UNIVERSITY OF COPENHAGEN
DEPARTMENT OF ECONOMICS

Seminar: Topics in Financial Econometrics

Frederik Degn Pedersen & Frederik Bach Trier

Volatility Trading in FX Options Markets

Submission date: 1 June 2022

Keystrokes: 54,779

ECTS: 7.5

Volatility Trading in FX Options Markets

Frederik Degn Pedersen & Frederik Bach Trier

June 2022

University of Copenhagen

Department of Economics

Abstract

This paper investigates whether the use of a simple GARCH(1,1) model can provide positive investment returns by constructing portfolios of currency options based on the conditional volatility forecasts. We replicate the approach by Dunis and Huang (2002) and augment it by adding additional currency pairs and allowing for time-varying bid-ask spreads. Our analysis finds that the GARCH(1,1) forecasts are inferior to those priced into options by market participants for our sample of three developed market and three emerging market currencies. This implies that our trading strategy backtest cannot generally provide positive and reliable returns over the trading horizon, which supports the conclusion of Pilbeam and Langeland (2014), who find that option traders are better at forecasting volatility than the class GARCH models. We conjecture that this is due to traders discounting future events into current prices, while the GARCH(1,1) assumes a zero mean innovations process.

Individual contributions:

- Frederik Degn Pedersen: Sections 2.1-2, 3 (pp. 14-16), 4.1-2, 5.1-2
- Frederik Bach Trier: Sections 1, 2.3, 3 (pp. 12-13), 4.3, 5.3, 6

Contents

1 Introduction

This paper investigates whether the use of a simple GARCH(1,1) model can provide positive investment returns by constructing portfolios of currency options based on the conditional volatility forecasts. Our approach is based on that of Dunis and Huang (2002), who formulate a volatility trading strategy that exploits potential mispricing in FX options markets by modelling FX returns with a GARCH(1,1) model. If one can consistently forecast future volatility more accurately than what markets are currently pricing into FX options, one may profit by the trading strategy based on long and short option "straddles". This strategy is based on constructing a portfolio of call and put options such that the trading profit is not affected by the overall direction of the market, only the realized volatility.

Our investigation of three developed market and three emerging market currencies finds that, in general, the straddle strategy is not profitable when forecasting future volatility within the GARCH framework. Despite some calibrations of the trading strategy providing positive returns, the overall conclusion is that markets are better at forecasting future volatility than the GARCH(1,1). We conjecture that this is due to the forward-looking nature of market-implied volatility versus the backward-looking structure of the GARCH(1,1).

In April 2019, the international foreign exchange (FX) markets, which is one of the most liquid asset markets in the world, saw a daily turnover of USD 6.6 trillion with most of the volume constituted by swaps and outright spot trades (Bank for International Settlements, 2019). FX markets consist of various agents participating with different objectives. These agents, or market particpants, which range from retail clients to banks and large institutions, trade FX to, e.g., speculate on price fluctuations, while others trade mechanically, e.g., central banks carrying out monetary policy or corporations trading FX for operations and cash flow hedging purposes.

The majority of FX trading, and the trading of its derivatives, is conducted on an over-the-counter (OTC) basis via a network of dealers in banks rather than via centralized exchanges. The magnitude of the market combined with the absence of regulation would suggest that the FX markets resemble a close proxy for a "perfect" market. Such a market should, according to the efficient market hypothesis, incorporate all available information in prices, which implies that no available information can provide more accurate forecasts of future prices than what is currently reflected in the market prices. Thus, modelling the conditional variance and attempting to beat market participants in forecasting volatility should, on average, not be feasible.

There exists an extensive literature on the study of financial market volatility, which was pioneered with the work of Engle (1982), who formulated the autoregressive conditional heteroskedasticity (ARCH) model for explicit conditional variance modelling. Focusing on the class of non-linear conditional heteroskedastic time series models, we summarize some central conclusions from the literature on FX market volatility modelling in the following.

In an extensive study of generalized ARCH-type models (GARCH), Hansen and Lunde (2005) evaluate the out-of-sample forecasting performance of 330 different model specifications for the conditional variance of the US Dollar versus the Deustche Mark exchange rate. Their conclusion is that there is no evidence of the parsimonious GARCH(1,1) specification being inferior to more elaborate and convoluted models such as the IGARCH, NGARCH, etc. Moreover, in their study of exchange rate data ranging from 1987 through 1992, they also conclude that estimating the models with an imposed t-distribution for the innovation process yields more precise estimates. Interestingly, the conditional mean specification does not have any significant impact on the forecasting ability of the models.

Thus, if the GARCH(1,1) provides a good measure of the conditional variance, we should be able to the exploit the available information at time t to profit at time $t+1$ if market participants are not incorporating this information when forming expectations. One such way would, as suggested by Dunis and Huang (2002), appear if the GARCH conditional volatility forecast is a better predictor of future volatility than what is actually priced by derivatives markets. Therefore, they suggest constructing an FX options trading strategy that is market, or ”delta”, neutral but long or short the volatility, or ”vega”. In short, a trader can profit by buying (writing) a portfolio consisting of an at-the-money call and an at-the-money put option with the same strike K, if the realized volatility at expiry turns out to be higher (lower) than what is priced into the options at time of entering the trade.

To investigate whether one such strategy is feasible, Pilbeam and Langeland (2014) investigate four highly liquid developed market dollar parities, namely the USD against the

Japenese Yen (JPY), Swiss Franc (CHF), British Pound (GBP), and Euro (EUR). They find that for the period 2002 through 2012 the GARCH(1,1), EGARCH(1,1), and GJR-GARCH(1,1) models are all inferior in terms of in-sample forecasting precision versus the market implied volatilties. As a result, they conclude that such models are "not particularly useful for forecasting exchange rate volatility", and that markets efficiently price expected future volatility. However, what is not accounted for in their study, is that even though the market-implied volatilities are better predictors than GARCH-predicted volatilities on average, there may be few, but substantial occurences of positive profits from the strategy which may exceed the losses of other periods. That is, despite the market beating the GARCH(1,1) on average, there may be some dynamics that the GARCH(1,1) is better at capturing than the markets.

The conclusion that GARCH-type models are not useful for forecasting and trading exchange rate volatility is in contrast to the conclusions of the empirical application in Dunis and Huang (2002). They find that even though the GARCH(1,1) prediction is inferior to the market-implied volatilities in terms of forecasting accuracy, they are able to generate profit estimates of approximately 21.7 percent on average across trading strategy calibrations. However, as we will also discuss in Section 5, the profitability is difficult to measure in a precise way due to actual option quotes not being widely available on a historical and large scale basis.

Our paper is structured as follows. In Section 2, we outline the central theory with respect to option pricing and trading, as well as the econometric foundation for estimation of non-linear time series models for the conditional variance. In Section 3 we present the data that we use for our application along with descriptional statistics. In Section 4 we present and analyse the findings of the implemented strategy, which are subsequently discussed and concluded upon in Section 5 and 6.

2 Theory

2.1 Foreign Exchange Options as an Asset Class

Foreign exchange (FX) markets are a central element of international finance since it allows for the transfer of capital between countries with different currency denominations. Moreover, there exists a plethora of FX derivitives, e.g., options, forwards, and swaps, that market participants use to hedge against and speculate on various risk factors.

A European call (put) option is the right, but not the obligation, to buy (sell) an asset at some pre-specified date at an agreed upon price (the ”strike” price). The potential use cases of options on currencies are plenty. Consider, for instance, a large US based corporation which knows that it one month from now, at time T, will receive a cash flow, say 1 billion, from international operations denominated in some foreign currency, say Japanese Yen (JPY).

One way to hedge the FX risk that the cash flow implies for the balance sheet, is to enter an outright forward contract for exhanging 1 billion JPY to S_T billion USD one month from now, where S_T is the prevailing forward exchange rate. With such a forward contract in place, the company knows the exact cash flow, measured in USD, that it will receive one month from now.

However, what if the company is only interested in limiting the the downside FX risk, or what if it is neutral on market direction, but has a stance on the future volatility of the JPY versus the USD? In such cases, FX options are relevant, as they may be combined as options portfolios in a multitude of ways to customize the expected cash flow at expiry. For instance, has some private information which it expects to increase the future volatility of the USDJPY, but it does not have a stance on the direction, then it may combine a long call and a long put option. At expiry this would yield a positive payoff if the volatility of the USDJPY increased over the holding period, i.e., if the price ended sufficiently far away from the strike price K.

Naturally, these complex payoff structures attract not only companies attempting to hedge FX risk from operations, but also hedge funds and other speculative investors who formulate trading strategies in order to provide consistently positive returns.

2.2 Valuation of Foreign Exchange Options

The valuation of FX options is analogous to the formulae known from Black and Scholes (1973) extended by Garman and Kohlhagen (1983) to allow for a domestic-foreign interest rate differential. The underlying asset is the spot exchange rate, defined as S_0, i.e. the value of one unit of foreign currency in domestic currency. As the holder of foreign currency receives the the risk-free yield in the foreign currency, the asset can be thought of as being similar to a stock with a known dividend yield.[1]

Assuming that the exchange rate is a log-normal process following a geometric brownian motion with a drift equal to the domestic-foreign risk-free interest rate differential,

$$dS_t = (r_d - r_f)S_t dt + \sigma S_t dW_t,$$

the pricing formulae for European calls and puts are given by equations (1) and (2) respectively:

$$c = S_0 e^{-r_f T} N(d_1) - K e^{-r_d T} N(d_2) \tag{1}$$

$$p = K e^{-r_d T} N(-d_2) - S_0 e^{-r_f T} N(-d_1) \tag{2}$$

where

$$d_1 = \frac{\ln(S_0/K) + (r_d - r_f + \sigma^2/2)\,T}{\sigma\sqrt{T}} \tag{3}$$

$$d_2 = \frac{\ln(S_0/K) + (r_d - r_f - \sigma^2/2)\,T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T} \tag{4}$$

with K being the strike price, $N(x)$ being the normal cumulative distribution function and T being *time to maturity* calculated in accordance to relevant day count conventions.[2] To obtain the lower bounds for the price of the European currency option, say a call, we realize that the same probability distribution for the exchange rate at time T is obtained from 1) the exchange rate starting at S_0 providing a continuously compounded risk-free yield of r_f

[1] Hull (2021).

[2] The value of the strike must also be quoted in units of domestic currency per unit of foreign currency.

and 2) the exchange rate starting at $S_0 e^{-r_f T}$ paying no yield.[3] Consider two portfolios:

Portfolio 1 : One European call option and one cash-account worth $Ke^{-r_d T}$

Portfolio 2 : $e^{-r_f T}$ units of foreign currency with continuously reinvested yield

We know that Portfolio 1 is worth $\max\{S_0 e^{-r_f T}, K\}$ at time T, whereas Portfolio 2 is worth $S_0 e^{-r_f T}$ at time T. By a no-arbitrage argument, Portfolio 1 must also be worth at least as much as Portfolio 2 at time t and thus we obtain the following lower bound:

$$
\begin{aligned}
&c + Ke^{-r_d T} \geq S_0 e^{-r_f T} \\
&c \geq S_0 e^{-r_f T} - Ke^{-r_d T} \\
&c \geq \max\left\{S_0 e^{-r_f T} - Ke^{-r_d T}, 0\right\}
\end{aligned} \tag{5}
$$

with the last expression coming from the fact that the value of a call option expiring worthless cannot be negative. Analogously, the lower bound for a European put option is given by $p \geq \max\left\{Ke^{-rT} - S_0 e^{-rf}, 0\right\}$ and from the combination of the two, the put-call parity may be derived from a similar replication argument:

$$
c + Ke^{-r_d T} = p + S_0 e^{-r_f T}. \tag{6}
$$

This parity must hold for European currency calls and puts with same strike and maturity for there to be no arbitrage oppotunities.

2.2.1 Implied volatility

The extended Black-Scholes model outlined above provides a closed-form solution for the price of a European FX option per unit of notional given the input parameters.[4] All of the input parameters apart from future volatility are observable in the market, and hence implied volatility is calculated from inverting the model and solving for future volatility using the

[3] Hull (2021)

[4] Today, many industry practitioners use more sophisticated models like SABR or a wide range of local volatility models. Risk metrics can differ across models, but as they are mostly in line when it comes to prices, we just stick to keeping it simple.

known market prices for that specific option. As such, implied volatility can be seen as a price-based proxy for market risk, i.e. what the market expects the volatility of the price of the underlying to be over the life of the option. An option with a high implied volatility is gauged as having a greater probability of ending ”in-the-money” and therefore has a higher premium.

If the assumptions of the Black-Scholes model hold, the implied volatility would be identical for all European options on a given FX rate, i.e. over the diffent strikes and times to maturity. In reality, however, the believe is that returns are more extreme than a log-normal process would prescribe which gives rise to the well-known volatility surface. For FX options in particular, implied volatility as a function of ”moneyness” usually takes the form of a ”smile” due to the fact that the price is merely a ratio, e.g., $S_t = \mathrm{EUR/USD}$:

$$-\log(S_t) = -\log\left(\frac{\mathrm{EUR}_t}{\mathrm{USD}_t}\right) = \log\left(\frac{\mathrm{USD}_t}{\mathrm{EUR}_t}\right) = \log(1/S_t). \tag{7}$$

As such, high skews on one side will automatically be accompanied by a high skew on the other side. Although they do not have to be perfectly symmetric, ”smirks” to either side are generally unsustainable for longer time periods.

In practice, it is the convention that options are quoted in terms of implied (annualized) volatility. The industry standard of quoting in units of the common IV model input ensures that market participants with different models and inputs can agree on the terms of the trade. It is important to note that the implied volatility of an FX option depends on the numeraire of the purchaser, i.e. the currency in which the option is valued. An option on EURUSD gives a USD value linear in the pair using USD as the numeraire and a non-linear value in EUR due to the non-linearity of the inversion operation: $x \mapsto 1/x$.

2.3 Modelling and Forecasting Volatility

In this section, we outline the GARCH framework as well as the approach for forecasting conditional variances based on the GARCH model. The main idea is that, in the long run and under the assumption a mean zero innovations process, the conditional variance will converge to the unconditional variance of the underlying data generating process.

Thus, the complicated part of modelling the conditional variance is to 1) obtain a reasonable estimate for the current conditional variance and 2) obtain a reasonable estimate for how fast the conditional variance converges. The GARCH literature is enormous, and we therefore rely on a few of the most common specifications as extensions to the baseline model for robustness checks.

2.3.1 The GARCH framework

To forecast volatility, we apply the workhorse GARCH(1,1) model suggested by Bollerslev (1986). The model is defined as a joint model for the conditional mean and conditional variance equations as follows:

$$r_t = \varepsilon_t \tag{8}$$

$$\varepsilon_t = \sigma_t z_t \tag{9}$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{10}$$

with r_t being the log return at time t, x_0, σ_0^2 given, and the parameter restrictions $\omega > 0$ and $\alpha, \beta \geq 0$. In our application, the innovation z_t is an i.i.d. process drawn from the (scaled) Student's t-distribution with v degrees of freedom such that $z_t = \sqrt{\frac{v-2}{v}} z_t^*$ where $z_t^* \overset{iid}{\sim} t_v$. For the standard asymptotic properties of the GARCH(1,1) to apply, one needs to establish a law of large numbers and a central limit theorem, e.g., by choosing a drift function and showing for which parameters the drift criterion is fulfilled. The regularity conditions that the asymptotic properties require can be shown to hold for the sufficient, but not necessary, condition $\alpha + \beta < 1$, where the process is stationary, geometrically ergodic and has a finite second order moment, which we need for forecasting the variance.

We estimate the model in Equations (8)-(10) with the Maximum Likelihood Estimator (MLE) using the density for the Student's t distribution. This allows for fatter tails than that of the Gaussian distribution, which is desirable in this application. The log-likelihood contribution from observation t can thus be written as:

$$\ell_t(\theta) = \log \left[\frac{\Gamma(\nu)}{\sqrt{(\nu-2)\pi\sigma_t^2}} \left(1 + \frac{r_t^2}{(\nu-2)\sigma_t^2} \right)^{\frac{\nu+1}{2}} \right],$$

where $\Gamma(\cdot)$ is the gamma function and $\theta = (\omega, \alpha, \beta, \nu)$. Hence, the MLE, $\hat{\theta}$, is obtained by maximizing the log-likelihood function:

$$\hat{\theta}_T = \arg\max_{\theta \in \Theta} L_T(\theta) = \arg\max_{\theta \in \Theta} \sum_{t=1}^{T} \ell_t(\theta)$$

where $\Theta := \{\theta \in \mathbb{R}^4 | \omega > 0, \alpha, \beta \geq 0, \nu > 2\}$. Estimation is in practice done via numerical procedures due to the non-linearity of the optimization problem. An important note here is that we apply a rolling-window estimation procedure to fit the GARCH model on a daily basis throughout the trading period.

2.3.2 Model extensions

To allow for other types of return dynamics we consider an extension to the model for the conditional variance in the appendix, which thus serves as a robustness check for our results in the empirical analysis. Although Hansen and Lunde (2005) find no evidence that extensions to the GARCH(1,1) improves forecasting accuracy, it is interesting to compare the results against this model. Should the ex post trading profit from applying other models be higher, it would be inefficient not to consider these, as we seek to maximize out-of-sample profit, not in-sample goodness-of-fit.

The literature on conditional volatility modelling has led to what Bollerslev has pointed out to be a "perplexing alphabet-soup of acronyms and abbreviations" (Verbeek (2012) p.327), which is also evident in the paper by Hansen and Lunde (2005) who investigate 330 different specifications of the model. In particular, all model extensions are constructed based on the need for alternative dynamics and parameter restrictions. To circumvent the potential drawbacks from the symmetry of how past innovations affect current volatility in the GARCH(1,1), we consider the model by Glosten et al. (1993) known as the GJR-GARCH model. In this model, past negative shocks have a larger impact on current volatility than past positive shocks. In particular, consider the conditional variance specification given by:

$$\sigma_t^2 = \begin{cases} \omega + (\alpha+\kappa)\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, & \text{if } \varepsilon_{t-1} < 0 \\ \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2, & \text{if } \varepsilon_{t-1} \geq 0 \end{cases}$$

with $\omega > 0, \alpha+\kappa, \beta \geq 0$. This model was originally motivated for modelling equity volatility due to the so-called leverage effect: A negative return implies that the market value of equity decreases which, in turn, increases the leverage ratio. A higher leverage ratio implies a higher default risk and therefore the volatility of the stock price should increase. Note that for foreign exchange, there is no such default risk, and the model is motivated by its ability to capture asymmetry in the effects of past innovations on current volatility.

2.3.3 Out-of-sample volatility forecasts

Having obtained a vector of parameter estimates, one may conduct an out-of-sample volatility forecast using the recursive structure of the GARCH models. For the purpose of the exposition, recall that the GARCH(1,1) may be written as an ARMA(∞) model by recursive substitution:

$$\begin{aligned} \sigma_t^2 &= \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \\ &= \omega + \alpha\varepsilon_{t-1}^2 + \beta(\omega + \alpha\varepsilon_{t-2}^2 + \beta\sigma_{t-2}^2) \\ &\vdots \\ &= \frac{\omega}{1-\beta} + \alpha\sum_{j=0}^{\infty}\beta^{j-1}\varepsilon_{t-j}^2 \end{aligned}$$

Recall that the process is weakly mixing with finite second order moment if $\alpha+\beta < 1$, which implies that the unconditional variance is $\sigma^2 = \frac{\omega}{1-\alpha-\beta} > 0$. Rewriting the unconditional variance to $\omega = \sigma^2(1-\alpha-\beta)$, we may rewrite (10) with $p = 1, q = 1$ as:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

$$\Leftrightarrow \sigma_t^2 = \sigma^2(1-\alpha-\beta) + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

$$\Leftrightarrow \sigma_t^2 = \sigma^2 + \alpha(\varepsilon_{t-1}^2 - \sigma^2) + \beta(\sigma_{t-1}^2 - \sigma^2)$$

From the last equation, it is clear that the conditional variance is given by the unconditional variance plus two "noise" terms adjusted by α and β. Now, we may utilize a forward recursion of the conditional variance equation, as $E[\sigma_{T+1}^2|\mathcal{I}_T]$ can be calculated directly with the information at time T. Doing so yields the forecasting equation for period $T+h$, $h \in \mathbb{Z}$:

$$E[\sigma_{T+h}^2|\mathcal{I}_t] = \omega\sum_{i=0}^{h-1}(\alpha+\beta)^i + (\alpha+\beta)^{h-1}\hat{\sigma}_{T+1}^2 \tag{11}$$

where $\hat{\sigma}_{T+1}^2|\mathcal{I}_t = \hat{\omega} + \hat{\alpha}r_T + \hat{\beta}\sigma_T^2$. From the forecasting equation, we may interpret the parameter sum $\alpha+\beta$ as the speed of convergence towards to unconditional variance. Here, the weakly mixing assumption for the return process is crucial. If $\alpha+\beta \geq 1$, the sum does not converge, which implies that the volatility becomes explosive and the second order moment would be infinite. Moreover, we emphasize that for longer horizons, the GARCH model becomes uninformative as the forecast converges towards the unconditional variance, which is seen by letting h tend to infinity and using $\omega = \sigma^2(1-\alpha-\beta)$, such that we obtain:

$$\begin{aligned} E[\sigma_{T+\infty}^2|\mathcal{I}_t] &= \lim_{h\to\infty}\left[\omega\sum_{i=0}^{h-1}(\alpha+\beta)^i + (\alpha+\beta)^{h-1}\hat{\sigma}_{T+1}^2\right] \\ &= \sigma^2(1-\alpha-\beta)\frac{1}{1-\alpha-\beta} \\ &= \sigma^2 \end{aligned}$$

For the empirical analysis, we are interested in the forecasted annualized volatility 21 trading days ahead. This is therefore, based on equation (11) given by:

$$\sigma_t^* \equiv \sqrt{252}\hat{\sigma}_{T+21} = \sqrt{252\left[\omega\sum_{i=0}^{20}(\alpha+\beta)^i + (\alpha+\beta)^{20}\hat{\sigma}_{T+1}^2\right]} \tag{12}$$

The 21-day realized (annualized) volatility that we use to calculate the performance is given by the equation

$$\tilde{\sigma_t} \equiv \sqrt{252}\frac{1}{21}\sum_{t-20}^{t}|s_t|, \tag{13}$$

where we have used that returns have an unconditional mean of zero (see Table 1) such that the average of the absolute returns are equal to the standard deviation.

Finally, we note that all GARCH estimation and volatility forecasting is conducted within the `arch` library for Python (Kevin Sheppard, 2022). Moreover, misspecifcation analysis is conducted with a combination of the `arch` and the `statsmodels` (Kevin Sheppard, 2022) Python libraries.

3 Data

We use daily historical data for 6 different currencies measured against the US Dollar: Russian Rubles (RUB), South African Rands (ZAR), Brazilian Reals (BRL), Euros (EUR), British Pounds (GBP), and Japenese Yen (JPY). All data is sourced via Bloomberg, and the sample period is January 15, 2000 to February 22, 2022 for all currencies, however, the individual series are only modelled for the periods where both returns and implied volatility are accessible (see Table 1 for the specific ranges). We calculate the log return of each currency pair as:

$$x_t = (s_t - s_{t-1}) \times 100, \tag{14}$$

where $s_t = \log S_t$ with S_t being the nominal exchange rate recorded at the end of the day. Here, we note that we use the denomination that matches that of the implied volatility quotes. For example, GBP volatility is quoted in USD terms, while the ZAR volatility is is quoted in ZAR terms. However, since we evaluate profitability in terms of volatility points, the denomination is not important for calculating profit and loss, as long as the quotes of the spot rates match those of the implied volatilities.

For our implied volatility series, we source daily end-of-day quotes from Bloomberg for currency options with exercise day one month ahead, which we approximate by 21 trading days following Dunis and Huang (2002). These series are constructed by backing out the volatility term from the Garman-Kohlhagen currency option pricing formula by using observed market prices as described in Section 2.2. The implied volatility series are quoted in terms of annualized volatility.

To account for trading costs when evaluting trading profits, we extend the initial approach by Dunis and Huang (2002) and define a cost measure c_t given by the bid-ask spread as a percentage of the mid price as:

$$c_t = \frac{ask_t - bid_t}{mid_t} = \frac{ask_t - bid_t}{\frac{ask_t + bid_t}{2}} \tag{15}$$

Table 1 below presents defines the sample and presents the summary statistics.

Table 1: Summary statistics of the log return in USD

	RUB	ZAR	BRL	EUR	JPY	GBP
Mean return (%)	0.018	0.016	0.018	0.002	0.001	-0.003
Std. dev. (%)	0.784	1.076	1.047	0.591	0.602	0.572
Min return (%)	-17.346	-6.630	-10.344	-2.522	-3.782	-8.395
Max return (%)	17.001	15.496	7.112	3.451	5.504	3.001
Skewness	0.555	0.838	0.062	0.042	-0.045	-0.755
Exc. kurt.	85.732	6.938	2.813	-1.210	1.524	7.436
Mean bid-ask (%)	0.106	0.308	0.067	0.025	0.038	0.033
Observations	4429	5740	4800	5725	5740	5740
Sample start	2005-03-03	2003-02-23	2003-10-01	2000-03-15	2000-02-23	2000-02-23
Sample end	2022-02-22	2022-02-22	2022-02-22	2022-02-22	2022-02-22	2022-02-22

Note: EUR and GBP are measured as units of USD per unit of currency, while the opposite is true for the remaining pairs.

We note that mean returns are approximately equal to zero and that sample standard deviations range between 0.6% and 1%. Some large outliers, discussed further below, seem to affect the skewness and kurtosis of the distributions, in particular for the RUB, ZAR, and GBP, who all seem highly fat-tailed compared to the Gaussian distribution. Lastly, we note that the mean bid-ask spreads for the group of emerging market currencies are, on average, higher than those of the developed market currencies.

In Figure 1 below, we plot the individual daily return series as well as the observed market-implied volatilities, from which we get graphical evidence for returns being fat-tailed for all series. Moreover, volatility tends to cluster, which motivates modelling them within a conditionally heteroskedastic framework where the conditional variance is allowed to spike and exert a persistent retraction towards the unconditional variance.

The figure also indicates that currency risk is driven by both common and idiosyncratic factors. The common factor, measured by a general market price of risk, is particularly evident in times of market turmoil. In particular, the global financial crisis of 2008 seems to have driven the spike in conditional variances seen in that period. A key insight from Figure 1 is, however, that the market price of risk differs between emerging and developed markets. Comparing the implied volatility following the crisis between the top and bottom three countries in the figure shows that the level of implied volatility is about twice as high in emerging versus developed currencies. This may be explained by the depth of the markets, and the fact that emerging markets are heavily reliant on both their own and the developed markets business cycles, such that the implied volatility spread between emerging and developed markets may be interpreted as a risk premium for holding emerging versus developed market FX.

In terms of idiosyncratic currency risk, many events can be highlighted: The Russian annexation of Crimea in 2014 spiked USDRUB implied volatility to a level of 70% percent p.a., while the outcome of the Brexit referendum in 2016 spiked the GBPUSD implied volatility to almost 30% p.a. Moreover, since the sample ends just before the Russian invasion of Ukraine, the forward-looking component of implied volatility in the USDRUB pair may be upwards biased in the end of the sample compared to what a GARCH(1,1) would be able to capture. We return to this critical issue in Section 5.

Such rare events are obviously difficult to model ex ante, but allowing for a fat-tailed innovation distribution may alleviate some of the effects from these so-called tail events. Using the Student's t-distribution instead of the Gaussian distribution is supported by the plots in Figure 2 below, which shows the return distributions against the standard normal distribution.

Figure 1: Daily returns and 1-month market-implied volatility

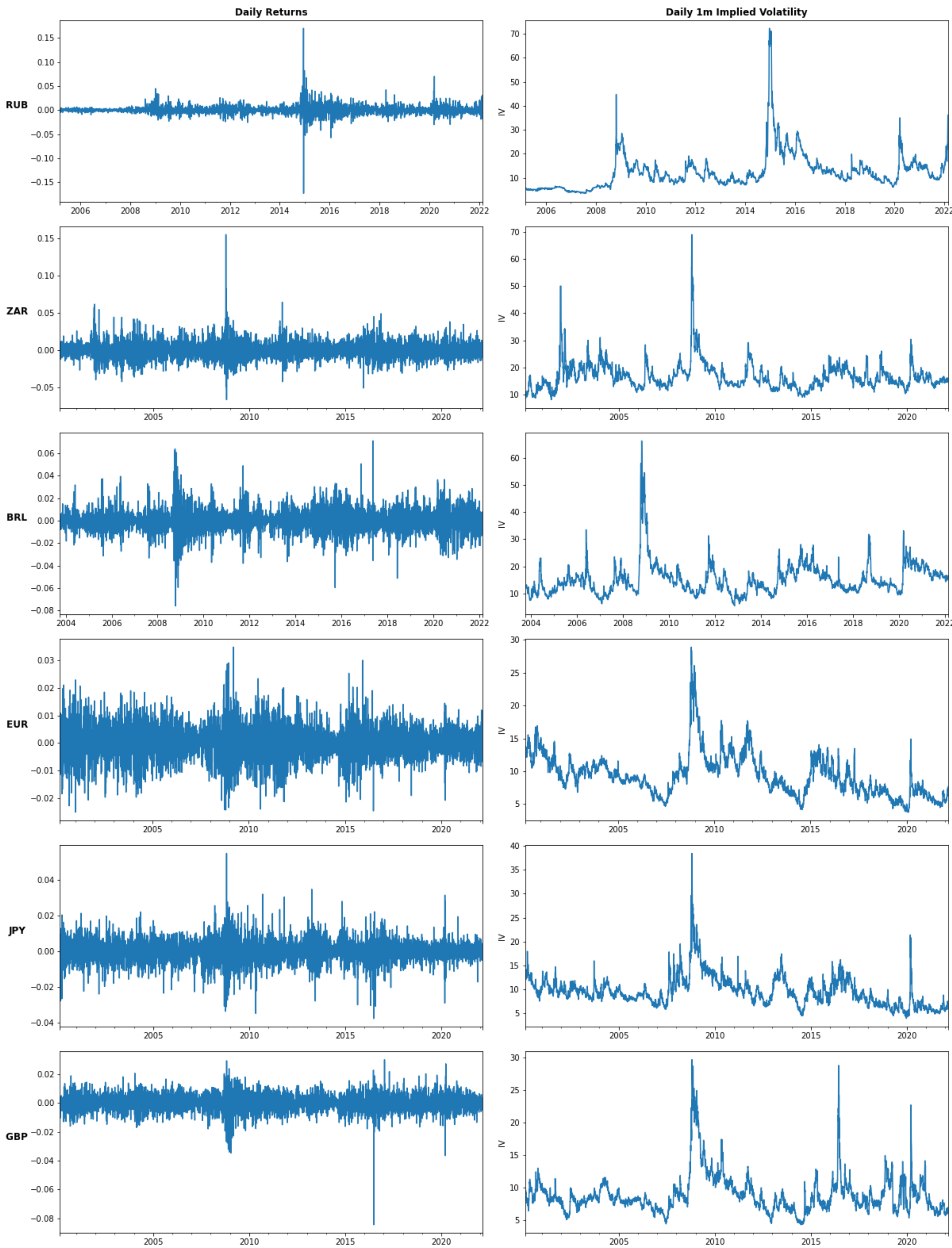
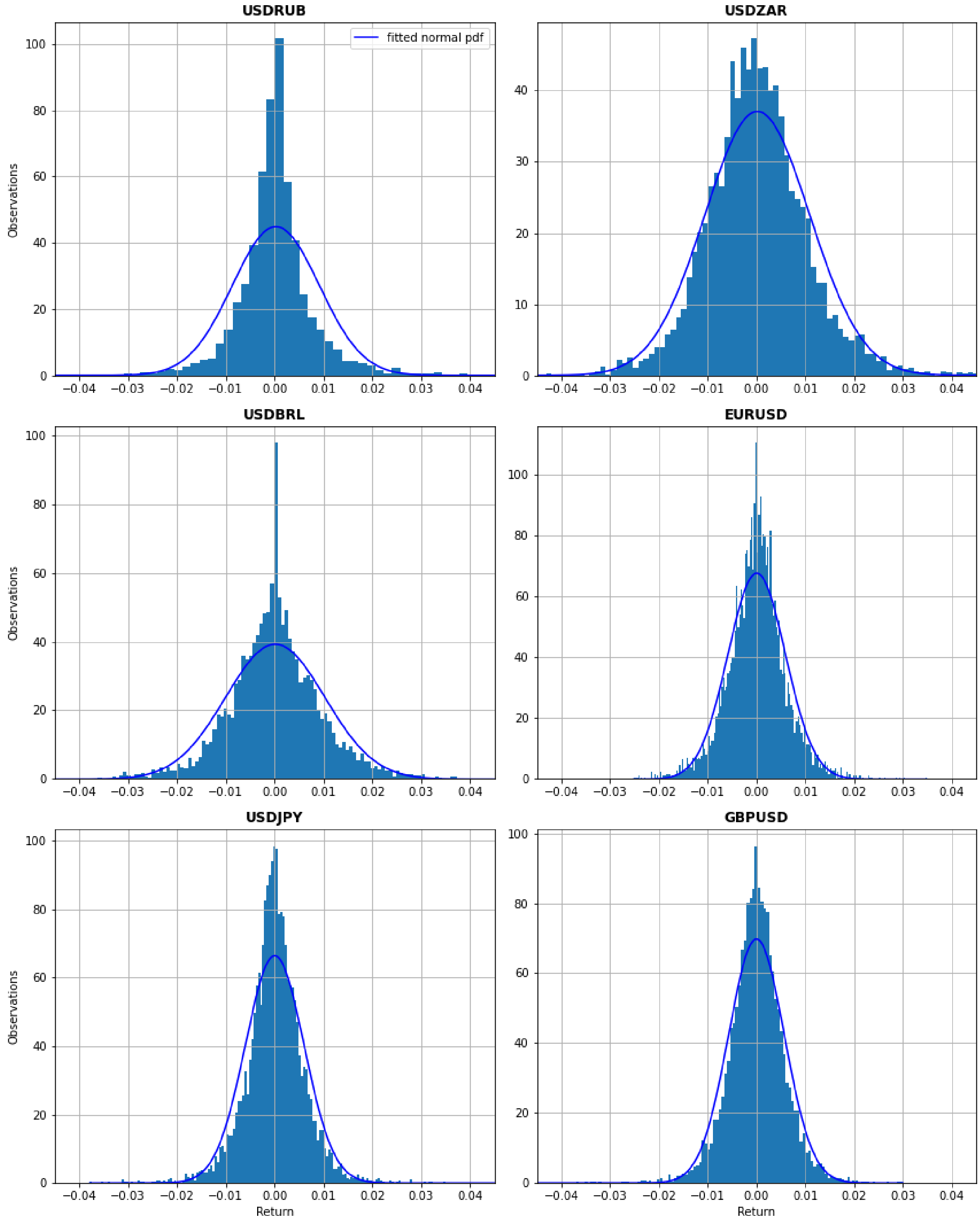


Figure 2: Empirical Return Distributions



Note: 250 bins.

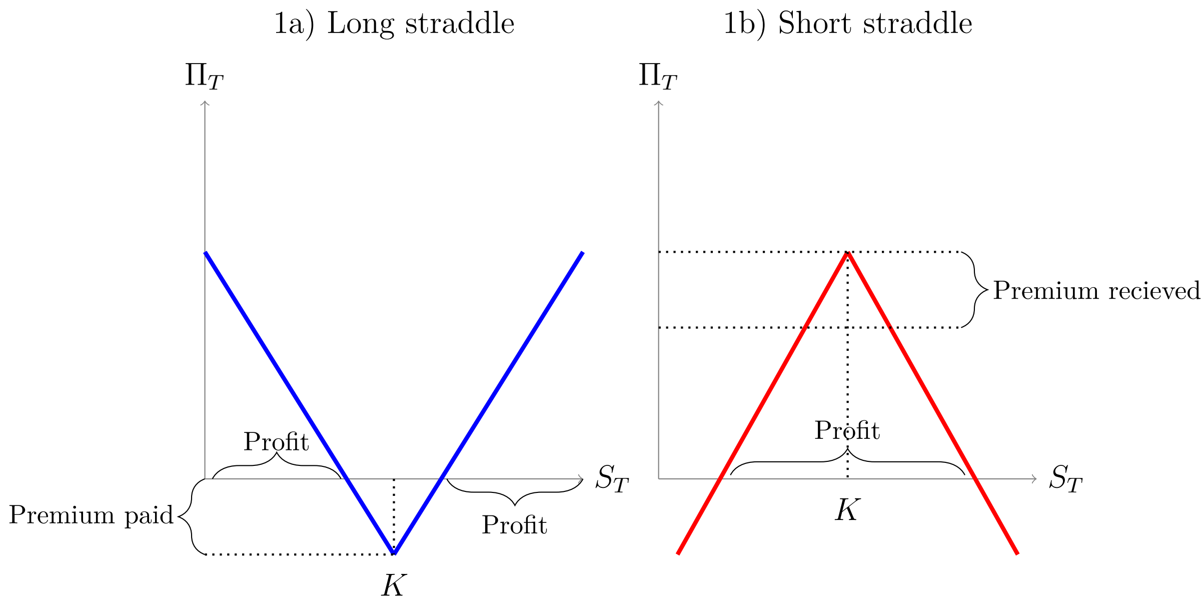
4 Implementation and Empirical Analysis

As established in the preceding sections, the determination of option prices hinges critically on expectations of future volatility. Kroner et al. (1995) point out that more accurate forecasts of volatility can help option traders identify rich and cheap options. This is the main idea behind the volatility trading strategy that we follow based on the methodology proposed by Dunis and Huang (2002): If a given volatility model can produce accurate forecasts of conditional volatility, and the market is inefficient in pricing in future volatility, then a profitable trading strategy emerges where the trader should go long volatility insofar the forecast of future volatility is higher than the prevailing implied volatility in the market and vice versa.

The volatility trading strategy should refrain from being affected by the trend of the underlying exchange rate and should therefore utilize at-the-money (ATM) straddles, i.e. using options that have their strike price equal to the prevailing exchange rate when initiating the trade. A straddle is a combination of a call and a put with the same strike so as to offset and neutralize the opposite deltas and thus hold no forward risk at inception. As established already from the put-call parity in Equation 6, the two ATM components have the same premium at inception. Figure 3 below shows the payoff structure from being long versus short a straddle.

As can be seen from the areas with positive profit in the payoff profiles, offset by the call (put) premium payed (received), a long (short) straddle profits from high (low) price volatility independent of the market direction. Also, as noted in Hull (2021), ATM calls and puts have the same gamma and vega sensitivities, i.e. sensitivities to changes in delta and volatility respectively, and the ATM straddle should therefore be free of directional bias. In addition, using ATM forward volatilities also avoids introducing bias caused from volatility smile effects into our trading strategy.

Figure 3: Payoff structures from long and short at-the-money straddles



4.1 Volatility Trading Methodology

To obtain the trading signals, we carry out 21-days ahead forecasts of conditional volatility on a rolling basis and compare with the prevaling 1-month implied volatility level.[5,6] For a given FX pair, we use the FX returns data until 2021-01-01 to fit the GARCH model for the first day, and then roll the window forward to obtain a total of 298 daily out-of-sample forecasts of volatility 1 month ahead, thus utilizing the recursive structure of the conditional variance equation. In this way, the length of the fitting period may vary between the different pairs based on data-availability, but the out-of-sample forecasting windows are identical.

To compare with the implied volatility, we annualize the conditional volatility forecasts and calculate the "forecast-to-implied"-ratio. To determine when the forecast discrepancy is wide enough to initiate a trade, we define thresholds, or confirmation filters, which we set symmetrically around 1 to act as triggers prior to trading. We analyze different sets of thresholds in the set of 5%, 10%, 20% and 30% away from forecasting parity, such that for, e.g., the 10% threshold a long straddle is entered when the ratio is at or above 1.1, while a short straddle position is initiated at or below 0.9. In this approach we differ from Dunis

[5]The volatility forecasts are obtained using the `arch` python library by Kevin Sheppard.
[6]Assuming 21 business/trading days per month.

and Huang (2002) as they measure forecast discrepancy as the difference in forecasted and implied volatility measured in absolute terms, not in relative terms as in this application. This model choice is based on the fact that the volatilities we model and forecast differ substantially in levels and forecast-to-implied ratios, hence a relative confirmation filter is a more coherent trading rule for our application.

When a trade is initiated it is held to maturity and no other trades are executed during that period. Volatility clustering would cause trading signals to cluster as well, so this restriction prevents the strategy from building up unreasonably large positions in either direction. As we acknowledge that holding to expiry might not be an optimal strategy due to the time-value (or so-called "theta") bleed over the life of the straddle, we also explore an exit rule of closing out the strategy by taking the opposite position and unwind at the prevailing implied volatility market rate after 5 trading days (1 week) adjusting for transaction costs. Naturally, this will allow the strategy to take on more trades during the out-of-sample period.

We note that the straddle position is only forward neutral at inception and that one should optimally adjust the gamma exposure on a continuous basis to keep the position neutral during the period to maturity. For simplicity, we refrain from rebalancing the initial straddle position and let it run for both calls and puts without adjustments noting the asymmetry in terms of loss between the two.

4.2 Performance Calculations

It is complicated to conduct an accurate backtesting procedure of a trading strategy for FX options, and thus to retrieve reliable historical performance measures. This is due to the decentralized structure of trading, that is, retrieving historical quotes across currency crosses, maturities, and strikes is infeasible on a large scale. Therefore, we approximate the return from each trade in terms of volatility points, following the approach of Dunis and Huang (2002). We approximate the net return on a long straddle as the realized volatility at expiry, given by Equation (13) minus the implied volatility at inception, given by the observed market prices, less transaction costs. Conversely, the return on a short straddle is given by the implied volatility at inception minus the realized volatility at expiry less transaction costs. In Section 5 we discuss the precision and potential pitfalls of this approach.

Within the market microstructure research, effective transaction costs in dealer markets are often modelled as functions of different factors, e.g. adverse selection and liquidity, but also price volatility. In Dunis and Huang (2002), all trades are penalized by 25 basis points (0.25%) regardless of the specific currency cross considered or the underlying volatility level. We choose to extend this approach by allowing for currency-specific and time-varying bid-ask spreads as they effectively incorporate the aforementioned factors.

In Figure B.1 in the Appendix, we have plot the individual bid-ask spread series. We note that our approach is an extension to Dunis and Huang (2002), however, it is not flawless; an options trader would not pay the spot market bid-ask spread directly, but the spread should be included in the option premia, hence this approximation should resemble actual trading costs.

In sum, the return at expiry T from each trade based on the straddle strategy is given by:

$$\text{Straddle return} \approx \begin{cases} \tilde{\sigma_T} - \sigma_{IV,T-21} - c_T, & \text{if long} \\ \sigma_{IV,T-21} - \tilde{\sigma_T} - c_T, & \text{if short,} \end{cases}$$

where $\sigma_{IV,T-21}$ is the implied volatility at the inception of the trade and $\tilde{\sigma_T}$ is defined as in Equation (13). In line with Dunis and Huang (2002), we also allow for gearing the position. If the model is accurate, increasing the position size along with the forecast discrepancy should increase expected profits. As such, when leverage is allowed, gross return is magnified by the gearing factor:

$$G^{buy} = \frac{\text{forecast/implied}}{\text{threshold}}, \quad G^{sell} = \frac{\text{threshold}}{\text{forecast/implied}}$$

4.3 Forecasting and Trading Results

4.3.1 Estimation and Forecasting Accuracy

Although the evaluation of the trading strategy should hinge mainly on its ability to generate profits after trading costs, it can be valuable to look at and compare different statistical measures for out-of-sample forecasting accuracy. Following the literature, we use point forecast error measures like the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) to compare 21-days ahead volatility forecasts with realized volatility in the out-of-sample period:

$$RMSE = \sqrt{\frac{1}{n}\sum_{\tau=t+1}^{t+n}(\sigma_t^* - \tilde{\sigma}_t)^2} \tag{16}$$

$$MAE = \frac{1}{n}\sum_{\tau=t+1}^{t+n}|\sigma_t^* - \tilde{\sigma}_t| \tag{17}$$

As is evident from Table 2 in which we report the estimation results, the GARCH(1,1) model is less accurate out of sample for forecasting volatility for emerging markets FX pairs. Noting that RMSE is scale-dependent and thus measured in the same unit as our forecast, our forecasts of volatility in the USDBRL pair is on average more than 6 percentage points away from what turns out realized. Moreover, it is worth noting that the group of emerging market FX pairs are those with the highest forecasting errors.

Table 2: Estimation results

GARCH(1,1)

	RUB	ZAR	BRL	EUR	JPY	GBP
	estimation					
ω	0.000 (2.613)	0.020 (3.484)	0.010 (3.131)	0.001 (2.017)	0.003 (2.932)	0.003 (2.611)
α	0.069 (7.362)	0.060 (6.608)	0.104 (8.357)	0.036 (9.354)	0.057 (7.223)	0.042 (6.185)
β	0.931 (96.456)	0.923 (76.389)	0.892 (72.100)	0.963 (245.854)	0.936 (102.809)	0.951 (115.230)
ν	5.731 (12.292)	10.307 (6.606)	7.680 (8.450)	10.002 (7.917)	5.444 (13.638)	8.8650 (7.744)
	log-likelihood and Akaike information criterion					
Log L	-3565.21	-6628.80	-5720.017	-4450.99	-4465.52	-4260.42
AIC	7138.43	13265.60	11448.03	8909.99	8939.05	8528.85
	out-of-sample forecasting accuracy					
RMSE	3.35	4.71	6.32	2.26	1.59	1.80
MAE	2.73	3.92	5.56	2.05	1.37	1.63

Note: Standard t-statistics in parentheses.
The estimation period is from the sample start date (see Table 1) until 2020-12-31.
We apply a rolling-window estimation, hence only the first window is reported here.

In Tables A.1 and A.2 in the Appendix, we report misspecification tests of no residual ARCH effects and no residual autocorrelation. The tests rejects the null of no ARCH effects at the 5 percent level for the USDRUB and GBPUSD series. Moreover, the null of no residual autocorrelation are rejected for the fifth lag for the USDBRL and the GBPUSD series. Thus, the misspecification tests suggest that there is some degree of residual ARCH and autocorrelation in some of the pairs. However, following Hansen and Lunde (2005), we continue our analysis with the zero mean GARCH(1,1) to consider the trading profitability of the most simple model specification.

Figure 4 illustrates the rolling volatility forecasts obtained with the GARCH(1,1) models against the prevailing realized volatility.[7] From the figure, we note that our GARCH(1,1) forecast seems to follow the market implied volatility closely for the USDZAR and the USD-BRL pairs. Interestingly, these are the two pairs with the highest out-of-sample forecasting errors, and thus the FX options market are approximately as imprecise as the GARCH(1,1) when evaluating the forecasts against ”realized” volatility. For the remaining pairs, the GARCH(1,1) forecast is generally lower than what the markets price into options, which may be explained by the forward-looking aspects of market pricing versus the backward-looking nature of the GARCH(1,1). In Section 4.3.2, we return to this issue with the example of the Russian Ruble.

[7] In order to align the backward-looking nature of realized volatility with the forward-looking volatility forecasts, the realized volatility series has been shifted backwards 21 trading days.

Figure 4: Out-of-sample Volatility Forecasts

4.3.2 Trading Profitability

Tables 3 and 4 present the profit and losses (PnL) and number of trades for the hold-to-expiry volatility strategy over different FX pairs and thresholds with and without gearing. Interestingly, the strategy almost consistently generates losss apart from in the USDRUB pair, where the returns seem extremely skewed to the positive side. This suggests that the market is at least more efficient in pricing in future volatility compared to the simple GARCH(1,1) model.

Table 3: Cumulative PnL from holding to maturity

Forecast discrepancy	RUB	ZAR	BRL	EUR	JPY	GBP
±5% (1.05/0.95)	83.96% (13)	-4.36% (12)	-15.13% (13)	-8.25% (12)	-21.74% (13)	-12.00% (12)
±10% (1.10/0.90)	87.39% (13)	-9.50% (7)	16.98% (10)	-8.51% (12)	-19.07% (13)	-9.24% (9)
±20% (1.20/0.80)	71.07% (11)	-0.74% (2)	-7.88% (4)	-6.18% (6)	-14.73% (9)	-5.80% (4)
±30% (1.30/0.70)	10.97% (4)	N/A (0)	2.99% (1)	-1.82% (1)	-7.94% (6)	-0.78% (1)

Note: Number of trades in parentheses.

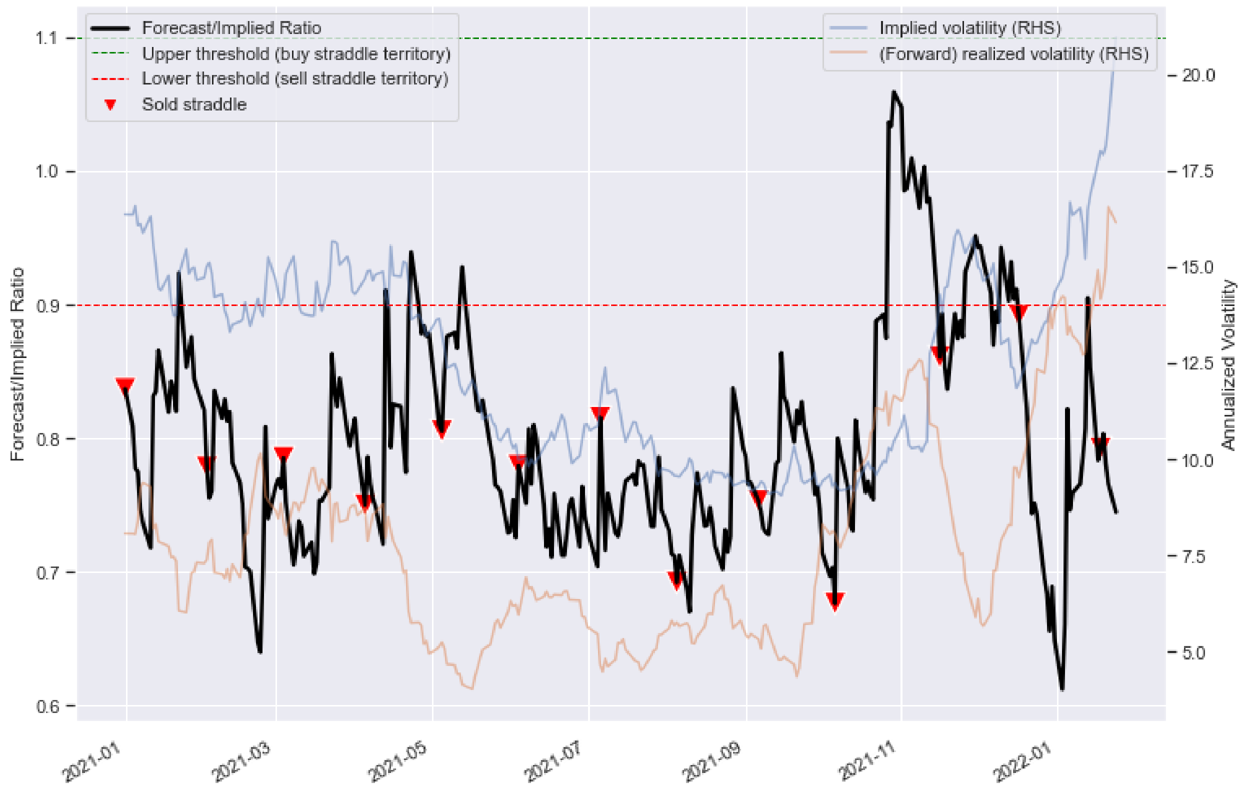
Table 4: Cumulative PnL from holding to maturity (with gearing)

Forecast discrepancy	RUB	ZAR	BRL	EUR	JPY	GBP
±5% (1.05/0.95)	106.73% (13)	-4.35% (12)	-15.81% (13)	-9.21% (12)	-24.74% (13)	-13.18% (12)
±10% (1.10/0.90)	110.64% (13)	-9.96% (7)	18.94% (10)	-9.71% (12)	-21.89% (13)	-10.43% (9)
±20% (1.20/0.80)	89.96% (11)	-0.66% (2)	-9.22% (4)	-7.30% (6)	-17.29% (9)	-6.74% (4)
±30% (1.30/0.70)	13.91% (4)	N/A (0)	3.74% (1)	-2.26% (1)	-9.75% (6)	-0.88% (1)

Note: Number of trades in parentheses.

In general, it seems that cumulative profits are less negative the larger the required forecasting discrepancy required for trading is. This would have been comforting if it was not from the fact that in general mean returns are actually slightly more negative, but compounded fewer times.

Figure 5: USDRUB forecast/implied-ratio and trading signals



Note: Hold-to-expiry strategy with forecast discrepancy of $\pm 10\%$. Arrows indicate the time of trade with the PnL coming from the difference in volatility points between implied and realized volatility (rhs).

It also appears that allowing the position to be geared according to the forecasting discrepancy, i.e. the strength of the signal, has the unintended effect of magnifying negative cumulative returns.

In order to understand why the strategy is so profitable on USDRUB, we investigate the signals ($\pm 10\%$ forecast discrepancy) on which trades were executed and compare against the realized volatility. As can be seen from Figure 5 above, the model forecasts of volatility are much lower than what is priced in the market and the strategy thus effectively sells volatility throughout the entire trading window. Figure 5 shows that realized volatility (shifted to match the forward looking nature of the 1m implied volatility), which the strategy effectively pays, consistently lies below the implied volatility (that is effectively received) and as such all of the 13 trades are profitable resulting in an impressive cumulative return.

Another result that stands out is the positive cumulative return on USDBRL for the $\pm 10\%$ forecast discrepancy threshold. Investigating the implied vs. realized volatility in

Figure 4, we see that it would have been profitable to sell volatility throughout the whole period. An investigation of the trading signals generated by the different thresholds suggests that the trading rule of not initiating any new trades in the period until expiry introduces some randomness in terms of timing. As such, the ±10% threshold arbitrarily results in selling volatility more often compared with the ±5% and ±20% thresholds, and for this threshold returns have fewer drawdowns caused by not buying volatility as many times.

To make sure the generally negative profits are not entirely driven by the time-decay from holding to expiry, we can compare the cumulative PnL to that of closing out the positions after 5 trading days by taking the opposite direction as displayed in tables 5 and 6.

Table 5: PnL from closing after 5 trading days

Forecast discrepancy	RUB	ZAR	BRL	USD	JPY	USD
±5% (1.05/0.95)	-4.61% (45)	-11.51% (31)	-10.34% (42)	-1.41% (39)	-6.84% (45)	-4.71% (40)
±10% (1.10/0.90)	-5.21% (44)	-7.12% (12)	-5.89% (28)	-2.37% (29)	-5.08% (44)	-3.72% (28)
±20% (1.20/0.80)	-5.28% (32)	-2.43% (2)	0.54% (7)	-0.47% (11)	-5.69% (29)	0.54% (5)
±30% (1.30/0.70)	-4.6% (6)	NA (0)	-0.62% (1)	0.46% (1)	-1.73% (13)	-0.07% (1)

Note: Number of trades in parentheses.

Table 6: PnL from closing after 5 trading days (with gearing)

Forecast discrepancy	RUB	ZAR	BRL	USD	JPY	USD
±5% (1.05/0.95)	-4.66% (45)	-11.11% (31)	-10.91% (42)	-0.88% (39)	-6.64% (45)	-4.25% (40)
±10% (1.10/0.90)	-5.59% (44)	-7.11% (12)	-6.46% (28)	-2.13% (29)	-4.75% (44)	-3.28% (28)
±20% (1.20/0.80)	-5.84% (32)	-2.69% (2)	0.67% (7)	-0.20% (11)	-5.84% (29)	0.78% (5)
±30% (1.30/0.70)	-5.58% (6)	NA (0)	-0.73% (1)	0.58% (1)	-1.46% (13)	0.03% (1)

Note: Number of trades in parentheses.

This strategy, allowing for even more trades during the trading period, is found to be unprofitable as well. The cumulative profits here also support the claim that it is indeed the hold-to-expiry trading rule that arbitrarily caused the cumulative profit divergence between triggers resulting from different thresholds.

5 Discussion

As shown in Section 4, the proposed trading strategy did not create consistent returns based on the proposed trading rule. However, in some instances, especially for the Russian Ruble, the cumulative return from holding the straddle trades to maturity were up to 87.39%. As briefly mentioned in Section 5, the market prices a significant premium into options versus what the GARCH(1,1) model would suggest. We interpret this result as market participants pricing options based on forward-looking expectations, not solely on backward-looking persistence in volatility as the GARCH(1,1) does. In particular, note that had the sample been extended a few weeks, then the Russian invasion of Ukraine would have been included in the sample, which yielded a large spike in USDRUB volatility. Thus, market participants may have priced the probability of invasion and full-scale war into options markets, which the GARCH(1,1) fails to capture.

With respect to the overall reliability of the obtained trading results, we note from Figure 4 that implied volatilities are generally significantly above the realized level of volatility, which has a direct effect on the profitability of the trading strategy: If the implied volatility σ_{IV} is always greater than "realized" volatility $\tilde{\sigma}_T$, then a profitable long straddle is inherently difficult to obtain, unless $\tilde{\sigma}_T$ increases sufficiently to offset the general level bias between realized and implied volatility. Thus, the market microstructure seems to affect to potential of the trading strategy in a real-world setting as compared to the case where no risk premium was present.

Dunis and Huang (2002) also make this remark with the possible explanation that "market makers are generally options sellers (whereas end users are more often option buyers): there is probably a tendency among option writers to include a 'risk premium' when pricing volatility. Kroner et al. (1995) suggest another two reasons: (i) the fact that if interest rates are stochastic, then the implied volatility will capture both asset price volatility and interest rate volatility, thus skewing implied volatility upwards, and (ii) the fact that if volatility is stochastic but the option pricing formula is constant, then this additional source of volatility will be picked up by the implied volatility". Thus, the spread between implied and realized volatility may be intepreted as a risk premium.

5.1 Approximating returns and transaction costs

We approximate the return of each long (short) straddle trade as the increase (decrease) in "observed" volatility over the holding period measured in percentage points. As Dunis and Huang (2002) point out, the approximation of returns by volatility points will overestimate the potential losses, since the long straddle has a lower payoff bound given by the premium paid. On the other hand, the losses on a short straddle have no lower bound, and Dunis and Huang (2002) argue that "in a real world environment with proper risk management controls", trades should be closed before losses become too high. This formulation is somewhat vague, however, the essence of their argument is that the directional bias of the approxmiation is mostly negative, such that the trading returns we found in Section 4 should be conservative.

Dunis and Huang (2002) approximate transaction costs for their three pairs of developed market currencies by imposing a constant penalization of 25 basis points per trade. From Figure B.1 in the Appendix, which shows historical bid-ask spreads, we note that the constant 25 basis points penalization is not an unrealistic estimate for EUR, JPY, and GBP, which they consider in their paper. However, their approach fails to capture events with large spikes in bid-ask spreads, which would affect the returns in a real-world setting. Thus, as noted in Section 4, we extend the approach by allowing for time-varying bid-ask spreads based on the prevailing bid and ask prices in the FX spot market at the time of trading the straddle.

A path for further research in this regard, which is however not in the scope of this paper, could be to model and forecast bid-ask spreads as a function of the forecasted volatility from the GARCH forecasting equation. Modelling the bid-ask spread as a function of volatility would be consistent with the market microstructure frictions research.[8] Given such volatility and bid-ask forecasts, the investor could incorporate the conditional expectation of the bid-

[8] See, e.g., Foucault et al. (2013) for an exposition on the topic.

ask spreads on an ex ante basis, and only trade if the forecasted volatility gain is sufficiently large net of transaction costs. Utilizing the transaction costs ex ante has, in other fields of empirical finance research, proven to provide gains in risk-adjusted returns, see inter alia Hautsch and Voigt (2019).

Moreover, calibrating and backtesting a trading strategy with historical data comes with the cost of ”lookback bias”, that is, we risk tuning the model choice, transaction cost definitions, and trading signal generation, in order to maximize ex post returns within the historical trading window. To circumvent this risk, we therefore follow Dunis and Huang (2002) who provide results for the trading strategy for different calibrations of the trading signal generation, i.e., how mispriced the options should be before trading on the signal. Since the return are generally similar in direction and magnitude across calibrations, we are confident that our results reflect the profitability of implementing the strategy in a real setting. However, as we have been trading the straddles in a stylized setting, further specifications of how one would trade in a real setting remain in order to provide more precise estimates.

5.2 Alternative GARCH specification

Besides the performance evaluation methodology discussed in Section 5.1, our results also rests on our choice of GARCH specification and estimation approach. Our argument for choosing the standard GARCH(1,1) is that the literature has yet to come up with a GARCH specification that can outperform the forecasting ability of the GARCH(1,1) as noted by Hansen and Lunde (2005). Nonetheless, the investor’s aim is not to minimize forecasting error, but to maximize returns. Therefore we have repeated the analysis on an alternative GARCH specification, namely the GJR-GARCH(1,1) to check if we can increase returns at the cost of forecasting accuracy. This asymmetric specification allows for negative past innovations to have a bigger impact on current volatility compared to positive innovations, which thereby implies an asymmetric so-called ”news impact curve”. The trading results from applying the model extension can be found in the Appendix.

With the alternative model, the main results are sustained, both in levels and directions for the returns. The only sigificant difference in trading results between the GARCH(1,1) and the GJR-GARCH(1,1), is that we can now obtain consistently positive results for the BRL straddles across forecast discrepancy calibrations. This is in contrast to what we found for the GARCH(1,1), where the returns were switching between positive and negative territory depending on the discrepancy choice.

5.3 Misspecification and evaluating the forecasting accuracy

As outlined in the theoretical section, we choose a zero conditional mean specification, which does not seem to fit all models in terms of the misspecification tests shown in Tables A.2 and A.1 in the Appendix. The zero conditional mean choice is based on the argument that exchange rates, like stock prices, are often argued to follow random walks such that a zero conditional mean for the return is the best predictor. Moreover, Hansen and Lunde (2005) find that no particular mean specification can consistently outperform the simple zero mean GARCH(1,1) in terms of forecasting.

The GJR-GARCH(1,1) suffers from some of the same misspecification flaws as we saw for the GARCH(1,1). Hence, despite the random walk hypothesis and the conclusions of Hansen and Lunde (2005), it may be relevant for further analysis to consider an autoregressive specification for the conditional means to mitigate the degree of autocorrelation present in the current models.

Lastly, we note that evaluating the forecasting ability of the GARCH(1,1) by use of the root mean squared errors, may not be appropriate in the context of forecasting the unobserved variance of returns. This is the case since GARCH models do not forecast future realizations that are observable, but rather non-observable and model-dependent future volatility. As such, the errors are not necessarily well-defined.

6 Conclusion

In this paper, we found that trading on mispricing in currency options markets by forecasting volatility with the GARCH(1,1) model was not a profitable trading strategy for our holding period of January 1, 2021 through February 22, 2022. We conducted a backtest on portfolios of options, known as straddles, for the Russian Ruble, South African Rand, Brazilian Real, Euro, Japanese Yen, and the British Pound, all against the US Dollar. Our conclusion stands in contrast to Dunis and Huang (2002) whose approach we have replicated and augmented by allowing for time-varying transaction costs and by considering four additional currency pairs. These authors apply the same trading strategy for the Japanese Yen and the British Pound, and find consistently positive returns across all model parametrizations.

Our main conclusion is not sensitive to the choice of a symmetric GARCH(1,1) conditional variance specification versus the asymmetric GJR-GARCH(1,1), however, the results for the BRL straddle become consistently positive with the alternative specification. This corroborates what Pilbeam and Langeland (2014) find in their study of the forecasting ability of GARCH-type models versus what market particpants are implicitly pricing into currency options markets. They find that the GARCH(1,1), as well as the GJR-GARCH(1,1) are signficantly inferior to option traders in forecasting volatility. We argue in our paper that the superiority of option traders to time series models stems from their ability to incorporate future volatility shocks in their forecasts, whereas the GARCH framework assumes a zero mean innovation process. This feature of the GARCH makes it inherently backward-looking, which may not be appropriate when discounting future events in FX markets.

References

Bank for International Settlements. Triennial Central Bank Survey - Foreign exchange turnover in April 2019. 2019. URL www.bis.org/statistics/rpfx19.htm.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–657, 1973. ISSN 1537534X. doi: 10.1086/260062.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 4 1986. ISSN 0304-4076. doi: 10.1016/0304-4076(86)90063-1.

C. L. Dunis and X. Huang. Forecasting and trading currency volatility: an application of recurrent neural regression and model combination. *Journal of Forecasting*, 21(5):317–354, 8 2002. ISSN 0277-6693. doi: 10.1002/for.833.

R. F. Engle. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987, 7 1982. ISSN 00129682. doi: 10.2307/1912773.

T. Foucault, M. Pagano, and A. Roell. *Market liquidity : Theory, Evidence, and Policy*. Oxford University Press, 2013. ISBN 0199936242. URL https://books.google.com/books/about/Market_Liquidity.html?id=VO6CtR3lvFkC.

M. B. Garman and S. W. Kohlhagen. Foreign currency option values. *Journal of International Money and Finance*, 2(3):231–237, 1983. ISSN 02615606. doi: 10.1016/S0261-5606(83)80001-1.

L. R. Glosten, R. Jagannathan, and D. E. Runkle. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, 48(5):1779–1801, 1993.

P. R. Hansen and A. Lunde. A Forecast Comparison of Volatility Models: Does Anything Beat a Garch(1,1)?, 12 2005.

N. Hautsch and S. Voigt. Large-scale portfolio allocation under transaction costs and model uncertainty. *Journal of Econometrics*, 212(1):221–240, 9 2019. ISSN 0304-4076. doi: 10.1016/J.JECONOM.2019.04.028.

J. C. Hull. *Options, futures and other derivatives*. Pearson, 11 edition, 2021.

Kevin Sheppard. arch, 2022.

K. F. Kroner, K. P. Kneafsey, and S. Claessens. Forecasting volatility in commodity markets. *Journal of Forecasting*, 14(2):77–95, 1995. ISSN 1099131X. doi: 10.1002/FOR.3980140202.

K. Pilbeam and K. N. Langeland. Forecasting exchange rate volatility: GARCH models versus implied volatility forecasts. *International Economics and Economic Policy 2014 12:1*, 12(1):127–142, 5 2014. ISSN 1612-4812. doi: 10.1007/S10368-014-0289-4. URL `https://link.springer.com/article/10.1007/s10368-014-0289-4`.

M. Verbeek. *A Guide to Modern Econometrics*. Wiley, 4th edition, 2012.

A Misspecification Tests

Table A.1: Engle's test for no ARCH effects in the residuals

GARCH(1,1)

	RUB	ZAR	BRL	EUR	JPY	GBP
Test statistic	24.962	9.390	10.651	7.180	2.089	14.933
P-value	0.000	0.094	0.059	0.208	0.837	0.011

GJR-GARCH(1,1)

$\mathcal{H}_0$	RUB	ZAR	BRL	EUR	JPY	GBP
Test statistic	16.414	6.985	10.552	6.807	2.003	13.560
P-value	0.006	0.222	0.061	0.235	0.849	0.019

Note: $\mathcal{H}_0$: No ARCH effects. The number of lags for the test is 5.
Under $\mathcal{H}_0$, the test statistics are $\chi^2(5)$.

Table A.2: Ljung-Box test for no residual autocorrelation

GARCH(1,1)

Lag	RUB	ZAR	BRL	EUR	JPY	GBP
1	0.484	0.309	0.0	0.370	0.006	0.012
5	0.073	0.528	0.0	0.439	0.090	0.039
7	0.080	0.512	0.0	0.606	0.149	0.091
10	0.000	0.074	0.0	0.392	0.139	0.038

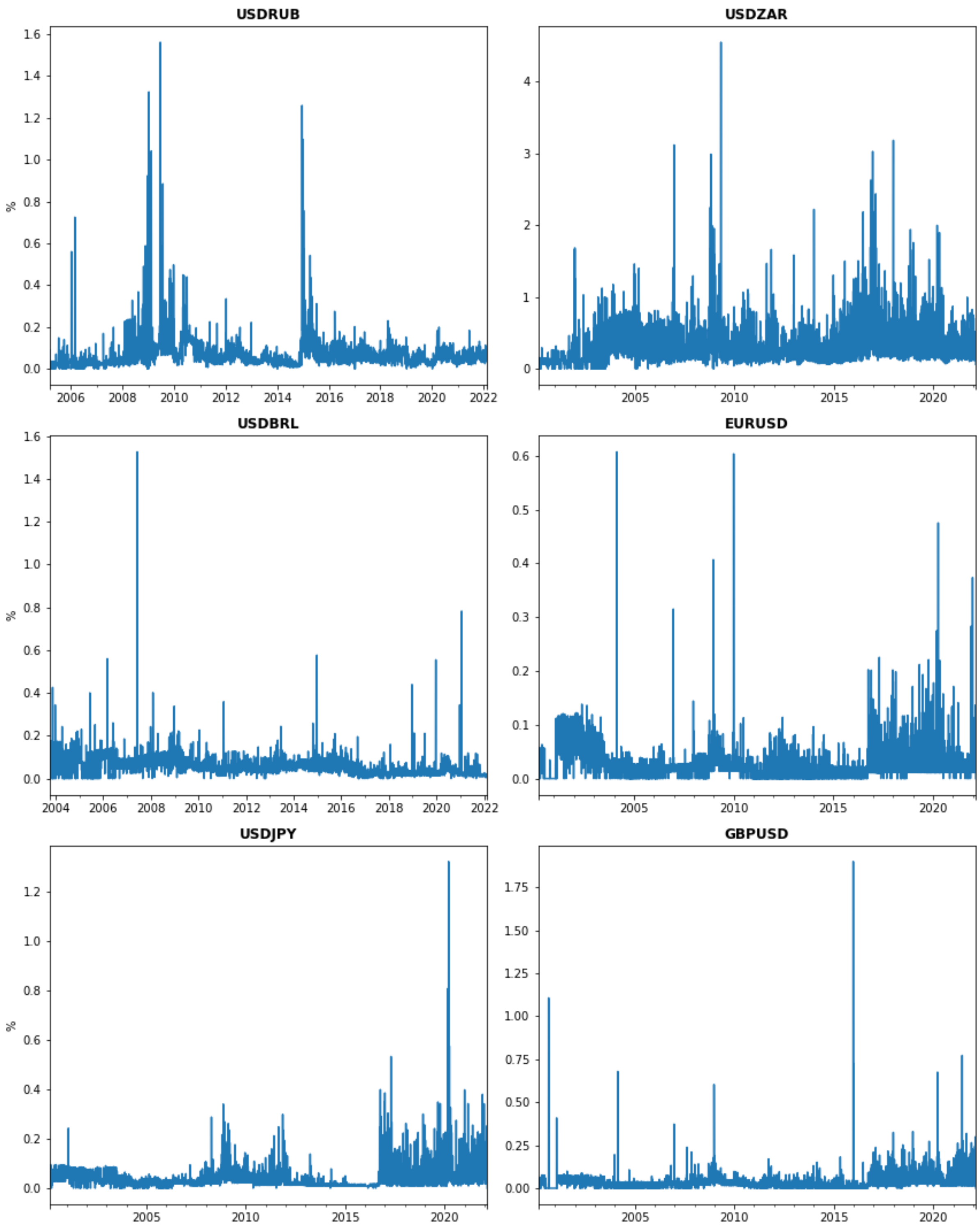
GJR-GARCH(1,1)

Lag	RUB	ZAR	BRL	EUR	JPY	GBP
1	0.484	0.309	0.0	0.370	0.006	0.012
5	0.073	0.528	0.0	0.439	0.090	0.039
7	0.080	0.512	0.0	0.606	0.149	0.091
10	0.000	0.074	0.0	0.392	0.139	0.038

Note: $\mathcal{H}_0$: No residual autocorrelation.
The table shows the p-values for each test and currency pair.

B Bid-ask spreads

Figure B.1: Daily normalized bid-ask spreads



C GJR-GARCH(1,1) estimation and trading results

Table C.1: Estimation results

	GJR-GARCH(1,1)					
	RUB	ZAR	BRL	EUR	JPY	GBP
	estimation					
ω	0.001 (3.117)	0.020 (3.076)	0.011 (3.538)	0.001 (2.011)	0.004 (2.945)	0.003 (2.591)
α	0.086 (6.504)	0.075 (5.723)	0.013 (8.064)	0.030 (5.910)	0.049 (6.705)	0.030 (4.091)
β	0.940 (94.099)	0.928 (68.294)	0.900 (74.743)	0.963 (238.513)	0.933 (95.881)	0.952 (109.142)
γ	-0.052 (-4.997)	-0.044 (-3.674)	-0.085 (-5.705)	0.011 (1.725)	0.020 (1.813)	0.020 (2.416)
ν	5.716 (11.471)	10.466 (6.603)	8.270 (7.842)	10.152 (7.842)	5.493 (13.542)	8.682 (7.610)
	log-likelihood and Akaike information criterion					
Log L	7114.36	13251.65	11413.02	8902.57	8936.99	8523.60
AIC	-3552.18	-6620.82	-5701.51	-4446.29	-4463.50	-4256.78
	out-of-sample forecasting accuracy					
RMSE	3.76	4.46	5.40	2.00	3.02	2.53
MAE	3.14	3.96	4.73	1.79	2.79	2.37

Note: Standard t-statistics in parentheses.
The estimation period is from the sample start date (see Table 1) until 2020-12-31.
We apply a rolling-window estimation, hence only the first window is reported here.

Table C.2: Cumulative PnL from holding to maturity

Forecast discrepancy	RUB	ZAR	BRL	USD	JPY	USD
±5% (1.05/0.95)	87.39% (13)	-8.28% (12)	55.36% (13)	-7.75% (12)	-21.12% (13)	-18.10% (12)
±10% (1.10/0.90)	87.68% (13)	-1.42% (7)	65.84% (12)	-9.17% (12)	-19.07% (13)	-10.28% (9)
±20% (1.20/0.80)	62.35% (11)	0.97% (2)	21.43% (6)	-6.43% (6)	-16.22% (10)	-4.55% (3)
±30% (1.30/0.70)	9.47% (4)	N/A (0)	2.99% (1)	-1.87% (1)	-7.90% (7)	-0.88% (1)

Note: Number of trades in parentheses.

Table C.3: Cumulative PnL from holding to maturity (with gearing)

Forecast discrepancy	RUB	ZAR	BRL	USD	JPY	USD
±5% (1.05/0.95)	107.84% (13)	-8.18% (12)	60.61% (13)	-8.75% (12)	-24.24% (13)	-19.44% (12)
±10% (1.10/0.90)	107.65% (13)	-1.33% (7)	74.46% (12)	-10.39% (12)	-21.92% (13)	-11.50% (9)
±20% (1.20/0.80)	77.61% (11)	1.22% (2)	25.84% (6)	-7.60% (6)	-19.03% (10)	-5.33% (3)
±30% (1.30/0.70)	11.96% (4)	N/A (0)	3.81% (1)	-2.34% (1)	-9.78% (7)	-0.98% (1)

Note: Number of trades in parentheses.