STATISTICS OF VARIABLES OBSERVED OVER OVERLAPPING INTERVALS

ULRICH A. MÜLLER

UAM.1993-06-18

November 30, 1993

A discussion paper by the O&A Research Group

Contents

Abstract

This working paper discusses the problem of overlapping intervals when computing statistical quantities for time series. For the simple case of the normal distribution, it is shown that the use of overlapping intervals does not improve the precision of the mean, or for that matter of any linear statistical estimator but provides a significant improvement of the precision of the variance or any non-linear statistical quantity.

The improvement of significance is computed in the case of the variance of a normal distribution. The error variance is shown to be reduced by 25% by using an overlap of 2 and at maximum by 33% for an infinite overlap. A formula for the effective number of observations as a function of the number of overlaps is derived from this computation. From this study of the "ideal case" of a normal distribution, recommendations are given for the use of overlapping intervals in empirical studies.

1 Introduction

In many statistical studies of time series, a special type of variables is observed: variables that belong to a *time interval* in the time series. A typical example is price change statistics, such as the analysis of the scaling law, see (Müller et al., 1990). The mean value of a variable, e. g. the absolute value of price changes over an interval of fixed length, has to be determined over a certain sample. The more observations in the sample, the smaller the random error of the mean and the more significant the result.

One way to increase the number of data points is to also consider *overlapping* time intervals, so the distance between the starting points of two subsequently analyzed intervals becomes smaller than the interval size. The obtained series of observations certainly exhibits serial dependence, we cannot regard the observations as independent. However, we intuitively feel that adding overlapping intervals to the sampling scheme might increase the precision of the result. Is this intuitive feeling justified by theory?

The exact answer depends on the statistical properties of the analyzed variable. The problem has already been discussed in (Hansen and Hodrick, 1980), where a method of estimating parameters and their significance limits from overlapping observations has been developed and applied. In (Dunis and Keller, 1993a) and (Dunis and Keller, 1993b), a "panel regression" technique is presented and applied: the overlapping observations are grouped in several non-overlapping series with phase-shifted starting points.

In this document, some simple results for "ideal" time series are presented. There is indeed a gain in significance due to interval overlapping which can be quantified. Under certain specified conditions, the use of our results obtained in an ideal case is recommended also for empirical studies, as an approximation.

2 Basic assumptions

For solving the overlap problem analytically, we use the simplest case of a random process: the Gaussian random walk. Let us assume an i. i. d. random variable x_i which is normally distributed with zero mean and variance one:

$$\mathrm{E}(x_i) \;=\; 0 \;, \quad \mathrm{E}(x_i^2) \;=\; 1 \;, \quad i \;=\; 1, 2, \ldots \tag{2.1}$$

x_i can be seen as the series of first differences of a regularly spaced time series X_i:

$$X_i \;=\; \sum_{j=1}^{i} x_j \;, \quad i \;=\; 0, 1, 2, \ldots \tag{2.2}$$

where X_0 = 0. In most empirical papers, we would use the notation Δx instead of x and x instead of X. For the computations of this document, the notation of eq. 2.2 is more convenient.

Now, an empirical researcher starts his analysis. This empirical researcher does not know the statistical nature of X_i (only *we* do); he just assumes that X_i is stationary. He has one focus of interest: the behavior of X_i *over intervals of a certain fixed size*. Let this size be a whole number m times the basic time interval of the series X_i. The variable analyzed by the researcher is thus

$$y_i \;=\; X_{m+i} - X_i \;=\; \sum_{j=1}^{m} x_{i+j} \;, \quad i \;=\; 0, 1, 2, \ldots \;, \quad m \;\geq\; 1 \tag{2.3}$$

This sum is the *aggregation* of m subsequent x_i values.

If the researcher decides to look at overlapping intervals with an *overlap factor* m, he will consider all observations y_0, y_1, y_2, ... If he rejects overlapping, he will take only y_0, y_m, y_{2m}, ...

3 Special case: the mean of the time series value change

The empirical researcher is determining the statistical properties of N time series observations y_i, where i = 0, 1, ... , $N-1$. One basic result to be computed is the *mean* value of y_i.

3.1 Using overlapping intervals

If the researcher uses overlapping intervals, he obtains the following mean:

$$\mathrm{E}_{\text{overlap}}(y_i) \;=\; \frac{1}{N} \sum_{i=0}^{N-1} y_i \tag{3.1}$$

Using our knowledge of the underlying process, eq. 2.3, we can express this equation in terms of x_i,

$$\mathrm{E}_{\text{overlap}}(y_i) \;=\; \frac{1}{N} \sum_{i=0}^{N-1} \sum_{j=1}^{m} x_{i+j} \tag{3.2}$$

This double sum can be reformulated:

$$\mathrm{E}_{\mathrm{overlap}}(y_i) \;=\; \frac{1}{N}\sum_{i=1}^{r-1}[i\,(x_i + x_{N+m-i})] \;+\; \frac{r}{N}\sum_{i=r}^{R} x_i \;, \tag{3.3}$$

with

$$r \;=\; \min(m,N)\;\;,\quad R \;=\; \max(m,N) \tag{3.4}$$

This empirical mean has a stochastic *error*, a deviation from the theoretical value 0 (which the empirical researcher, unlike us, does not know). The expected variance of this error is

$$\mathrm{E}\{[\mathrm{E}_{\mathrm{overlap}}(y_i) - 0]^2\} \;=\; \mathrm{E}\left\{\{\frac{1}{N}\sum_{i=1}^{r-1}[i\,(x_i + x_{N+m-i})] \;+\; \frac{r}{N}\sum_{i=r}^{R} x_i\}^2\right\} \tag{3.5}$$

Here, a sum over many terms has to be squared. All cross products between *different* terms vanish. They have an expectation of zero, because the x_i with different i are defined to be independent with zero mean. The expectation value of the square terms follows from eq. 2.1: $\mathrm{E}(x_i^2) = 1$. We obtain

$$\mathrm{E}\{[\mathrm{E}_{\mathrm{overlap}}(y_i) - 0]^2\} \;=\; \frac{2}{N^2}\sum_{i=1}^{r-1} i^2 \;+\; \frac{r^2}{N^2}\,(R - r + 1) \tag{3.6}$$

The first term has a sum over i^2 which can be expressed as $(r-1)r(2r-1)/6$. This can be looked up in (Gradshteyn and Ryzhik, 1980), eq. 0.121.2. Inserting this, we obtain the final form of the expected error variance of the mean:

$$\mathrm{E}\{[\mathrm{E}_{\mathrm{overlap}}(y_i) - 0]^2\} \;=\; \frac{r}{N^2}\,(r\,R - \frac{r^2-1}{3}) \tag{3.7}$$

This result has to be compared with the error variance obtained *without* interval overlapping.

3.2 Comparison with non-overlapping intervals

If the researcher rejects overlapping, he can make only n observations, where

$$n \;=\; [(N-1)\ \mathrm{div}\ m] + 1 \tag{3.8}$$

(The div operator returns the maximum integer number that does not exceed the quotient of the two arguments). The mean of y_i without overlapping is

$$\mathrm{E}_{\text{no overlap}}(y_i) \;=\; \frac{1}{n}\sum_{i=0}^{n-1} y_{mi} \;=\; \frac{1}{n}\sum_{i=1}^{mn} x_i \tag{3.9}$$

The last form of this equation is obtained by inserting eq. 2.3.

This empirical mean has also a stochastic *error*, a deviation from the theoretical value 0, whose variance is as follows:

$$\mathrm{E}\{[\mathrm{E}_{\text{no overlap}}(y_i) - 0]^2\} \;=\; \mathrm{E}[(\frac{1}{n}\sum_{i=1}^{mn} x_i)^2] \tag{3.10}$$

Again, a sum over many terms has to be squared; all cross products between different terms vanish, and the expectation of x_i^2 is 1. We obtain

$$\mathrm{E}\{[\mathrm{E}_{\text{no overlap}}(y_i) - 0]^2\} \;=\; \frac{m}{n} \tag{3.11}$$

The two error variances with and without overlapping, eqs. 3.7 and 3.11, can now be compared. The direct comparison becomes especially clear under the (frequently occurring) condition of a not too small sample: $N \gg m$. In this case, $r = m$ and $R = N$ (see eq. 3.4), and eq. 3.8 for n (which contains a "div" operation) can be approximated by

$$n \;\approx\; \frac{N}{m}\,(1 + \frac{m}{2\,N} - \frac{1}{2\,N})\;,\quad \frac{1}{n} \;\approx\; \frac{m}{N}\,(1 - \frac{m}{2\,N} + \frac{1}{2\,N}) \tag{3.12}$$

This approximation *averages* over different cases: sometimes, the whole sample can be exactly covered by non-overlapping intervals, but sometimes the last few x_i cannot be covered by non-overlapping intervals (when $N-1$ is no integer multiple of m). Eq. 3.7 adopts the following form:

$$\mathrm{E}\{[\mathrm{E}_{\text{overlap}}(y_i) - 0]^2\} \;=\; \frac{m}{N^2}(m\,N - \frac{m^2-1}{3}) \;=\; \frac{m^2}{N}\,(1 - \frac{m}{3\,N} + \frac{1}{3\,m\,N}) \tag{3.13}$$

and eq. 3.11 becomes

$$\mathrm{E}\{[\mathrm{E}_{\text{no overlap}}(y_i) - 0]^2\} \;\approx\; \frac{m^2}{N}\,(1 - \frac{m}{2\,N} + \frac{1}{2\,N}) \tag{3.14}$$

Both error variances are essentially m^2/N. The difference between the two results with and without overlapping is tiny, only in the smaller terms which tend to vanish for really large samples. For determining the mean of y_i, the method of interval overlapping only marginally affects the precision of the result.

There is a deeper reason behind this fact: the two operations of *aggregation* ($x_i \rightarrow y_i$) and *taking the mean* ($y_i \rightarrow \mathrm{E}(y_i)$) are essentially identical: both imply a summation over the time series elements. This becomes apparent in the double sum of eq. 3.2. Since summation is a *associative* operation, the two operations collapse into one. This can be illustrated in the simple case $m = 2$, $N = 5$. In this case, eq. 3.2 for the mean *with* overlapping takes the form $[(x_1 + x_2)+ (x_2 + x_3)+ (x_3 + x_4)+ (x_4 + x_5)+ (x_5 + x_6)]/5$. By simple regrouping of the terms, we can write this mean in form of a *non-overlapping* mean: $2/5\,[(x_1 + x_2)+ (x_3 + x_4)+ (x_5 + x_6)]- (x_1 + x_6)/5$. Only the last term with the two edge observations x_1 and x_6 marks the difference.

If determining the mean was the only statistical analysis, the method of overlapping would be superfluous. A reason to apply it arises only when some other statistical variables are also computed. In that case, overlapping is recommended also for computing the mean, in order to be consistent in the sampling of all statistical variables.

4 Exemplary case: the variance of the time series value change

Another basic statistical variable of the time series y_i is its *variance* (about zero or also the sample mean). In the scaling law statistics of price changes (Müller et al., 1990), the variance about zero is directly used to determine the scaling law of the RMS values of price changes. In forecasting, we also determine some variances, e. g. the variance of the forecasting error.

The empirical determination of time series variances is thus a typical problem in our work. An alternative to be studied would be the determination of mean *absolute* values. The statistics of absolute values is, however, known for its bad analytical tractability. There are indications that the results obtained for the significance of empirical variances can also be used for empirical means of absolute values. The analysis of empirical variances is thus regarded as an exemplary case.

4.1 Using overlapping intervals

The variance of y_i about zero *with overlapping* is defined as follows:

$$\mathrm{E}_{\mathrm{overlap}}(y_i^2) \;=\; \frac{1}{N} \sum_{i=0}^{N-1} y_i^2 \;=\; \frac{1}{N} \sum_{i=0}^{N-1} (\sum_{j=1}^{m} x_{i+j})^2 \tag{4.1}$$

where the sample is the same as in eq. 3.1. The expectation value of this empirical variance is, using the fact that all the cross products terms of the squared sum have zero expectation and the square terms an expectation of one,

$$\mathrm{E}[\mathrm{E}_{\mathrm{overlap}}(y_i^2)] \;=\; \frac{1}{N} \sum_{i=0}^{N-1} \sum_{j=1}^{m} x_{i+j}^2 \;=\; m \tag{4.2}$$

The *theoretical* result for the variance of y_i is also m, the overlap factor or degree of aggregation. This ensues from the scaling law of the Gaussian random walk. The empirical result m is thus unbiased.

The empirical variance, however, has a *stochastic error*. Analogously to eq. 3.5, we formulate the variance of this error:

$$\mathrm{E}\{[\mathrm{E}_{\text{overlap}}(y_i^2) - m]^2\} \;=\; \mathrm{E}\{[\mathrm{E}_{\text{overlap}}(y_i^2)]^2\} - 2\,m\,\mathrm{E}[\mathrm{E}_{\text{overlap}}(y_i^2)] + m^2 \;= \tag{4.3}$$

$$\frac{1}{N^2}\,\mathrm{E}\{[\sum_{i=0}^{N-1}(\sum_{j=1}^{m} x_{i+j})^2]^2\} \;-\; m^2$$

where the last form is attained by inserting eqs. 4.1 and 4.2. The further computation of this expression is somewhat tedious because of the two squares. If these squares are explicitly formulated, a long sum of terms is obtained. Each term is a constant times four factors of the type x_i. The expectation values of these terms can be determined according to the following rules:

- x_i^4 has an expectation of 3 for any i. This is the theoretical fourth moment of a Gaussian random variable with zero mean and unit variance.
- x_i^2 has an expectation of 1, the unit variance. This is also true for $x_i^2\,x_j^2$ (if $i \neq j$), thanks to the independence of x_i and x_j.
- Each term that contains a factor x_i to an odd power (x_i or x_i^3) has the expectation zero, for reasons of symmetry and independence of the different x_i. Example: $\mathrm{E}(x_i^2\,x_j\,x_k) = 0$ if $j \neq k$.

By using these rules, the computation becomes tractable.

In the simple, special case $m = 1$, we obtain

$$\mathrm{E}\{[\mathrm{E}_{\text{overlap}}(y_i^2) - m]^2\} \;=\; \frac{1}{N^2}\,\mathrm{E}[(\sum_{i=0}^{N-1} x_{i+1}^2)^2] - 1 \;=\; \frac{3}{N} + \frac{N-1}{N} - 1 \tag{4.4}$$

The first term, $3/N$, summarizes the terms x_i^4 and the second one, $(N-1)/N$, the terms $x_i^2\,x_j^2$. The result is

$$\mathrm{E}\{[\mathrm{E}_{\text{overlap}}(y_i^2) - m]^2\} \;=\; \frac{2}{N}\;, \quad \text{for } m = 1 \tag{4.5}$$

For larger m, the computation is more tedious. For a small number of observations, $N < m$, a different regrouping of terms is needed than for the much more relevant case $N \geq m$. For this latter case, the following formula can be derived:

$$\mathrm{E}\{[\mathrm{E}_{\text{overlap}}(y_i^2) - m]^2\} \;=\; \frac{2\,m\,(2\,m^2+1)}{3\,N}\,[1 - \frac{m\,(m^2-1)}{2\,(2\,m^2+1)\,N}]\;, \quad \text{for } N \;\geq\; m \tag{4.6}$$

This formula has been found with long hand calculations. A study with the *Mathematica* software, conducted by Rakhal D. Davé, has confirmed the result for many values of m and N.

The case $N < m$ is not important enough in practice to be fully analyzed in this document. In a formal publication, however, it should be treated for the sake of completeness.

4.2 Comparison with non-overlapping intervals

If the researcher rejects overlapping, he can make only n observations, where n is defined by eq. 3.8. The empirical variance is then

$$\mathrm{E}_{\text{no overlap}}(y_i^2) = \frac{1}{n} \sum_{i=0}^{n-1} y_{mi}^2 \tag{4.7}$$

This empirical variance has also a stochastic error, a deviation from the theoretical value m, whose variance is as follows:

$$\mathrm{E}\{[\mathrm{E}_{\text{no overlap}}(y_i^2) - m]^2\} = \frac{1}{n^2} \mathrm{E}\{[\sum_{i=0}^{n-1} y_{mi}^2]^2\} \tag{4.8}$$

in analogy to eq. 4.3. This can be computed by inserting eq. 2.3. There is, however, a more direct computation. The case of no overlapping becomes identical to the case $m = 1$ if we take n observations instead of N and account for the fact that the *aggregated* time series y_i has m times the theoretical variance of the series x_i. If we regard an underlying time series with a variance of m instead of 1, the *variance of the error of the variance* is multiplied with the *squared* factor, m^2. By taking eq.4.5, replacing N by n, and multiplying the correction factor m^2, we obtain

$$\mathrm{E}\{[\mathrm{E}_{\text{no overlap}}(y_i^2) - m]^2\} = \frac{2\, m^2}{n} \tag{4.9}$$

The two error variances with and without overlapping, eqs. 4.6 and 4.9, can again be compared. As in the case of the empirical mean, the comparison becomes clear in the frequently occurring case of a not too small sample: $N \gg m$. We use once more the approximative eq. 3.12 to relate n with N. Inserting this, eq. 4.9 adopts the following form:

$$\mathrm{E}\{[\mathrm{E}_{\text{no overlap}}(y_i^2) - m]^2\} \approx \frac{2\, m^3}{N} [1 - \frac{m-1}{2\, N}] \ , \quad \text{for } N \gg m \tag{4.10}$$

This is compared with eq. 4.6. If we neglect the small second term in the square bracket of both equations, the error variance with overlapping is $2\, m\, (2\, m^2 + 1)/(3\, N)$ and that without overlapping $2\, m^3/N$. The ratio of these expressions can be termed the *error variance reduction factor*

$$\frac{\mathrm{E}\{[\mathrm{E}_{\text{overlap}}(y_i^2) - m]^2\}}{\mathrm{E}\{[\mathrm{E}_{\text{no overlap}}(y_i^2) - m]^2\}} \approx \frac{2}{3} + \frac{1}{3\, m^2} \ , \quad \text{for } N \gg m \tag{4.11}$$

Thus, the variance can be reduced to 2/3 of its value without overlapping in the limit of large m. With an overlap factor $m = 2$, the error variance is reduced to 75% of its value without overlapping.

Reducing the error variance of the empirically determined variance to two thirds of its "non-overlapping" value at maximum is a considerable, though not enormous advantage of the method of interval overlapping.

5 The effective number of observations

The comparison between the results with and without overlapping, eqs. 4.6 and 4.10, can be interpreted in another way. We determine the number of non-overlapping intervals necessary to reach the *same* error variance as for N overlapping intervals. This number, the *effective* number of observations, n_{eff}, follows from eqs. 4.6 and 4.9:

$$n_{\text{eff}} \;=\; \frac{3\,m\,N}{2\,m^2+1}\,[1 - \frac{m\,(m^2-1)}{2\,(2\,m^2+1)\,N}] \;,\quad \text{for } N \,\geq\, m \tag{5.1}$$

Having N overlapping observations is equivalent to having n_{eff} non-overlapping observations, as far as the error of the empirical variance is concerned.

The reason for using n_{eff} rather than directly computing the error variance formula is a conjecture: the n_{eff} variable of eq. 5.1 might be far more generally applied than the specific error variance formulas. This n_{eff} might be used also for the error of mean absolute value changes (used in the scaling law of price changes), the covariance or the correlation between *different* time series. The size of the stochastic noise of the continuous UBF forecast optimization (Dacorogna et al., 1992) might also be related to n_{eff}.

The relevance of n_{eff} for covariances between different time series is more than a conjecture, as to be shown below.

6 Confirmation: the covariance of the value changes of two time series

Another important statistical operator is the *covariance* between two time series, y_i and v_i. The test series y_i and its underlying process x_i have been defined by eqs. 2.1 and 2.2. We assume another Gaussian random walk process u_i with the same properties as x_i, but independent. v_i is aggregated from u_i as y_i from x_i, see eq. 2.2. Therefore, y_i and v_i are also independent and their theoretical covariance is zero.

The empirical covariance, using overlapping intervals, can be written

$$\mathrm{E}_{\text{overlap}}(y_i\;v_i) \;=\; \frac{1}{N}\sum_{i=0}^{N-1} y_i\;v_i \;=\; \frac{1}{N}\sum_{i=0}^{N-1}[(\sum_{j=1}^{m} x_{i+j})\;(\sum_{j=1}^{m} u_{i+j})] \tag{6.1}$$

in analogy to eqs. 3.1 and 4.1.

The variance of its stochastic error is

$$\mathrm{E}\{[\mathrm{E}_{\text{overlap}}(y_i\;v_i) - 0]^2\} \;=\; \frac{1}{N^2}\,\mathrm{E}\left\{\sum_{i=0}^{N-1}[(\sum_{j=1}^{m} x_{i+j})\;(\sum_{j=1}^{m} u_{i+j})]^2\right\} \tag{6.2}$$

The empirical covariance and its stochastic error *without* overlapping can be computed from the solution with overlapping by inserting $m = 1$, as it has been done in the derivation of eq. 4.9.

The results for the error variance have been explicitly computed for $m = 1$, 2, and 3. In all these cases, with and without overlapping, the error variance of empirical *covariances* turned out to be exactly *half* that of the empirical *variances* (see eqs. 4.5, 4.6, 4.9, and 4.10). This result might be analyzed and almost certainly confirmed also for larger m.

Thus, the error variance formulas for the variances are not directly usable for covariances. Nevertheless, the formula 5.1 for the effective number n_{eff} of observations remains valid. This is a first example for the stability and usefulness of n_{eff}.

The formula for the reduction of the error variance, eq. 4.11 for $N \gg m$, is also confirmed.

7 Conclusion and recommendation for the usage in practice

After analyzing the behavior of a few statistical operators under "ideal" assumptions, we have found that interval overlapping is a method of reducing the error variance. However, the attainable error reduction is limited even in the case of infinitely dense overlapping.

An exception is the *mean* time series value change, where the error variance cannot be reduced. We have explained this by the fact that aggregating time series value changes from sub-intervals is based on the *same* operation as computing the mean: taking the sum. Thus, using overlapping intervals (which just means a different aggregation scheme) cannot add any information.

It is not clear whether the simple results of this paper have been published before, although the general problem of overlapping observations has found its first solution as early as 1980 (Hansen and Hodrick, 1980). A related problem is addressed by the literature on the *jackknife* method, for example (Yang and Robinson, 1986). This method is based on strongly overlapping *resamples* of the original sample.

The beneficial effect of overlapping is probably lower than in the "ideal" example if there is already considerable serial dependence in the underlying x_i, e. g. if the time series x_i is a computed series that already, implicitly used some overlapping in its computation formula. If, on the other hand, the x_i are known to have negative autocorrelation, as for instance the empirical price changes over intervals of 1 minute, we expect to gain more by using overlapping intervals than in the Gaussian case.

The effect of heteroskedastic and/or leptokurtic times series x_i on the statistics of overlapping intervals is unknown. It is just intuition if we say that the result for the Gaussian case can also be approximately adopted for *moderately* heteroskedastic and leptokurtic processes as they frequently occur in economics and finance.

For time series with a complicated generation process, we can use the method of *Monte-Carlo simulation* to assess the error variance with and without overlapping. Of course, this generation process should be known at least as an approximation.

The use of interval overlapping is generally recommended. In the worst case, we lose nothing, in the best case, we gain an error variance reduction of about one third of the value without overlapping. The choice of the overlap factor m is a trade-off between the error variance reduction, see eq. 4.11 for $N \gg m$, and the computation time that is linearly increasing with m. In the Unbiased Forecaster (UBF) of the OIS, presented in (Dacorogna et al., 1992), our heuristic experience made us choose an overlap factor of $m = 6$ for the continuous model optimization, although the computation times are quite long.

Another, independent reason for interval overlapping arises in *real-time* applications. There, we not only want small errors (low noise) in the statistical average, but we want to be *up to date*. The indicators of the Historical Analysis service of the OIS and their real-time updates, for example, are based on overlapping intervals.

Confidence limits can be derived from error variances, under certain assumptions. We frequently use a 95% confidence range of $\pm 1.96 \sqrt{\text{error variance}}$. The discussion of the problems of confidence limits is not the subject of this document. Once a confidence limit without overlapping is known, the variable n_{eff} of eq.5.1 can be used to compute the corresponding confidence limit *with* overlapping.

References

Dacorogna M. M., Gauvreau C. L., Müller U. A., Olsen R. B., and Pictet O. V., 1992, *Short term forecasting models of foreign exchange rates*, Presentation at the IBM Summer Research Institute in Oberlech Austria on "Advanced Applications in Finance, Investment and Banking" held from July 27 to July 31, 1992. MMD.1992-05-12, Olsen & Associates, Seefeldstrasse 233, 8008 Zürich, Switzerland.

Dunis C. and Keller A., 1993a, *Implied versus historical volatility: an empirical test of the efficiency of the currency options market using non-overlapping data*, Paper presented at the Conference on Financial Markets Dynamics and Forecasting in Paris, 2-4 Sep 1993.

Dunis C. and Keller A., 1993b, *Implied versus historical volatility: an empirical test of theefficiency of the currency options market using non-overlapping data*, Paper presented at the XXXIXth International AEA Conference on Real Time Econometrics in Luxembourg, 14-15 Oct 1993.

Gradshteyn I. S. and Ryzhik I. M., 1980, *Table of integrals, series, and products*, Academic Press, London.

Hansen L. P. and Hodrick R. J., 1980, *Forward exchange rates as optimal predictors of future spot rates: an econometric analysis*, Journal of Political Economy, **88**(51), 829–853.

Müller U. A., Dacorogna M. M., Olsen R. B., Pictet O. V., Schwarz M., and Morgenegg C., 1990, *Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis*, Journal of Banking and Finance, **14**, 1189–1208.

Yang M. C. K. and Robinson D. H., 1986, *Understanding and Learning Statistics by Computer*, volume 4 of *Series in Computer Science*, World Scientific Publishing Co Pte Ltd, Singapore.