Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER Journal of Economic Dynamics & Control 27 (2003) 909–935
JOURNAL OF Economic Dynamics & Control
www.elsevier.com/locate/econbase

Foreign exchange trading models and market behavior

Ramazan Gençay[a,*], Michel Dacorogna[b], Richard Olsen[c], Olivier Pictet[d]

[a] *Department of Economics, University of Windsor, 401 Sunset Avenue, Windsor, Ont., Canada N9B 3P4*
[b] *Converium, Switzerland*
[c] *Olsen & Associates, Switzerland*
[d] *Dynamic Asset Management, Switzerland*

Abstract

The contributions of this paper are twofold. First, the performance of a widely used commercial real-time trading model is compared with a simple exponential moving average model. Second, the trading models are used as diagnostic tools to evaluate the statistical properties of foreign exchange rates.

The results presented in this paper have a general message to the standard paradigm in econometrics. It is not sufficient to develop sophisticated statistical processes and choose an arbitrary data frequency (e.g. one week, one month, annual, etc.) claiming afterwards that this particular process does a "good job" of capturing the dynamics of the data generating process. In financial markets, the data generating process is a complex network of layers where each layer corresponds to a particular frequency. A successful characterization of such data generating processes should be estimated with models whose parameters are functions of *intra* and *inter* frequency dynamics. In other fields, such as in signal processing, paradigms of this sort are already in place. Our understanding of financial markets would be increased with the incorporation of such paradigms into financial econometrics. Our trading models, within this perspective, help to observe this subtle structure as a diagnostic tool. © 2002 Elsevier Science B.V. All rights reserved.

JEL classification: G14; C45; C52; C53

Keywords: Real-time trading models; Exponential moving averages; Robust kernels; Technical trading

* Corresponding author. Tel.: +519-253-3000 ext. 2382; fax: +519-973-7096.
E-mail address: gencay@uwindsor.ca (R. Gençay).

0165-1889/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0165-1889(02)00049-0

1. Introduction

Recently, the skepticism among academics towards the possibility of developing profitable trading models has decreased with the publication of many papers that document profitable trading strategies in financial markets even when including transaction costs. 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 III (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. Dacorogna et al. (1995) 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. 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.

The contributions of this paper are twofold. First, the performance of a widely used commercial real-time trading model is compared with a simple exponential moving average trading model. Second, the trading models are used as diagnostic tools 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 from January 1, 1990 to December 31, 1996. From its launch in 1989 until the end of 1996, the trading models had not been re-optimized and was running on the original set of parameters estimated with data prior to 1989. 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

[1] In the earlier literature, Meese and Rogoff (1983) show that a simple random walk model forecasts as well as most linear exchange rate models. Using nonparametric kernel regression, Diebold and Nason (1990) were not able to improve upon a simple random walk in the out-of-sample prediction of ten major dollar spot rates in the post-1973 float period. Gençay (1999) shows that past buy-sell signals of simple moving average rules provide statistically significant sign predictions for modeling 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 modeling the conditional mean dynamics in the nonparametric models.

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

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[3] to control for data snooping effects.

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.

The results presented in this paper have a general message to the standard paradigm in econometrics. It is not sufficient to develop sophisticated statistical processes and choose an arbitrary data frequency (e.g. one week, one month, annual, etc.) claiming afterwards that this particular process does a "good job" of capturing the dynamics of the data generating process. In financial markets, the data generating process is a complex network of layers where each layer corresponds to a particular frequency. A successful characterization of such data generating processes should be estimated with models whose parameters are functions of *intra* and *inter* frequency dynamics. In other fields, such as in signal processing, paradigms of this sort are already in place. Our understanding of financial markets would be increased with the incorporation of such paradigms into financial econometrics. Our trading model, within this perspective, helps to observe this subtle structure as a diagnostic tool.

This paper is organized as follows. In Section 2, the trading models are explained. In Section 3, performance measures are presented. The simulation models and the simulation methodology are presented in Section 4. We discuss the empirical results in Section 5. We conclude afterwards.

2. 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

[3] In other words, there is no socially determined co-evolutionary relationship between our data set and the technical strategies used in implementing our specification tests.

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.

In general, 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 for utilizing 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, the high frequency data in these models lets us learn the underlying heterogeneous market microstructure properties of the foreign exchange markets.

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 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.

There are two trading models that we study here. The first one is a highly sophisticated O&A real-time trading model (RTT). The second model is based on an single exponential moving average (EMA) indicator. The specifications of these two models are summarized below.

2.1. 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.[4] 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. The RTT model goes neutral only to save profits or when a stop-loss is reached. Its profit objective is typically at 3%. 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[5] for the RTT is

$$g_t(I_p) = \text{sign}(I_p(t)) f(|I_p(t)|) c(I(t)),$$

where

$$I_p(t) = p_t - EMA(\tau = 20),$$

where p_t is the logarithmic price at time t, $EMA(\tau = 20)$ refers to 20-day equally weighted iterative exponential moving average and

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

and

$$c_t(I) = \begin{cases} +1 & \text{if } |I_p(t)| < d, \\ -1 & \text{if } |I_p(t)| > d \text{ and } g_I \cdot \text{sign}(I_p(t)) > 0 \text{ and } r_1 > P, \end{cases}$$

[4] The Parker Systematic Index 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 USD–DEM, USD–CHF, USD–FRF and DEM–JPY is constructed. This portfolio is leveraged so that it reaches the same standard deviation as the Parker Systematic Index (in other words, the borrowing amount is such that the two portfolio has similar risk characteristics) over the sample period. 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% management fee on assets is deducted monthly and a performance fee of 20% on net profits is deducted yearly to simulate actual fund environment. The results indicate 12.65% and 10.76% for the RTT and Parker Systematic Index performances, respectively. Therefore, the RTT model exhibits a realistic performance relative to the performance of the systematic currency traders.

[5] A gearing calculator provides the trading model with its intelligence and the ability to capitalize on movements in the markets by controlling the frequency of dealing and the circumstances under which positions may be entered. The possible exposures (gearings) are $\pm\frac{1}{2}$, ± 1 (full exposure) or 0 (no exposure).

where $a < b < d$ and r_1 is the return of the last deal, P is the profit objective and g_I is the gearing of the previous position. 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 60–70 times. The hit rate of the contrarian strategy is of about 75%.

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

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

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:

$$a(t) = \begin{cases} a & \text{if } |p_{\mathrm{e}} - p_t| > v, \\ 10a & \text{if } |p_{\mathrm{e}} - p_t| < v, \end{cases}$$

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.

The overall structure and data flow of a simple real-time trading model is depicted in Fig. 1. An extensive description of this model can be found in Dacorogna et al. (2001).

2.2. A simple exponential average (EMA) model

The EMA model indicator is a momentum based indicator consisting of a difference between two exponential moving averages of range $\tau = 0.5$ and 20 days. The gearing function for the EMA model is

$$g(I_p) = \mathrm{sign}(I_p) f(|I_p|),$$

where

$$I_p = EMA(\tau = 0.5) - EMA(\tau = 20)$$

and

$$f(|I_p|) = \begin{cases} \text{if } |I_p| > a & 1, \\ \text{if } |I_p| < a & 0, \end{cases}$$

where $a > 0$. The model is subject to the open–close and holiday closing hours. The model has maximum stop-loss and profit objectives which are set to the same values as in the RTT model.

3. 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

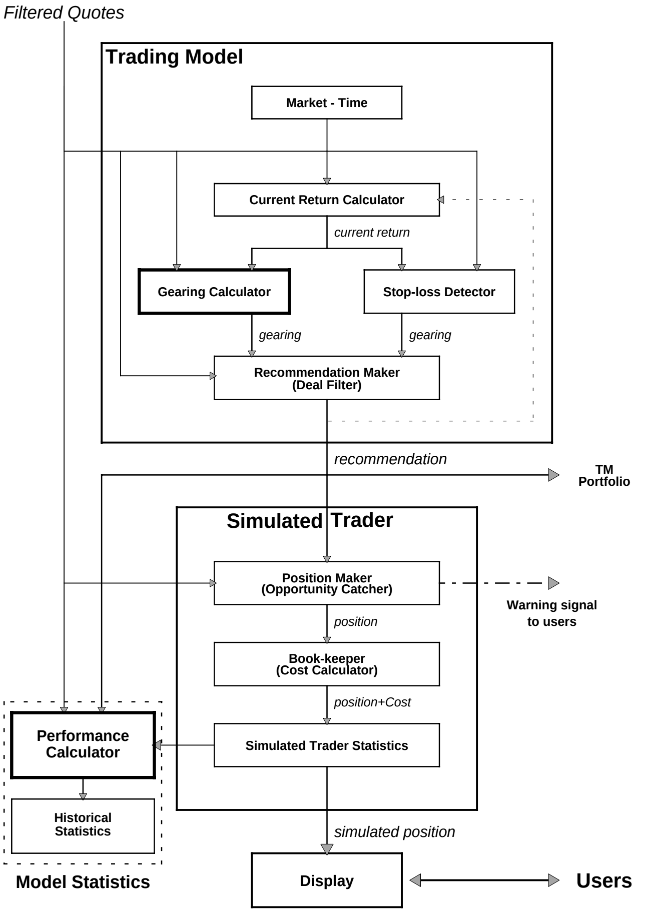

Fig. 1. Data flow of prices and deal recommendations within a real-time trading model.

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 model over a 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 and j is the jth transaction and r_j is the return from the jth transaction. The total return expresses the amount of profit made by a trader always investing up to his initial capital or credit

[6] The performance measures of this paper are also used in Pictet et al. (1992) and Dacorogna et al. (2000).

limit in his home currency.[7] The annualized return, $\bar{R}_{T,\mathrm{A}}$, 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.[8]

The *maximum drawdown*, D_T, over a certain period $T = t_\mathrm{E} - t_0$, is defined by

$$D_T \equiv \max(R_{t_a} - R_{t_b} | t_0 \leqslant t_a \leqslant t_b \leqslant t_\mathrm{E}), \tag{2}$$

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.

The trading model performance needs to account for a high total return; a smooth, almost linear increase of the total return over time; a small clustering of losses and no bias towards low frequency trading models. 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 become later 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, a trading model return variable $\tilde{R}$ is defined to be the sum of the total return R_T and the unrealized current return r_c.[9] The variable $\tilde{R}$ reflects the additional risk due to unrealized returns. Its change over a time interval Δt is

$$X_{\Delta t} = \tilde{R}_t - \tilde{R}_{t-\Delta t}, \tag{3}$$

where t expresses the time of the measurement. In this paper, Δt is allowed to vary from seven days to 301 days.

[7] The example below demonstrates the accounting of a trading model of USD–CHF, where CHF is the *home* (*numeraire*) currency and USD is the foreign *exchanged* currency. The trading model is played with a limit of 100 CHF. The usual practice for the capital flow in FX trading is that it is started from a capital of zero with credit limit. This is what is assumed here. All of our return calculations are expressed in terms of the home currency. In other words, the returns are calculated in terms of DEM for USD–DEM, CHF for USD–CHF, FRF for USD–FRF and JPY for DEM–JPY.

Time	Gearing	Current pos. in CHF	Current pos. in USD	FX rate
0	0	0	0	Do not care
1	0.5	− 50	35.71	1.4
2	1	− 100	69.04 (33.33 more)	1.5
3	0	10.46 (69.04*1.6 more)	0	1.6

This means that the trading lots in CHF are always 50 (half gearing step) or 100 (full gearing step) when increasing the (long or short) position, whereas decreasing the position means selling the full current USD amount (when going to neutral) or half the current USD amount (when going from gearing 1 to 1/2).

[8] If it is the annualization of one particular return (for one trade going from Neutral to Neutral), one simply needs to multiply the return by the ratio one year in days/(time interval from Neutral to Neutral). Usually, we only annualize the total return which means that this is the return achieved after having done all the trades during a whole year. There it is simply the sum of all trade returns not annualized during the whole year. If at the end of the year there is an open position, the current return of your open position is added to the total return.

[9] The current return, r_c, is the unrealized return of the current position when off the equilibrium ($g_i \neq 0$).

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 $\bar{X}_{\Delta t}$ 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 $\bar{u} = u(\bar{X}_{\Delta t})\exp(\alpha^2\sigma^2_{\Delta t}/2)$, where $\sigma^2_{\Delta t}$ is the variance of $X_{\Delta t}$. The expected utility can be transformed back to the *effective return*, $X_{\text{eff}} = -\log(-\bar{u})/\alpha$ where

$$X_{\text{eff}} = \bar{X}_{\Delta t} - \frac{\alpha\sigma^2_{\Delta t}}{2}. \tag{4}$$

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

$$\sigma^2_{\Delta t} = \frac{n}{n-1}(\overline{X^2_{\Delta t}} - \bar{X}^2_{\Delta t}). \tag{5}$$

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.

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

$$X_{\text{eff,ann},\Delta t} = A_{\Delta t}X_{\text{eff}} = \bar{X} - \frac{\alpha}{2}A_{\Delta t}\sigma^2_{\Delta t}, \tag{6}$$

where $\bar{X}$ 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 $X_{\text{eff,ann}}$ is computed with n different time horizons Δt_i, and thus takes advantage of the fact that annualized $X_{\text{eff,ann}}$ can be directly compared

$$X_{\text{eff}} = \frac{\sum_{i=1}^{n} w_i X_{\text{eff,ann},\Delta t_i}}{\sum_{i=1}^{n} w_i}, \tag{7}$$

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 0.10.

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 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 et al., 1993; Dacorogna et al., 2000) where R_{eff} has high risk aversion in the zone of negative returns and a low one in the zone of profits whereas $X_{\rm eff}$ 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 $R_{\rm eff}$ 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)

$$\alpha = \begin{cases} \alpha_+ & \text{for } \Delta\tilde{R}_t \geqslant 0, \\ \alpha_- & \text{for } \Delta\tilde{R}_t < 0, \end{cases} \tag{8}$$

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 participants as the distribution of losses (drawdowns). Therefore, α_+ is quite smaller than α_- and we assume that $\alpha+=\alpha_-/4$ and $\alpha_-=0.20$. These values are under the assumption of the return measured in *percent*. They have to be multiplied by 100 if the returns are not expressed as percent figures. The *risk aversion* α associated with the utility function $u(\Delta\tilde{R})$ is defined in Keeney and Raiffa (1976) as follows:

$$\alpha = -\frac{\mathrm{d}^2u/[\mathrm{d}(\Delta\tilde{R})]^2}{\mathrm{d}u/\mathrm{d}(\Delta\tilde{R})}. \tag{9}$$

The utility function is obtained by inserting Eq. (8) into Eq. (9) and integrating twice over $\Delta\tilde{R}$

$$u = u(\Delta\tilde{R}) = \begin{cases} -\dfrac{\mathrm{e}^{-\alpha_+\Delta\tilde{R}}}{\alpha_+} & \text{for } \Delta\tilde{R} \geqslant 0, \\ \dfrac{1}{\alpha_-} - \dfrac{1}{\alpha_+} - \dfrac{\mathrm{e}^{-\alpha_-\Delta\tilde{R}}}{\alpha_-} & \text{for } \Delta\tilde{R} < 0. \end{cases} \tag{10}$$

The utility function $u(\Delta\tilde{R})$ is monotonically increasing and reaches its maximum 0 in the case $\Delta\tilde{R} \to \infty$ (infinite profit). All other utility values are negative. (The absolute level of u is not relevant; we could add and/or multiply all u values with the same constant factor(s) without affecting the essence of the method). The inverse formula computes a return value from its utility

$$\Delta\tilde{R} = \Delta\tilde{R}(u) = \begin{cases} -\dfrac{\log(-\alpha_+ u)}{\alpha_+} & \text{for } u \geqslant -\dfrac{1}{\alpha_+}, \\ -\dfrac{\log(1 - \frac{\alpha_-}{\alpha_+} - \alpha_- u)}{\alpha_-} & \text{for } u < -\dfrac{1}{\alpha_+}. \end{cases} \tag{11}$$

The more complicated nature of the new utility definition, Eq. (10), makes the derivation of a formula for the *mean* utility quite hard and with no analytical solution. Moreover, the $R_{\rm eff}$ is dominated by the drawdowns which are in the *tail* of the distribution, not in the center. The assumption of a Gaussian distribution, which may be acceptable for the distribution as a whole, is insufficient in the tails of the distribution, where the stop-loss, the leptokurtic nature of price changes, and the clustering of market conditions such as volatility cause very particular forms of the distribution. Therefore, the use of *explicit* utilities is suggested in the R_{eff} algorithm. The end results, R_{eff} and the effective returns for the individual horizons, will, however, be transformed back with the help of Eq. (11) to a return figure directly comparable with the annualized return and X_{eff}. The utility of the jth observation for a given time interval Δt is

$$u_{\Delta t,j} = u(\tilde{R}_{\Delta t,j} = u(\tilde{R}_{t_j} - \tilde{R}_{t_j-\Delta t}). \tag{12}$$

The total utility is the sum of the utility for each observation

$$u_{\Delta t} = \frac{\sum_{j=1}^{N_j} v_j u_{\Delta t,j}}{\sum_{j=1}^{N_j} v_j}. \tag{13}$$

In this formula, N_j is the number of observed intervals of size Δt that overlaps with the total sampling period of size T and the weight v_j is the ratio of the amount of time during which the jth interval coincides with the sampling period over its interval size Δt. This weight is generally equal to one except for the first observation(s) which can start before the sample start and the last one(s) which can end after the sample end. To obtain a lower error in the evaluation of the mean utility, different regular series of overlapping interval of size Δt can be used. The use of overlapping intervals is especially important when the interval size Δt is large compared to the full sample size T. Another argument for overlapping is the high, overproportional impact of *drawdowns* on R_{eff}. The higher the overlap factor, the higher the precision in the coverage of the worst drawdowns. The mean utility $u_{\Delta t}$ can be transformed back to an *effective* return value by applying Eq. (11)

$$\Delta\tilde{R}_{\mathrm{eff},\Delta t} = \Delta\tilde{R}(u_{\Delta t}). \tag{14}$$

This $\Delta\tilde{R}_{\mathrm{eff},\Delta t}$ is the typical, effective return for the horizon Δt, but it is not yet annualized. As in the case of the X_{eff}, an annualization is necessary for a comparison between R_{eff} values for different intervals. The annualization factor, $A_{\Delta t}$, is the ratio of the number of Δt in a year divided by the number of non-overlapping Δt's in the full sample of size T. We have

$$\tilde{R}_{\mathrm{eff,ann},\Delta t} = A_{\Delta t}\Delta\tilde{R}_{\mathrm{eff},\Delta t}. \tag{15}$$

To achieve a measure that simultaneously considers a wide range of horizons, we define R_{eff} as a weighted mean over all the n horizons

$$R_{\mathrm{eff}} = \frac{\sum_{i=1}^{n} w_i R_{\mathrm{eff,ann},\Delta t_i}}{\sum_{i=1}^{n} w_i}, \tag{16}$$

where the weights w_i are chosen according to the relative importance of the time horizons Δt_i and may differ for trading models with different trading frequencies. In the case of the trading models studied here, we have choosen a weighting function

$$w_i = w(\Delta t_i) = \frac{1}{2 + (\log(\Delta t_i/90\ \text{days}))^2} \tag{17}$$

with the maximum for a Δt of 90 days. 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 has heavier penalty.

4. Simulation methodology

The distributions of the performance measures under various null processes will be 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 $p_{\text{ask},j}$ and the bid price $p_{\text{bid},j}$. 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 indicators which are used in the real-time trading model do not directly analyze raw bid or ask prices, but rather the logarithmic *middle prices*, x_j, defined as

$$x_j = (\log(p_{\text{ask},j}) + \log(p_{\text{bid},j}))/2.0.$$

Similarly we define the relative bid-ask spreads, s_j, as

$$s_j = \log(p_{\text{ask},j}) - \log(p_{\text{bid},j}).$$

In our trading model simulations, we use 5 min prices to limit the required computational time for simulations. This selection is done to mimic the real-time trading model results generated from all ticks. The main information used by a trading model to update its indicators are 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 theta [10] time between these two ticks is

$$\Delta\theta_j = \theta_j - \theta_{j-1}.$$

By construction, in the sampled time series, the average elapsed theta time between two ticks, $\overline{\Delta\theta}$, is nearly 5 min.

[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 the business time scale and utilizes the average volatility to represent the activity of the market. The ϑ-time method is based on three geographical markets namely the East Asia, Europe and the North America. A more detailed exposition of the ϑ methodology is presented in Dacorogna et al. (1993). The seasonality correction method is required here for proper estimation and the simulation of the process equations.

For the trading model simulations, multiple time series from a given theoretical price generation processes need to be generated. To keep the impact of special events like the data holes in the model behavior, we have 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 5 min 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 of the observed returns. This is done by generating returns, $\hat{r}_j$, corresponding to a 5 min interval in theta time. In the case of a random walk process, the returns $\hat{r}_j$ are computed with

$$\hat{r}_j = \alpha + \varepsilon_j,$$

where $\varepsilon_j \sim N(0, \sigma^2)$. When the effective elapsed time between two ticks, $\Delta\theta_j$, is not exactly 5 min 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 5 min. 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 theta 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, i.e.

$$p'_{\text{ask},j} = \exp\left(x'_j + \frac{\bar{s}}{2} \right)$$

and

$$p'_{\text{bid},j} = \exp\left(x'_j - \frac{\bar{s}}{2} \right).$$

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 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 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 from these processes.

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

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

$$\sigma_t^2 = c_0 + \sum_{j=1}^{n} c_j \sigma_{t_j}^2, \quad \sigma_{t_j}^2 = \mu_j \sigma_{t-1,j}^2 + (1-\mu_j)\left(\sum_{i=1}^{k_j} r_{t-i}\right)^2, \tag{18}$$

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 min, 10 min up to 10 days. The weight parameter μ_j is chosen such that $\mu_j = \mathrm{e}^{-2/(k_{j+1}-k_j)}$. This choice guarantees that the center of the weight is in the middle of a given data interval.[11] 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. AR(p)-HARCH(n) processes together with the simulated residuals are used to generate the simulated returns from these processes.

5. Empirical findings

The simulated data is the five minutes[12] ϑ-time series 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 intra-day seasonalities and requires deseasonalization. This paper uses the deseasonalization methodology advocated in Dacorogna et al. (1993) named as the ϑ-time seasonality correction method.[13] The ϑ-time method uses the business time scale and utilizes the average volatility to represent the activity of the market. The ϑ-time method is based on three geographical

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

[12] The real-time system uses tick-by-tick data for its trading recommendations. The simulations in this paper are carried out with 5 min 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 min series are exactly compatible with performance of the real-time trading model which utilize the tick-by-tick data. Therefore, there is no loss of generality from the usage of 5 min frequency for the simulations instead of the tick-by-tick feed.

[13] The seasonality correction method is required here for proper estimation and the simulation of the process equations.

markets namely the East Asia, Europe and the North America. A more detailed exposition of the ϑ methodology is presented in Dacorogna et al. (1993). The simulations for each process are done for 1000 replications.

5.1. Random walk process

5.1.1. The RTT model

The performance for the RTT model is reported in Table 1. 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%, 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% and 4%. For the USD–CHF, the X_{eff} and R_{eff} are -1.68% and -4.23% reflecting the weakness of its performance.

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.[14] This tells us whether the historical realization is likely to be generated from this particular distribution or not. A small p-value ($< 5\%$) 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 generated 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 the representative of the data generating process.

The p-values[15] 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%, respectively. For the USD–DEM and USD–FRF, the p-values are less than the 2% level and it is about 2% for the DEM–JPY. In the case of the USD–CHF, the p-value for the annualized return is 8.9 which is well above the 5% level. The p-values of X_{eff} and R_{eff} are 0.0, 0.0% for USD–DEM; 0.7% and 0.6% for USD–CHF; 0.2% and 0.1% for USD–FRF and 0.2% and 0.1% for DEM–JPY. The p-values for the X_{eff} and R_{eff} are all $< 1\%$ rejecting the null

[14] 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.

[15] The p-value represents a decreasing index of the reliability of a result. The higher the p-value, the less we can believe that the observed relation between variables in the samples is a reliable indicator of the relation between the respective variables in the population. Specifically, the p-value represents the probability of error that is involved in accepting our observed result as valid, that is, as *representative of the population*. For example, the p-value of 0.05 indicates that there is a 5% probability that the relation between the variables found in our sample is a *fluke*. In other words, assuming that in the population there was no relation between those variables whatsoever, and by repeating the experiment, we could expect that approximately every 20 replications of the experiment there would be one in which the relation between the variables in question would be equal or stronger than ours. In many areas of research, the p-value of 5% is treated as a *borderline acceptable* level.

Table 1
Historical performance of the RTT and EMA models

	USD–DEM	USD–CHF	USD–FRF	DEM–JPY
RTT model				
Annual return	9.63	3.66	8.20	6.43
X_{eff}	3.78	− 1.68	4.80	3.81
R_{eff}	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				
X_{eff}	3.47	− 2.96	3.18	1.87
R_{eff}	1.80	− 4.81	1.97	0.58
Horizon: 29 days				
X_{eff}	3.27	− 4.10	4.41	2.12
R_{eff}	2.16	− 8.97	3.99	0.19
Horizon: 117 days				
X_{eff}	4.07	− 0.67	3.83	4.90
R_{eff}	5.10	− 1.77	3.77	5.17
Horizon: 301 days				
X_{eff}	4.62	1.94	6.35	5.39
R_{eff}	6.83	1.71	7.38	5.56
EMA model				
Annual return	3.33	4.40	6.01	7.09
X_{eff}	− 0.67	− 0.46	2.25	2.63
R_{eff}	− 2.48	− 3.00	2.02	2.97
Max drawdown	17.73	17.10	14.52	13.06
Deal frequency	0.77	0.85	0.70	0.75
Horizon: 7 days				
X_{eff}	− 2.09	− 2.86	1.36	1.81
R_{eff}	− 4.29	− 5.66	− 0.07	0.57
Horizon: 29 days				
X_{eff}	− 2.81	− 3.45	1.15	1.94
R_{eff}	− 7.18	− 7.26	− 0.91	1.47
Horizon: 117 days				
X_{eff}	− 1.03	0.39	1.92	3.43
R_{eff}	− 3.96	− 2.53	2.57	4.26
Horizon: 301 days				
X_{eff}	1.65	2.09	4.06	3.76
R_{eff}	1.75	2.17	5.17	4.31

hypothesis that the random walk process is consistent with the data generating process of exchange rate returns.

The maximum drawdowns for the USD–DEM, USD–CHF, USD–FRF and DEM–JPY are 11.02%, 16.08%, 11.36% and 12.03%. 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/week for the four currency pairs from the actual data. The deal frequencies indicate that the RTT model trades on average no more than two trades/week although the data feed is at the 5 min frequency. The mean simulated deal frequencies are 2.46, 1.98, 1.65 and 3.08 which are significantly larger than the actual deal frequencies.

The values for the maximum drawdown and the deal frequency indicate that the random walk simulation yield 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 the 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. The summary statistics of the simulated performance measures have negligible skewness and statistically insignificant excess kurtosis. This indicates that the distribution of the performance measures are symmetric and do not exhibit thick tails.

The behavior of the performance measures across 7 day, 29 day, 117 day and 301 day horizons are 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's 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% for all horizons. Overall, the multi-horizon analysis indicates the rejection of the random walk process as a representative data generating process for the foreign exchange returns.

5.1.2. The EMA model

The performance results for the EMA model are presented in Table 1. The annualized returns for the USD–DEM, USD–CHF, USD–FRF and DEM–JPY are 3.33%, 4.40%, 6.01% and 7.09%, respectively. Relative to the annualized return performance of the RTT model, the EMA model's return performance is lower for the USD–DEM and the USD–FRF. On the other hand, the annualized returns of the EMA model for USD–CHF and DEM–JPY are slightly higher. One noticeable difference is that the EMA model has higher maximum drawdowns for all four currency pairs relative to the RTT model. The EMA model has also a smaller deal frequency relative to the RTT model. The reason for the difference between the deal frequencies between the two models is that the RTT model has a contrarian strategy whereas the EMA model does not. The second reason is that the long indicator of the RTT model is 16 days whereas the EMA model has a long indicator of 20 days. A relatively longer average allows the EMA model to engage lower number of deals.

In Table 2, the *p*-values for the EMA model are presented. The *p*-values for the annualized returns are 12.6%, 8.8%, 6.4% and 6.6% for the four currency pairs. All four *p*-values are greater than the 5% level of confidence. Therefore, it is not possible to reject the null hypothesis that the data generating process is likely to be generated by a random walk process with the EMA model under the simulated annualized return

Table 2
Random walk simulations (1990–1996, 5 min frequency)

	USD–DEM	USD–CHF	USD–FRF	DEM–JPY
RTT model				
Annual return	0.3	8.9	1.2	2.1
X_{eff}	0.0	0.7	0.2	0.2
R_{eff}	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				
X_{eff}	0.0	1.2	0.2	0.3
R_{eff}	0.0	0.4	0.2	0.2
Horizon: 29 days				
X_{eff}	0.0	2.2	0.2	0.3
R_{eff}	0.0	2.3	0.1	0.3
Horizon: 117 days				
X_{eff}	0.0	0.9	0.2	0.2
R_{eff}	0.0	0.9	0.1	0.2
Horizon: 301 days				
X_{eff}	0.0	0.5	0.2	0.1
R_{eff}	0.0	0.7	0.2	0.3
EMA model				
Annual return	12.6	8.8	6.4	6.6
X_{eff}	2.8	0.9	0.6	1.4
R_{eff}	2.3	0.4	0.4	1.0
Max drawdown	98.5	99.9	99.6	99.6
Deal frequency	100.0	100.0	100.0	99.9
Horizon: 7 days				
X_{eff}	4.8	2.2	1.1	2.0
R_{eff}	3.3	1.1	0.4	1.3
Horizon: 29 days				
X_{eff}	6.1	2.7	0.3	1.6
R_{eff}	7.4	1.8	1.0	0.6
Horizon: 117 days				
X_{eff}	4.3	0.7	0.9	0.9
R_{eff}	4.8	1.1	0.6	0.9
Horizon: 301 days				
X_{eff}	2.0	1.2	0.7	1.1
R_{eff}	3.2	1.3	0.6	1.0

distributions. The examination of the p-values for the X_{eff} and R_{eff} indicate that these values are 2.8% and 2.3% for the USD–DEM; 0.9% and 0.4% for the USD–CHF; 0.6% and 0.4% for the USD–FRF and 1.4% and 1.0% for the DEM–JPY. Under these more stringent performance measures, the null hypothesis that the random walk process is consistent with the data generating process for the foreign exchange returns is rejected.

The p-values for the maximum drawdown and the deal frequency indicate that the EMA model's historical performance stays well below the ones generated under the random walk simulations. In other words, the random walk simulations always generate larger drawdowns relative to the historical drawdowns from the four currency pairs. In fact, the mean simulated drawdowns for the USD–DEM, USD–CHF, USD–FRF and DEM–JPY are 49.70%, 60.72%, 47.51% and 44.28% which are at least three times larger than the historical drawdowns. Similarly, the mean of the simulated deal frequencies for the random walk process stays approximately 20% above the historical realizations. The multi-horizon analysis of the EMA model indicate that the model's performance improves in longer horizons. This is mostly due to the low dealing frequency. The p-values of the four currency pairs for X_{eff} and R_{eff} remain $< 3.3\%$ for the 301 day horizon.

The overall evaluation of the EMA model is that this simple technical model generates net positive returns for all currencies after taking the transaction costs into account. The simulated probability distributions of the performance measures also indicate that the null hypothesis of whether the foreign exchange returns can be characterized by random walk process is rejected. Both the RTT model as well the EMA model are able to generate positive annualized returns (after taking the transaction costs into account) and the performance of the RTT model is superior to the EMA model with higher returns and smaller drawdowns. The EMA model has smaller deal frequency per week relative to the RTT model.

5.2. GARCH(1,1) process

5.2.1. The RTT model

In Table 3, the RTT model simulations 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 therefore leading to the failure of the rejection of 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. 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's 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.

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 0.9%, respectively. All four currency pairs except USD–CHF yield p-values which are $< 1\%$. The X_{eff} and R_{eff} are 0.1% and 0.0% for USD–DEM; 1.4% and 0.9% for USD–CHF; 0.1% and 0.1% for USD–FRF and 0.4% and 0.4% for DEM–JPY.

The historical maximum drawdown and deal frequency of the RTT model is small relative to the ones generated from the simulated data. The maximum drawdowns for

Table 3
GARCH(1,1) simulations (1990–1996, 5 min frequency)

	USD–DEM	USD–CHF	USD–FRF	DEM–JPY
RTT model				
Annual return	0.4	8.4	0.9	1.2
X_{eff}	0.1	1.4	0.1	0.4
R_{eff}	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				
X_{eff}	0.2	1.6	0.3	0.6
R_{eff}	0.0	1.0	0.1	0.6
Horizon: 29 days				
X_{eff}	0.2	2.6	0.1	0.5
R_{eff}	0.1	2.8	0.0	0.7
Horizon: 117 days				
X_{eff}	0.1	0.8	0.4	0.2
R_{eff}	0.1	0.7	0.2	0.2
Horizon: 301 days				
X_{eff}	0.2	0.9	0.1	0.4
R_{eff}	0.3	1.5	0.2	0.5
EMA model				
Annual return	14.1	9.6	4.6	2.8
X_{eff}	2.3	0.7	0.6	0.5
R_{eff}	1.8	0.5	0.4	0.2
Max drawdown	99.2	99.9	99.7	99.8
Deal frequency	100.0	100.0	100.0	97.8
Horizon: 7 days				
X_{eff}	3.8	1.7	0.8	1.2
R_{eff}	2.8	1.2	0.6	0.6
Horizon: 29 days				
X_{eff}	5.8	2.7	0.9	0.6
R_{eff}	7.2	1.5	0.9	0.3
Horizon: 117 days				
X_{eff}	3.1	0.7	0.7	0.4
R_{eff}	3.8	1.5	0.5	0.3
Horizon: 301 days				
X_{eff}	2.1	1.3	0.8	3.6
R_{eff}	2.4	1.5	0.7	2.4

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 or 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 difference 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 indicate that the historical realizations of these two measures stay outside of the 5% level of simulated distributions of these two performance measures.

The multi-horizon examination of the equity curve with the X_{eff} and R_{eff} performance measures indicate 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.

5.2.2. The EMA model

The simulation performance of the EMA model under the GARCH(1,1) process is presented in Table 3. The p-values for the annualized returns for the four currency pairs are 14.1%, 9.6%, 4.6% and 2.8%, respectively. Based on the annualized return performance, the null hypothesis that the GARCH(1,1) process is consistent with the data generating process of foreign exchange returns cannot be rejected for the USD–DEM and USD–CHF. The other currency pairs stay below the 5% level but relatively close to the 5% level providing a weak level of confidence. In comparison with the RTT model, the p-values of the EMA model are substantially higher indicating the relative weakness of this model as a model of foreign exchange dynamics.

The p-values of the X_{eff} and R_{eff} are 2.3%, 1.8% for USD–DEM; 0.7%, 0.5% for USD–CHF; 0.6%, 0.4% for USD–FRF and 0.5%, 0.2% for the DEM–JPY. Relative to the annualized return performance, the p-values of the X_{eff} and R_{eff} remain statistically significant as the largest values are not $> 2.3\%$. As indicated earlier, the annualized return does not utilize the entire equity curve. Rather, only the first and the last points of the equity curve are used in the calculation of the annualized returns. Therefore, a straight line equity curve as well as an equity curve which is subject to extreme variations can have the same annualized returns. Based on the X_{eff} and R_{eff} p-values, it can be concluded that the GARCH(1,1) process be rejected as a representative data generating process of foreign exchange returns with the EMA model.

5.3. AR(4)-GARCH(1,1) process

A further direction is to investigate whether a 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 Ljung–Box statistic indicate no serial correlation in the normalized residuals. The variance of the normalized residuals are near one. There is no evidence of skewness but the excess kurtosis remains large for the residuals.

5.3.1. The RTT model

The p-values of the annualized returns are presented in Table 4 which are 0.1%, 3.7%, 0.3% and 0.5% for the USD–DEM, USD–CHF, USD–FRF and DEM–JPY. The results indicate that the AR(4)-GARCH(1,1) process is rejected under the RTT model as a representative data generating process of foreign exchange returns. One possible

Table 4
AR(4)-GARCH(1,1) simulations (1990–1996, 5 min frequency)

	USD–DEM	USD–CHF	USD–FRF	DEM–JPY
RTT model				
Annual return	0.1	3.7	0.3	0.5
X_{eff}	0.1	1.9	0.2	0.1
R_{eff}	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				
X_{eff}	0.1	2.8	0.2	0.5
R_{eff}	0.2	2.4	0.2	0.5
Horizon: 29 days				
X_{eff}	0.1	4.6	0.0	0.4
R_{eff}	0.1	9.7	0.0	0.4
Horizon: 117 days				
X_{eff}	0.1	1.3	0.3	0.1
R_{eff}	0.0	1.4	0.3	0.1
Horizon: 301 days				
X_{eff}	0.1	0.8	0.2	0.0
R_{eff}	0.1	1.1	0.2	0.2
EMA model				
Annual return	10.0	5.9	3.2	0.8
X_{eff}	3.8	1.7	0.7	0.5
R_{eff}	3.7	2.4	0.6	0.4
Max drawdown	96.9	99.0	98.8	99.8
Deal frequency	75.1	95.1	98.8	61.1
Horizon: 7 days				
X_{eff}	6.8	4.8	1.6	1.3
R_{eff}	6.8	4.9	1.2	0.9
Horizon: 29 days				
X_{eff}	9.4	6.8	1.5	0.9
R_{eff}	17.4	7.0	1.7	0.4
Horizon: 117 days				
X_{eff}	5.8	1.5	1.1	0.5
R_{eff}	8.9	2.9	0.7	0.5
Horizon: 301 days				
X_{eff}	1.8	1.6	0.5	0.2
R_{eff}	1.8	2.4	0.3	0.6

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 min frequency but the model's dealing frequency is between one or two deals per week. Therefore, the model picks up the high frequency serial correlation as a 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. The p-values of the X_{eff} and R_{eff} are 0.1%, 0.0% for USD–DEM; 1.9%, 2.3% for USD–CHF; 0.2%, 0.1% for USD–FRF and 0.1%, 0.1% 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 almost all replications the AR(4)-GARCH(1,1) generates higher maximum drawdowns and deal frequencies.

5.3.2. The EMA model

The performance of the EMA model with the AR(4)-GARCH(1,1) process are presented in Table 4. The p-values are 10.0%, 5.9%, 3.2% and 1.8% for the USD–DEM, USD–CHF, USD–FRF and DEM–JPY. The p-values for the USD–DEM and USD–CHF are higher than the 5% level. The results with the X_{eff} and R_{eff}, on the other hand, indicate that the p-values remain under the 5% levels. In fact, the p-values for the X_{eff} and R_{eff} are 3.8%, 3.7% for the USD–DEM; 1.7%, 2.4% for the USD–CHF; 0.7%, 0.6% for the USD–FRF and 0.5%, 0.4% for the DEM–JPY. In comparison with the p-values of the same process for the RTT model, the p-values of the EMA model remain high. This is mostly due to the simplistic nature of the EMA model which cannot capture the dynamics of the foreign exchange returns as successfully as the RTT model. The multi-horizon dynamics of the EMA model are weak with the USD–DEM series, as the p-values of the X_{eff} and R_{eff} over the 5% level for all horizons except the 301 day horizon. The USD–CHF series also behave similarly in the first two horizons. The performance however goes against the AR(4)-GARCH(1,1) process in longer horizons. For the USD–FRF and DEM–JPY, all horizons have p-values less than 5% indicating the rejection of the AR(4)-GARCH(1,1) as a representative data generating process of the foreign exchange returns.

Overall, both the RTT model and the EMA model generate net positive annualized returns for the seven years of high frequency analysis of the four currencies studied here. The performance of the RTT model dominates the EMA model for all currencies. The RTT model also yields smaller drawdowns implying less volatile equity curves. The simulation results indicate that the random walk, the GARCH(1,1) and the AR(4)-GARCH(1,1) cannot successfully characterize the dynamics of the foreign exchange returns. In particular, the results indicate that the temporal aggregation properties of the GARCH(1,1) and AR(4)-GARCH(1,1) processes fail to match the temporal aggregation properties of the actual foreign exchange returns.

5.4. AR(4)-HARCH(9) process

The conditional means of the foreign exchange returns are estimated with nine lags of returns and fourth order HARCH parametrization. The additional lags did not lead to substantial increases in the likelihood value. The Ljung–Box statistic indicate no serial correlation in the normalized residuals. The variance of the normalized residuals are near one. There is no evidence of skewness but the excess kurtosis remains large for the residuals.

Table 5
AR(4)-HARCH(9) USD–DEM simulations (1990–1996, 5 min frequency)

	RTT	EMA
Annual return	0.1	7.9
X_{eff}	0.1	2.6
R_{eff}	0.0	3.5
Max drawdown	100.0	97.3
Deal frequency	100.0	29.3
Horizon: 7 days		
X_{eff}	0.1	5.5
R_{eff}	0.2	9.3
Horizon: 29 days		
X_{eff}	0.1	8.8
R_{eff}	0.1	26.0
Horizon: 117 days		
X_{eff}	0.1	3.6
R_{eff}	0.0	9.9
Horizon: 301 days		
X_{eff}	0.1	0.9
R_{eff}	0.1	1.5

5.4.1. The RTT model

In a GARCH process, the conditional heteroskedasticity exists in 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 multiple frequency heteroskedastic structure may possibly be more in line with the heterogeneous structure of the foreign exchange markets.

The HARCH findings with the USD–DEM [16] are presented in Table 5 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) indicate that 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

[16] AR(4)-HARCH(9) process yields the largest likelihood value among other AR-HARCH parametrizations that we studied.

memory, it results in large sudden jumps in the simulation followed by series of large movements.

In the presence of statistically significant four conditional mean parameters, the HARCH model is a successful characterization of the data dynamics at the 5-min frequency. However the important message that we are trying to convey here is that 5-min 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-min 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-min 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.

5.4.2. The EMA model

In Table 5, p-values for the annual return, X_{eff}, R_{eff} are presented. These values are 7.9%, 2.6% and 3.5% for the EMA model and 0.1%, 0.1% and 0% for the RTT model. Similar performance difference also prevails at 7, 29, 117 and 301 day horizons and the p-values of the EMA model remain high. This is mostly due to the simplistic nature of the EMA model which cannot capture the dynamics of the foreign exchange returns as successfully as the RTT model.

6. Conclusions

Both the RTT model and the EMA model generate net positive annualized returns for the seven years of high frequency analysis of the four currencies studied here. The performance of the RTT model dominates the EMA model for all currencies. The RTT model also yields smaller drawdowns implying less volatile equity curves. The simulation results indicate that the random walk, the GARCH, AR-GARCH and AR-HARCH cannot successfully characterize the dynamics of the foreign exchange returns. In particular, the results indicate that the temporal aggregation properties of the GARCH, AR-GARCH and AR-HARCH processes fail to match the temporal aggregation properties of the actual foreign exchange returns.

The simulation results also indicate that the foreign exchange series may possess a multi-frequency conditional mean and conditional heteroskedastic dynamics. The traditional heteroskedastic models fail to capture the entire dynamics by only capturing a slice of this dynamics at a given frequency. Therefore, a more realistic process for foreign exchange returns should consider the scaling behavior of returns at different frequencies. This scaling behavior should be taken into account in the construction and the estimation of a representative process.

Acknowledgements

Ramazan Gençay thanks the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada for financial support.

References

Andersen, T.G., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 4 (2–3), 115–158.

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

Brock, W., Lakonishok, J., LeBaron, B., 1992. Simple technical trading rules and the stochastic properties of stock returns. The Journal of Finance 47 (5), 1731–1764.

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 FX market. Journal of International Money and Finance 12 (4), 413–438.

Dacorogna, M.M., Müller, U.A., Jost, C., Pictet, O.V., Olsen, R.B., Ward, J.R., 1995. Heterogeneous real-time trading strategies in the foreign exchange market. The European Journal of Finance 1, 243–253.

Dacorogna, M.M., Gençay, R., Müller, U.A., Pictet, O.V., 2000. Effective return, risk aversion and drawdowns. Physica A 289, 66–88.

Dacorogna, M., Gençay, R., Müller, U., Olsen R., Pictet, O., 2001. An Introduction to High Frequency Finance. Academic Press, San Diego, CA.

Diebold, F.X., Nason, J.A., 1990. Nonparametric exchange rate prediction? Journal of International Economics 28, 315–332.

Gençay, R., 1999. Linear, non-linear and essential foreign exchange rate prediction with simple technical trading. Journal of International Economics 47 (3), 91–107.

Keeney, R.L., Raiffa, H., 1976. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Wiley, New York.

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

LeBaron, B., 1997. Technical trading rules and regime shifts in foreign exchange. In: Acar, E., Satchell, S. (Eds.), Advanced Trading Rules. Butterworth-Heinemann, Stoneham, MA.

Levich, R.M., Thomas III, L.R., 1993. The significance of technical trading-rule profits in the foreign exchange market: a bootstrap approach Journal of International Money and Finance 12 (5), 451–474.

Meese, R.A., Rogoff, J., 1983. Empirical exchange rate models of the seventies, do they fit out of sample? Journal of International Economics 14, 3–24.

Müller, U.A., Dacorogna, M.M., Pictet, O.V., 1993. A trading model performance measure with strong risk aversion against drawdowns. Internal document UAM 1993-06-03, Olsen & Associates, Seefeldstrasse 233, 8008 Zürich, Switzerland.

Müller, U.A., Dacorogna, M.M., Davé, R.D., Olsen, R.B., Pictet, O.V., von Weizsäcker, J.E., 1997. Volatilities of different time resolutions—analyzing the dynamics of market components. Journal of Empirical Finance 4 (2–3), 213–239.

Pictet, O.V., Dacorogna, M.M., Müller, U.A., Olsen, R.B., Ward, J.R., 1992. Real-time trading models for foreign exchange rates. Neural Network World 2 (6), 713–744.
Sharpe, W.F., 1966. Mutual fund performance. Journal of Business 39 (1), 119–138.
Sharpe, W.F., 1994. The sharpe ratio. Journal of Portfolio Management 21, 49–59.
Sullivan, R., Timmermann, A., White, H., 1999. Data-snooping, technical trading rule performance and the bootstrap. The Journal of Finance 54 (5), 1647–1692.