Realized Covariance Tick-by-Tick in Presence of Rounded Time Stamps and General Microstructure Effects

Fulvio Corsi and Francesco Audrino

Januar 2008 Discussion Paper no. 2008-04

Department of Economics University of St. Gallen

Editor: Prof. Jörg Baumberger
University of St. Gallen
Department of Economics
Bodanstr. 1
CH-9000 St. Gallen
Phone +41 71 224 22 41
Fax +41 71 224 28 85
Email joerg.baumberger@unisg.ch
Publisher: Department of Economics
University of St. Gallen
Varnbüelstrasse 19
CH-9010 St. Gallen
Phone +41 71 224 23 25
Fax +41 71 224 22 98
Electronic Publication: http://www.vwa.unisg.ch

Realized Covariance Tick-by-Tick in Presence of Rounded Time Stamps and General Microstructure Effects

Fulvio Corsi and Francesco Audrino

Author's address: Dr. Fulvio Corsi
Swiss Finance Institute
Via Buffi 13
CH-6904 Lugano
Tel. +41 58 666 4243
Fax +41 58 666 46 47
Email fulvio.corsi@lu.unisi.ch

Author's address: Prof. Dr. Francesco Audrino
Institute of Mathematics and Statistics
Bodanstrasse 6
9000 St. Gallen
Tel. +41 71 224 2431
Fax +41 71 224 2894
Email francesco.audrino@unisg.ch
Website www. istituti.usilu.net/audrinof/web/

Abstract

This paper presents two classes of tick-by-tick covariance estimators adapted to the case of rounding in the price time stamps to a frequency lower than the typical arrival rate of tick prices. We investigate, through Monte Carlo simulations, the behavior of such estimators under realistic market microstructure conditions analogous to that of the financial data studied in the empirical section; that is, non-synchronous trading, general ARMA structure for microstructure noise, and true lead-lag cross-covariance. Simulation results show the robustness of the proposed tick-by-tick covariance estimators to time stamps rounding, and their overall performance superior to competing covariance estimators under empirically realistic microstructure conditions.

Keywords

High frequency data; Realized covariance; Market microstructure; Bias correction.

JEL Classification

C13; C22; C51; C53

1 Introduction

Asset returns covariance plays a prominent role in many important theoretical as well as practical problems in finance. Analogous to the realized volatility approach (Andersen, Bollerslev, Diebold, and Labys, 2001, 2003,; Barndorff-Nielsen and Shephard, 2001, 2002a, 2002b, 2005; Comte and Renault, 2001), the idea of employing high frequency data in the computation of daily (or lower frequency) covariance between two assets leads to the concept of *realized covariance* (or covariation).

The standard way to compute realized covariance is first to choose a time interval, construct an artificially regularly spaced time series by using an interpolation scheme, and then take the contemporaneous sample covariance of those regularly spaced returns. However, simulations and empirical studies indicate that such a covariance measure is biased toward zero, which rapidly increases with the reduction of the time length of the fixed interval chosen. Similar to the construction of realized volatility, the presence of market microstructure can induce significant bias in the standard realized covariance measure, but the microstructure effects responsible for this bias are different. Bid-ask bouncing, the major source of bias for the realized volatility, will merely increase the variance of the covariance estimator, but it will not induce any bias. On the contrary, the so-called non-synchronous trading effect (Lo and MacKinlay, 1990) strongly affects the estimation of the realized covariance and correlation. Because the sampling from the underlying stochastic process is different for different assets, assuming that two time series are sampled simultaneously (when the sampling is non-synchronous) gives rise to the non-synchronous trading effect. As a result, covariances and correlations measured with high frequency data will possess a bias toward zero which increases as the sampling frequency increases. This dramatic drop of the absolute value of correlations among stocks when increasing the sampling frequency was first reported by Epps (1979). Since then, this effect has been confirmed on real data and simulations by many other authors, such as Dacorogna and Lundin (1999), Renó (2003), and Martens (2004), among others.

This paper instead proposes an unbiased tick-by-tick estimator of realized covariance, and formally analyzes it under the assumption of no microstructure noise (Hayashi and Yoshida, 2005[1]. This estimator was recently investigated by Griffin and Oomen (2006), Palandri (2006),

[1]This also independently appeared (in a much less formalized version) in Martens (2004) as a simple, more

Sheppard (2006), and Voev and Lunde (2007) .

The application of the Hayashi and Yoshida estimator requires precise knowledge of the time stamps of every tick. Unfortunately, the time stamps of tick-by-tick data are often rounded at some minimum time frequency (one second or one minute). Depending on the liquidity of the asset, it is possible (and very likely – especially for the one-minute rounding) that a sequence of successive prices will be recorded with identical time stamps. For example, important databases that use this type of format include the index, bond gold and oil futures from Price-data.com (an example of which is shown in Table 1) and Tickdata.com, and individual stocks, index futures, and options from the Nikkei Needs tick database.

Insert Table 1 about here

For tick-by-tick realized volatility computation, this time stamp rounding does not pose any problems as long as the true order of the prices is preserved. However, for tick-by-tick realized covariance, even if the order of the prices is kept for each asset, the rounding of the time stamps precludes the knowledge of the correct time ordering among the ticks of the two series inside the minimum time interval, which is a necessary condition for the application of the Hayashi and Yoshida tick-by-tick estimator.

This paper proposes two tick-by-tick covariance estimators adapted to the case of rounding in the price time stamps: the first–last and the needlework estimators. We investigate, through Monte Carlo simulations, the behavior of the so modified tick-by-tick estimators under market microstructure conditions analogous to that of the financial data studied in the empirical section of this paper. In particular, we consider non-synchronous trading, general ARMA structure for the microstructure noise, and true lead-lag cross-covariance. Our main finding is that the proposed tick-by-tick estimators clearly outperform several different alternatives proposed in the literature.

The remainder of the paper is organized as follows: Section 2 describes the classical tick-by-tick realized covariance estimator introduced in the literature. Section 3 defines the modified tick-by-tick estimators adapted to the case of rounded time stamps. Section 4 shows the results of realistic Monte Carlo simulations, and Section 5 presents an empirical application to

efficient version of the De Jong and Nijman estimator in the absence of true lead-lag cross-covariance.

a bivariate series of S&P 500 and 30-year US treasury bond futures tick-by-tick data. Section 6 summarizes and concludes.

2 Realized covariance tick-by-tick

Contrary to standard approaches, the tick-by-tick realized covariance estimator does not rely on the construction of a regular grid, because it is based on the whole tick-by-tick raw data series. This approach has the twofold advantage of exploiting all the information available in the data and the ability to avoid the bias toward zero of the realized covariance. The non-synchronous trading effect produces a bias in the usual covariance measure as a consequence of the synchronization of the two series; that is, as a consequence of the construction of a regular grid in physical time.

The bias of the covariance estimator based on fixed interval returns can be seen as arising from two distinct effects. First, the absence of trading on one asset in a certain interval produces a zero return for that interval, and then artificially imposes a zero value to the cross-product of returns producing a bias toward zero in the realized covariance (which, in its standard version, is simply the sum of those cross-products). Second, the construction of a regular grid, depending on the frequency of tick arrivals, affects the computation of the realized covariance. For more liquid assets with higher average arrival rates, the last tick to fall in a certain grid interval is typically much closer to the end point of the grid compared to that of a less liquid asset. Any difference in the time stamps between these last ticks in the grid for the two assets will correspond to a portion of the cross-product returns, which will not be accounted for in the computation of the covariance. This occurs because, for the more liquid asset, the (unobserved) returns corresponding to this time difference will be imputed to the current grid interval; while for the less liquid asset, this portion of returns will be ascribed to the next grid interval, so that the two will no longer be matched and their contribution to the cross-products sum will be lost (see Figure 1). This lost portion of covariance in each interval also produces a downward bias in the realized covariance computed with a regular grid; a bias which will also increase with the number of intervals and hence with the frequency.

Insert Figure 1 about here

Under the assumption of efficient markets (that is, markets that have no leads and lags cross-covariance between the true latent efficient price of the two assets) and no microstructure effects, Hayashi and Yoshida (2005) formally proved that an unbiased and consistent covariance estimator can be computed simply by summing all the cross-products of returns which have a non-zero overlapping of their respective time span. In other words, a given tick-by-tick return on one asset is multiplied with any other tick-by-tick return of the other asset that has a non-zero overlap in time; that is, which share (even for a very small fraction) the same time interval.

Analytically, given a standard continuous time process for asset j:

$$dp_j(t) = \mu_j(t)dt + \sigma_j(t)dW_j(t) \tag{1}$$

and discrete price observations $\{p_j(n_{j,q})\}_{q=0,1,2,\cdots,M}$ with associated tick returns:

$$r_{j,q} = p_j(n_{j,q}) - p_j(n_{j,q-1}) \tag{2}$$

the Hayashi and Yoshida realized covariance estimator for two asset i and j, and a given time interval t (for example one day), is defined as:

$$RC_t = \sum_{s=1}^{M_{i,t}} \sum_{q=1}^{M_{j,t}} r_{i,s}\, r_{j,q}\, I(\delta_{q,s} > 0) \tag{3}$$

with $M_{i,t} + 1$, and $M_{j,t} + 1$ the total number of ticks on time interval t for asset i and j respectively, $I(\cdot)$ the indicator function, and

$$\delta_{q,s} = \max(0, \min(n_{i,s}, n_{j,q}) - \max(n_{i,s-1}, n_{j,q-1})) \tag{4}$$

the overlap in time between any two tick returns $r_{i,s}$ and $r_{j,q}$.

This estimator is unbiased because no portion of covariance will be lost; while the portion of cross-product that does not overlap will have zero mean. Avoiding the noise and the discarding of price observations caused by the regular grid interpolation, will considerably reduce the variance of the estimator. Nevertheless, in the presence of a fixed amount of market microstructure noise and under the assumption of no cross-correlated noise structures, the estimator in this form will not be consistent because, although unbiased, its variance will diverge as the number of observations tends to infinity[2].

[2]For a simple adjustment, based on sub-sampling and averaging, this estimator is consistent (see Palandri, 2006, and Voev and Lunde, 2007).

3 Realized covariance with rounded time stamps

We propose simple modified tick-by-tick estimators designed to overcome or alleviate the imprecise time stamps problem discussed above that renders the application of the Hayashi and Yoshida realized covariance estimator unfeasible. In order to define it, we first establish the following notation. Time stamps are discretized on a regular grid $\tau = \{h\Delta\}_{h=1,2,\ldots,H}$ where Δ is the rounding frequency (one minute on our data):

$$n_{j,\tau}^{(F)} \equiv \min_q \{n_{j,q} : n_{j,q} \geq \tau\} \quad \text{and} \quad n_{j,\tau}^{(L)} \equiv \max_q \{n_{j,q} : n_{j,q} \leq \tau\} \tag{5}$$

$$p_{j,\tau}^{(F)} \equiv p_j\left(n_{j,\tau}^{(F)}\right) \quad \text{and} \quad p_{j,\tau}^{(L)} \equiv p_j\left(n_{j,\tau}^{(L)}\right) \tag{6}$$

so that $p_{j,\tau}^{(F)}$ is the first and $p_{j,\tau}^{(L)}$ is the last price inside the time interval τ with identical time stamps.

For each asset we also define two tick-types of return series:

$$r_{j,\tau}^{(F)} = p_{j,\tau}^{(F)} - p_{j,\tau-1}^{(F)} \text{ and } r_{j,\tau}^{(L)} = p_{j,\tau}^{(L)} - p_{j,\tau-1}^{(L)} \tag{7}$$

one is constructed only from $p_\tau^{(F)}$ (for any τ) and the other one is constructed only using $p_\tau^{(L)}$. These two returns series can be seen as the returns of two different regular grid interpolation schemes: one using next-tick interpolation $r_{j,\tau}^{(F)}$ and the other employing previous-tick interpolation $r_{j,\tau}^{(L)}$.

3.1 The first-last estimator

The first estimator we propose to overcome the problem due to rounded time stamps in the data combines the two different interpolation schemes (previous and next tick), introduced above, in the following way:

$$FL_t = \frac{1}{2}\left(\sum_{s=1}^{M_{i,t}^{(F)}} \sum_{q=1}^{M_{j,t}^{(L)}} r_{i,s}^{(F)}\, r_{j,q}^{(L)}\, I(\delta_{q,s} > 0) + \sum_{s=1}^{M_{i,t}^{(L)}} \sum_{q=1}^{M_{j,t}^{(F)}} r_{i,s}^{(L)}\, r_{j,q}^{(F)}\, I(\delta_{q,s} > 0)\right). \tag{8}$$

It can be seen as the average of two Hayashi and Yoshida-type of estimators: one applied to the return series constructed with next-tick interpolation (that is, taking the first tick) for the first series, and previous-tick interpolation (which considers the last tick) for the second series (see Figure 2), and doing exactly the contrary for the second estimator. We call this realized covariance estimator the first–last tick-by-tick covariance estimator.

Insert Figure 2 about here

The reason we only consider the first $p_\tau^{(F)}$ and last $p_\tau^{(L)}$ price tick inside each bin is because among these type of ticks we can most reasonably ascertain their correct order (by assuming only that $n_{i,\tau}^{(F)} < n_{j,\tau}^{(L)}$ and $n_{j,\tau}^{(F)} < n_{i,\tau}^{(L)}$) and maximizing the time overlap between the corresponding return intervals.

As long as $n_{i,\tau}^{(F)} < n_{j,\tau}^{(L)}$ and $n_{j,\tau}^{(F)} < n_{i,\tau}^{(L)}$, and assuming no correlated noise structures, the two Hayashi and Yoshida-type of estimators in the first–last are unbiased. Hence, being the average of two unbiased estimators, the first–last estimator is also unbiased. But, although no problems are apparent in the empirical application, in theory this can occur: $n_{i,\tau}^{(F)} > n_{j,\tau}^{(L)}$ or $n_{j,\tau}^{(F)} > n_{i,\tau}^{(L)}$. In this case, the first–last estimator may still suffer a downward bias. A simple solution would be to skip one time stamp. Consequently, however, we would further reduce the frequency and the number of employed returns, diminishing the precision of the estimator. A more efficient, alternative approach to overcome this problem is proposed in the next section.

3.2 The needlework estimator

Consider the "cross-bins" tick returns:

$$r_{j,\tau+1}^{(LF)} = p_{j,\tau+1}^{(F)} - p_{j,\tau}^{(L)} \tag{9}$$

and the standard last-tick interpolation returns $r_{j,\tau}^{(L)}$ to compute:

$$NW_t = \sum_{\tau=1}^{M} r_{i,\tau}^{(L)}\ r_{j,\tau}^{(L)} + r_{i,\tau}^{(L)}\ r_{j,\tau+1}^{(LF)} + r_{j,\tau}^{(L)}\ r_{i,\tau+1}^{(LF)} \tag{10}$$

We call this realized covariance estimator the needlework tick-by-tick covariance estimator. Exactly the same idea can be applied to the first-tick interpolation.

The intuition behind this construction is (see Figure 3):

Insert Figure 3 about here

The lost portion of covariance (given by the time difference of the last ticks in the two series) is considered in one of the two cross-product with the cross-bin returns. The other cross-product will be ineffective, but since we do not know the time order between the last two returns of the two series, both cross-bin returns must be included to ensure coverage of the lost portion of covariance induced by the interpolation. This method may be seen as a general way of correcting standard interpolation scheme, and could then be used whenever one wants to compute standard realized covariance at any given time interval.

Using a sub-sample of the total number of ticks employed by the Hayashi and Yoshida estimator, we can expect the needlework tick-by-tick estimator to be less efficient in the absence of microstructure noise. Nevertheless, this estimator represents a form of sub-sampling of the Hayashi and Yoshida one, which could help in the presence of market microstructure noise (as suggested by Palandri, 2006, and Voev and Lunde, 2007). As our simulation results summarized in the next section show, it could help to correct for the significant downward bias empirically found in the Hayashi and Yoshida estimator for highly liquid assets (see Griffin and Oomen, 2006) produced by the empirical presence of lead-lag cross-covariance (assumed to be zero in Hayashi and Yoshida, 2005).

The efficiency of the proposed estimators in the presence of microstructure noise will depend on the characteristics of the data. Thus, their asymptotic and finite sample properties will be hard to be computed analytically. In order to assess the efficiency of the modified tick-by-tick estimators on empirical data, we perform a simulation study in which the data generating process (DGP) mimics (as close as possible) the econometric properties of the two empirical series we investigate in our real data application. The parameters are chosen to match (as close as possible) the empirical observation frequencies, level of volatilities, noise structures, and intensities.

4 Monte Carlo simulations

In this section we evaluate the performance of different covariance estimators in two different simulation environments. In the first, no lead and lag cross-covariance is allowed (standard setting); in the second, we generalize the simulation conditions to take into account lead-lag cross-covariance also.

4.1 Standard setting

The data generating process we consider here is a Lo and MacKinlay (1990) non-synchronous trading model with a heteroskedastic factor and microstructure noise, calibrated on our empirical data. Our data is a 14 years tick-by-tick bivariate series of S&P 500 and 30-year US treasury bond futures with time stamps rounded to one minute.

The Lo and MacKinlay model defines the true return of asset i as given by a single factor model. Considering two assets, the two return series are then given by:

$$r_{i,t} = \mu_i + \beta_i f_t + \epsilon_{i,t} \qquad i = 1, 2 \tag{11}$$

where β_i is the factor loading of asset i, $\epsilon_{i,t}$ represents the idiosyncratic noise of asset i, and f_t is the zero mean common factor.

Assuming that the idiosyncratic noises $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are mutually uncorrelated, and that both are uncorrelated with the common factor f_t, the true covariance between the two assets is:

$$\sigma_{1,2,t} = \beta_1 \beta_2 \sigma^2_{f,t} \tag{12}$$

where $\sigma^2_{f,t}$ is the variance of the common factor f_t.

In the Lo and MacKinlay model the common factor f is assumed to be a simple homoskedastic process; hence the variance of f is a constant σ^2_f. Consequently, the true covariance also remains constant. In the version adopted here, however, in order to increase the dynamics and realism of the DGP, the common factor f is assumed to follow the stochastic volatility model of Heston (1993), so that the true covariance will also dynamically change over time.

Therefore, the dynamics of the common factor is given by the following continuous time processes:

$$df_t = \left(\mu - \frac{v_t}{2}\right) dt + \sigma_{f,t}\, dB_t \tag{13}$$

$$dv_t = k(\alpha - v_t)dt + \gamma v_t^{1/2} dW_t \tag{14}$$

where $v_t = \sigma^2_{f,t}$ and the initial value v_0 is drawn from the unconditional gamma distribution of v.

In the Lo and MacKinlay model the prices are assumed to be observed with a certain probability $1 - \pi_i$, where π_i is the so-called non-trading probability. We found it more convenient to express the frequency of the price observations in terms of the corresponding average intertrade duration between ticks τ_i.[3]

The values of the parameters are chosen to match (as close as possible) the statistical properties observed in the empirical data. Therefore, with asset 1 mimicking the S&P and asset 2 the US bond, the following configuration of the parameters are chosen: $\tau_1 = 8$ seconds, $\tau_2 = 18$ seconds, an average annualized volatility of about 20 percent for asset 1 and 10 percent for asset 2 and a correlation of 30 percent between the two assets. Time stamps of the observed prices are rounded at the one-minute level.

Each time a price is observed we simulate market microstructure effects by adding a stationary noise component independent from the price process. From the empirical study of the tick-by-tick series of those assets, we found significant departure from the standard independent and identically distributed random variables (i.i.d.) assumption on the structure of the market microstructure noise. In studying the autocorrelation of the tick returns of those series, more complex structures than those of a simple MA(1) expected under the standard i.i.d. assumption were found. We suggest that such autocorrelation patterns of the tick returns could be explained by assuming a more complex ARMA structure for the microstructure noise. The noise structure is closely reproduced (see Figure 4) by introducing an MA(2) for the asset 1, mimicking the S&P 500 with $\theta_1 = 0.85$, $\theta_2 = 0.25$ and a noise-to-signal ratio of 0.45, and a strong oscillatory AR(1) with $\phi_1 = -0.65$ and noise-to-signal ratio of 0.6 for the asset 2 corresponding to the US bond.

Insert Figure 4 about here

In addition to the Hayashi and Yoshida and the two modified tick-by-tick estimators (first–last and needlework), other covariance measures are included for comparison in the simulation study, namely:

[3]For example, a non-trading probability of 90 percent corresponds to an exponential distribution of the intertrade duration with a mean value of 10 seconds. In our data the average intertrade duration is eight seconds for the S&P return series and 18 seconds for the US bond return series.

- The standard realized covariance computed with an interpolated regular grid of one-minute returns.

- The standard realized covariance computed with a fixed return time interval of five minutes.

- The Scholes and William (1977) covariance estimator, which adds to the contemporaneous sample covariance of fixed interval returns, one lead-lag cross-covariance. To improve the performance of this estimator we chose the frequency of the fixed interval returns which provided the best results in terms of the root mean squared errors (RMSE). In our simulation set-up such an optimal frequency was approximately one minute.

- The estimator proposed by Cohen et al. (1983), which is a simple generalization of the Scholes and Williams estimator, where more than one lead and lag are considered. As in Bollerslev and Zhang (2003), we compute the Cohen et al. estimators with 12 leads and lags at the 10-second frequency, which was the better performing choice given our simulation set-up.

- The Lo and MacKinlay estimator, given by:

$$\hat{\sigma}_{1,2} = \frac{1 - \hat{\pi}_1 \hat{\pi}_2}{(1 - \hat{\pi}_1)(1 - \hat{\pi}_2)} \text{Cov}\left[r^s_{1,t}, r^s_{2,t}\right] \tag{15}$$

  where $\text{Cov}\left[r^s_{1,t}, r^s_{2,t}\right]$ is the covariance between the observed one-second returns $r^s_{i,t}$. Contrary to the highly noisy non-trading probability estimation proposed by Lo and MacKinlay, where $\hat{\pi}_1 = \text{Cov}\left[r^s_{1,t}, r^s_{2,t+1}\right] / \text{Cov}\left[r^s_{1,t}, r^s_{2,t}\right]$ ($\hat{\pi}_2$ is defined in an analogous way), we estimate those probabilities by counting the observed number of ticks in each day and dividing that by the total number of seconds in the day.

Figure 5 and Table 2 report the results of the 25,000 simulations.

Insert Table 2 about here

Insert Figure 5 about here

With these observation frequencies the one-minute realized covariance is slightly biased. In contrast, the five-minute realized covariance is unbiased, but has a larger variance. Despite the direct estimation of the non-trading probabilities, the Lo and MacKinlay estimator (though unbiased) is extremely inaccurate, probably because of the significant presence of market microstructure noise. With their carefully chosen frequencies both the Scholes and Williams and the Cohen et al. estimators are almost unbiased and reasonably accurate. However, the better ones are the tick-by-tick covariance estimators with no bias and the smallest dispersion among the estimators considered. With this moderate level of microstructure noise, the first–last and needlework estimators turn out to be less precise than the Hayashi and Yoshida one, which is unfeasible in the presence of rounded time stamps. This loss of efficiency is a direct consequence of the lower number of ticks employed in constructing the estimator. Nevertheless, differences with the Hayashi and Yoshida estimator are small. The first–last and the needlework estimators remain superior to all other feasible classical covariance estimators.

4.2 Generalized setting

An important empirical feature recently found by Griffin and Oomen (2006) is that financial high frequency data (especially more liquid data) show the presence of positive and significant lead and lag cross-covariance. As suggested by these authors, cross-dependence between non-overlapping returns can be due to non-instantaneous price adjustment, meaning that more trades are necessary before prices fully incorporate the new available information. In such cases, the Hayashi and Yoshida estimator becomes downward biased because it neglects portions of cross-dependence which extend beyond the overlapping interval (dependence is assumed to be zero in the Hayashi and Yoshida derivation of their covariance estimator). The non-instantaneous price adjustment interpretation suggests that lead and lags cross-covariance might be better modeled by introducing delay adjustment in the true underlying price process compared to the introduction of cross-dependence in the microstructure noise (which will also suffer from the problem of linking a cross-dependence, which, arguably, exists in physical time with a microstructure noise, which is instead observed under trading time).

In order to reproduce in our data-generating process, we generate leads and lags cross-dependence between the true return processes by introducing strong persistence in the dynamics of the common factor. To this end, the Heston model for f_t is replaced by a simple AR(1) process with a large and positive autoregressive parameter ($\phi_1 = 0.85$). In this simple way, our DGP can now produce a lead lag cross-covariance structure which mimics closely those empirically found in Griffin and Oomen (2006) (see Figure 3).

Insert Figure 3 about here

Table 3 and Figure 4 report the estimation results of the different covariance estimators.

Insert Table 3 about here

Insert Figure 4 about here

In the presence of lead lag cross-covariance, the Lo and MacKinlay estimator, the standard one-minute estimator, and the Hayashi and Yoshida estimator are all severely downward biased. The Hayashi and Yoshida estimator has the smallest dispersion, but the presence of such significant bias substantially increase its RMSE. In contrast, the modified tick-by-tick estimators proposed remain both virtually unbiased and achieve the lowest RMSE values among all the competing estimators (also including the Hayashi and Yoshida one).

4.3 Summarizing the simulation results

The proposed first–last tick-by-tick estimator is the best performing estimator (closely followed by the needlework estimator) among the feasible ones, being the Hayashi and Yoshida estimator unfeasible on tick-by-tick data with rounded time stamps. It also performs favorably compared to the Scholes and Williams and the Cohen et al. estimators, even if their return frequency was chosen according to the simulation settings to give the best results. When lead-lag cross-covariance is introduced, both proposed estimators remain unbiased and clearly outperform even the Hayashi and Yoshida estimator, which under these conditions, becomes severely downward biased.

5 Empirical application

5.1 Data

We now apply the proposed first–last covariance estimator to the tick-by-tick bivariate series of S&P 500 and 30-year US treasury bond futures. The data are from the Price-data.com database. Time stamps are rounded at the one-minute level (see Table 1). The period considered covers January 1990 to October 2003. The time series of the realized daily stock–bond covariances is plotted in Figure 8.

Insert Figure 8 about here

To better appreciate the remarkable difference between the daily realized covariances measured using tick-by-tick data and the standard cross-products of daily returns (the usual, inaccurate proxy for daily covariances in standard multivariate volatility models), both measures are plotted together on the same scale. These differences are of central importance when assessing the fitness of a model for volatilities or correlations (see, for example, Patton, 2006). Using inaccurate proxies for daily covariance may lead to the choice of a less accurate model for forecasting second-order dynamics.

6 Conclusions

We adapted the approach of computing realized covariance with tick-by-tick data to the case where price time stamps are rounded to a frequency lower than the typical arrival frequency of the asset encountered in the real world. Monte Carlo simulations calibrated on realistic conditions such as non-synchronous trading, general ARMA structure for the microstructure noise and true lead-lag cross covariance show that the first–last and needlework tick-by-tick covariance estimators proposed turn out to be the best performing among the realized covariance estimators which can be feasibly computed in presence of rounded time stamps. Because they also remaining largely unbiased in the presence of significant lead-lag cross-covariance, they can even outperform (in terms of RMSE) the Hayashi and Yoshida tick-by-tick estimator, which is very sensitive to cross-dependence and shows substantial downward bias under such conditions.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P.: 2001, The distribution of exchange rate volatility, *Journal of the American Statistical Association* **96**, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P.: 2003, Modeling and forecasting realized volatility, *Econometrica* **71**(2), 579–625.

Barndorff-Nielsen, O. E. and Shephard, N.: 2001, Non-gaussian ornstein-uhlembech-based models and some of their uses in financial economics, *Journal of the Royal Statistical Society, Series B* **63**, 167–241.

Barndorff-Nielsen, O. E. and Shephard, N.: 2002a, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society, Series B* **64**, 253–280.

Barndorff-Nielsen, O. E. and Shephard, N.: 2002b, Estimating quadratic variation using realized variance, *Journal of Applied Econometrics* **17**(5), 457–477.

Barndorff-Nielsen, O. E. and Shephard, N.: 2005, How accurate is the asymptotic approximation to the distribution of realized volatility?, *in* D. W. F. Andrews and J. H. Stock (eds), *Identification and Inference for Econometric Models. A Festschrift in Honour of T.J. Rothenberg*, Cambridge University Press, pp. 306–331.

Bollerslev, T. and Zhang, B. Y. B.: 2003, Measuring and modeling systematic risk in factor pricing models using high-frequency data, *Journal of Empirical Finance* **10**(5), 533–558.

Cohen, K., Hawawini, G. A., Maier, S. F., Schwartz, R. and Whitcomb, D.: 1983, Friction in the trading process and the estimation of systematic risk, *Journal of Financial Economics* **12**, 263–278.

Comte, F. and Renault, E.: 2001, Long memory in continuous time stochastic volatility models, *Mathematical Finance* **8**, 291–323.

Dacorogna, M. and Lundin, M.: 1999, Correlation of high-frequency financial time series, *in* P. Lequeux (ed.), *Financial Markets Tick by Tick*, Wiley & Sons.

Epps, T.: 1979, Comovements in stock prices in the very short run, *Journal of the American Statistical Association* **74**(366), 291–298.

Griffin, J. E. and Oomen, R. C. A.: 2006, Covariance measurement in the presence of non-synchronous trading and market microstructure noise. Unpublished manuscript.

Hayashi, T. and Yoshida, N.: 2005, On covariance estimation of non-synchronously observed diffusion processes, *Bernoulli* **11**(2), 359–379.

Heston, S.: 1993, A closed-form solution for options with stochastic volatility with applications to bonds and currency options, *Review of Financial Studies* **6**, 327–343.

Lo, A. and MacKinlay, A.: 1990, An econometric analysis of nonsynchronous trading, *Journal of Econometrics* **45**, 181–212.

Martens, M.: 2004, Estimating unbiased and precise realized covariances. Social Science Research Network Electronic Library.

Palandri, A.: 2006, Consistent realized covariance for asynchronous observations contaminated by market microstructure noise. Unpublished manuscript.

Patton, A.: 2006, Volatility forecast comparison using imperfect volatility proxies. Unpublished manuscript, London School of Economics.

Renó, R.: 2003, A closer look at the epps effect, *International Journal of Theoretical and Applied Finance* **6**(1), 87–102.

Scholes, M. and Williams, J.: 1977, Estimating betas from nonsynchronous data, *Journal of Financial Economics* **5**, 181–212.

Sheppard, K.: 2006, Realized covariance and scrambling. Unpublished manuscript.

Valeri, V. and Lunde, A.: 2007, Integrated covariance estimation using high-frequency data in the presence of noise, *Journal of Financial Econometrics* **5**, 68–104.

S&P 500 data sample

Date	Time	Name	Price
20030724	1322	SP_03U	994.3
20030724	1322	SP_03U,	994.5
20030724	1322	SP_03U	994.7
20030724	1322	SP_03U	994.8
20030724	1322	SP_03U	994.9
20030724	1322	SP_03U	995.0
20030724	1322	SP_03U	995.2
20030724	1322	SP_03U	995.1
20030724	1322	SP_03U	995.0
20030724	1323	SP_03U	994.8
20030724	1323	SP_03U	994.7
20030724	1323	SP_03U	994.6
20030724	1323	SP_03U	994.5
20030724	1324	SP_03U	994.3
20030724	1324	SP_03U	994.2
20030724	1324	SP_03U	994.0

Table 1: Small data sample example for the S&P 500 index future on the 24 July 2003 (20030724). The data are from the Price-data.com database. Time is written in the format hours minutes (hhmm).

Calibrated S&P US bond simulation results: standard setting			
	bias	std	RMSE
1-min no correction	-0.1764	0.2732	0.3252
5-min no correction	-0.0374	0.3392	0.3413
Scholes and Williams Cov	-0.0073	0.2592	0.2593
10-sec Cohen 12 leads-lags	-0.0057	0.2719	0.2719
first–last tick	-0.0032	0.1973	0.1973
needlework	-0.0042	0.2060	0.2060
Hayashi and Yoshida	0.0002	0.1602	0.1602
Lo and MacKinlay	-0.0009	0.8090	0.8090

Table 2: Mean, standard deviation and root mean squared errors (RMSE) of the estimation errors on the annualized covariance for a simulation set-up which reproduces the statistical properties of the S&P 500 and US bond future data.

Calibrated S&P US bond simulation results: generalized setting

	bias	std	RMSE
1-min no correction	-0.4698	0.1710	0.5000
5-min no correction	-0.0802	0.4405	0.4478
Scholes and Williams Cov	0.0111	0.2945	0.2947
10-sec Cohen 12 leads-lags	0.0110	0.3158	0.3159
first–last tick	0.0018	0.1938	0.1938
needlework	-0.0081	0.2046	0.2047
Hayashi and Yoshida	-0.2077	0.1513	0.2570
Lo and MacKinlay	-1.3238	0.7719	1.5324

Table 3: Mean, standard deviation and root mean squared errors (RMSE) of the estimation errors on the annualized covariance for a simulation set up which reproduces the statistical properties of the S&P 500 and US bond future data, including lead-lag cross-dependence.

Graphical illustration of the Epps effect

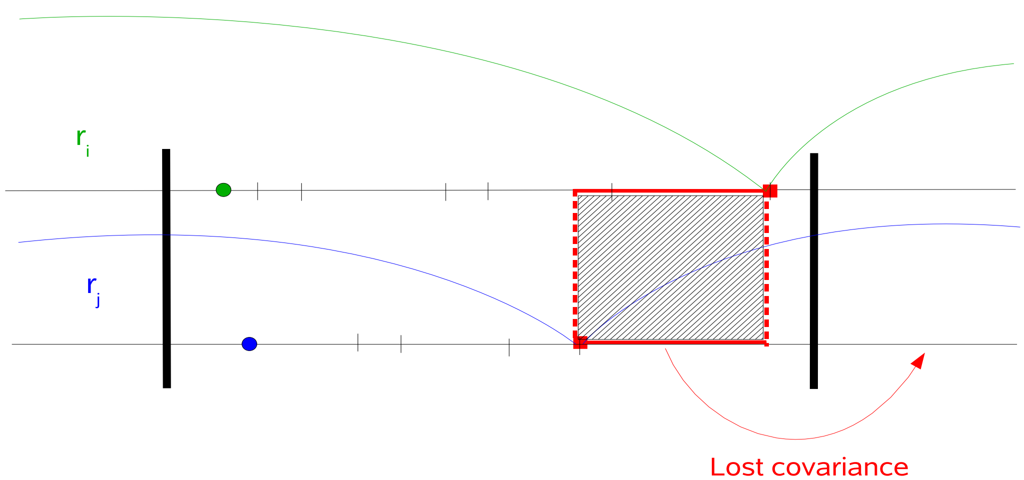

Figure 1: Graphical representation of the Epps effect arising from standard covariance estimators when returns are interpolated on a regular grid (vertical lines) using previous tick interpolation, that is considering the last price tick of the interval (represented with a small square) as the price prevailing at the end of the grid. The shaded area is the portion of covariance lost because of the wrong imputation of a portion of the second asset return to the next interval.

Graphical illustration of the first–last covariance estimator

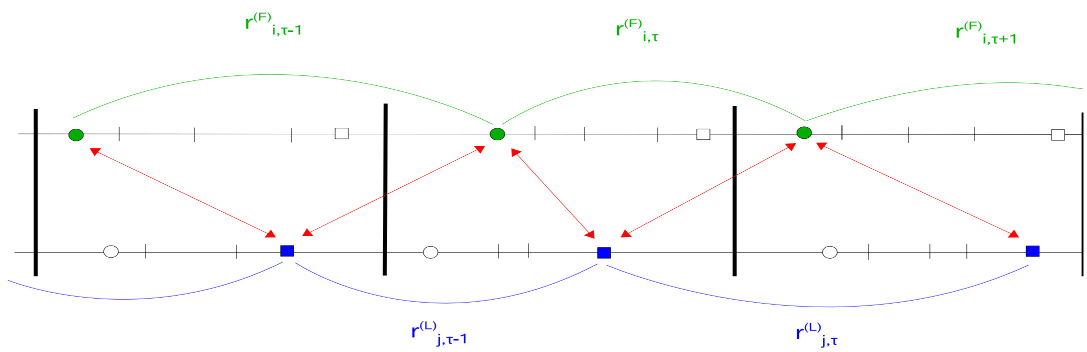

Figure 2: Graphical representation of the first component $\sum_{s=1}^{M_{i,t}^{(F)}} \sum_{q=1}^{M_{j,t}^{(L)}} r_{i,s}^{(F)} r_{j,q}^{(L)} I(\delta_{q,s} > 0)$ of the first–last covariance estimator, where the first asset i is interpolated with next-tick interpolation (small circles represent the first ticks) and asset j with previous tick interpolation (with last ticks being the small squares).

Graphical illustration of the needlework covariance estimator

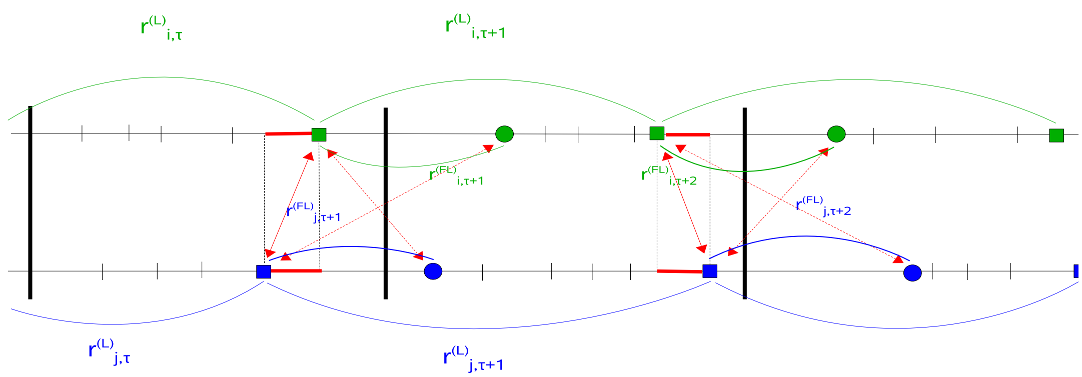

Figure 3: Graphical illustration of the needlework covariance estimator with the two cross-bins returns $r^{(LF)}_{i,\tau+1}$ and $r^{(LF)}_{j,\tau+1}$ (small inner arch) multiplying the last-tick interpolated returns of the other asset in order to correct for the lost portion of covariance induced by previous tick interpolation.

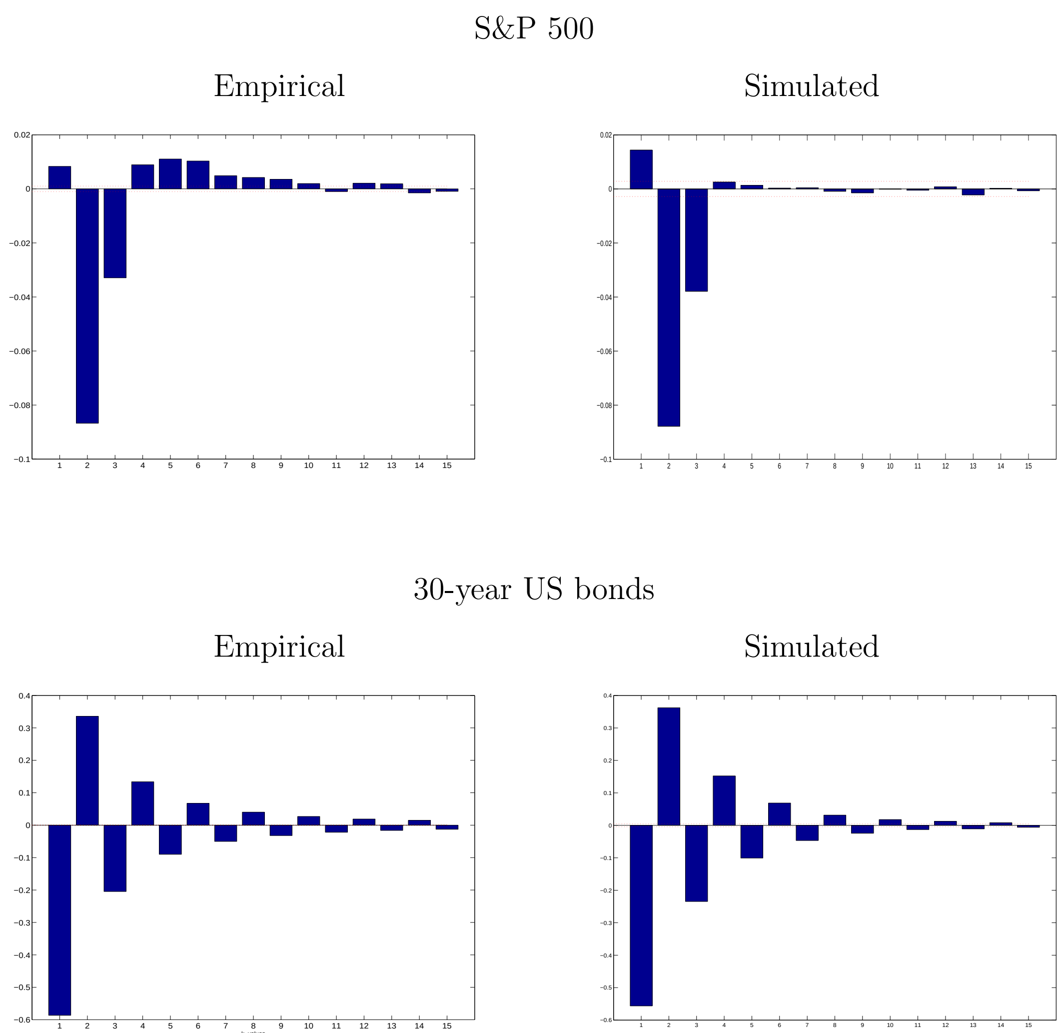

Figure 4: Tick-by-tick autocorrelations of S&P 500 (top) and US bond (bottom) empirical (left) and simulated (right) returns. In order to mimic the dynamics of the S&P 500, an MA(2) microstructure noise has been simulated with $\theta_1 = 0.85$, $\theta_2 = 0.25$ and noise to signal of 0.45. For the same reason, for the US bonds the microstructure noise is an AR(1) with $\phi_1 = -0.65$ and noise to signal of 0.6.

Covariance estimation errors for empirically calibrated simulations

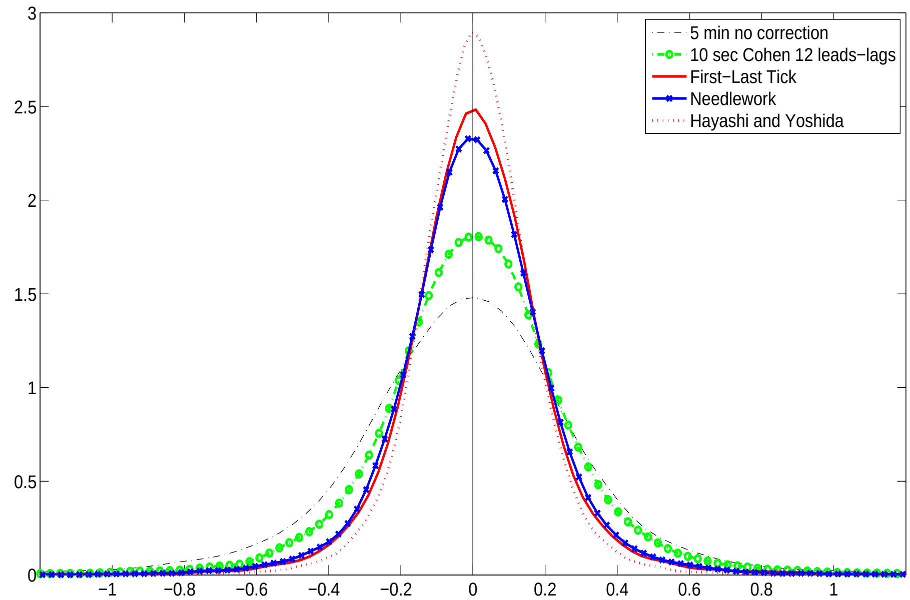

Figure 5: Comparison of the probability density function of the covariance estimation errors for a simulation set-up which reproduces the statistical properties of the S&P 500 and US bond future data. Only the best (in terms of MSE) five estimators are plotted.

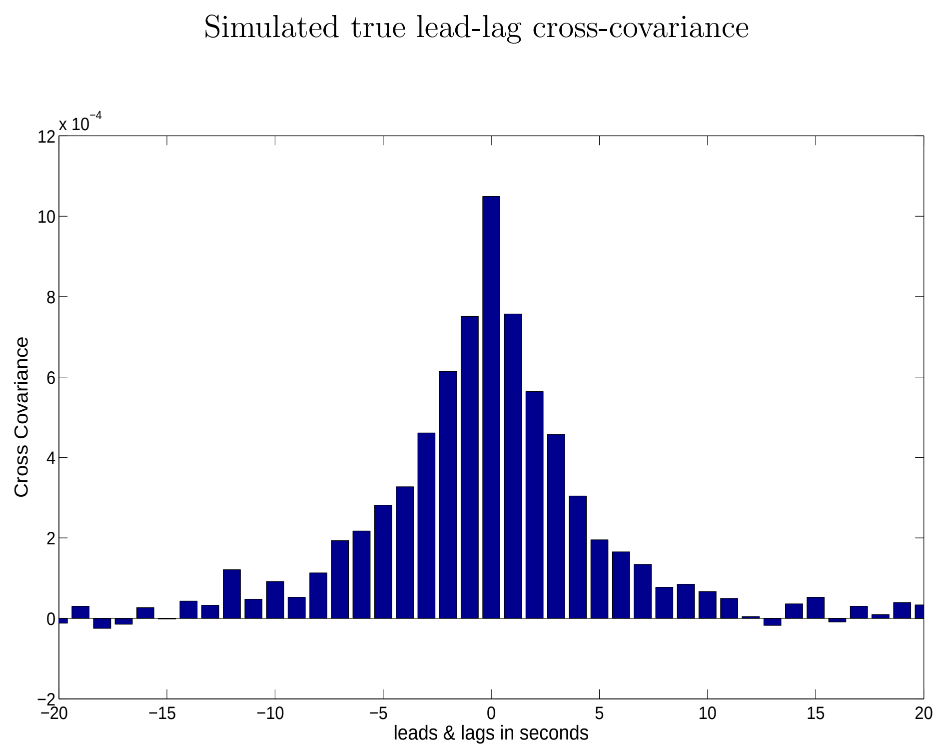

Figure 6: Lead-lag cross-covariances of simulated data obtained from a Lo and MacKinlay (1990) non-synchronous trading model with market microstructure noise and common factor following an AR(1) process with $\phi_1 = 0.85$.

Covariance estimation errors with true lead-lag cross-covariance

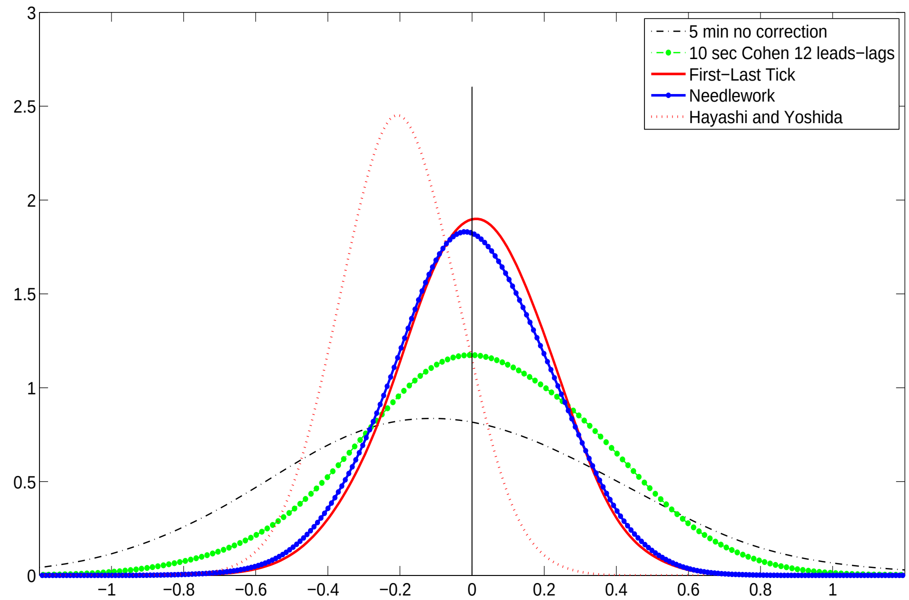

Figure 7: Comparison of the probability density function of the covariance estimation errors for a simulation set-up which reproduces the statistical properties of the S&P 500 and US bond future data and the presence of true lead-lag cross-covariance between the two series. Only the best (in terms of MSE) five estimators are plotted.

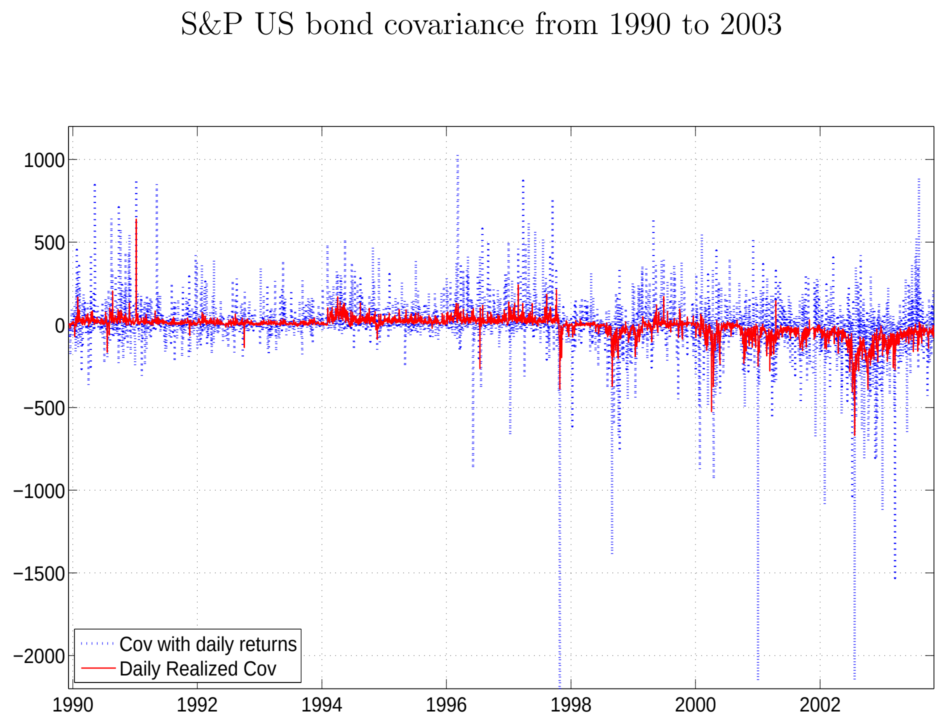

Figure 8: Time series of daily realized covariances constructed using tick-by-tick data (solid line) superimposed on the daily cross-product returns (dotted line) of S&P 500 and 30-year US treasury bonds from 1990 to 2003.