1993), Guillaume et al. (1997) and Andersen and Bollerslev (1997) have found strong daily and weekly seasonal patterns in the volatility of exchange rates. Their empirical evidence also indicates that the three exchange rates have distinct activity patterns.

observations. These covariances contain all information needed to analyse lead–lag relations and to estimate more structural models of the information flow between two markets. The covariance functions can be used to identify the order and structure of a multivariate time series model. The main technical obstacle in modelling the interactions among the different markets consists of the irregular arrival of new quotes. We present a moment estimator for estimating (auto-)covariances between returns from irregularly observed data.

An important feature of our method is that time is measured as real (calendar) time or clock-time. We treat the irregular spacing of the price data by assuming that there exists a regular underlying discrete time price process from which observations are recorded irregularly. We want to measure the correlations between changes of these underlying, but partly unobserved, price processes.

Various sampling schemes have been adopted in the literature. The simplest way to circumvent the irregularly spaced data problem would be by assuming that the most recent quote remains valid if no new quote arrives in some interval. Assuming zero returns if there has been no quote revision will usually lead to a smaller estimate of the variance and much smaller autocorrelations. The consequences of filling the gaps with standing quotes are explored in the Appendix.

An alternative sampling scheme is employed by Harris et al. (1995). They estimate lead–lag relations between markets with different activity patterns, recording a new observation only if prices on all markets have moved. The most recent observation from each series is then used as the observation for that particular series. This method preserves the causal ordering of the data, but implicitly assumes that all markets operate on a ‘tick’-time determined by the least active market. Both approaches contrast to our approach, where we explicitly put a missing value at the intervals during which no new quotes arrived and use an estimator which takes these missing observations explicitly into account.

Our method is based on a moment estimator.² We consider the case where the returns have a zero mean and we assume that there are no additional deterministic components like seasonals.³ We introduce the following notation. Let p_t and q_t denote the logarithm of the price level at time t in the two markets under consideration. Denote the covariance function of the underlying returns by⁴

²Another approach to deal with the missing observations would be the Kalman filter. However, the Kalman filter method is computationally quite intensive and therefore not attractive for the very large data set under consideration. Moreover, the Kalman filter requires a fully specified model for all three exchange rates. Among other things, this requires one to decide or test whether there is more than one cointegrating relation between the three exchange rate series. Of course, if the usual normality assumptions are also satisfied, the Kalman filter approach will increase the efficiency of the estimator. Herwartz (1996) reports efficiency comparisons between the Kalman filter approach, the moment estimator we develop in this article and some further sampling schemes. He finds that the moment estimator developed in this article is only slightly less efficient compared to a full ML procedure based on the Kalman filter.

³It is not difficult to introduce deterministic components, but it considerably complicates the notation. In the empirical analysis we correct for the drift in the exchange rates by centering all second moments using the average return computed from the first and last observation in the sample.

⁴The exposition is for the computation of cross-covariances, but the method can be applied with slight modifications for autocovariances, imposing the linear restrictions $\gamma(-k) = \gamma(k)$.

$$\gamma(k) = \text{Cov}(\Delta p_t, \Delta q_{t-k} | S_t), \quad k = -K, \dots, K. \tag{1}$$

The covariance is conditional on some variable S_t , which in our case of exchange rate data indicates the time of the day, where we divide the day in different segments related to business hours in Japan, Europe and the US. Since we measure returns in discrete time, we need to assume a basic unit observation interval. The length of this interval can be arbitrarily short; in the empirical analysis it is taken to be either 1 minute or 30 seconds. If we sample at such a high frequency there will be many missing values. A missing value for a particular time is recorded, if no new quote has arrived in the last interval. The graph below shows a typical bivariate sample.

Let us index the observations on p_t by the index i and the observations on q_t by the index j . The return over a time interval that spans several sampling intervals can be expressed as the sum of the underlying unobserved one-period returns,

$$p_{t_{i+1}} - p_{t_i} = \sum_{t=t_i+1}^{t_{i+1}} \Delta p_t, \tag{2}$$

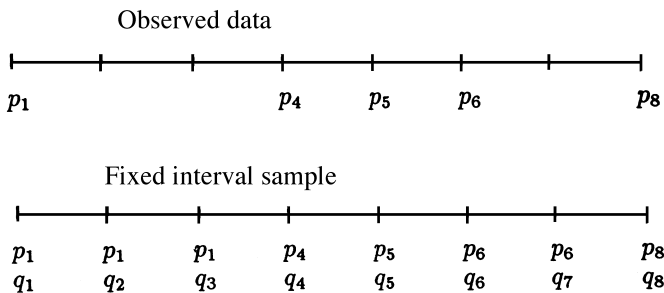
where t_i denotes the calendar time index of the i th observation. The cross product of prices on two markets can thus be written as

$$y_{ij} = (p_{t_{i+1}} - p_{t_i})(q_{t_{j+1}} - q_{t_j}) = \sum_{t=t_i+1}^{t_{i+1}} \sum_{s=t_j+1}^{t_{j+1}} \Delta p_t \Delta q_s. \tag{3}$$

The expectation of this linear combination of cross-products is a linear combination of the covariances $\gamma(k)$ of the underlying process,

$$E[y_{ij}] = E \left[\sum_{t=t_i+1}^{t_{i+1}} \sum_{s=t_j+1}^{t_{j+1}} \Delta p_t \Delta q_s \right] = \sum_{t=t_i+1}^{t_{i+1}} \sum_{s=t_j+1}^{t_{j+1}} \gamma(t-s). \tag{4}$$

If all covariances of order $k > K$ are assumed to be zero, the summation of $\gamma(k)$ is truncated between $-K$ and K . By changing the index of the summation we can write $E[y_{ij}]$ as a linear combination of the covariances $\gamma(k)$,



$$E[y_{ij}] = \sum_{k=-K}^K x_{ij}(k)\gamma(k), \quad (5)$$

where $x_{ij}(k) = \max[0, \min(t_{i+1}, t_{j+1}) - \max(t_i, t_j + k)]$. For the estimator of the covariances we regard (Eq. (5)) as a regression equation with the unknown covariances $\gamma(k)$ as parameters and the coefficients x_{ij} as explanatory variables. Under assumptions explicated in De Jong and Nijman (1997) the covariances can be estimated consistently by ordinary least squares on the observations of y_{ij} and the constructed x_{ij} 's.

As a refinement of the covariance estimators we employ a weighted least squares estimator with weights proportional to the inverse of the square root of the length of the time interval between two observations. For the return regression (Eq. (5)) we assume weights given by

$$w_{ij} = [(t_{i+1} - t_i)(t_{j+1} - t_j)]^{-1/2}. \quad (6)$$

The weighting scheme has two advantages for our dataset. First it is a simple heteroskedasticity correction, assuming that the error variance for an observation is proportional to the number of moments involved. The correction would be exact if both p_t and q_t are a random walk with constant innovation variances.

The second aim of the weighting scheme is to reduce the influence of observations that are far apart. In the exchange rate data there are almost no observations on Sundays and very few in the early hours of the night (between the end of business hours in Los Angeles and the opening of Tokyo). Cross-products involving these long spans contain many high order covariances, which are collectively set to zero due to the truncation at lag K . Since the sum of the high order covariances will presumably be different from zero due to daily and weekly seasonals, including the observations with long spans will introduce an omitted variables bias in the regression. By downweighting these observations the omitted variables bias is greatly reduced.⁵

Since the regression models deal with estimating second moments, the covariance matrix of the estimator involves the fourth moment structure of the time series. De Jong and Nijman (1997) discuss the details of inference on the parameters $\gamma(k)$. However, the results of De Jong and Nijman (1997) are computationally demanding for large datasets like the HFDF-93 database. They are even more demanding for inference on functions that depend on the autocovariances of several different series, like for example the implied yen/dmark exchange rate, since these also require the covariances between the estimators for the individual covariance functions. Therefore we do not report standard errors. A conservative

⁵An alternative to reduce the omitted variables bias is to assume a regular pattern for all high order covariances. For example, under the assumption that the correlations decline geometrically we could specify $\gamma(k) = \phi_0 \phi_1^k$ for $|k| > K$, where ϕ_1 is pre-specified. The parameter ϕ_0 then enters linearly in (Eqs. (4) and (5)). These further refinements appeared not important for our dataset.

estimate of standard errors for autocorrelations would be $1/\sqrt{N_1}$, where N_1 is the number of 30-seconds or 1-minute intervals. For our data this is always less than 0.013.

Our interest is in the interaction between the actual yen/dmark rate and the implied yen/dmark rate which can be obtained through the two dollar rates. The bid price of the implied yen/dmark rate is defined as the bid price of the yen/dollar rate divided by the ask price of the dmark/dollar rate. In logarithms this gives the definition

$$y_{J/G}^b = x_{J/U}^b - x_{G/U}^a, \quad (7)$$

where x denotes the actual (log –) exchange rate, y the implied (log –) exchange rate, the superscripts a and b denote ask and bid and the subscripts J , G and U stand for Japanese yen, German mark and US dollar, respectively. The implied ask price of the yen/dmark is defined analogously.

The autocovariance function of the actual yen/dmark exchange rate will be denoted by $\gamma_{xx}(k)$ and can be computed directly using the estimator described above. Using the identity (Eq. (7)) the autocorrelation function of the bid quotes for the implied yen/dmark rate is computed from the moments of the bid quotes of the yen/dollar rate and the ask quotes of the dmark/dollar as

$$\gamma_{yy}(k) = \gamma_{J/U}^b(k) - \gamma_{J/U,G/U}^{ba}(k) - \gamma_{J/U,G/U}^{ba}(-k) + \gamma_{G/U}^a(k), \quad (8)$$

where $\gamma_{J/U,G/U}^{ba}(k)$ denotes the covariance between the return of the yen/dollar bid quote and the return of the dmark/dollar ask quote leading k periods.⁶ The same identity (Eq. (7)) also provides the covariance between the actual and implied rates:

$$\gamma_{xy}(k) = \gamma_{J/U,J/G}^{bb}(k) - \gamma_{D/U,J/G}^{ab}(k). \quad (9)$$

Since the 1-year sample is too short for meaningful formal cointegration tests, we assume that the direct yen/dmark exchange rate and the rate implied by the yen/dollar and dmark/dollar rate are cointegrated. The last assumption must hold in order to avoid arbitrage, which would be profitable if the actual and implied rate are allowed to drift too far apart. Assuming cointegration, the deviation between the actual and implied yen/dmark exchange rates also has a well-behaved autocorrelation function. Because the deviations deal with levels instead of returns, the estimator described above does not apply without modification. However, for level variables the moments can be estimated straightforwardly on complete observations only. Complete observations are defined as those observations for which all three exchange rates are available, i.e. the intersection of the set of observations

⁶Note that we are comparing the level of the bid (or ask) of the actual and implied yen/dmark rate. To be consistent with that formulation we compute returns solely from the bids (or asks). This is not a return available in the market, where a realized return would compare a bid with an ask.

for each of the three individual exchange rate series.⁷ For the covariance between returns and the error correction term we also confine ourselves to complete observations on z_t . Let Δp_t be a returns series and p_t and q_t be levels time series. The error correction term is defined as $z_t = q_t - p_t$. The covariance of the returns with lags of the error correction term can be written as

$$\begin{aligned} E[\Delta p_t z_{t-k}] &= E\left[\Delta p_t \left(z_t - \sum_{j=0}^{k-1} \Delta z_{t-j}\right)\right] \\ &= \delta_{pz} - \sum_{j=0}^{k-1} (\gamma_{pp}(j) - \gamma_{pq}(j)), \end{aligned} \quad (10)$$

where $\delta_{pz} = E(\Delta p_t z_t)$ is the only unknown parameter. All the other moments are already available from the (auto-)covariance functions of the returns Δp_t and Δq_t . Suppose we have observations on prices p_{t_i} and error correction term z_{t_j} (with $t_i = t_j$ because we have complete observations on z). The cross-product becomes

$$y_{ij} = (p_{t_i} - p_{t_{i-1}})z_{t_j} = \sum_{t=t_{i-1}+1}^{t_i} \Delta p_t z_{t_j}. \quad (11)$$

Using (Eq. (10)) the expectation of the right hand side of (Eq. (11)) is

$$\begin{aligned} E[y_{ij}] &= (t_i - t_{i-1})\delta_{pz} - \sum_{t=t_{i-1}+1}^{t_i} \sum_{k=1}^{t-t_{i-1}} (\gamma_{pp}(k) - \gamma_{pq}(k)) \\ &= (t_i - t_{i-1})\delta_{pz} + \sum_{k=1}^{t_i-t_{i-1}} ((t_i - t_{i-1}) - k + 1)(\gamma_{pp}(k) - \gamma_{pq}(k)). \end{aligned} \quad (12)$$

Since all the $\gamma(k)$ moments have been estimated before, the second term can be moved to the left hand side. This leaves a regression model with δ_{pz} as the single unknown parameter and observations t_i ($i = 1, \dots, N$). The regression is run with weighted least squares just as for the return regressions, with the weights w_{ij} defined analogously to (Eq. (6)) as $w_{ij} = (t_i - t_{i-1})^{-1/2}$. Given δ_{pz} and δ_{qz} all other moments and cross-moments can be computed.

⁷In our data set this does not entail much loss in efficiency, since there are many more observations on the dollar rates than on the yen/dmark rate. There are relatively few cases in which the yen/dmark rate is observed and the corresponding dollar rate is not observed. Instead of adjusting the method for this situation, we decided to use only the periods for which there are observations on all series (yen/dmark and the dollar rates). Given the size of our data set, this does not have any noticeable influence on the estimates. By using complete observations on $z_t = p_t - q_t$ we avoid working with ill-defined moments like $E[p_t \Delta p_t]$ that involve I(1) variables. Also, in dealing with the irregular spacing in the sample we would have to use a triple summation with terms $x_{J/U}(t_i)$, $x_{D/U}(t_j)$ and $x_{J/G}(t_k)$, which is computationally demanding.

2. Auto- and cross-correlation functions

Our data consist of exchange rate quotes from the Reuters screens over the period October 1, 1992 to September 30, 1993. This HFDF-93 dataset has been made available by Olsen and Associates. The dataset is described extensively in a series of empirical studies from Olsen and Associates (see for example Dacorogna et al., 1993 and Guillaume et al., 1997). We have sampled the bid and ask quotes of the yen/dollar, yen/dmark and dmark/dollar rates at fixed calendar time intervals of either 30 seconds or 1 minute (last quote of the interval) and transformed the quotes by taking logarithms. If for some interval there is no new quote, we record a missing value. Since there is not much difference between the correlation structures of the bid and the ask prices, we only report results for the bid prices, i.e. the implied yen/dmark rate is computed as in (Eq. (7)).

Table 1 contains some summary statistics on the numbers of observations depending on the sampling frequency. We have split the trading day in four different segments according to business hours in different parts of the world. The first segment corresponds to Japanese business hours and runs from 0.00 to 08.00 h GMT (08.00–16.00 h in Japan). The second segment covers the early European business hours from 08.00 to 14.00 h GMT (09.00–15.00 h in Germany).⁸ The third segment contains the overlap between European and US business hours, going from 14.00 to 20.00 h GMT (14.00–19.00 h in the United Kingdom, 09.00–14.00 h US Eastern time). The last segment covers the rest of the day, covering part of US west coast business hours: 20.00–24.00 h GMT. The two dollar markets are much more active than the market for the cross-rate, for which we get many more missing values. The highest percentage of missing values occurs in the first segment, where we only obtain new quotes for the yen/dmark in approximately 15% of the minutes. Best covered is the European segment. Quotes for the dmark/dollar in this segment are so dense that even the 30-seconds sampling frequency provides close to 60% of the maximum possible observations. This may reflect high trading intensity during these hours, or a high market share of the Reuters system relative to competitors in Europe.

All (auto-)covariances and correlations have been computed using the methods in Sec. 1. For all series we simultaneously estimated the autocovariances up to 30 minutes. For cross-covariances we included leads and lags up to order plus and minus 30 min (a total of 61 and 121 parameters for the 1-minute and 30-seconds intervals, respectively). The moments have been estimated both for the full sample and for the four subsamples with the different segments of the trading day. Figure 1 shows the estimated autocorrelation functions for the three series of observed exchange rate returns, the implied yen/dmark returns and also the correlation between the actual and implied yen/dmark rate, using the full sample.

⁸All local times refer to winter time. We did not adjust the segments for the shift to daylight saving time in some countries.

Table 1
Number of observations

| | Segment | | | | | | | |
|--------------------------|---------|--------|---------|--------|-----------|--------|--------|--------|
| | Japan | | Europe | | Europe/US | | US | |
| Thirty seconds intervals | | | | | | | | |
| Yen/dmark | 31 087 | (8.9) | 46 552 | (15.2) | 49 250 | (16.1) | 21 725 | (9.9) |
| Yen/dollar | 109 673 | (31.3) | 101 500 | (33.1) | 83 911 | (27.4) | 50 595 | (23.3) |
| Dmark/dollar | 143 727 | (41.0) | 179 484 | (58.5) | 123 525 | (40.3) | 70 319 | (32.1) |
| One minute intervals | | | | | | | | |
| Yen/dmark | 27 051 | (15.4) | 44 198 | (28.8) | 42 571 | (27.8) | 18 898 | (17.3) |
| Yen/dollar | 70 438 | (40.2) | 68 078 | (44.4) | 54 354 | (35.5) | 39 949 | (36.5) |
| Dmark/dollar | 82 522 | (47.1) | 91 466 | (59.7) | 68 032 | (44.4) | 53 192 | (48.7) |

Notes: entries indicate the number of observations with a sampling frequency of either 30 s or 1 min. In parentheses the number of observations is given as a percentage of the potential number of observations in a segment. The Japanese segment is defined as 0–8 h GMT, Europe is 8–14 h GMT, US/Europe is 14–20 h GMT and the US segment is defined as 20–24 h GMT. The counting includes all Sundays and holidays.

At the 1-min frequency all return series show a strong negative first order autocorrelation and almost zero higher order autocorrelations. This drop of the autocorrelations to zero is most pronounced for the dollar exchange rates. The autocorrelations for the observed yen/dmark rate fluctuate a bit more; this is probably only sampling error due to the much smaller number of observations of this series relative to the dollar rates (see Table 1). The autocorrelation functions for the four subsamples are very similar, with the same highly negative first order autocorrelation.⁹ These autocorrelation functions would identify an MA(1) time series model. The implications of that model will be analysed in detail in the next section.

The magnitude of the first order autocorrelations is much larger than is commonly reported in the literature. At the 1-min frequency Guillaume et al. (1997) find autocorrelations of approximately -0.10 . The differences are entirely due to our estimator, which takes into account the irregular spacing. The differences are further explored in the Appendix.

Figure 1 also shows the cross-correlation function for the actual and implied yen/dmark rates. The contemporaneous correlation is only 0.20 for most of the day and even less during the late US segment. However, there are also strong positive correlations between the current return of the actual exchange rate and lagged returns of the implied yen/dmark rate. In fact, the cross-correlation peaks at the 1-min lag, i.e. $\text{Cov}(\Delta x_t, \Delta y_{t-1})$. The leading cross-correlations of the implied yen/dmark rate are an indication that the dollar markets are more important than

⁹We only show the correlation functions for the full sample. The figures for the different segments of the day are very similar.

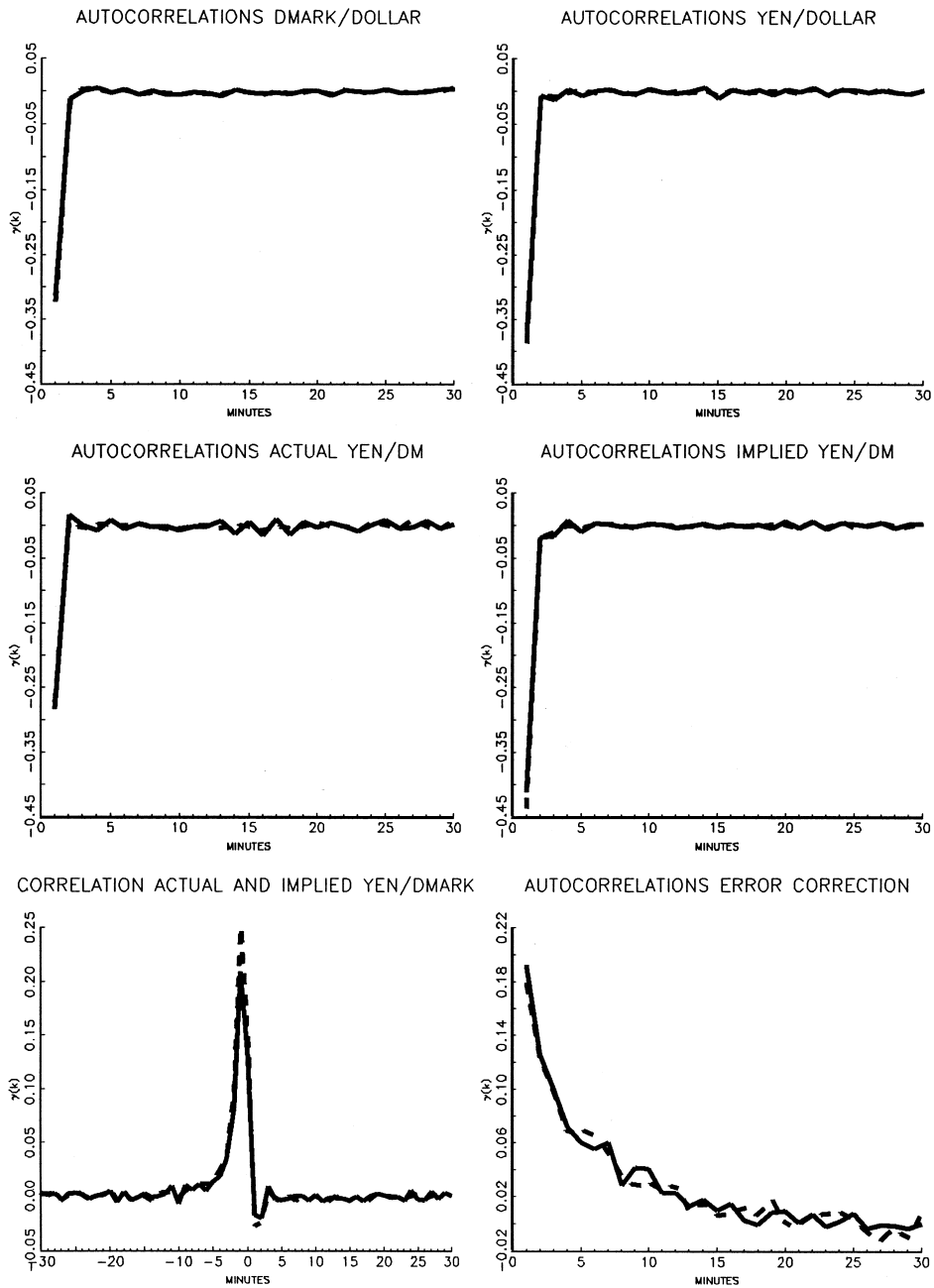


Fig. 1. This figure shows the correlation functions of 30-seconds (dashed line) and 1-minute (solid line) returns using the moment estimator for irregularly spaced observations for the full sample.

Table 2
Return variances

| Variable | Full | Japan | Europe | Eur./US | US |
|--------------------------|-------|-------|--------|---------|-------|
| Thirty seconds intervals | | | | | |
| Actual | 9.05 | 10.35 | 8.00 | 9.61 | 7.32 |
| Implied | 23.00 | 22.14 | 21.16 | 28.77 | 21.12 |
| One minute intervals | | | | | |
| Actual | 9.20 | 10.41 | 8.31 | 9.77 | 7.35 |
| Implied | 23.18 | 22.35 | 21.26 | 29.32 | 20.58 |

Entries show the variance of 1-min returns of the yen/dmark exchange rate for various segments of the trading day. Units are $(\% \text{ per minute})^2$ multiplied by 10 000. See Table 1 for the definition of the various segments of the day.

the actual market for the yen/dmark cross-rate. The direct yen/dmark rate adjusts with a lag to changes induced by the two dollar rates. There is no response of the implied rate to lagged changes in the actual yen/dmark cross-rate. This pattern appears in all segments of the day.

The autocorrelations of the error correction term $(x - y)$ are shown in the last panel of Fig. 1. The first order autocorrelation is approximately 0.20, after which the autocorrelations quickly drop off to zero. The error correction moments support the impression that convergence of the actual and implied yen/dmark rate is completed in a few minutes, but not immediate.

Table 2 provides summary statistics on the volatilities of the actual and implied yen/dmark rate. It is evident that the volatility is lowest in the late US trading hours (the last data segment). Also, because the contemporaneous correlation between yen/dollar and dmark/dollar is close to zero at the 1-min frequency, the variance of the implied yen/dmark rate is close to the sum of the two dollar variances. The variance of the implied rate is more than twice as high as the variance of the direct yen/dmark rate at the 1-minute frequency. This means that the ratio of the two dollar rates contains far more noise about the yen/dmark rate than the actual yen/dmark rate itself.

The results for the 30-seconds sample are hardly distinguishable from the 1-minute samples. To compare the autocovariances we converted the (auto-)covariance function at the 30-seconds interval to the implied 1-min function.¹⁰ The only difference is for the autocovariance function of the error correction term, which arises because these estimates only use complete observations. The main reason for comparing the 1-min and 30-s sampling intervals is to investigate the problem of 'stale quotes'. The stale quote problem occurs, if the last quote in a particular minute is situated near the beginning of the minute and possibly just a few seconds away from the previous quote. Because the estimates appear robust with respect to

¹⁰To be precise, let r_t be the return over an interval of 30 s and let $\gamma(k)$ be the autocovariance function of r_t . Then the implied 1-min autocorrelation function is given by $\tilde{\gamma}_{60}(k) = \gamma(2k - 1) + 2\gamma(2k) + \gamma(2k + 1)$ for $k > 0$ and $\tilde{\gamma}_{60}(0) = 2\gamma(0) + 2\gamma(1)$.

the sampling frequency, we conclude that the problem of stale quotes is not very important for our combination of data and estimator. For the remainder of the empirical analysis we only report the 1-min results.

3. Random walk plus noise model

Bollerslev and Domowitz (1993) and Guillaume et al. (1997) provide several explanations for the negative autocorrelation in exchange rate returns. Possible reasons are that certain banks systematically set higher bid/ask spreads, or that traders have diverging opinions, or order imbalances. All these explanations are consistent with a time series model, in which the observed quote x_t has two components: the true underlying value x_t^* and a random noise u_t . Hasbrouck (1995) refers to x_t^* as the *efficient price*, which he defines as the permanent component of an observed price time series. The *efficient* exchange rate x_t^* is a random walk, but not directly observable. This random walk plus noise or measurement error model has been used in Hasbrouck (1993) for a univariate price series. We extend the univariate model to a bivariate case. In our model both the *actual* and the *implied* yen/dmark exchange rate are a random walk plus noise, sharing the same efficient price x_t^* . We thus obtain the cointegrated random walk plus noise model

$$x_t = x_t^* + u_t, \quad (13)$$

$$y_t = x_t^* + v_t, \quad (14)$$

$$x_t^* = x_{t-1}^* + e_t. \quad (15)$$

We associate the transitory shocks with microstructure effects. We assume that the shocks u_t , v_t and e_t are serially uncorrelated and have mean zero.¹¹ The covariance matrix of the three shocks has variances σ_u^2 , σ_v^2 and σ_e^2 and covariances σ_{eu} , σ_{ev} and σ_{uv} . Correlation between the noises (u_t , v_t) and the information shock e_t can be interpreted as evidence of transactions costs or as lagged adjustment to information (see Hasbrouck, 1993). The model is what Harvey (1989) refers to as a structural time series model. Hence we will call the parameters in (Eqs. (13–15)) the ‘structural’ parameters. Our interest is in the relative magnitudes of the variance of the random walk component σ_e^2 relative to the variances of the noise components σ_u^2 and σ_v^2 .

The data provide 10 useful moments for estimation of the parameters: the contemporaneous covariance matrix of the returns (three moments), the covariances of returns with lagged returns (four moments), the variance of the error correction term $z_t = x_t - y_t$ (one moment) and the contemporaneous covariances of z_t with the returns (two moments). All other data moments should be equal to

¹¹ We implicitly assume that there is a constant term, because we take $(x - y)$ in deviation of its sample mean.

zero in the model, consistent with the individual series being MA(1). The moments of the random walk plus noise model (Eqs. (13–15)) are all linear in the parameters. Straightforward calculation of variances and covariances yields a system of 10 moment equations as a function of the six structural parameters.

$$\begin{pmatrix} \gamma_{xx}(0) \\ \gamma_{yy}(0) \\ \gamma_{xy}(0) \\ \gamma_{xx}(1) \\ \gamma_{yy}(1) \\ \gamma_{xy}(1) \\ \gamma_{xy}(-1) \\ \gamma_{zz}(0) \\ \gamma_{xz}(0) \\ \gamma_{yz}(0) \end{pmatrix} = \begin{pmatrix} 1 & 2 & 0 & 2 & 0 & 0 \\ 1 & 0 & 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 & 1 & 2 \\ 0 & -1 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 1 & 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} \sigma_e^2 \\ \sigma_u^2 \\ \sigma_v^2 \\ \sigma_{eu} \\ \sigma_{ev} \\ \sigma_{uv} \end{pmatrix}. \quad (16)$$

Not all parameters are identified. For instance, try the reparameterization

$$\tilde{\sigma}_{eu} = \sigma_{eu} + \sigma_{uv},$$

$$\tilde{\sigma}_{ev} = \sigma_{ev} + \sigma_{uv},$$

$$\tilde{\sigma}_u^2 = \sigma_u^2 - \sigma_{uv},$$

$$\tilde{\sigma}_v^2 = \sigma_v^2 - \sigma_{uv}.$$

Substituting the four ‘tilde’ parameters for their original parameters, the covariance σ_{uv} drops out of the moment equations. Without loss of generality we can thus impose the identifying restriction $\sigma_{uv} = 0$. Subject to this assumption all other parameters are identified. The only parameter in (Eqs. (13–15)) that is always identified, regardless of σ_{uv} , is the random walk variance σ_e^2 . We lose one overidentifying restriction due to the model-free moment identity $\text{Cov}(\Delta x_t, z_t) - \text{Cov}(\Delta y_t, z_t) = \text{Var}(\Delta z_t)/2$.

The parameters can be estimated either by using GMM, or by choosing five moments from which they can be recovered exactly. GMM is not really feasible here, since we can not construct a weighting matrix.¹² Estimation of the structural

¹²The problem is that we just have the data moments and not the individual time series observations themselves. For a proper GMM estimator we must go back to the moment estimator developed in Sec. 2 and replace the $\gamma(i)$ moments by functions of the structural parameters and estimate the structural parameters directly from the quote data. Even then the construction of a GMM weighting matrix is computationally demanding. The problem is that the y_{ij} elements in the regression (Eq. (5)) have no natural time series ordering, so that the standard non-parametric Newey–West type estimator does not apply.

parameters is just a linear transformation of the data moments. For the empirical analysis we estimate the structural parameters using the variances of the actual and implied yen/dmark exchange rate returns, their covariance and the two lagged cross-covariances. The estimator is linear in the data moments and follows immediately from inverting the relevant submatrix in (Eq. (16)):

$$\begin{pmatrix} \sigma_u^2 \\ \sigma_v^2 \\ \sigma_e^2 \\ \sigma_{eu} \\ \sigma_{ev} \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 & -1 & 1 \\ 0 & 1 & -1 & 1 & -1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \gamma_{xx}(0) \\ \gamma_{yy}(0) \\ \gamma_{xy}(0) \\ \gamma_{xy}(-1) \\ \gamma_{xy}(1) \end{pmatrix}. \tag{17}$$

The covariance between the microstructure noise u_t and the information shock e_t is equal to the lagged cross-covariance $\text{Cov}(\Delta x_{t-1}, \Delta y_t)$, consistent with the interpretation of σ_{eu} as measuring lagged adjustments.

Two of the overidentifying restrictions come from the cointegration restriction. Cointegration [with cointegrating vector (1 -1)] implies that the long run variances of Δx_t and Δy_t are equal and also equal to the long run covariance:

$$\gamma_{xx}(0) + 2\gamma_{xx}(1) = \sum_{j=-1}^1 \gamma_{xy}(j) = \gamma_{yy}(0) + 2\gamma_{yy}(1). \tag{18}$$

From Eq. (18) the first order autocovariance can be written as a function of the cross-covariances. The other over-identifying restrictions are related to first order dynamics.

Table 3 reports the parameter estimates and examines the overidentifying restrictions. The estimates σ_e^2 measure the information intensity over the various segments of the day. The main result of the structural model is that the information related volatility (σ_e^2) peaks in the late Europe/early US segment. This is a time zone when it is (late) afternoon in Germany and when the Japanese business day is over. One would thus not expect much news about the underlying yen/dmark exchange rate in this segment of the trading day. The Europe/US segment also shows by far the largest noise variance for the dollar implied exchange rate. The late US hours are very different: there is very little genuine news, but the noise level in the implied rate remains at the same level as the rest of the day. Subject to the identification restriction $\sigma_{uv} = 0$, the strong positive covariance between the actual return and the lagged implied return translates into the negative covariance σ_{eu} between the innovation of the effective exchange rate and the microstructure noise in the direct yen/dmark rate. At the 1-min sampling interval the trading noise dominates the news in the effective exchange rate. Increasing the sampling interval will reverse this inequality, as the random walk component will eventually be responsible for almost all the variance in the returns.

The first order autocorrelations fit the restrictions remarkably well. Although

Table 3
Random walk plus noise model

| Structural parameters | | Full | Japan | Europe | Eur./US | US |
|--|---------|-------|-------|--------|---------|-------|
| σ_u^2 | | 5.32 | 6.20 | 5.06 | 4.96 | 4.00 |
| σ_v^2 | | 9.00 | 8.34 | 8.44 | 11.53 | 8.83 |
| σ_e^2 | | 4.67 | 4.47 | 4.61 | 6.32 | 1.97 |
| σ_{eu} | | -3.06 | -3.23 | -3.21 | -3.23 | -1.31 |
| σ_{ev} | | 0.26 | 0.60 | -0.12 | -0.02 | 0.47 |
| Overidentified moments | | | | | | |
| Corr($\Delta x_t, \Delta x_{t-1}$) | Data | -0.28 | -0.31 | -0.26 | -0.24 | -0.33 |
| | Implied | -0.25 | -0.29 | -0.22 | -0.18 | -0.37 |
| Corr($\Delta y_t, \Delta y_{t-1}$) | Data | -0.41 | -0.40 | -0.39 | -0.41 | -0.46 |
| | Implied | -0.40 | -0.40 | -0.39 | -0.39 | -0.45 |
| Var($x_t \Delta - y_t$) | Data | 17.32 | 16.35 | 16.89 | 20.78 | 15.08 |
| | Implied | 14.32 | 14.53 | 13.50 | 16.48 | 12.83 |
| Corr($\Delta x_t, x_t \Delta - y_t$) | Data | 0.18 | 0.24 | 0.15 | 0.18 | 0.33 |
| | Implied | 0.18 | 0.19 | 0.19 | 0.14 | 0.23 |
| Corr($\Delta y_t, x_t \Delta - y_t$) | Data | -0.68 | -0.56 | -0.57 | -0.54 | -0.46 |
| | Implied | -0.61 | -0.68 | -0.68 | -0.67 | -0.65 |

Notes: The upper panel reports the estimates of the 'structural' parameters of the model under the identifying assumption $\sigma_{uv} = 0$. σ_e^2 is the variance of the random walk innovation, σ_u^2 and σ_v^2 are the variances of the microstructure noise associated with the actual and implied yen/dmark, respectively. The lower panel examines the overidentified moments, that have not been used for estimation. The first line gives the data moments; on the second line are the moments that are implied by the structural parameters. See Table 1 for the definition of the different segments of the day.

not used in the estimation, the actual and implied autocorrelations are often almost identical, lending credibility to the stylized model. The large negative first order autocorrelations are entirely consistent with the cross-correlations of the two series. Still the autocorrelations in the data are generally larger than implied by the model.¹³ The implied variance of the error correction term is of the same order of magnitude as observed in the data, although the data show a larger error correction variance for every subsample. The same holds for the correlation of the returns and the deviation between the levels. Due to the large sample size the implications will surely be rejected at usual significance levels in a formal hypothesis test. Summarising, the implied exchange rate has predictive power for the actual yen/dmark rate, but the information is very noisy due to the large volatility.

From the structural model it is not clear which variable contains the most information about the *efficient* exchange rate. All that we observe are the two

¹³This small discrepancy in the estimated moments creates an unfortunate problem for estimating different time series models such as a Vector AutoRegression. Since the autocorrelations are negative, the long run variance $\lim_{k \rightarrow \infty} \text{Var}(x_{t+k} - x_t)/k = \gamma_{xx}(0) + 2\gamma_{xx}(1)$ (and similar for y_t) will be slightly smaller than the long run covariance $\lim_{k \rightarrow \infty} \text{Cov}(x_{t+k} - x_t, y_{t+k} - y_t)/k = \gamma_{xy}(-1) + \gamma_{xy}(0) + \gamma_{xy}(1)$. As a result the long horizon covariance matrix is not positive definite. This remains true if we add additional higher order covariances. Technically it means that we can not estimate an unrestricted VAR with a large lag length from the estimated data moments. The estimation routine will break down at inversion of the moment matrix.

exchange rates themselves; the efficient price is a latent variable. For inference on the efficient exchange rate we need to derive the conditional expectation of x_t^* given observations x_{t-i} and y_{t-i} ($i \geq 0$). The conditional moments of x_t^* are most easily computed from the reduced form of the structural model (Eqs. (13–15)). The reduced form is a bivariate first order MA process. The general form of a first order VMA is

$$\begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} + \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} \epsilon_{t-1} \\ \eta_{t-1} \end{pmatrix}, \tag{19}$$

where c_{ij} are parameters to be determined and where ϵ_t and η_t are the innovations of x_t and y_t , respectively. The variances of ϵ_t and η_t are denoted by σ_ϵ^2 and σ_η^2 ; the covariance will be denoted $\sigma_{\epsilon\eta}$. Since the two series x_t and y_t are cointegrated the long-term impact matrix of the VMA will be of rank one. Moreover, since the cointegrating vector equals $(1 \ -1)$, we have the restrictions:

$$1 + c_{11} - c_{21} = 0, \tag{20}$$

$$1 + c_{22} - c_{12} = 0. \tag{21}$$

Define $\phi = c_{21}$ and $\theta = c_{12}$. This gives the representation

$$\begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} \phi & \theta \\ \phi & \theta \end{pmatrix} \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} + \begin{pmatrix} (1 - \phi) & -\theta \\ -\phi & (1 - \theta) \end{pmatrix} \begin{pmatrix} \Delta \epsilon_t \\ \Delta \eta_t \end{pmatrix}. \tag{22}$$

The first term determines the long-run impact of the shocks. All terms with $\Delta \epsilon_t$ and $\Delta \eta_t$ describe the transitory dynamics.

The transformation from the five structural parameters to the five VMA parameters $(\theta, \phi, \sigma_\epsilon^2, \sigma_\eta^2, \sigma_{\epsilon\eta})$ is non-linear. The parameters ϕ and θ are the roots of a fourth degree polynomial. In general this has multiple solutions, from which we choose the unique solution for which (Eq. (22)) is an invertible MA process.

With the VMA already written in the form (Eq. (22)), the common trends representation (see Stock and Watson, 1988), identifying the permanent and transitory parts of the process, follows immediately as

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tau_t + \begin{pmatrix} 1 - \phi & -\theta \\ -\phi & 1 - \theta \end{pmatrix} \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix}, \tag{23}$$

$$\tau_t = \tau_{t-1} + \phi \epsilon_t + \theta \eta_t, \tag{24}$$

where τ_t is the common trend. Since the random walk component is uniquely identified, the common trend can be interpreted as the conditional expectation of the efficient price x_t^* given current and lagged quotes. The random walk innovation $\phi \epsilon_t + \theta \eta_t$ is a linear combination of the two innovations. Our interest is in the relative contribution of the two innovations. How much of a shock in the actual or implied yen/dmark rate is related to a change in the efficient exchange rate? In

other words, how much of a shock is permanent? The importance of each of the two innovations is measured by the variance decomposition of (Eq. (24)):

$$\sigma_{\text{tot}}^2 = \phi^2 \sigma_{\epsilon}^2 + 2\phi\theta\sigma_{\epsilon\eta} + \theta^2 \sigma_{\eta}^2 = \sigma_{\epsilon}^2. \quad (25)$$

If the covariance $\sigma_{\epsilon\eta} = 0$, the information share (see Hasbrouck, 1995) of x and y is defined as

$$k_{xx} = \frac{\phi^2 \sigma_{\epsilon}^2}{\sigma_{\epsilon}^2} \quad \text{and} \quad k_{yy} = \frac{\theta^2 \sigma_{\eta}^2}{\sigma_{\epsilon}^2}. \quad (26)$$

However, with a non-zero covariance between the innovations the information shares are not unambiguously defined; they depend on how the covariance term is split over k_{xx} and k_{yy} .

The volatilities σ_{ϵ} and σ_{η} reported in Table 4 show the same seasonal pattern as in the structural model. Volatility is highest in the period where European and US business hours overlap. It is in this most active period of the day that the direct yen/dmark obtains its highest information share. This time zone is the only part of the day that the actual yen/dmark rate has a higher information share than the implied rate. The information share of the actual yen/dmark rate is especially low during Japanese business hours. An explanation might be that during these hours the foreign exchange activity is lower than when European and/or US markets are open. When activity is low there is simply no room for a second vehicle currency.

The information share due to the covariance term in (Eq. (24)) is always approximately 0.3, implying that contemporaneous movements are already strong even at the high 1-min frequency.

Two data moments are especially important for determining the information

Table 4
Cointegrated first order moving average process

| VMA parameters | Full | Japan | Europe | Eur./US | US |
|-------------------------|-------|-------|--------|---------|-------|
| ϕ | 0.43 | 0.33 | 0.47 | 0.56 | 0.35 |
| θ | 0.34 | 0.38 | 0.35 | 0.29 | 0.24 |
| σ_{ϵ}^2 | 6.87 | 7.20 | 6.20 | 8.10 | 5.29 |
| σ_{η}^2 | 17.18 | 17.00 | 15.77 | 21.08 | 13.69 |
| $\sigma_{\epsilon\eta}$ | 4.86 | 4.84 | 4.24 | 6.34 | 3.07 |
| Information shares | | | | | |
| k_{xx} | 0.28 | 0.18 | 0.29 | 0.41 | 0.34 |
| k_{xy} | 0.30 | 0.27 | 0.30 | 0.32 | 0.26 |
| k_{yy} | 0.42 | 0.55 | 0.41 | 0.27 | 0.40 |

Notes: The VMA parameters are obtained from the structural parameters. The information shares are a variance decomposition of the random walk component of the VMA, with k_{xx} and k_{yy} being the part of the variance of the random walk innovation due to the variance of the innovations in x and y , respectively and with k_{xy} the part of the variance related to the covariance of x and y .

shares. First, the implied rate on average leads the actual rate [$\gamma_{xy}(1) > \gamma_{xy}(-1)$]. This normally leads to a high information share for the leading market, in this case the dollar markets. Second, the volatility of the actual rate is less than that of the implied rate [$\gamma_{xx}(0) < \gamma_{yy}(0)$]. This reduces the information advantage of the implied rate, since in the VMA(1) model a high variance implies that the series is less correlated with the random walk component and contains more noise. The covariance effect dominates in most parts of the day. However, in the Europe/US segment the noise in the dollar rates is so large that the direct yen/dmark quotes become a more reliable signal.

For impulse response analysis the innovations are transformed to orthogonal shocks μ_t and ν_t . Since the error covariance matrix can be decomposed in many different ways, it is necessary to have an additional identifying restriction. In the context of the cointegrated system the most useful decomposition is the one proposed by Blanchard and Quah (1989), which is closely related to the common trends representation. In the Blanchard–Quah decomposition ν_t is defined as the normalized random walk innovation:

$$\nu_t = \frac{\phi\epsilon_t + \theta\eta_t}{\sigma_\epsilon} \tag{27}$$

The second innovation, μ_t is defined as being orthogonal to ν_t with unit variance. Because ν_t by definition contains the full permanent impact, the second innovation μ_t will have only temporary effects; hence it is called the transitory shock. We associate the transitory shock with microstructure dynamics. The information shares measure the percentage of the variance of ν_t that can be attributed to the innovation of a particular variable.¹⁴

Table 5 shows the variance decomposition of the actual and implied yen/dmark rate. This decomposition only considers the variance of the shocks and not the total 1-min variance, which also includes the lagged effect of the noise. The permanent shock accounts for most of the innovation variance of both series.

Table 5
Blanchard–Quah variance decomposition

| Permanent fraction | Full | Japan | Europe | Eur./US | US |
|--------------------|------|-------|--------|---------|------|
| Actual | 0.66 | 0.56 | 0.66 | 0.79 | 0.65 |
| Implied | 0.78 | 0.86 | 0.76 | 0.69 | 0.71 |

Notes: The innovation variances of actual (σ_ϵ^2) and implied (σ_η^2) yen/dmark are decomposed in fractions attributed to the permanent and transitory components in a Blanchard–Quah decomposition. The entries show the fraction of the innovation variance attributed to the permanent shock.

¹⁴The information shares from the Blanchard–Quah decomposition are of course the same as in (Eq. (26)), since ν_t is nothing but the random walk innovation in (Eq. (23)).

For the full sample the Blanchard–Quah representation of the model is

$$\begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tau_t + \begin{pmatrix} 1.53 & -0.03 \\ -1.93 & 1.50 \end{pmatrix} \begin{pmatrix} \mu_t \\ v_t \end{pmatrix}, \quad (28)$$

$$\tau_t = \tau_{t-1} + 2.16v_t. \quad (29)$$

Impulse responses from this system will show that the two series react in opposite directions to the temporary shock μ_t . Also, the implied yen/dmark rate will ‘overreact’ with respect to the permanent shock v_t , with the initial impulse being $2.16 + 1.50$ and the long-run impact equal to 2.16 . The actual yen/dmark rate does not exhibit such ‘overreaction’. Because the model is a first order MA, all dynamic effects are fully incorporated after one period. The impulse responses for the different subsamples are qualitatively similar: opposite signs on the temporary shock and overreaction of the implied rate with respect to the permanent shock. The interpretation of the ‘overreaction’ in a microstructure model is not clear.

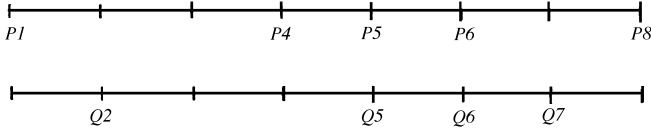
4. Conclusions

In this article we used quotes from the Reuters screens at 30-seconds and 1-minute intervals to investigate the dynamic interactions between the observed yen/dmark exchange rate and the implied rate obtained as the ratio of the yen/dollar and dmark/dollar rates. The analysis has been carried out using a moment estimator designed to account for the irregular spacing of high frequency data. The main empirical regularities are that (i) all time series of exchange rate quotes show very strong negative first order autocorrelations; (ii) microstructure noise accounts for much of the variance of both the actual as well as the implied quotes; (iii) the implied rate has predictive power for the actual rate; (iv) the implied yen/dmark rate is much noisier than the actual rate; (v) part of price discovery about the yen/dmark rate takes place through the actual yen/dmark cross-market with at least 25% of the fundamental news entering directly; (vi) also at least 25% of the fundamental news in the yen/dmark exchange rate arrives through the dollar markets; and (vii) the actual yen/dmark obtains its strongest role as an independent market during the busiest part of the day, when European and US business hours overlap.

Appendix 1

Bias in alternative sampling scheme

An alternative sampling scheme in calendar time would be to record every minute the current quote. If over the last minute there have been no quote revisions this means that last minute’s observation is repeated. The following graph illustrates this sampling process.



Because there are no new quotes between times 1 and 4, the observation p_1 is repeated at times 2 and 3. Returns in these periods are thus set to zero. The motivation behind this sampling scheme is that if there has not been any news, then the current quote might still be valid. Its main advantage is that missing value problems are avoided. Its drawback is a bias in the estimator of the autocovariance function. Below we will derive the bias for this *fill-in* sampler.

Let q_t be the data from the *fill-in* sample, i.e. $q_2 = p_1$ in the example in the graph above. With T observations $q_t(t = 1, \dots, T)$ the variance of returns with the fill-in sample is

$$\tilde{\gamma}(0) = \frac{1}{T} \sum_{t=2}^T (\Delta q_t)^2. \tag{30}$$

The tildes will generally denote a sample moment computed from the *fill-in* sample. The expected value of a term like $(\Delta q_4)^2 = (p_4 - p_1)^2$ depends on the autocovariances of the returns series: $E[(\Delta q_4)^2] = 3\gamma(0) + 2\gamma(1) + \gamma(2)$. In general,

$$E[(p_{t_{i+1}} - p_{t_i})^2] = \sum_{j=1}^{s_i} j\gamma(s_i - j), \tag{31}$$

where $s_i = t_{i+1} - t_i$ is the time between observations i and $i + 1$. Returns for periods without quote revisions are zero, so they do not contribute to the sum of squares in (Eq. (30)). We will calculate the expected value of the estimator $\tilde{\gamma}(0)$. Let N be the number of observations p_{t_i} , ($i = 1, \dots, N$). Further define f_s as the fraction of intervals of length s in the original p_{t_i} sample. These fractions satisfy the properties $\sum_{s=1}^K f_s = 1$ and $\sum_{s=1}^K s f_s = T/N$, where T/N is the average length of an interval and where K is the maximum interval length. Then:

$$\begin{aligned} E[\tilde{\gamma}(0)] &= \sum_{s=1}^K \frac{N f_s}{T} \sum_{j=1}^s j\gamma(j) \\ &= \frac{N}{T} \sum_{l=0}^{K-1} \sum_{i=l+1}^K f_i(i-l)\gamma(l). \end{aligned} \tag{32}$$

In the special case of no missing observations all interval lengths are equal to one, hence $f_1 = 1$ and all other f_i are zero. Then (Eq. (32)) reduces to $E[\tilde{\gamma}(0)] = \gamma(0)$ and there is no bias. Also, if $\gamma(l) = 0$ for $l > 0$, the time series is a random walk and again $E[\tilde{\gamma}(0)] = \gamma(0)$. In the more likely case that the price series is approximately a random walk plus noise, i.e. returns are an MA(1) process, the bias term amounts to

$$E[\tilde{\gamma}(0)] = \gamma(0) + \left(1 - \frac{N}{T}\right)\gamma(1), \tag{33}$$

where $(1 - N/T)$ is the fraction of missing values. Since in the empirical application $\gamma(1) < 0$, the variance will be underestimated. The first order autocovariance of the fill-in sample is computed as:

$$\tilde{\gamma}(1) = \frac{1}{T} \sum_{t=2}^T \Delta q_t \Delta q_{t-1}. \quad (34)$$

Its expected value can be computed analogously. The product $\Delta q_t \Delta q_{t-1}$ will only be non-zero if both Δq_t and Δq_{t-1} are non-zero. This only happens if there are new quotes at times t as well as $t - 1$. The summation in (Eq. (34)) can thus be rewritten as

$$\tilde{\gamma}(1) = \frac{1}{T} \sum_{i \in I_1} (p_{t_{i+1}} - p_{t_i})(p_{t_i} - p_{t_{i-1}}), \quad (35)$$

where $I_1 = \{i: t_{i+1} - t_i = 1\}$ is the set of indices that selects the Nf_1 observations with a 1-min interval. The expected value for each non-zero product in (Eq. (35)) is $\sum_{j=1}^{s_i} \gamma(j)$. The expected value of $\tilde{\gamma}(1)$ follows as

$$\begin{aligned} E[\tilde{\gamma}(1)] &= \frac{1}{T} \sum_{i \in I_1} \sum_{j=1}^{s_i} \gamma(j) \\ &= \frac{Nf_1}{T} \sum_{j=1}^K (1 - F_{1j}) \gamma(j), \end{aligned} \quad (36)$$

where F_{1j} is the fraction of intervals that are of length less than j , given that the previous interval was of length one ($F_{11} = 0, F_{1K} = 1$). The last line of the expression is obtained by changing the order of summation in the first line. The special case of an MA(1) process is of empirical interest. With only $\gamma(0)$ and $\gamma(1)$ non-zero we obtain

$$E[\tilde{\gamma}(1)] = \frac{Nf_1}{T} \gamma(1). \quad (37)$$

Since $(Nf_1/T < 1)$ the estimator $\tilde{\gamma}(1)$ will be biased towards zero. In large samples the first order autocorrelation $\tilde{\rho}(1)$ converges to

$$\tilde{\rho}(1) \rightarrow \frac{Nf_1}{T} \frac{\gamma(1)}{\gamma(0) + (1 - N/T)\gamma(1)} = \frac{Nf_1}{T} \frac{\rho(1)}{1 + (1 - N/T)\rho(1)}. \quad (38)$$

The first order autocorrelation from the *fill-in* sample will be biased towards zero, $|\tilde{\rho}(1)| < |\rho(1)|$ if $f_1 N/T < 1 + (1 - N/T)\rho(1)$. Since $\rho(1) > -1$, this inequality is always satisfied. The bias can be very large. For example, for the yen/dmark we find a first order autocorrelation $\rho(1) = -0.31$ (see Table 3 in Sec. 3). With a 1-min sampling interval the HFDF-93 dataset contains $(1 - N/T) = 44\%$ missing values and $f_1 = 28\%$ intervals of length one. Substituting these numbers in (Eq. (38)) we find that under the assumption of an MA(1) process $\tilde{\rho}(1) = -0.03$. This example highlights the potential vast bias in irregularly spaced samples. It also shows that our extremely negative first order autocorrelations are consistent with much smaller autocorrelations reported elsewhere in the literature (see Goodhart and Figliuoli, 1991 and Guillaume et al., 1997).

Similar, though much more complicated, bias expressions can be derived for all the autocovariances. The derivations will be facilitated by some assumptions on the distributions of the interval lengths. Most convenient would be a Markov process, which gives the probability distribution of the duration of the next interval given the current duration. An example is the Autoregressive Conditional Duration (ACD) model introduced by Engle and Russell (1997).

References

- Andersen, T.G., Bollerslev, T., 1997. DM Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies. *J. Fin.*, forthcoming.
- Blanchard, O.J., Quah, D., 1989. The dynamic effects of aggregate demand and supply disturbances. *Am. Econ. Rev.* 79, 655–673.
- Bollerslev, T., Domowitz, I., 1993. Trading patterns and prices in the interbank foreign exchange market. *J. Fin.* 48, 1421–1443.
- Dacorogna, M.M., Müller, U.A., Nagler, R.J., Olsen, R.B., Pictet, O.V., 1993. A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *J. Int. Mon. Fin.* 12, 413–438.
- De Jong, F., Nijman, T., 1997. High frequency analysis of lead-lag relationships between financial markets. *J. Emp. Fin.* 4, 259–277.
- Engle, R.F., Russell, J.R., 1997. Forecasting transaction rates: the autoregressive conditional duration model. *J. Emp. Fin.* 4, 187–212.
- Goodhart, C.A.E., Figliuoli, L., 1991. Every minute counts in financial markets. *J. Int. Mon. Fin.* 10, 23–52.
- Guillaume, D.M., Dacorogna, M.M., Dave, R.R., Müller, U.A., Olsen, R.B., Pictet, O.V., 1997. From the Bird's Eye to the Microscope: A Survey of New Stylized Facts of the Intra-Daily Foreign Exchange Markets, *Finance and Stockastics 1*. Olsen and Associates, Zurich, 95–129.
- deB Harris, F.H., McNish, T.H., Shoesmith, G.L., Wood, R.A., 1995. Cointegration, error correction, and price discovery on informationally linked security markets. *J. Fin. Quant. Anal.* 30, 563–579.
- Hartmann, P., 1996. Vehicle Currencies in the Foreign Exchange Market, Working Paper. LSE Financial Markets Group.
- Harvey, A.C., 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.
- Hasbrouck, J., 1993. Assessing the quality of a security market: a new approach to transaction-cost measurement. *Rev. Fin. Stud.* 6, 191–212.
- Hasbrouck, J., 1995. One security, many markets: determining the contributions to price discovery. *J. Fin.* 50, 1175–1199.
- Herwartz, H., 1996. Foreign Exchange Analysis with Missing Data: Comparing a Moment Estimator and the Kalman Filter, Working Paper. Humboldt Universitat, Berlin.
- Krugman, P., 1980. Vehicle currencies and the structure of international exchange. *J. Mon. Cr. Banking* 12, 503–526.
- Krugman, P., 1984. The international role of the dollar: theory and prospect. In: Bilson, J., Marston, R. (Eds.), *Exchange Rate Theory and Practice*. University of Chicago Press, Chicago.
- Müller, U.A., Dacorogna, M.M., Olsen, R.B., Pictet, O.V., Schwarz, M., Morgenege, C., 1990. Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. *J. Banking Fin.* 14, 1189–1208.
- Stock, J.H., Watson, M.W., 1988. Testing for common trends. *J. Am. Stat. Assoc.* 83, 1097–1107.