Real-Time Trading Models and the Statistical Properties of Foreign Exchange Rates

Ramazan Gençay* Giuseppe Ballocchi Michel Dacorogna
Richard Olsen Olivier Pictet

This Version: July 2000

Abstract

Real-time trading models use high frequency live data feeds and their recommendations are transmitted to the traders through data feed lines instantaneously. The contributions of this paper are twofold. First, the performance of a widely used commercial real-time trading model is compared with the performance of systematic currency traders. Second, the real-time trading model is used to evaluate the statistical properties of foreign exchange rates. The out-of-sample test period is seven years of high frequency data on three major foreign exchange rates against the US Dollar and one cross rate.

The trading model yields positive annualized returns (net of transaction costs) in all cases. Performance is measured by the annualized return, two measures of risk corrected annualized return, deal frequency and maximum drawdown. Their simulated probability distributions are calculated with the four well-known processes, the random walk, GARCH, AR-GARCH and AR-HARCH. The null hypothesis of whether the real-time performances of the foreign exchange series are consistent with these traditional processes is tested under the probability distributions of the performance measures. The results from the real-time trading model are not consistent with these processes.

JEL No: G14, C45, C52, C53

Key Words: Real-time trading models, exponential moving averages, robust kernels, technical trading.

*We are grateful to Bernard Hechinger for the Parker Systematic Index calculations reported in this paper. This paper was written while Gençay visited Olsen & Associates as a research scholar. He is grateful to them for their generosity and for providing an excellent research platform for high frequency finance research. Gençay also thanks the SSRHC and the NSERC for financial support. Address for correspondance: Ramazan Gençay, Department of Economics, University of Windsor, 401 Sunset Avenue, Windsor, Ontario, Canada. Email: gencay@uwindsor.ca, Tel: (519) 253 3000 ext. 2382, Fax: (519) 973 7096. Giuseppe Ballocchi, Michel Dacorogna and Richard Olsen are affiliated with Olsen & Associates, Switzerland. Olivier Pictet is with Dynamic Asset Management, Switzerland.

1. Introduction

The foreign exchange market is the largest financial market worldwide. It involves actors in different geographical locations, time zones, working hours, time horizons, home currencies, information access, transaction costs, and other institutional constraints. The time horizons vary from intraday dealers, who close their positions every evening, to long-term investors and central banks. In this highly complex structure, the market participants are faced with different constraints and use different strategies to reach their financial goals, which differ because of their heterogeneous risk profiles. The main essence behind the real time trading models is that they are designed to capture the conditional mean dynamics of the return process under this complex heterogeneous structure.

Real-time trading models are based on around-the-clock collection and treatment of foreign exchange market makers quotes at the tick-by-tick frequency level. The models use live data feeds and their recommendations are transmitted to the traders through telephone lines instantaneously. The most important reason for using high frequency data in these models is to be able to exploit the underlying heterogeneity of the foreign exchange market. In low frequency data, the intraday market microstructure properties are already averaged out so that only certain properties of the data generating process can be studied. This framework provides a general setting where all properties of the market ranging from intraday to long term are embedded in our econometric methodology.

The contributions of this study are twofold. First, the performance of a widely used commercial real-time trading model is compared to the performance of currency traders using technical analysis models. Second, the real-time trading model is used as a specification test to uncover the performance of well-known statistical processes for the price generating process. Moreover, the rejections of the models by this test provide directions for further research since the trading models capture medium to long-term mean dynamics of financial markets by indicating positive performance over long periods.

In the earlier literature, simple technical indicators for the securities market have been tested by Brock et al. (1992). Their study indicates that patterns uncovered by technical rules cannot be explained by simple linear processes or by changing the expected returns caused by changes in volatility[1]. LeBaron (1992,1997) and Levich and Thomas (1993) follow the methodology of Brock et al. (1992) and use bootstrap simulations to demonstrate the statistical significance of the technical trading rules against well-known parametric null models of exchange rates.

[1]In Gençay (1998), the DJIA data set of Brock et al. (1992) is studied with simple moving average indicators within the nonparametric conditional mean models. The results indicate that nonparametric models with buy-sell signals of the moving average models provide more accurate sign and mean squared prediction errors (MSPE) relative to random walk and GARCH models. Gençay (1999) shows that past buy-sell signals of simple moving average rules provide statistically significant sign predictions for modelling the conditional mean of the returns for the foreign exchange rates. The results in Gençay (1999) also indicate that past buy-sell signals of the simple moving average rules are more powerful for modelling the conditional mean dynamics in the nonparametric models.

In Sullivan et al. (1999), an extensive study of the trading rule performance is examined by extending the Brock et al. (1992) data for the period of 1987-1996. They show that the trading rule performance remains superior for the time period that Brock et al. (1992) studied, however, these gains disappear in the last ten years of the Dow Jones Industrial Average (DJIA) series. Overall, the scope of the most recent literature supports the technical analysis but it is limited to simple univariate technical rules. One particular exception is the study by Dacorogna et al. (1995) which examines real-time trading models of foreign exchanges under heterogeneous trading strategies. They conclude that it is the identification of the heterogeneous market microstructure in a trading model which leads to an excess return after adjusting for market risk. An extensive examination of the statistical properties and modelling methods of high frequency financial markets can be found in Dacorogna et al. (2000).

One important aspect of our analysis is that the real-time trading model had not been optimized during the eight years between 1989-1996 which we study. This gives us a unique platform to control for the data-snooping problems pointed out in Sullivan et al. (1999). The Brock et al. (1992) study has been mentioned frequently in the popular press[2] and this may have caused the patterns discussed in this paper to vanish post 1986, coupled with the low-cost and fast computing power. In principle, it is impossible to avoid data snooping unless one is in a real-time context. Even Brock et al. (1992) tried to control for it by selecting their strategy by studying what traders actually do before looking at the data. This type of strategy may be considered as data dependent such that the data set and the strategy inevitably *coevolved*. In our framework, the model has not been re-optimized during the eight years of real-time feed and this gives us a unique advantage. There is no socially determined co-evolutionary relationship between our data set and the technical strategies used in implementing our specification tests[3].

The results of this paper indicate that the four currencies pairs, USD-DEM (US Dollar-Deutsche Mark), USD-CHF (US Dollar-Swiss Franc), USD-FRF (US Dollar-French Franc) and DEM-JPY (Deutsche Mark-Japanese Yen) yield 9.63, 3.66, 8.20 and 6.43 percent annualized returns in the sampling period in consideration. These annualized returns are unleveraged, net of transaction costs and above the risk free interest rate. The Parker Systematic Index reports cumulated returns net of fees and interest for a class of systematic currency managers[4]. It therefore represents a good benchmark for the real-time trading model we study here both in terms of asset class and methodology. A detailed comparison with the RTT model performance shows that the RTT models perform on average 2 percent per year better than the Parker Systematic Index in the sampling period. Therefore, the RTT model performance is consistent with the performance of the systematic currency traders.

In the second part of this paper we use simulated data to obtain probability distributions of RTT model performances. The simulated probability distributions of the performance measures are calculated with the random walk, GARCH(1,1) and AR(4)-GARCH(1,1), AR(4)-

[2] In particular by Hulbert Financial Digest (`www.hulbertdigest.com`)) which rates investment newsletters and is widely read by investors.

[3] We thank a referee for his stimulating comments on this issue.

[4] In their terminology, "systematic currency managers" means traders who follow technical analysis models.

HARCH(9) processes at the 5-minute data frequency. The null hypothesis of whether the real-time performances of the foreign exchange series are consistent with these traditional processes is rejected under the probability distributions of the performance measures.

Since the trading frequency of the model is less than two deals per week, the trading model does not pick up the five minute level heteroskedastic structure at the weekly frequency. Rather, the heteroskedastic structure behaves as if it is measurement noise where the model takes positions and this leads to the rejection of the GARCH(1,1) as a data generation process of the foreign exchange series. A similar explanation holds for the rejection of the AR(4)-GARCH(1,1). The model picks up the high frequency serial correlation as a noise and this short-term behavior works against the process. This cannot be treated as a failure of the real-time trading model. Rather, this strong rejection is evidence of the failure of the temporal aggregation properties of the AR(4)-GARCH(1,1) process over lower frequencies.

The conditional mean dynamics of the AR(4)-HARCH(9) model does not help the trading model capture possible trends because the movements in the prices are too small for the model to take a position. On the other hand, the simulated series are more volatile due to the fully absorbed tail dynamics and the long memory effect. In the absence of a weekly conditional mean information (this information is left to the residuals and is unconditional now), the trading model cannot interpret these large moves and ends up taking consecutive wrong positions. Therefore, one implication of our results is that a realistic volatility model should take the scaling properties[5] of the data into account in *estimation*. This may, for instance, be accomplished by carrying out the estimation both in the time as well as in the frequency domain.

This paper is organized as follows. In section two, the performance measures are explained. In section three, the comparison to the Parker Systematic Index is reported. The simulation models and the simulation methodology are presented in section four. The technical indicators and their robustness properties are explained in section five. The trading models are described in section six. We discuss the empirical results in section seven. We conclude afterwards.

2. Performance Measures

Evaluating the performance of an investment strategy generally gives rise to many debates. This is due to the fact that the performance of any financial asset cannot be measured only by the increase of capital but also by the risk incurred during the time to reach this increase. Returns and risk must be evaluated together to assess the quality of an investment. In this section we describe the performance measures[6] used to evaluate the trading models in this paper. The *total return*, R_T, is a measure of the overall success of a trading strategy over a

[5]The scaling properties of foreign exchange returns and volatilities are studied in Müller et al. (1990) and Andersen et al. (1999). LeBaron (1999) demonstrates that a non self-similar process, such as stochastic volatility, may also exhibit scaling behaviour. Brock (1999) has an extensive and stimulating discussion of the problem of scaling in economics.

[6]The performance measures of this paper are also used in Pictet et al. (1992) and Dacorogna et al. (1999).

period T, and defined by

$$R_T \equiv \sum_{j=1}^{n} r_j \tag{1}$$

where n is the total number of transactions during the period T, j is the jth transaction and r_j is the return from the jth transaction. The total return expresses the amount of profit (or loss) made by a trader always investing up to his initial capital or credit limit in his home currency. The annualized return, $\bar{R}_{T,A}$, is calculated by multiplying the total return with the ratio of the number of days in a year to the total number of days in the entire period. The *maximum drawdown*, D_T, over a certain period $T = t_E - t_0$, is defined by $D_T \equiv \max(R_{t_a} - R_{t_b} \mid t_0 \leq t_a \leq t_b \leq t_E)$ where R_{t_a} and R_{t_b} are the total returns of the periods from t_0 to t_a and t_b, respectively.

In order to achieve a high performance and good acceptance among investors, investment strategies or trading model performance should fulfill a few conditions: provide high total return, a smooth increase of the equity curve over time and a small clustering of losses. The fulfilment of these conditions would account for a high return and low risk investment. In addition, a performance measure should present no bias towards low frequency models by always including the unrealized return of the open position and not only the net result after closing the position. Sharpe (1966) introduced a measure of mutual funds performance which he called at that time a *reward-to-variability ratio*. This performance measure was to later become the industry standard in the portfolio management community under the name of Sharpe ratio, Sharpe (1994).

Unfortunately, the Sharpe ratio is numerically unstable for small variances of returns and cannot consider the clustering of profit and loss trades. As the basis of a risk-sensitive performance measure, we define a cumulative variable $\tilde{R}_t$, at time t, as the sum of the total return R_T of equation (1) and the unrealized current return of the open position. This quantity reflects the current value of the investment and includes not only the results of previously closed transactions but also the value of the open position (mark-to-market). This means that $\tilde{R}_t$ is measuring the risk independently of the actual trading frequency of the model. Similar to the difference between price and returns, the variable of relevance for the utility function is the change of $\tilde{R}$ over a time interval Δt: $X_{\Delta t} = \tilde{R}_t - \tilde{R}_{t-\Delta t}$ where t expresses the time of the measurement. In this paper, Δt is allowed to vary from seven days to 301 days.

A risk-sensitive measure of trading model performance can be derived from the utility function framework (Keeney and Raiffa (1976)). Let us assume that the variable $X_{\Delta t}$ follows a Gaussian random walk with mean $\overline{X}_{\Delta t}$ and the risk aversion parameter α is constant with respect to $X_{\Delta t}$. The resulting utility $u(X_{\Delta t})$ of an observation is $-\exp(-\alpha X_{\Delta t})$, with an expectation value of $\overline{u} = u(\overline{X}_{\Delta t}) \exp(\alpha^2 \sigma^2_{\Delta t}/2)$, where $\sigma^2_{\Delta t}$ is the variance of $X_{\Delta t}$. The expected utility can be transformed back to the *effective return*, $X_{\mathit{eff}} = -\log(-\overline{u})/\alpha$ where $X_{\mathit{eff}} = \overline{X}_{\Delta t} - \frac{\alpha \sigma^2_{\Delta t}}{2}$. The risk term $\alpha \sigma^2_{\Delta t}/2$ can be regarded as a risk premium deducted from the original return where $\sigma^2_{\Delta t}$ is computed by $\sigma^2_{\Delta t} = \frac{n}{n-1} \left(\overline{X^2_{\Delta t}} - \overline{X}^2_{\Delta t} \right)$. Unlike the Sharpe ratio, this measure is numerically stable and can differentiate between two trading models with a straight line behavior ($\sigma^2_{\Delta t} = 0$) by choosing the one with the better average return[7].

The measure X_{eff} still depends on the size of the time interval Δt. It is hard to compare X_{eff} values for different intervals. The usual way to enable such a comparison is through the annualization factor, $A_{\Delta t}$[8],

$$X_{\mathit{eff},ann,\Delta t} \;=\; A_{\Delta t}\, X_{\mathit{eff}} \;=\; \overline{X} - \frac{\alpha}{2}\, A_{\Delta t}\, \sigma^2_{\Delta t} \tag{2}$$

where $\overline{X}$ is the annualized return and it is no longer dependent on Δt. The factor $A_{\Delta t}\sigma^2_{\Delta t}$ has a constant expectation, independent of Δt. This annualized measure still has a risk term associated with Δt and is insensitive to changes occurring with much longer or much shorter horizons. To achieve a measure that simultaneously considers a wide range of horizons, a weighted average of several $X_{\mathit{eff},ann}$ is computed with n different time horizons Δt_i, and thus takes advantage of the fact that annualized $X_{\mathit{eff},ann}$ can be directly compared $X_{\mathit{eff}} = \frac{\sum_{i=1}^{n} w_i X_{\mathit{eff},ann,\Delta t_i}}{\sum_{i=1}^{n} w_i}$ where the weights w are chosen according to the relative importance of the time horizons Δt_i and may differ for trading models with different trading frequencies. In this paper, α is set to $\alpha = 0.1$ when the returns are expressed in percent. If they are expressed in numbers, α would be equal to 10.

The risk term of X_{eff} is based on the volatility of the total return curve against time, where a steady, linear growth of the total return represents the zero volatility case. This volatility measure of the total return curve treats positive and negative deviations symmetrically, whereas foreign exchange dealers become more risk averse in the loss zone and do hardly care about the clustering of positive profits. A measure which treats the negative and positive zones asymmetrically is defined to be R_{eff} (Müller, Dacorogna and Pictet (1993) and Dacorogna et al. (1999)) where R_{eff} has a high risk aversion in the zone of negative returns and a low one in the zone of profits whereas X_{eff} assumes constant risk aversion. A high risk aversion in the zone of negative returns means that the performance measure is dominated by the large drawdowns. The R_{eff} has two risk aversion levels: a low one, α_+, for positive $\Delta\tilde{R}_t$ (profit intervals) and a high one, α_-, for negative $\Delta\tilde{R}_t$ (drawdowns)

$$\alpha \;=\; \begin{cases} \alpha_+ & for \quad \Delta\tilde{R}_t \geq 0 \\ \alpha_- & for \quad \Delta\tilde{R}_t < 0 \end{cases}$$

where $\alpha_+ < \alpha_-$. The high value of α_- reflects the high risk aversion of typical market participants in the loss zone. Trading models may have some losses but, if the loss observations strongly *vary* in size, the risk of very large losses becomes unacceptably high. On the side of the positive profit observations, a certain regularity of profits is also better than a strong variation in size. However, this distribution of positive returns is never as *vital* for the future of market

[7] An example for the limitation of the Sharpe ratio is its inability to distinguish between two straight line equity curves with different slopes.

[8] $A_{\Delta t}$ is the ratio of the number of Δt in a year divided by the number of Δt's in the full sample.

participants as is the distribution of losses (drawdowns). Therefore, α_+ is much smaller than α_-. In this paper, we assume that $\alpha+ = \alpha_-/4$ and $\alpha_- = .20$, when the quantities are expressed in percent.

Both X_{eff} and R_{eff} are quite natural measures. They treat risk as a discount factor to the value of the investment. In other words, the performance of the model is discounted by the amount of risk that was taken to achieve it. In the X_{eff} case the risk is treated similarly both for positive or negative outcome while in the case of R_{eff}, negative performance is more penalized.

Amongst annualized returns, X_{eff} and R_{eff}, the last two performance measures are more stringent since they examine the entire equity curve contrary to the annualized total return. The annualized return, on the other hand, leaves large degrees of freedom to an infinite number of equity curve paths by only considering the beginning and the end points of the equity curve performance.

It is also important to point out that a buy-and-hold strategy, as a benchmark performance measure, in foreign exchange markets is not appropriate as there is no trend for the main rates (the Peso effect is only existing for emerging market currencies). This is due to the fact that foreign exchange markets are, by nature, symmetric and the direction of a buy and hold strategy is completely arbitrary. The buy-and-hold strategy is more appropriate for the equity markets.

3. Parker Systematic Index and the RTT Model

The Parker Systematic Index[9] reports cumulated returns net of fees and interest for a class of systematic currency managers. It therefore represents a good benchmark for the real-time trading (RTT) model we study here, both in terms of asset class and methodology. To compare the trading model performance with the Parker Systematic Index, an equally weighted portfolio of monthly returns of the USD-DEM, USD-CHF, USD-FRF and DEM-JPY models is constructed. We define the raw return of the portfolio by $r_i = \frac{1}{4}\sum_{j=1}^{4} r_{ij}$ where r_{ij}'s are the individual monthly returns of the respective RTT models. This portfolio is leveraged so that it reaches the same standard deviation as the Parker Systematic Index (in other words, comparing two portfolios with the same risk characteristics) over the sample period. This normalization is obtained by multiplying each monthly raw return by the ratio of the standard deviation of the Parker Index monthly returns σ_{PI} and the standard deviation of the monthly returns of our unleveraged portfolio, σ_{RTTP}, $\tilde{r}_i = \frac{\sigma_{PI}}{\sigma_{RTTP}} r_i$. To compute the cumulated return, we simply

[9]The Parker Systematic Index is a performance-based benchmark that measures both the reported and the risk adjusted returns of global currency managers. The Index tracks the performance, or value-added, which managers have generated from positioning long or short foreign currencies. It is equally weighted, as opposed to capitalization weighted, to preclude very large managers from swaying the performance in a direction that may not be representative of the currency manager universe. As reported by the Parker Global Strategies, the Parker Systematic Index currently includes programs managing over $9.4 billion in currency assets. Interested readers are referred to the web site of the Parker Global Strategies company (`http://www.parkerglobal.com/`) which provides these indices.

	Cumulative Returns	Annualized Cumulative Return
Parker Systematic Index	104.49%	10.76%
Portfolio of Equally Weighted RTT Models:		
Leveraged to equal std deviation	218.01%	17.97%
After mgmt fees of 2%	168.32%	15.15%
mgmt fees + performance fees	130.24%	12.65%

Table 1: Performance comparison between the Parker Systematic index and a portfolio of the RTT models during the same period from 1.1.1990 to 31.12.1996.

apply the following relation

$$C_{RTTP} = \prod_{i=1}^{N}(1+\tilde{r}_i) = \prod_{i=1}^{N} C_i \tag{3}$$

where N is the number of months in our sample and C_i is the capital achieved in the i-th month.

The trading model raw returns are net of interest but the performance needs to be adjusted for asset management and performance fees. Although the studied real-time trading model has not been sold on a performance fee basis, we corrected the returns with the standard performance and management fees of currency fund managers: 2 percent management fee on assets is deducted monthly in the following way $c_i = 1 + \tilde{r}_i - f \cdot c_{i-1}$ where c_i is the capital achieved the i-th month net of management fees and $f = \frac{1}{12} \cdot 0.02$, which is the monthly percentage for an annualized 2 percent. This net monthly capital is then multiplied by the previous one to get the cumulative return as in equation (3). In addition, we deduct an annual performance fee of 20 percent on net profits. This is done yearly by reducing the capital accumulated over the year by 20 percent of the net profit as it is customary in actual fund management environments. The results shown in Table 1 indicate that the yearly trading model performance is almost 2 percent above the Parker Systematic Index per year in the sampling period. This is true even after deduction of all management and performance fees and leads to a capital that is 25 percent larger at the end of the sampling period for the RTT models.

4. Simulation Methodology

The distributions of the performance measures under various null processes are calculated by using a simulation methodology. The recorded prices in the database are composed of three quantities. These quantities are the time t_j at which the price has been recorded, the ask price

$p_{ask,j}$ and the bid price $p_{bid,j}$. The sequence of the tick recording times t_j is unequally spaced. The majority of these ticks are concentrated in the periods of high market activity. The real-time trading model indicators do not directly analyze raw bid or ask prices, but rather the logarithmic *middle prices*, x_j, are utilized. x_j is defined as $x_j = (log(p_{ask,j}) + log(p_{bid,j}))/2$. Similarly we define the relative bid-ask spreads, s_j, as $s_j = log(p_{ask,j}) - log(p_{bid,j})$.

In our trading model simulations, we use a 5 minute interval sampling of the prices in order to keep the computation within manageable bounds. It is a good compromise between efficient computation and realistic behavior when compared to the real-time trading model results generated from all ticks. The main information used by a trading model to update its indicators is the logarithmic price changes or *returns*. The return between two consecutive selected ticks at time t_{j-1} and t_j is defined as $r_j = x_j - x_{j-1}$ and the corresponding elapsed ϑ-time[10] between these two ticks is $\Delta\theta_j = \theta_j - \theta_{j-1}$. By construction, the average elapsed theta time between two ticks, $\overline{\Delta\theta}$, is nearly five minutes. Multiple time series from a given theoretical price generation process need to be generated. To keep the impact of special events like the data holes in the model behavior, we decided to replace the different bid/ask price values but always keep the recorded time values. As the different ticks are not exactly regularly spaced, even in theta time, the average return corresponding to a five minute interval needs to be calculated. This is calculated by rescaling the observed return values, $r_j^* = r_j \left(\frac{\Delta\theta_j}{\overline{\Delta\theta}}\right)^{1/E}$ where the exponent $1/E$ is called the drift exponent and it is set to 0.5 under the random walk process. To obtain meaningful results, a simulated time series should have the same average drift α and average variance σ^2 as the observed returns. This is done by generating returns, $\hat{r}_j$, corresponding to a five minute interval in ϑ-time. In the case of a random walk process, the returns $\hat{r}_j$ are computed with, $\hat{r}_j = \alpha + \epsilon_j$ where[11] $\epsilon_j \sim N(0, \sigma^2)$. When the effective elapsed time between two ticks, $\Delta\theta_j$, is not exactly five minutes, we scale again the generated return using the same scaling formula, $r'_j = \hat{r}_j \left(\frac{\Delta\theta_j}{\overline{\Delta\theta}}\right)^{1/2}$ where $\overline{\Delta\theta}$ is five minutes. If there is a data hole, the sum of the generated return $\hat{r}_i$ is computed until the sum of the added five minute intervals is larger than the size of the data hole measured in ϑ-time. The sum of the returns is scaled with the same technique, i.e., $\sum_{i=1}^{n} \Delta\theta_i n \ \overline{\Delta\theta}^{1/2}$. The simulated logarithmic prices, x'_j, are computed by adding the generated returns r'_j to the first real logarithmic price value x_0. The bid/ask prices are computed by subtracting or adding half the average spread.

The GARCH(1,1) process is written as

$$r_t = \gamma_0 + \epsilon_t \tag{4}$$

[10] The high frequency data inherits intraday seasonalities and requires deseasonalization. This paper uses the deseasonalization methodology advocated in Dacorogna et al. (1993) named as the ϑ-time seasonality correction method. The ϑ-time method uses a business time scale and utilizes the average volatility combined with its scaling behavior to represent the activity of the market. The activity is divided into three geographical markets namely East Asia, Europe and the North America. A more detailed exposition of the ϑ methodology is presented in Dacorogna et al. (1993).

[11] In the simulations, ϵ is specified to be normally distributed. We also explored bootstrapping the residuals of the studied models. The main findings of the paper remain unchanged between these two approaches.

where $\epsilon_t = h_t^{1/2} z_t$, $z_t \sim$ N(0,1) and $h_t = \alpha_0 + \alpha_1 h_{t-1} + \beta_1 \epsilon_{t-1}^2$. GARCH specification (Bollerslev (1986)) allows the conditional second moments of the return process to be serially correlated. This specification implies that periods of high (low) volatility are likely to be followed by periods of high (low) volatility. GARCH specification allows the volatility to change over time and the expected returns are a function of past returns as well as volatility.

The AR(p)-GARCH(1,1) process is written as

$$r_t = \gamma_0 + \sum_{i=1}^{p} \gamma_i r_{t-i} + \epsilon_t \tag{5}$$

where $\epsilon_t = h_t^{1/2} z_t$ $z_t \sim$ N(0,1) and $h_t = \alpha_0 + \alpha_1 h_{t-1} + \beta_1 \epsilon_{t-1}^2$.

The parameters and the normalized residuals of the GARCH(1,1) process are estimated using the maximum likelihood procedure. The simulated returns are generated from the simulated normalized residuals and the estimated parameters. The estimated parameters of the AR(p)-GARCH(1,1) processes together with the simulated residuals are used to generate the simulated returns for this process. As before, half of the average spread is subtracted (added) from the simulated price process to obtain the simulated bid (ask) prices.

The AR(p)-HARCH(n) process, Müller et al. (1997) is written as

$$r_t = \gamma_0 + \sum_{i=1}^{p} \gamma_i r_{t-i} + \sigma_t \varepsilon_t,$$

$$\sigma_t^2 = c_0 + \sum_{j=1}^{n} c_j \sigma_{t_j}^2, \qquad \sigma_{t_j}^2 = \mu_j \sigma_{t-1,j}^2 + (1-\mu_j)\left(\sum_{i=1}^{k_j} r_{t-i}\right)^2 \tag{6}$$

where $k_j = 4^{j-2} + 1$ for $j > 1$ and $k_1 = 1$. k_j is chosen so that each horizon corresponds to meaningful time horizons. In this case, $k_j = 1, 2, \ldots, 9$ corresponds to 5 minutes, 10 minutes up to 10 days. The weight parameter μ_j is chosen such that $\mu_j = e^{-2/(k_{j+1}-k_j)}$. This choice guarantees that the center of the weight is in the middle of a given data interval[12]. The details of this model can be found in Müller et al. (1997).

The heterogeneous set of relevant interval sizes leads to the process name HARCH for "Heterogeneous Autoregressive Conditional Heteroskedasticity". The HARCH process belongs to the wide ARCH family but differs from all other ARCH-type processes in the unique property of considering the volatilities of price changes measured over different interval sizes. Due to this property, the HARCH process is able to capture the hyperbolic decay of the volatility autocorrelations.

For each replication we start by generating the simulated data a year before the model is tested. This year is 1989 and it is used to create the history dependency in returns and to initialize the different trading model indicators.

[12] In the literature, this model is referred to as AR(p)-EMAHARCH(n) as well.

5. Exponential Moving Averages with Robust Kernels

We turn now to the description of the techniques used to build the trading models. One basic element is the moving average operator. Indicators based on moving averages are used to summarize the past behavior of a time series at a given point in time. In many cases, they are used in the form of a *momentum or differential*, the difference between two moving averages. The moving averages can be defined with their *weight* or *kernel* function. The choice of the kernel function has an influence on the behavior of the moving average indicator. A particular type of moving average called *exponential average* plays an important role in the technical analysis literature. Exponential moving average (EMA) operator is a simple average operator with

$$w_{ema}(t;\tau) = \frac{e^{-t/\tau}}{\tau}$$

an exponential decaying kernel. τ determines the range of the operator and t indexes the time. An EMA is written as

$$EMA_p(\tau,t) = \int_{-\infty}^{t} w_{ema}(t-t')p(t')dt'$$

where

$$w_{ema}(t-t';\tau) = \frac{e^{-(t-t')/\tau}}{\tau} \ .$$

In Figure 1, we present the kernel function of an exponential moving average with $\tau = 0.5, 20$ and their differential kernel. The sequential computation of exponential moving averages is simple with the help of a recursion formula and it is more efficient than the computation of any differently weighted moving averages.

This basic exponential average kernel can be iterated to provide a family of *iterated exponential moving average* kernels (Müller (1989, 1991), Zumbach and Müller (2000))

$$w_{iema}(t;\tau,n) = \frac{1}{(n-1)!}\frac{e^{-t/\tau}}{\tau}(t/\tau)^{n-1} \ .$$

The larger the n, the more weight is allocated towards the middle range of the kernel. In the limit as n goes to infinity, the iterated exponential average behaves like a bell-shaped curve. This implies that the center of the weight is placed in the middle range of the kernel rather than the most recent past. For instance, a second iterative EMA, EMA^2 is written as

$$EMA_p^2(\tau,t) = \int_{-\infty}^{t} \frac{(t-t')}{\tau} w_{ema}(t-t')p(t')dt'$$

where

$$w_{ema}(t-t';\tau) = \frac{e^{-(t-t')/\tau}}{\tau} \ .$$

In general, an n iterative EMA, EMA^n is written as

$$EMA_p^n(\tau,t) = \frac{1}{(n-1)!}\int_{-\infty}^{t} \frac{(t-t')^{n-1}}{\tau^{n-1}} w_{ema}(t-t')p(t')dt'$$

where

$$w_{ema}(t-t';\tau) = \frac{e^{-(t-t')/\tau}}{\tau} \ .$$

In Figure 2, the iterative moving averages for $n = 1, 2$ and $n = 4$ are plotted. It makes it clear that as n gets larger the center of the weight distribution moves to the middle part of the kernel function. *A simple arithmetic moving average* of length m has a rectangular kernel which makes it very sensitive to the observations leaving the average when the average moves over time. More robust classes of kernels that remedy this sensitivity are those that assign exponentially decaying weights to the observations in the more distant past. These classes of robust kernels are obtained from the simple arithmetic average of the iterated exponential average kernels

$$w_{ma}(t;\tau,n) = (1/n)\sum_{j=1}^{n} w_{iema}(t;\tau',j) \tag{7}$$

where $\tau' = 2\tau/(n+1)$ so that the range, r, is independent of n. The robust exponential moving average is written as

$$MA_p^n(\tau,t) = \int_{-\infty}^{t} w_{ma}(t-t',\tau,n)p(t')dt' \ . \tag{8}$$

This is a special case where all weights assigned to each iterative kernel are the same in equation (7). Examples of these robust kernels are presented in Figure 3 where equally weighted iterative exponential moving average kernels are plotted up to $n = 8$. The property of this kernel is a plateau in the function before it asymptotically declines to zero and it is robust against extreme variations leaving the average by assigning exponentially decaying weights. Contrary to simple exponential average, which is very sensitive to the most recent history, it assigns relatively uniform weights to such new information. Therefore, a robust kernel preserves only the desirable robustness properties of the simple average and exponential average kernels but ignores their highly noisy unrobust properties. In Figure 4, a robust differential kernel is presented which is based on the difference between the exponential moving average with $\tau = 1$ and a robust kernel with $w_{ma}(\tau = 1, n = 8)$. By construction, the area under the kernel[13] sums to zero. The differential kernel assigns positive weights to the recent past and negative weights to the distant past. The real-time trading model of this paper uses a similar robust differential kernel in the construction of the gearing function.

[13]For a full description of this methodology and its theoretical foundations, the reader is referred to Zumbach and Müller (2000).

6. Trading Models

A distinction should be made between a price change forecast and an actual trading recommendation. A trading recommendation naturally includes a price change forecast, but it must also account for the specific constraints of the dealer of the respective trading model because a trading model is constrained by its past trading history and the positions to which it is committed. A price forecasting model, on the other hand, is not limited to similar types of constraints. A trading model thus goes beyond predicting a price change such that it must decide if and at what time a certain action has to be taken.

Trading models offer a real-time analysis of foreign exchange movements and generate explicit trading recommendations. These models are based on the continuous collection and treatment of foreign exchange quotes by market makers around-the-clock at the tick-by-tick frequency level. There are important reasons to utilize high frequency data in the real-time trading models. The first one is that the model indicators acquire robustness by utilizing the intraday volatility behavior in their build-up. The second reason is that any position taken by the model may need to be reversed quickly although these position reversals may not need to be observed often. The stop-loss objectives need to be satisfied and the high frequency data provides an appropriate platform for this requirement. Third, the customer's trading positions and strategies within a trading model can only be replicated with a high statistical degree of accuracy by utilizing high frequency data in a real-time trading model. More importantly for this study, the high frequency data in these models lets us learn the underlying heterogeneous market microstructure properties of the foreign exchange markets.

The trading models imitate the trading conditions of the real foreign exchange market as closely as possible. The deals are announced by the system three minutes before the execution so that a human foreign exchange dealer can monitor a specific trade. After this announcement, the model picks a price[14] from the real-time data feed, executes a deal and calculates its return. Since these prices are selected in real-time they are not subject to slippage. In order to imitate real-world trading accurately, the models take transaction costs into account, they do not trade outside market working hours except for executing stop-loss and they avoid trading at a frequency which cannot be followed by a human trader. In short, these models act realistically in a manner which a human dealer can easily follow.

Every trading model is associated with a local market that is identified with a corresponding geographical region. In turn, this is associated with generally accepted office hours and public holidays. The local market is defined to be open at any time during office hours provided it

[14]The execution price does not rely on a single price and is determined by the median price in the last few minutes of trading.

Slippage is not a factor in a liquid foreign exchange rate if trades are executed when the market is fully open, as the RTT model does. For instance, several hundred million dollars in USD-DEM can be traded without incurring slippage. Slippage may occur with the cross rates such as DEM-JPY, where sometimes the frequency of quote update is low and the actual price in the market is different from what the trading model receives from the data supplier.

is neither a weekend nor a public holiday. The O&A trading models presently support the Zurich, London, Frankfurt, Vienna and New York markets. Typical opening hours for a model are between 8:00 and 17:30 local time, the exact times depending on the particular local market.

The central part of a trading model is the analysis of the past price movements which are summarized within a trading model in terms of indicators. The indicators are then mapped into actual trading positions by applying various rules. For instance, a model may enter a long position if an indicator exceeds a certain threshold. Other rules determine whether a deal may be made at all. Among various factors, these rules determine the timing of the recommendation. A trading model thus consists of a set of indicator computations combined with a collection of rules. The former are functions of the price history. The latter determine the applicability of the indicator computations to generating trading recommendations. The model gives a recommendation not only for the direction but also for the amount of the exposure. The possible exposures (gearings) are $\pm\frac{1}{2}$ (half exposure), ± 1 (full exposure) or 0 (no exposure). The trading model works with a maximum gearing of 1. If the model is fully exposed it cannot buy or sell anymore but can revert its position.

6.1. The Real Time Trading (RTT) Model

The real-time trading model studied in this paper is classified as a one-horizon, high risk/high return model. The RTT is a trend-following model and takes positions when an indicator crosses a threshold. The indicator is momentum based, calculated through specially weighted moving averages with repeated application of the exponential moving average operator. In the case of extreme foreign exchange movements, however, the model adopts an overbought/oversold (contrarian) behavior and recommends taking a position against the current trend. The contrarian strategy is governed by rules that take the recent trading history of the model into account[15]. The RTT model goes neutral only to save profits or when a stop-loss is reached. Its profit objective is typically at three percent. When this objective is reached, a gliding stop-loss prevents the model from losing a large part of the profit already made by triggering its going neutral when the market reverses.

At any point in time t, the gearing function for the RTT is

$$g_t(I_p) = sign(I_p(t))\ f(|I_p(t)|)\ c(I(t))$$

where

$$I_p(t) = p_t - MA_p^4(\tau\ =\ 20)$$

[15]Hong and Stein (1999) demonstrate that if information diffuses gradually across a population, prices underreact in the short run. The underreaction means that the momentum traders can profit by trend-chasing. However, if they can only implement simple univariate strategies, their attempts at arbitrage must lead to overreaction at longer horizons.

where p_t is the logarithmic price at time t, τ refers to 20 day equally weighted iterative moving average and

$$f(|I_p(t)|) \quad = \quad \begin{cases} if & |I_p(t)| > b & 1 \\ if & a < |I_p(t)| < b & 0.5 \\ if & |I_p(t)| < a & 0 \end{cases}$$

and

$$c_t(I) \quad = \quad \begin{cases} +1 & if \quad |I_p(t)| < d \\ -1 & if \quad |I_p(t)| > d \; and \; g_{t-1} \cdot sign(I_p(t)) > 0 \; and \; r_l > P \end{cases}$$

where $a < b < d$ and r_l is the return of the last deal and P the profit objective. The function, $f(|I_p(t)|)$, measures the size of the signal at time t and the function, $c(|I_p|)$, acts as a contrarian strategy. The model will enter a contrarian position only if it has reached its profit objective with a trend following position. In a typical year, the model will play against the trend 2 to 3 times while it deals roughly 60 to 70 times. The hit rate of the contrarian strategy is of about 75 percent.

The parameters a and b depend on the position of the model:

$$a(t) \quad = \quad \begin{cases} a & if \quad g_{t-1} \neq 0 \\ 2a & if \quad g_{t-1} = 0 \end{cases}$$

and $b = 2a$. The thresholds are also changed if the model is in a position $g_t \neq 0$ and the volatility of the price has been low, in the following way:

$$a(t) \quad = \quad \begin{cases} a & if \quad |p_e - p_t| > v \\ 10a & if \quad |p_e - p_t| < v \end{cases}$$

where p_e is the logarithmic entry price of the last transaction and v is a threshold, generally quite low $< 0.5\%$. This means that the model is only allowed to change position if the price has significantly moved from the entry point of the deal.

Since X_{eff} and R_{eff} are implicit functions of the gearing, the optimization of the RTT model is based on the X_{eff} and R_{eff} performance. The parameters subject to optimization are: τ, a, d and v. There are two other auxiliary parameters: S the stop loss at which an open position is automatically closed and the profit objective P. These parameters are only optimized at the end once the others have been found and they are also not allowed to vary all the way since maximum stop-loss and maximum gain limits are set by the environment[16]. The model is subject to the open-close and holiday closing hours of the Zurich market.

[16] For more details on the optimization procedure the reader is referred to Pictet *et al.* (1992).

7. Empirical Results

The simulated data is the five minute ϑ-time series[17] from January, 1, 1990 to December 31, 1996 for the three major foreign exchange rates, USD-DEM, USD-CHF (Swiss Franc), USD-FRF (French Franc), and the cross-rate DEM-JPY (Deutsche Mark - Japanese Yen). The high frequency data inherits intraday seasonalities and requires deseasonalization. As explained in section 4, we use the Dacorogna et al. (1993) deseasonalization methodology which is based on the change of a time scale. Our data set contains 671,040 observations per currency. The simulations for each currency and process are done for 1000 replications.

The summary of the trading model performance with actual data is presented in Table 2. After the transaction costs, actual data with the USD-DEM, USD-CHF, USD-FRF and DEM-JPY yield an annualized total return of 9.63, 3.66, 8.20 and 6.43 percent, respectively. The USD-CHF has the weakest performance relative to the other three currencies. The X_{eff} and R_{eff} performance of the USD-DEM, USD-FRF and DEM-JPY are all positive and range between 3-4 percent. For the USD-CHF, the X_{eff} and R_{eff} are -1.68 and -4.23 percent reflecting the weakness of its performance.

7.1. Random Walk Process

The results for the random walk process are reported in Table 3 for the p-values[18] of all currency pairs. The methodology of this paper places a historical realization in the simulated distribution of the performance measure under the assumed process and calculates its one-sided p-value[19]. This indicates whether the historical realization is likely to be generated from this particular distribution or not. More importantly, it indicates whether the historical performance is likely to occur in the future. A small p-value (less than 5 percent) indicates that the historical performance lies in the tail of the distribution and the studied performance distribution is not representative of the data generating process assuming that the trading model is a good one. If the process which generates the performance distribution is close to the data generating process of the foreign exchange returns, the historical performance would lie within two standard deviations of the performance distribution, indicating that the studied process may be retained as representative of the data generating process.

The p-values of the annualized return for the USD-DEM, USD-CHF, USD-FRF and DEM-JPY are 0.3, 8.9, 1.2 and 2.1 percent, respectively. For the USD-DEM and USD-FRF, the

[17]The real-time system uses tick-by-tick data for its trading recommendations. The simulations in this paper are carried out with 5 minute data, as it is computationally expensive to use the tick-by-tick data for the simulations. The historical performance of the currency pairs from the 5 minute series are within a few tenths of a percent for all performance measures with the performance of the real-time trading models which utilize the tick-by-tick data. Therefore, there is no loss of generality from the usage of 5 minute frequency for the simulations instead of the tick-by-tick feed.

[18]The p-value represents the fraction of simulations generating a performance measure larger than the original.

[19]p-value calculations reported in this paper are the *simulated* p-values obtained from the distribution of one thousand replications of a given performance measure. For brevity, we simply refer to it as p-value in the text.

p-values are less than the 2 percent level and it is about 2 percent for the USD-CHF. In the case of the USD-CHF, the p-value for the annualized return is 8.9 which is well above the 5 percent level. As indicated in Section 2, the annualized return only utilizes two points of the equity curve leaving a large degrees of freedom to infinitely many paths that would be compatible with a given total return. X_{eff} and R_{eff} are more stringent performance measures which utilize the entire equity curve in their calculations. The p-values of X_{eff} and R_{eff} are 0.0; 0.0 percent for USD-DEM; 0.7 and 0.6 percent for USD-CHF; 0.2 and 0.1 percent for USD-FRF and 0.2 and 0.1 percent for DEM-JPY. The p-values for the X_{eff} and R_{eff} are all less than one percent rejecting the null hypothesis that the random walk process is consistent with the data generating process of exchange rate returns.

As presented in Table 2, the maximum drawdowns for the USD-DEM, USD-CHF, USD-FRF and DEM-JPY are 11.02, 16.08, 11.36 and 12.03 percent. The mean maximum drawdowns from the simulated random walk processes are 53.79, 63.68, 47.68 and 53.49 for the USD-DEM, USD-CHF, USD-FRF and DEM-JPY, respectively. The mean of the simulated maximum drawdowns are three or four times larger than the actual maximum drawdowns. The deal frequencies are 1.68, 1.29, 1.05, 2.14 per week for the four currency pairs from the actual data. The deal frequencies indicate that the RTT model trades on average no more than 2 trades per week although the data feed is at the 5 minute frequency. The mean simulated deal frequencies are 2.46, 1.98, 1.65 and 3.08 which are significantly larger than the actual ones.

The values for the maximum drawdown and the deal frequency indicate that the random walk simulation yields larger maximum drawdown and deal frequency values relative to the values of these statistics from the actual data. In other words, the random walk simulations deal more frequently and result in more volatile equity curves on average relative to the equity curve from the actual data. Correspondingly, the p-values indicate that the random walk process cannot be the representative of the actual foreign exchange series under these two performance measures[20].

The simulation results with the random walk process demonstrate that the real-time trading model is a consistent model. In other words, a process with no mean and a homoskedastic variance should only perform to generate an average return which would match the mean transaction costs. This consistency property is an essential ingredient of a trading model and the real-time trading model passes this consistency test. The means of the simulations indicate that the distributions are correctly centered at the average transaction costs which is expected under the random walk process. For instance, the mean simulated deal frequency of the USD-DEM is 2.46 deals per week or 127.92 (2.46×52) deals per year. The percentage spread for the USD-DEM is 0.00025 which in turn indicates an average transaction cost of -3.20 percent per year. Given that the mean of the simulated annualized return is -3.44, we can conclude that the mean of the simulated annualized return distribution is centered around the mean transaction

[20]Although it is not reported in the tables, the summary statistics of the simulated performance measures have negligible skewness and statistically insignificant excess kurtosis. This indicates that the distributions of the performance measures are symmetric and do not exhibit fat tails.

cost.

The behavior of the performance measures across 7 day, 29 day, 117 day and 301 day horizons is also investigated with X_{eff} and R_{eff}. The importance of the performance analysis at various horizons is that it permits a more detailed analysis of the equity curve at the predetermined points in time. These horizons correspond approximately to a week, a month, four months and a year's performance. The X_{eff} and R_{eff} values indicate that the RTT model performance improves over longer time horizons. This is in accordance with the low dealing frequency of the RTT model. In all horizons, the p-values for the X_{eff} and R_{eff} are less than a half percent for USD-DEM, USD-FRF and DEM-JPY. For USD-CHF, the p-values are less than 2.4 percent for all horizons. Overall, the multi-horizon analysis indicates that the random walk process is not consistent with the data generating process of the foreign exchange returns.

7.2. GARCH(1,1) Process

A more realistic process for the foreign exchange returns is the GARCH(1,1) process which allows for conditional heteroskedasticity. The GARCH(1,1) estimation results are presented in Table 4. The numbers in parentheses are the robust standard errors and the GARCH(1,1) parameters are statistically significant at the 5 percent level for all currency pairs. The Ljung-Box statistic is calculated up to 12 lags for the standardized residuals and it is distributed with χ^2 with 12 degrees of freedom. The Ljung-Box statistics indicate no serial correlation for the USD-DEM and USD-FRF but the USD-FRF and DEM-JPY remain serially correlated. The variances of the normalized residuals are near one. There is no evidence of skewness but the excess kurtosis remains large for the residuals.

In Table 6, the results with the GARCH(1,1) process are presented. Since GARCH(1,1) allows for conditional heteroskedasticity, it is expected that the simulated performance of the RTT model would yield higher p-values and retain the null hypothesis that GARCH(1,1) is consistent with the data generating process of the foreign exchange returns. The results, however, indicate smaller p-values which is in favor of a stronger rejection of this process relative to the random walk process.

One important reason for the rejection of the GARCH(1,1) process as a representative data generating process of foreign exchange returns is the aggregation property of the GARCH(1,1) process[21]. The GARCH(1,1) process behaves more like a homoskedastic process as the frequency is reduced from high to low frequency. Since the RTT model trading frequency is less than two deals per week, the trading model does not pick up the five minute level heteroskedastic structure at the weekly frequency. Rather, the heteroskedastic structure behaves as if it is measurement noise where the model takes positions and this leads to the stronger rejection of the GARCH(1,1) as a candidate for the foreign exchange data generating process.

[21] Guillaume et. al (1995) show that the use of an alternative time scale can eliminate the inefficiencies in the estimation of a GARCH model caused by intraday seasonal patterns. However, the temporal aggregation properties of the GARCH models do not hold at the intraday frequencies, revealing the presence of several time-horizon components.

In a GARCH process, the conditional heteroskedasticity is captured at the frequency that the data has been generated. As it is moved away from this frequency to lower frequencies, the heteroskedastic structure slowly dies away leaving itself to a more homogeneous structure in time. More elaborate processes, such as the multiple horizon ARCH models (as in the HARCH process of Müller et al. (1997)) possess conditionally heteroskedastic structure at all frequencies in general. The existence of a multiple frequency heteroskedastic structure can be more in line with the heterogeneous structure of the foreign exchange markets.

The p-values of the annualized return for the USD-DEM, USD-CHF, USD-FRF and DEM-JPY are 0.4, 8.4, 0.9 and 1.2 percent, respectively. All four currency pairs except USD-CHF yield p-values which are smaller than 1.3 percent. The X_{eff} and R_{eff} are 0.1 and 0.0 percent for USD-DEM; 1.4 and 0.9 percent for USD-CHF; 0.1 and 0.1 percent for USD-FRF and 0.4 and 0.4 percent for DEM-JPY.

The historical maximum drawdown and deal frequency of the RTT model is smaller than those generated from the simulated data. The maximum drawdowns for the USD-DEM, USD-CHF, USD-FRF and DEM-JPY are 11.02, 16.08, 11.36 and 12.03 for the four currencies. The mean simulated drawdowns are 53.33, 60.58, 46.00 and 48.77 for the four currencies. The mean simulated maximum drawdowns are three to four times larger than the historical ones. The historical deal frequencies are 1.68, 1.29, 1.05 and 2.14. The mean simulated deal frequencies are 2.39, 1.87, 1.59 and 2.66 for the four currencies. The differences between the historical deal frequencies and the mean simulated deal frequencies remain large. Therefore, the examination of the GARCH(1,1) process with the maximum drawdown and the deal frequency indicates that the historical realizations of these two measures stay outside of the 5 percent level of simulated distributions of these two performance measures.

The mean simulated deal frequency for the USD-DEM is 2.39 trade per week. In annual terms, this is approximately 124.28 deals per year. The half spread for the USD-DEM series is about 0.00025 and this yields 3.11 percent when multiplied with the number of deals per year. The -3.11 percent return would be the annual transaction cost of the model. For the model to be profitable, it should yield more than 3.11 percent per year. Table 2 indicates that the RTT model generates an excess annual return of 9.63 percent whereas the mean of the annualized return from the GARCH(1,1) process stays at the -3.27 percent level.

The multi-horizon examination of the equity curve with the X_{eff} and R_{eff} performance measures indicates that the GARCH(1,1) process as a candidate for the data generation mechanism is strongly rejected at all horizons from a seven day horizon to a horizon as long as 301 days. The overall picture coming out of the test is not very different for the GARCH(1,1) than that of the random walk process.

7.3. AR(4)-GARCH(1,1) Process

A further direction is to investigate whether conditional mean dynamics with GARCH(1,1) innovations would be a more successful characterization of the dynamics of the high frequency foreign exchange returns. The conditional means of the foreign exchange returns are estimated with four lags of these returns. The additional lags did not lead to substantial increases in the likelihood value. The results of the AR(4)-GARCH(1,1) are presented in Table 5. The numbers in parentheses are the robust standard errors and all four lags are statistically significant at the 5 percent level. The negative autocorrelation is large and highly significant for the first lag of the returns. This is consistent with the high frequency behavior of the foreign exchange returns and is also observed in Dacorogna et. al (1993). The Ljung-Box statistics indicate no serial correlation in the normalized residuals. The variances of the normalized residuals are near one. There is no evidence of skewness but the excess kurtosis remains large for the residuals.

The p-values of the annualized returns are presented in Table 7. They are 0.1, 3.7, 0.3 and 0.5 percent for the USD-DEM, USD-CHF, USD-FRF and DEM-JPY. The results indicate that the AR(4)-GARCH(1,1) process is also rejected under the RTT model as a representative data generating process of foreign exchange returns. Here again, a possible explanation of this failure is the relationship between the dealing frequency of the model and the frequency of the simulated data. The AR(4)-GARCH(1,1) process is generated at the 5 minutes frequency but the model dealing frequency is between one or two deals per week. Therefore, the model picks up the high frequency serial correlation as noise and this serial correlation works against the process. This cannot be treated as a failure of the RTT model. Rather, this strong rejection is evidence of the failure of the temporal aggregation properties of the AR(4)-GARCH(1,1) process at lower frequencies.

The rejection of the AR(4)-GARCH(1,1) process with the X_{eff} and R_{eff} are even stronger and very much in line with the results for the random walk and the GARCH(1,1). The p-values of the X_{eff} and R_{eff} are 0.1, 0.0 percent for USD-DEM; 1.9, 2.3 percent for USD-CHF; 0.2, 0.1 percent for USD-FRF and 0.1, 0.1 percent for DEM-JPY. The p-values remain low at all horizons for the X_{eff} and R_{eff}. The p-values of the maximum drawdown and the deal frequency also indicate that in almost all replications the AR(4)-GARCH(1,1) generates higher maximum drawdowns and deal frequencies.

7.4. AR(4)-HARCH(9) Process

The HARCH findings with the USD-DEM[22] are presented in Tables 8 and 9 which indicate the rejection of this model as well. The HARCH model is designed to capture the asymmetry in volatility dynamics measured at different frequencies and the long memory of the data generating process. The comparison of the HARCH model with the GARCH model as in Müller et al. (1997) indicates that the HARCH process can extract the tail information more successfully than the GARCH model. One implication of this is that the simulated series from the HARCH process will have more extreme observations due to the well-approximated tail dynamics. Combined with the long memory, it results in large sudden jumps in the simulation followed by series of large movements.

[22] AR(4)-HARCH(9) process yields the largest likelihood value among other AR-HARCH parametrizations that we studied. For space limitations, the results with the other currencies are not presented here. These results are similar to the USD-DEM findings.

In the presence of statistically significant four conditional mean parameters, the HARCH model is a successful characterization of the data dynamics at the 5-minute frequency. However, the important message that we are trying to convey here is that the 5-minute dynamics is one particular frequency where we observe the data generating process. The underlying dynamics is different at different frequencies (see, for instance Müller et al. (1997) and Andersen and Bollerslev (1997)), and the 5-minute data frequency is a particular slice of this layered interconnected dynamics. Although the HARCH model is designed to capture the volatility dynamics at all frequencies, the *mean dynamics* is only present at the highest frequency.

Since the trading frequency of the models is typically 2-3 times a week, small movements of prices in the 5-minute data frequency are not significant from the trading model perspective. Therefore, the conditional mean dynamics of the AR-HARCH model does not help the trading model to capture possible trends because the movements in the prices are too short lived for the model to take a position. On the other hand, the simulated series are more volatile due to the fully absorbed tail dynamics and the long memory effect. In the absence of weekly conditional mean information (this information is left to the residuals and is unconditional now), the trading model cannot interpret these large moves and ends up taking consecutive wrong positions. Due to this reason the skewness and the kurtosis of the trading model statistics, such as X_{eff} and R_{eff}, are larger than with the other processes.

8. Conclusions

We analysed real-time technical trading model behavior in the foreign exchange market by means of high frequency data and theoretical process simulations. This extensive analysis of real-time trading models with high frequency data suggests two main conclusions. First, technical trading models can generate excess returns which are explained neither by traditional theoretical processes nor by luck. It should be noted that this study was conducted over 7 years of five minutes data, which represents more than 500,000 independently sampled observations. This represents an unusually long out-of-sample period that guarantees the robustness of our conclusions. Since the underlying data generating process for returns is not known, we have used the largest ex-ante intraday data available to avoid any sample bias. Furthermore, the trading strategies were also fixed ex-ante to minimize any model selection bias. Under such a setting, the results indicate that the trading strategies have successfully generated net positive returns. Whether these results continue to hold for future years needs to be studied separately in another study, although our other work in progress indicates that our findings in this paper continue to hold.

Second, the foreign exchange rates contain conditional mean dynamics that are not present in the random walk, GARCH(1,1), AR-GARCH(1,1) and AR(4)-HARCH(9) processes. Moreover, the GARCH and HARCH type models are not able to capture the aggregation properties of the data which are essential in this case since we are using 5 minute price changes for trading models which trade on average once a week. Therefore, it is not sufficient to develop sophisticated statistical processes and study them with data measured at an arbitrary frequency (e.g. one day, one week, one month, annual etc.). In financial markets, the data generating process is a complex network of layers where each layer corresponds to a particular frequency (Müller et al. (1997)). A successful characterization of such data generating processes should be estimated with models whose parameters are functions of *intra* and *inter* frequency dynamics.

References

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys, 1999, The distribution of exchange rate volatility, Stern School of Business, Working Paper FIN-99-059.

Andersen, T. G. and T. Bollerslev, 1997, Intraday periodicity and volatility persistence in financial markets, *Journal of Empirical Finance*, 4, 115-158.

Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* 31, 307-327.

Brock, W. A., J. Lakonishok and B. LeBaron, 1992, Simple technical trading rules and the stochastic properties of stock returns, *Journal of Finance* 47, 1731-1764.

Brock, W. A., 1999, Scaling in economics, A reader's guide, *Industrial and Corporate Change*, 8, 409-446.

Dacorogna, M. M., R. Gençay, U. A. Müller, O. V. Pictet and R. B. Olsen, 2000, An Introduction to High Frequency Finance, *Academic Press*, forthcoming.

Dacorogna, M. M., U. A. Müller, R. J. Nagler, R. B. Olsen, and O. V. Pictet, 1993, A geographical model for the daily and weekly seasonal volatility in the foreign exchange market, *Journal of International Money and Finance*, 12, 413-438.

Dacorogna, M. M., U. A. Müller, C. Jost, O. V. Pictet, R. B. Olsen and J. R. Ward, 1995, Heterogeneous real-time trading strategies in the foreign exchange market, *The European Journal of Finance*, 1, 383-404.

Dacorogna, M. M., R. Gençay, U. A. Müller, O. V. Pictet, 1999, Effective return, risk aversion and drawdowns, Olsen Research Institute Discussion Paper, Zurich, Switzerland.

Gençay, R., 1998, The predictability of security returns with simple technical trading rules, *Journal of Empirical Finance*, 5, 347-359.

Gençay, R., 1999, Linear, nonlinear and essential foreign exchange prediction with simple technical rules, *Journal of International Economics*, 47, 91-107.

Guillaume D. M., Dacorogna M. M., and Pictet O. V., 1994, On the intra-daily performance of GARCH processes, *Paper presented at the First International Conference on High Frequency Data in Finance, Zürich.*

Hong, H. and J. Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *Journal of Finance*, 54, 2143-2184.

Keeney, R. L. and H. Raiffa, 1976, Decision with multiple objectives: Preferences and value tradeoffs, John Wiley & Sons, New York.

LeBaron, B., 1992, Do moving average trading rule results imply nonlinearities in foreign exchange markets?, Social Systems Research Institute, University of Wisconsin-Madison.

LeBaron, B., 1997, Technical trading rules and regime shifts in foreign exchange, forthcoming in *Advanced Trading Rules*, edited by E. Acar and S. Satchell, Butterworth-Heinemann.

LeBaron, B., 1999, Volatility persistence and apparent scaling laws in finance, Graduate School of International Economics and Finance, Brandeis University.

Levich, R. M. and L. R. Thomas, 1993, The significance of technical trading-rule profits in the foreign exchange market: A bootstrap approach, *Journal of International Money and Finance*, 12, 451-474.

Müller, U. A., 1989, Indicators for trading systems, Olsen Research Institute Discussion Paper, Zurich, Switzerland.

Müller, U. A., M. M. Dacorogna, R. B. Olsen, O. V. Pictet, M. Schwarz and C. Morgenegg, 1990, *Journal of Banking and Finance*, 14, 1189-1208.

Müller, U. A., 1991, Specially weighted moving averages with repeated application of the EMA operator, Olsen Research Institute Discussion Paper, Zurich, Switzerland.

Müller, U. A., M. M. Dacorogna and O. V. Pictet, 1993, A trading model performance measure with strong risk aversion against drawdowns, Olsen Research Institute Discussion Paper, Zurich, Switzerland.

Müller, U. A., M. M. Dacorogna, R. D. Davé, Richard B. Olsen, O. V. Pictet and J E. von Weizsaӧcker, 1997, Volatilities of different time resolutions - Analyzing the dynamics of market components, *Journal of Empirical Finance*, 4, 213-240.

Pictet, O. V., M. M. Dacorogna, U. A. Müller, R. B. Olsen and J. R. Ward, 1992, Real-time trading models for foreign exchange rates, *Neural Network World*, 6, 713-744.

Sharpe, W. F., 1966, Mutual fund performance, *Journal of Business*, 39, 119-138.

Sharpe, W. F., 1994, The Sharpe ratio, *Journal of Portfolio Management*, 21, 49-59.

Sullivan, R., A. Timmermann and H. White, 1999, Data-snooping, technical trading rules and the bootstrap, *Journal of Finance*, 54, 1647-1692.

Zumbach, G. and U. A. Müller, 2000, Operators on inhomogeneous time series, accepted for publication in *International Journal of Theoretical and Applied Finance*.

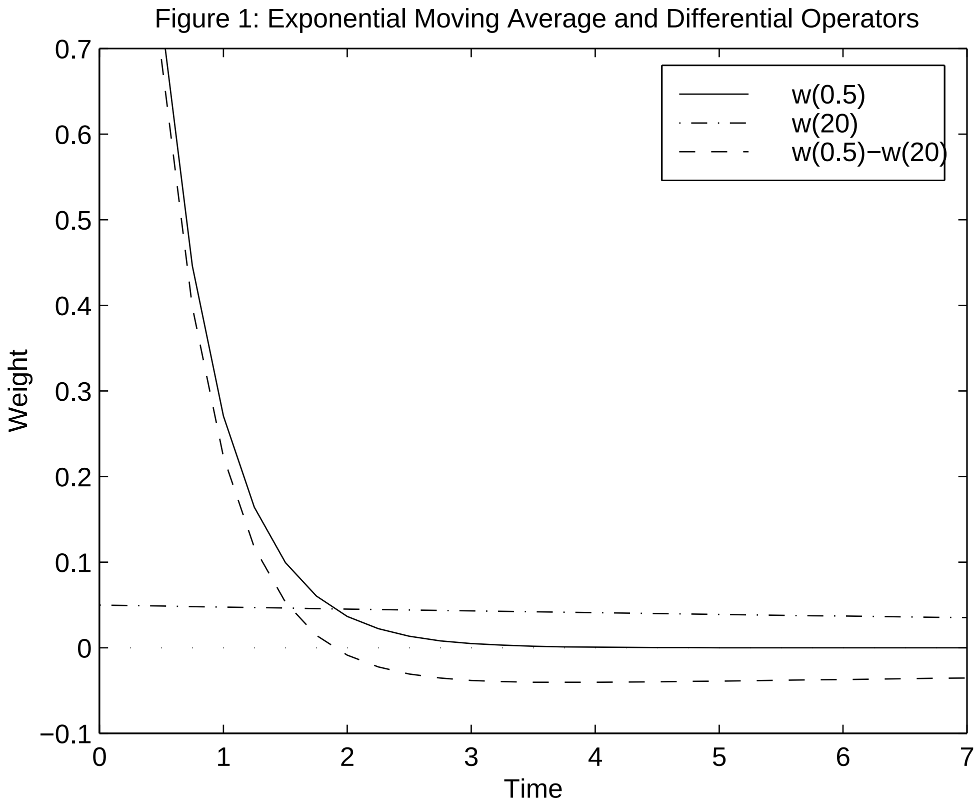

Figure 1: Exponential moving average (EMA) operator is a simple average operator with

$$w_{ema}(t;\tau) = \frac{e^{-t/\tau}}{\tau}\ ,$$

an exponential decaying kernel. τ determines the range of the operator and t indexes the time. An EMA is written as

$$EMA_p(\tau, t) = \int_{-\infty}^{t} w_{ema}(t - t')p(t')dt'$$

where $w_{ema}(t - t'; \tau) = \frac{e^{-(t-t')/\tau}}{\tau}$.

The figure above demonstrates the kernel function of an exponential moving average with $\tau = 0.5$ and $\tau = 20$ and their differential kernel. The sequential computation of exponential moving averages is simple with the help of a recursion formula and it is more efficient than the computation of any differently weighted moving averages.

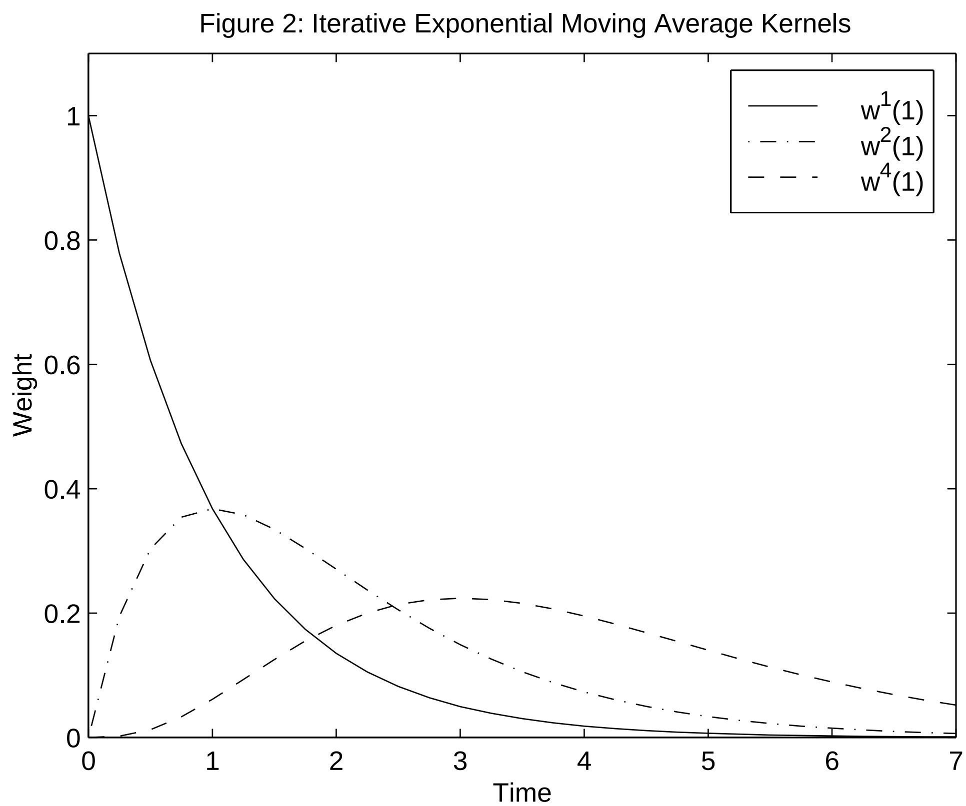

Figure 2: An n iterative EMA, EMA^n is written as

$$EMA_p^n(\tau, t) = \frac{1}{(n-1)!}\int_{-\infty}^{t}\frac{(t-t')^{n-1}}{\tau^{n-1}}w_{ema}(t-t')p(t')dt'$$

where $w_{ema}(t-t';\tau) = \frac{e^{-(t-t')/\tau}}{\tau}$.

In the figure above, the iterative moving averages for $n = 1, 2$ and $n = 4$ are plotted which indicate that as n gets larger the center of the weight distribution moves to the middle part of the kernel function.

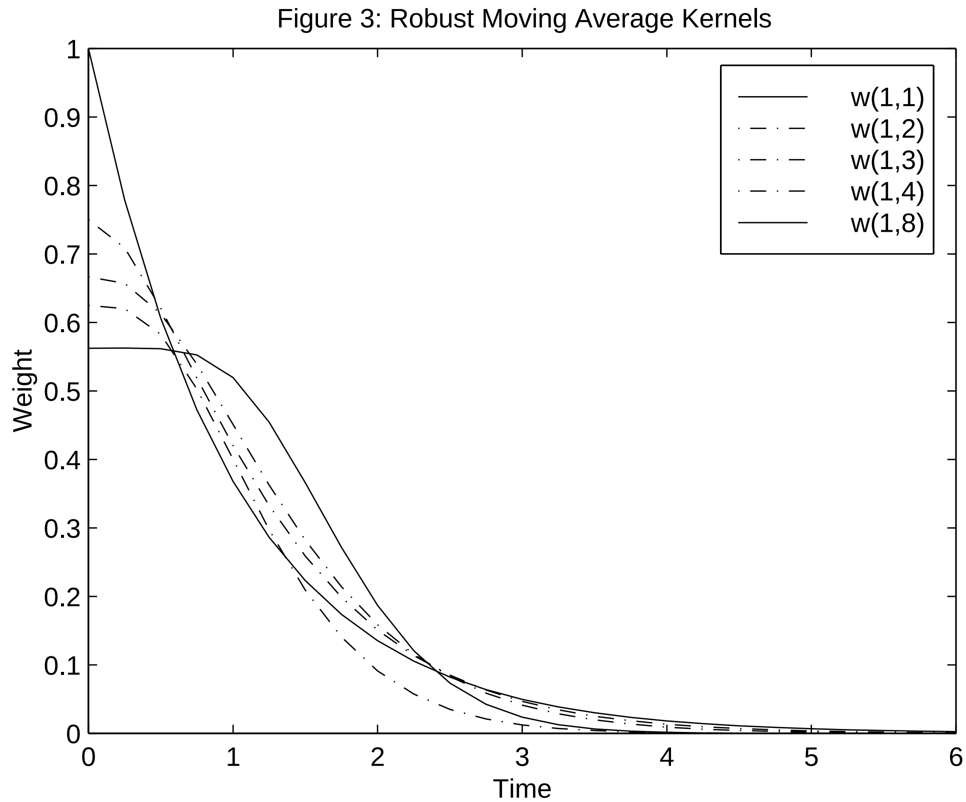

Figure 3: The robust exponential moving average is written as

$$MA_p^n(\tau, t) = \int_{-\infty}^{t} w_{ma}(t - t', \tau, n) p(t') dt'.$$

This is a special case where all weights assigned to each iterative kernel are the same as in equation (7). In the figure above, the examples of these robust kernels are plotted with $n = 1, 2, 3, 4$ and $n = 8$. The property of this kernel is that its function has a plateau before it asymptotically declines to zero and it is robust to the extreme variations leaving the average by assigning exponentially decaying weights. It also has the property that it assigns relatively uniform weights to the most recent history whereas a simple exponential average would be very sensitive with such new information. Therefore, a robust kernel has the property that it preserves only the desirable robustness properties of the simple average and exponential average kernels but ignores their highly noisy unrobust properties.

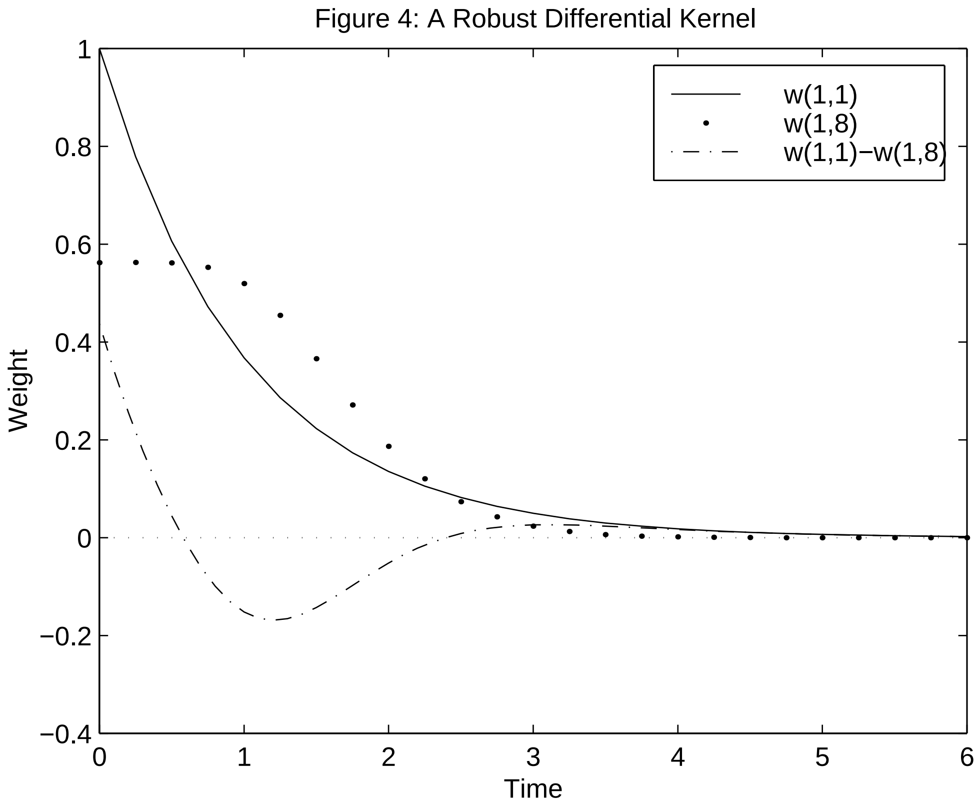

Figure 4: A robust differential kernel is presented which is based on the difference between the exponential moving average with $\tau = 1$ and a robust kernel with $w_{ma}(\tau = 1, n = 8)$. By construction, the area under the kernel sums to zero. The differential kernel assigns positive weights to the recent past and negative weights to the distant past. The real-time trading model of this paper uses a similar robust differential kernel in the construction of the gearing function.

Table 2
Historical Realizations

Description	Historical Realizations			
	USD-DEM	USD-CHF	USD-FRF	DEM-JPY
Annual Return	9.63	3.66	8.20	6.43
Xeffective	3.78	-1.68	4.80	3.81
Reffective	4.43	-4.23	4.95	3.45
Max Drawdown	11.02	16.08	11.36	12.03
Deal frequency	1.68	1.29	1.05	2.14
Horizon: 7 days				
Xeffective	3.47	-2.96	3.18	1.87
Reffective	1.80	-4.81	1.97	0.58
Horizon: 29 days				
Xeffective	3.27	-4.10	4.41	2.12
Reffective	2.16	-8.97	3.99	0.19
Horizon: 117 days				
Xeffective	4.07	-0.67	3.83	4.90
Reffective	5.10	-1.77	3.77	5.17
Horizon: 301 days				
Xeffective	4.62	1.94	6.35	5.39
Reffective	6.83	1.71	7.38	5.56

Notes: *Historical Realizations* present the performance of the trading model with the actual series from January 1, 1990 until December 31, 1996 with 5 minute frequency.

Table 3

Random Walk Simulations

1990-1996, 5 minute frequency

Description	p-value (in %)			
	USD-DEM	USD-CHF	USD-FRF	DEM-JPY
Annual Return	0.3	8.9	1.2	2.1
Xeffective	0.0	0.7	0.2	0.2
Reffective	0.0	0.6	0.1	0.1
Max Drawdown	100.0	100.0	100.0	100.0
Deal frequency	100.0	100.0	100.0	100.0
Horizon: 7 days				
Xeffective	0.0	1.2	0.2	0.3
Reffective	0.0	0.4	0.2	0.2
Horizon: 29 days				
Xeffective	0.0	2.2	0.2	0.3
Reffective	0.0	2.3	0.1	0.3
Horizon: 117 days				
Xeffective	0.0	0.9	0.2	0.2
Reffective	0.0	0.9	0.1	0.2
Horizon: 301 days				
Xeffective	0.0	0.5	0.2	0.1
Reffective	0.0	0.7	0.2	0.3

Notes: The results under columns p-value present the values of these statistics from 1000 replications with the random walk process. The random walk estimation involves the regression of the actual USD-DEM returns on a constant. A simulation sample for the random walk series with drift is obtained by sampling from the Gaussian random number generator with the mean and the standard deviation of the residual series. The simulated residuals are added to the conditional mean defined by $\hat{\alpha}$, to form a new series of returns. The new series of the returns has the same drift in prices, the same variance and the same unconditional distribution. From the new series of returns, the simulated price process is recovered recursively by setting the initial price to the true price at the beginning of the sample. The trading models use the bid and ask prices as inputs. Half of the average spread is subtracted (added) from the simulated price process to obtain the simulated bid and ask prices. The p-values are reported in percentage terms (e.g. 8.4 refers to 8.4 percent).

Table 4

GARCH(1,1) Parameter Estimates, 1990-1996, 5 minute frequency

	USD-DEM	USD-CHF	USD-FRF	DEM-JPY
α_0	4.95 (4.23)	0.11 (0.12)	9.38 (7.09)	2.97 (4.03)
α_1	0.1111 (0.0005)	0.1032 (0.0007)	0.1572 (0.0007)	0.0910 (0.0005)
β_1	0.8622 (0.0007)	0.8578 (0.0009)	0.8137 (0.0009)	0.8988 (0.0006)
LL	6.45	6.17	6.29	6.34
$Q(12)$	4810	4201	4256	3089
$\hat{\epsilon}_{\sigma^2}$	1.04	1.03	1.07	1.05
$\hat{\epsilon}_{sk}$	-0.07	-0.03	-0.05	0.16
$\hat{\epsilon}_{ku}$	11.73	7.28	22.93	27.73

Table 5

AR(4)-GARCH(1,1) Parameter Estimates, 1990-1996, 5 minute frequency

	USD-DEM	USD-CHF	USD-FRF	DEM-JPY
α_0	3.90 (3.40)	8.19 (9.03)	7.28 (5.80)	2.92 (3.93)
α_1	0.099 (0.0005)	0.0874 (0.0006)	0.1349 (0.0007)	0.088 (0.0005)
β_1	0.8796 (0.0006)	0.8833 (0.0007)	0.8411 (0.0008)	0.9008 (0.0006)
γ_1	-0.176 (0.001)	-0.208 (0.001)	-0.200 (0.002)	-0.130 (0.002)
γ_2	-0.011 (0.001)	-0.031 (0.002)	-0.025 (0.002)	-0.090 (0.002)
γ_3	0.003 (0.001)	-0.001 (0.002)	-0.005 (0.002)	-0.005 (0.002)
γ_4	-0.004(0.001)	-0.002 (0.001)	-0.008 (0.002)	-0.010 (0.002)
LL	6.46	6.19	6.30	6.35
$Q(12)$	623	531	492	374
$\hat{\epsilon}_{\sigma^2}$	1.04	1.03	1.07	1.05
$\hat{\epsilon}_{sk}$	-0.07	-0.04	-0.05	0.15
$\hat{\epsilon}_{ku}$	12.29	7.86	21.84	27.98

Notes: LL is the average log likelihood value. $Q(12)$ refer to the Ljung-Box portmanteau test for serial correlation and it is distributed χ^2 with 12 degrees of freedom. The $\chi^2_{0.05}(12)$ is 21.03. $\hat{\epsilon}_{\sigma^2}$, $\hat{\epsilon}_{sk}$ and $\hat{\epsilon}_{ku}$ are the variance, skewness and the excess kurtosis of the residuals.

Table 6
GARCH(1,1) Simulations
1990-1996, 5 minute frequency

Description	p-value (in %)			
	USD-DEM	USD-CHF	USD-FRF	DEM-JPY
Annual Return	0.4	8.4	0.9	1.2
Xeffective	0.1	1.4	0.1	0.4
Reffective	0.0	0.9	0.1	0.4
Max Drawdown	100.0	100.0	99.9	100.0
Deal frequency	100.0	100.0	100.0	100.0
Horizon: 7 days				
Xeffective	0.2	1.6	0.3	0.6
Reffective	0.0	1.0	0.1	0.6
Horizon: 29 days				
Xeffective	0.2	2.6	0.1	0.5
Reffective	0.1	2.8	0.0	0.7
Horizon: 117 days				
Xeffective	0.1	0.8	0.4	0.2
Reffective	0.1	0.7	0.2	0.2
Horizon: 301 days				
Xeffective	0.2	0.9	0.1	0.4
Reffective	0.3	1.5	0.2	0.5

Notes: The results under columns p-value present the values of these statistics from 1000 replications with the GARCH(1,1) process. The parameters and the normalized residuals are estimated from the foreign exchange returns using the maximum likelihood procedure. The simulated returns for the GARCH(1,1) process are generated from the simulated residuals and the estimated parameters. From the new series of returns, the simulated price process is recovered recursively by setting the initial price to the true price at the beginning of the sample. The trading models use the bid and ask prices as inputs. Half of the average spread is subtracted (added) from the simulated price process to obtain the simulated bid and ask prices. The p-values are reported in percentage terms (e.g. 8.4 refers to 8.4 percent).

Table 7

AR(4)-GARCH(1,1) Simulations

USD-DEM, 1990-1996, 5 minute frequency

Description	p-value (in %)			
	USD-DEM	USD-CHF	USD-FRF	DEM-JPY
Annual Return	0.1	3.7	0.3	0.5
Xeffective	0.1	1.9	0.2	0.1
Reffective	0.0	2.3	0.1	0.1
Max Drawdown	100.0	99.7	99.9	100.0
Deal frequency	100.0	99.9	100.0	100.0
Horizon: 7 days				
Xeffective	0.1	2.8	0.2	0.5
Reffective	0.2	2.4	0.2	0.5
Horizon: 29 days				
Xeffective	0.1	4.6	0.0	0.4
Reffective	0.1	9.7	0.0	0.4
Horizon: 117 days				
Xeffective	0.1	1.3	0.3	0.1
Reffective	0.0	1.4	0.3	0.1
Horizon: 301 days				
Xeffective	0.1	0.8	0.2	0.0
Reffective	0.1	1.1	0.2	0.2

Notes: The results under columns p-value present the values of these statistics from 1000 replications with the AR(4)-GARCH(1,1) process. The parameters and the normalized residuals are estimated from the foreign exchange returns using the maximum likelihood procedure. The simulated returns for the AR(4)-GARCH(1,1) process are generated from the simulated residuals and the estimated parameters. From the new series of returns, the simulated price process is recovered recursively by setting the initial price to the true price at the beginning of the sample. The trading models use the bid and ask prices as inputs. Half of the average spread is subtracted (added) from the simulated price process to obtain the simulated bid and ask prices. The p-values are reported in percentage terms (e.g. 8.4 refers to 8.4 percent).

Table 8

AR(4)-HARCH(9) Parameter Estimates, 1990-1996, 5 minute frequency

	USD-DEM
I_0	2.252 (5.241)
I_1	0.1157 (0.0085)
I_2	0.1670 (0.0147)
I_3	0.1925 (0.0108)
I_4	0.1056 (0.0071)
I_5	0.0469 (0.0045)
I_6	0.0767 (0.0060)
I_7	0.1165 (0.0088)
I_8	0.0921 (0.0067)
I_9	0.0421 (0.0042)
γ_1	-0.1957 (0.0093)
γ_2	-0.0373 (0.0057)
γ_3	-0.0064 (0.0016)
γ_4	-0.0087 (0.0023)
LL	6.48
$Q(12)$	726
$\hat{\epsilon}_{\sigma^2}$	1.04
$\hat{\epsilon}_{sk}$	-0.022
$\hat{\epsilon}_{ku}$	12.56

Notes: $I_j = k_j c_j$ which is the impact of each time horizon. I_0 value is 10^{-8}. The numbers in parentheses are the standard errors. The standard error of I_0 is 10^{-9}. LL is the average log likelihood value. $Q(12)$ refer to the Ljung-Box portmanteau test for serial correlation and it is distributed χ^2 with 12 degrees of freedom. The $\chi^2_{0.05}(12)$ is 21.03. $\hat{\epsilon}_{\sigma^2}$, $\hat{\epsilon}_{sk}$ and $\hat{\epsilon}_{ku}$ are the variance, skewness and the excess kurtosis of the residuals.

Table 9
AR(4)-HARCH(9) Simulations
DEM-JPY, 1990-1996, 5 minute frequency

Description	Historical Realization	p-value (in%)
Annual Return	6.43	0.2
Xeffective	3.81	0.2
Reffective	3.45	0.1
Max Drawdown	12.03	100.0
Deal frequency	2.14	88.0
Horizon: 7 days		
Xeffective	1.87	0.2
Reffective	0.58	0.0
Horizon: 29 days		
Xeffective	2.12	0.2
Reffective	0.19	0.0
Horizon: 117 days		
Xeffective	4.90	0.1
Reffective	5.17	0.1
Horizon: 301 days		
Xeffective	5.39	0.1
Reffective	5.56	0.1

Notes: The column under *Historical Realization* presents the performance of the trading model with the actual DEM-JPY series from January 1, 1990 until December 31, 1996 with 5 minute frequency. The results under columns p-value present the values of these statistics from 1000 replications with the AR(4)-HARCH(9) process. The parameters and the normalized residuals are estimated from the foreign exchange returns using the maximum likelihood procedure. The simulated returns for the AR(4)-HARCH(9) process are generated from the simulated residuals and the estimated parameters. From the new series of returns, the simulated price process is recovered recursively by setting the initial price to the true price at the beginning of the sample. The trading models use the bid and ask prices as inputs. Half of the average spread is subtracted (added) from the simulated price process to obtain the simulated bid and ask prices. The p-values are reported in percentage terms (e.g. 8.4 refers to 8.4 percent).