Extremal Forex Returns in Extremely Large Data Sets

MICHEL M. DACOROGNA  
Zurich Insurance Company, Reinsurance  

ULRICH A. MÜLLER  
Olsen and Associates, Research Institute for Applied Economics  

OLIVIER V. PICTET  
Dynamic Asset Management SA  

CASPER G. DE VRIES*  
Erasmus University Rotterdam and Tinbergen Institute  
Email: cdevries@few.eur.nl  

[Received March 24, 2000; Revised and Accepted October 29, 2001]  

Abstract. Exciting information for risk and investment analysis is obtained from an exceptionally large and automatically filtered high frequency data set containing all the forex quote prices on Reuters during a ten-year period. It is shown how the high frequency data improve the efficiency of the tail risk cum loss estimates. We demonstrate theoretically and empirically that the heavy tail feature of foreign exchange rate returns implies that position limits for traders calculated under the industry standard normal model are either not prudent enough, or are overly conservative depending on the time horizon.  

Key words. extreme value theory, regular variation, large data sets, position limit, foreign exchange rates  

AMS Subject Classification. Primary—62G32  62G20  

1. Introduction  

Asset data sets covering a few thousand price observations per contract are commonly used for financial analysis. Over the past decade high frequency data sets containing tick by tick observations have become available, see Baillie and Dacorogna (1997) who devoted an entire journal issue to the topic of high frequency data in finance. Most of the studies in this area focus on facets related to the center of the distribution and the central limit law. This paper, instead, characterizes the distribution of the outliers at the very highest frequency level in an exceptionally large data set, amounting to over 10 million foreign exchange rate price quotes. Handling such a sizeable data set requires techniques which are novel to economics and finance. The size of the data set enables one to

* Corresponding author. Department of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.
accurately determine the probability mass in the tails and demonstrate the uses of extreme value analysis for economic data. We discuss the effects of portfolio diversification under the alternative stochastic models and find that the heavy tails of the typical return distribution induce better diversification properties against the largest risks, than if normality prevailed. Our estimates of the tails are used to calculate reliable overnight position limits for foreign exchange traders over different time horizons. We show theoretically and empirically that extending the time horizon increases the Value-at-Risk (VaR) more rapidly under the normal than under the heavy tail hypothesis, so that the normal model underestimates the short term risk and overestimates longer term risk.\(^1\)

The high frequency data yield a considerable improvement in the efficiency of the position limit estimates. Because outliers by their very nature are rare events, a finely sampled period yields more extreme outcomes (corrected for scale) than if a coarse grid is applied. For a sample of \(n\) random variables \(\varepsilon_i\), we have that

\[
\max\left\{\frac{\varepsilon_1 + \varepsilon_2}{2}, \ldots, \frac{\varepsilon_{n-1} + \varepsilon_n}{2}\right\} \leq \max\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{n-1}, \varepsilon_n\}.
\]

This suggests that there could be a benefit from increasing the sampling frequency for information concerning the tailshape. A well known counter-example is the linear estimator of the average return per unit time interval, which does not benefit from an increase in frequency. Because the tail shape estimator is nonlinear in the underlying data one might conjecture that using higher frequency data improves the fit. In the theory section we give conditions and a formal proof to this conjecture; but we also provide counter-examples.\(^2\) The size of the data set is conducive to implementing the Danielsson et al. (2000) subsample bootstrap procedure for selecting the number of tail area observations which must be used in the estimation.

The structure of the paper is further as follows. The next section starts by deriving the implications of heavy tails for portfolio diversification and time aggregation. The statistical benefits from using high frequency data are discussed in Section 3. Section 4 provides a description of the data and data handling. Empirical results on the shape of the tail, the position limits and time aggregation effects are given in Section 5. The last section concludes.

### 2. Heavy tails and portfolio diversification

We briefly review the statistical properties of the log-returns of foreign exchange rates. The log-return is defined as the natural logarithm of the price ratio \(p_{t+1}/p_t\), where \(p_t\) denotes the spot price of one unit of a foreign currency in terms of the domestic currency units. Under continuous discounting, the log-return is the appropriate measure rather than discrete return \(p_{t+1}/p_t - 1\); moreover for the log-return Jensen’s inequality is not relevant, so that domestic and foreign agents can agree on the moments of the returns. A first property of foreign exchange rate returns (abbreviated as forex returns) is that the mean is approximately zero (when corrected for interest gains\(^3\)). This fair game property is