The Telescoping Overlap Problem in Options Data

Bent Jesper Christensen
School of Economics and Management
University of Aarhus, Denmark
Tel: +45 8942 1547

Charlotte Strunk Hansen
School of Economics and Management
University of Aarhus, Denmark
Tel: +45 8942 1548

Nagpurnanand R. Prabhala
Robert H. Smith School of Business
University of Maryland, MD
Tel: (301) 405 2165

First Version: November 4, 1999
Current Version: January 16, 2002

We are grateful to Peter Phillips, Bjorn Tuybens, Doron Avramov (discussant) and participants at the 2001 European Finance Association meeting in Barcelona for useful comments, and to the Danish Social Science Research Council for research support. The data were made available by the Center for Analytical Finance, Aarhus. The paper was formerly titled: 'Accounting for the Overlapping Data Problem in the Implied-Realized Volatility Regression.' To be presented at the 2002 American Finance Association meeting in Atlanta, GA.

Corresponding author: Charlotte S. Hansen, Mailing address: Stern School of Business, New York University, 44 West 4th Street, Suite 7-174, New York, NY 10012, Email: chansen@stern.nyu.edu.
Abstract

In analyzing the relation between implied and realized volatility, researchers confront samples that have a high degree of overlap, which “telescopes” as option maturities are reached. With volatility samples of this nature, we show that the regression coefficients are disperse even in the limit, the $t$-statistics diverge, the Durbin-Watson statistic converges to zero, and the regression $R^2$ converges to a positive random variable in the limit. We develop an alternative asymptotic theory that accounts for both the high degree of overlap and its telescoping nature, and illustrate it empirically with an application to S&P 100 (OEX) index options. Our theory reconciles the seemingly contradictory results from overlapping and non-overlapping samples, and suggests that option markets aggregate volatility information efficiently, as suggested by the (statistically consistent) results from non-overlapping samples.

Keywords: Implied volatility; S&P 100 index options; Market efficiency; Overlapping data

JEL Classifications: C15; C22; C53; G13; G14
1 Introduction

Several papers in the options literature analyze the relation between the implied volatility of an option and the subsequent realized volatility over the option’s life. These studies typically regress subsequent realized volatility on implied volatility and a measure of past realized volatility. A positive and statistically significant coefficient for implied volatility would suggest that implied volatility bears information about subsequent realized volatility beyond that contained in past volatility. Additionally, if options markets are informationally efficient, implied volatility should subsume the information content of past volatility, whose regression coefficient should be indistinguishable from zero. There are perhaps a few dozen published studies that analyze the information content of implied volatility in this framework, examining options on stock indexes, currencies, futures, and individual stocks among other assets, and new evidence continues to appear.1 Virtually all of these studies use samples that have a high degree of overlap. In this paper, we analyze the asymptotic properties of regressions when the underlying volatility series are sampled with a high degree of overlap. We demonstrate that the consequences can be potentially severe: for instance, the regression coefficients do not converge to a fixed number even in the limit, and $t$-statistics diverge. We develop an alternative asymptotic theory for implied volatility regressions in this context and illustrate with an empirical application to the S&P 100 (OEX) index options.

The use of overlapping data in options markets is necessary because of the availability of only a few option maturities per year. For instance, the most liquid option contracts in the US, the OEX index options, expire once a month (on the third Friday), or 12 times a year. Thus, we can obtain at most 12 observations per year if volatility data must be sampled in a non-overlapping manner, with no overlap in time periods covered by successive volatility observations. If more than 12 observations per year are used, the data will involve overlapping samples, with the degree of overlap determined by the frequency of sampling. The more frequent the sampling, the greater the degree of overlap in successive observations. In typical studies where option prices are sampled once a day, overlap can be quite severe.

To illustrate the degree of overlap in daily options data, consider an option expiry date of October 19, 2001 (date $\tau$) and an implied volatility sampled on September 17, 2001 (date $t$). There are 24 trading days between $t$ and $\tau$. The square of the dependent variable (realized volatility) for date $t$, say $\sigma^2_{r,t}$, is to a first approximation the mean of

---

the 24 squared daily returns between \( t \) and \( \tau \)

\[
\sigma_{\tau,t}^2 = \frac{1}{24} \sum_{i=1}^{24} r_{t+i}^2.
\]

The next day squared dependent variable, \( \sigma_{\tau,t+1}^2 \), is the average of 23 of these squared returns from date \((t + 1)\) to date \(\tau\). Thus,

\[
\sigma_{\tau,t+1}^2 = \frac{1}{23} \sum_{i=2}^{24} r_{t+i}^2.
\]

23 out of the 24 squared returns used for \( \sigma_{\tau,t}^2 \) are used for the next observation in the series, i.e., \( \sigma_{\tau,t+1}^2 \). Likewise, the next date squared dependent variable, \( \sigma_{\tau,t+2}^2 \), uses 22 out of the 23 observations used in computing \( \sigma_{\tau,t+1}^2 \) and 22 out of 24 observations used in computing \( \sigma_{\tau,t}^2 \). Thus, the degree of overlap between successive realized volatility observations is quite extreme, as each realized volatility shares \( n \) of the \( n + 1 \) returns used for the previous day estimate, and the difference between two successive observations is only of the order of \( \frac{1}{n} \) times a single day’s squared return. The overlap is telescoping because the number of return observations used to estimate ex-post volatility (\( n \)) gets lower as \( t \to \tau \). A similar extreme degree of overlap characterizes the time series of past volatility, though the telescoping feature is absent here because of fixed time intervals (e.g., 60 days) used to compute these volatilities.\(^2\)

While the use of overlapping volatility series in implied volatility regressions is pervasive, and is indeed necessitated because of the nature of options data, its econometric consequences have not been analyzed. We offer an asymptotic distribution theory for implied volatility regressions with overlapping data, paying particular attention to the peculiar telescoping structure of the overlap and the time series dependence that is so characteristic of financial volatility series. We show that while the use of overlapping data produces more observations for estimating the implied volatility regression, which may increase efficiency, this can cause severe econometric problems as the regression coefficients are no longer consistent. The asymptotic distributions of the regression coefficients do not converge to a fixed value and remain dispersed even in the limit, converging in distribution to a limiting random vector. We characterize the statistical distribution of this vector in terms of functionals of Brownian motion.

We also show that the regression statistics commonly used for statistical inference are ill-specified in the implied-realized regressions. The \( t \)-statistics based on OLS standard errors as well as others used in the literature, such as those based on Newey and West (1987), diverge. The Durbin-Watson statistics converge to zero, and the regression \( R^2 \) converges to a random variable whose limiting distribution is also characterized in terms

\(^2\)The telescoping overlap problem arises in all empirical options research including also tests of particular volatility functions and implied binomial trees, e.g. Rubinstein (1994), Bakshi, Cao and Chen (1997), Campa and Chang (1998), Dumas, Fleming and Whaley (1998), and Jackwerth (2000).
of Brownian motion functionals. While the usual procedures for statistical inference are unreliable in light of our results, we show how to draw correct inferences based on the new theory. We illustrate the asymptotic theory through an empirical application to the OEX index options market. The results reconcile the apparent discrepancy between the daily and monthly results reported in this literature.

Besides the literature on implied volatility, our analysis also generalizes results on stock return predictability using overlapping data. Richardson and Stock (1989) – henceforth RS – analyze the regression of a long horizon stock return on its lagged value when the horizon covered by stock returns (a few years) well exceeds the interval between successive returns (up to one year). We generalize the RS analysis along three important dimensions. First, while stock returns are serially uncorrelated under the null hypothesis in RS, our framework allows for temporal dependence in the underlying series. Absence of correlation is natural in the stock return context studies in RS, as at least some models of market efficiency suggest that stock returns should be serially uncorrelated. However, lack of temporal dependence in volatility is neither required theoretically, nor is it justified empirically given the extensive evidence on time varying volatility from the ARCH/GARCH literature. Accordingly, we develop the asymptotic theory allowing for potential dependence in the underlying series even under the null. A second difference is that while RS examine regressions of long horizon returns on their lagged counterparts, we allow for other predictive variables such as implied volatility that enter as regressors in our context. We also allow for the additional complication that past realized (“historical”) volatility is not just the lagged version of subsequent realized volatility. This is unlike the model studied in RS, where the independent variable is the lagged dependent variable. Finally, our framework allows for and explicitly models the peculiar telescoping structure of the overlap in implied volatility regressions.

The remainder of the paper is organized as follows. Section 2 introduces the nature of the overlap problem and develops the necessary distribution theory for inference in the presence of overlap and temporal dependence in the underlying series. Section 3 illustrates the apparent discrepancies in inferences between implied volatility regressions with daily overlapping data and monthly non-overlapping data, and applies our asymptotic theory for the correct inferences with such data. Section 4 offers conclusions.

2 Asymptotic Theory for Overlapping Options Data

In this section, we establish the required asymptotic theory for the typical overlapping options data sample. For concreteness, we consider the following stylized sampling scheme. Let the time index \( t = 0, 1, ..., T \) indicate the observation times of daily returns, e.g. \( t \) is measured in days for daily sampling, with \( t = 0 \) corresponding to the beginning of the
sampling period, and sample size $T + 1$. Let $T_i$, $i = 1, 2, \ldots$ indicate the option expiration dates in the sample. For example, for OEX options, $T_i$ is the third Friday in month $i$. Assume for simplicity that the intervals between expiration dates are of equal length $K$, i.e. $T_i - T_{i-1} = K$ for all $i$. In the OEX case, $K$ is one month.

Now consider sampling an option price every period $t$ between expiration dates $T_{i-1}$ and $T_i$. More specifically, assume that the price of the near term option expiring at $T_i$ is sampled. As $t$ runs from $T_{i-1}$ to $T_i - 1$, the terms to maturity of the sampled options decline from $K$ at $T_{i-1}$ to 1 at $T_i - 1$. The next option, sampled at $T_i$, again has term to maturity $K$, and between $T_i$ and $T_{i+1} - 1$, a new set of options are sampled, with maturities declining from $K$ to 1. Any two options sampled between $T_{i-1}$ and $T_i$ have lives that overlap, and the period of overlap coincides with the life span of the shorter of the two options. Over the sampling period, option maturity telescopes between $K$ and 1, with a linear decline between expiration dates $T_i$ and a jump to $K$ at each such $T_i$. As the term to maturity telescopes from $K$ to 1, it measures not only the length of the life span of the current option, but also the overlap of this with all other options sampled since the previous expiration date.\(^3\)

Let $J(t)$ denote the term to maturity of the option sampled on date $t$, and hence also the overlap of this option’s life span with those of previous options sampled since the most recent expiration date. Then $J(t)$ is clearly equal to $K$ less the time elapsed since the most recent expiration date. We always initiate sampling on an expiration date ($t = 0$). Hence, if there are $p$ expiration dates in the sample before $t$, then the elapsed time since the last expiration date is $t - pK$, so the time until the next expiration date is $J(t) = K - (t - pK)$. It is useful to note that the number of previous expiration dates in the sample is $p = [t/K]$, where $[\cdot]$ denotes the greatest lesser integer. For example, if $K = 30$ and $t = 140$, then there are $[140/30] = 4$ previous expiration dates in the sample, and $J(t) = 30 - (140 - 4 \cdot 30) = 10$ days to expiration for the current option. The general definition of the time to expiration (and hence overlap) function is therefore

$$J(t) = \left(1 + \left[\frac{t}{K}\right]\right)K - t, \text{ for } t = 1, \ldots, T. \quad (1)$$

Figure 1 exhibits the function $J(t)$.

The one period return $r_t$ on the underlying asset is also sampled in each time period $t$. Realized volatility is computed from squared returns over the life of the current option, and historical volatility from past squared returns (prior to $t$). Implied volatility is backed out from the current option price and represents an ex-ante measure of volatility over the life of the option, as opposed to the ex-post realized volatility. We assume that squared

\(^3\)Our theory may easily be modified to accommodate other cases. For instance options may be sampled with longer term to maturity than the interval $K$ between expiration dates, or where only options with at least a fixed minimum term to maturity greater than one (say, 7 days) are sampled.
returns have a constant mean $\mu$, a random component $\omega_t$ which is anticipated in previous option prices, and a random innovation $\nu_t$ which is only realized at $t$ and is unanticipated by option market participants. Thus, we have

$$r_t^2 = \mu + \omega_t + \nu_t, \quad t = 0, 1, ..., T$$

and we take the random variables $\omega_t$ and $\nu_t$ to be independent, with zero means, variances $\sigma^2_\omega$ and $\sigma^2_\nu$, respectively, and support $(-\mu, \infty)$ for the sum of the two, as squared returns are positive. Since $\nu_t$ is the unanticipated component, we take this to be serially uncorrelated.

To a first approximation, the realized return variance is simply the average of the squared returns over the life of the relevant option,

$$\sigma^2_{r,t} = \frac{1}{J(t)} \sum_{i=1}^{J(t)} r_{t+i}^2.$$ 

Leaving out the mean-correction corresponds to the practice in the high-frequency realized volatility literature (see e.g. Andersen, Bollerslev and Lange (1998)). In particular, as the sampling frequency increases beyond bounds, $J(t)\sigma^2_{r,t}$ converges to integrated variance. To emphasize the dependence on the time $J(t)$ until the next expiration date, we write $\sigma^2_{r,t}(J(t))$, so we have

$$\sigma^2_{r,t}(J(t)) = \mu + \frac{1}{J(t)} \sum_{i=1}^{J(t)} (\omega_{t+i} + \nu_{t+i}).$$

Implied volatility at time $t$ is backed out of the relation equating the observed option price at $t$ with the theoretical counterpart, for a given option pricing formula (e.g. the Black and Scholes (1973) formula) and observed values of the other arguments, including term to maturity $J(t)$, strike price, value of the underlying, and a suitable interest rate. Assuming that the resulting implied variance $\sigma^2_{i,t}(J(t))$ (the square of implied volatility) has a constant mean $\mu_i$, we model its stochastic behavior as

$$\sigma^2_{i,t}(J(t)) = \mu_i + \frac{1}{J(t)} \sum_{i=1}^{J(t)} \omega_{t+i}.$$ 

Thus, $\omega_t$ is the part of the volatility process which is anticipated by option markets, and $\nu_t$ is the subsequently realized volatility innovation. We allow for the possibility that $\mu_i$ differs from $\mu$, i.e. that implied and realized variances have different means. Note that implied and realized variance share the telescoping overlap feature, and we retain $J(t)$ in the notation also for the implied measure $\sigma^2_{i,t}(J(t))$. In the time series, term to maturity $J(t)$ starts at $K$ for $t = 0$, then declines until $t$ reaches an option expiration
date, at which time it bounces back to $K$ and starts declining again. Throughout, $J(t)$ in addition indicates the overlap of both the implied and realized volatility estimates with previous estimates associated with the same expiration date.

We define the historical or past variance, $\sigma^2_{p,t-H}$, as the realized variance over the $H$ periods prior to $t$,

$$\sigma^2_{p,t-H} = \mu + \frac{1}{H} \sum_{i=0}^{H-1} (\omega_{t-i} + \nu_{t-i}),$$

(5)

where $H$ is a fixed interval length (say, 60) for all $t$. Thus, historical variance exhibits a fixed rather than a telescoping overlap feature.

Our asymptotic theory is relevant for testing the null that past volatility does not contain any information about future volatility beyond that contained in implied volatility, and for establishing the asymptotic distributions of estimators under this null.\(^4\)

We allow the returns to be generated by a process from a very general class that includes ARCH/GARCH models as special cases (See Appendix A). While $\nu_t$ is itself an innovation sequence, the process $\omega_t$ that drives implied volatility may move smoothly through time. Let $e_t$ denote the innovations in the process $\omega_t$, and write $\sigma^2_\nu$ and $\sigma^2_e$ for the innovation variances.

**Assumption 1** The processes $\nu_t$ and $e_t$ are independent and weakly stationary, with zero mean, $E(\nu_t^2) = \sigma^2_\nu > 0$, $E(e_t^2) = \sigma^2_e > 0$, and $\omega_t$ has the (possibly infinite) moving average (MA) representation

$$\omega_t = C(L)e_t,$$

(6)

where $C(L) = \sum_{j=0}^{\infty} c_j L^j$ is a one-sided MA polynomial in the lag operator $L$, with $C(1) \neq 0$. Further, $\nu_t$, $e_t$, and $C(L)$ satisfy either of the following:

(a) $\{\nu_t\}$ and $\{e_t\}$ are i.i.d., $\sigma^2_\nu < \infty$, $\sigma^2_e < \infty$, and $C(L)$ satisfies $\sum_1^{\infty} j^2 c_j^2 < \infty$ or $\sum_1^{\infty} j^{1/2}|c_j| < \infty$.

(b) $\{\nu_t\}$ and $\{e_t\}$ are i.i.d. with $E|\nu_t|^p < \infty$, $E|e_t|^p < \infty$ for some $p$ satisfying $2 < p < \infty$, and $\sum_1^{\infty} j|c_j| < \infty$.

(c) $\{\nu_t\}$ and $\{e_t\}$ are martingale difference sequences and strongly uniformly integrable (see e.g. Hamilton, 1994 pp 189 for definitions), with $E|\nu_t|^p < \infty$, $E|e_t|^p < \infty$ for some $p$ satisfying $2 < p < \infty$, $T^{-1}\sum_1^{T} E(\nu_t^2|\mathcal{F}_{t-1}) \to_{a.s.} \sigma^2_\nu$, and $\sum_1^{\infty} j^2 c_j^2 < \infty$.

Under this assumption, a functional central limit theorem (FCLT) applies to $\omega_t$ as $T \to \infty$, i.e. for $\lambda \in [0, 1]$ we have

$$\frac{1}{\sqrt{T}} \sum_{j=1}^{[\lambda T]} \omega_i \Rightarrow \sigma_\omega W^0(\lambda)$$

\(^4\)Testing for a bias involves examining whether the difference $\mu_i - \mu$ equals zero. Following the implied volatility literature, we do not focus on this issue.
where \( \Rightarrow \) denotes convergence in distribution, \( W^0 \) is a standard Brownian motion restricted to the unit interval and \( \sigma^2_0 = C(1)^2 \sigma^2_e \) is the long-run variance of \( \omega_i \).

Modelling returns rather than their squares, RS use a condition that is similar to our assumption (c) on the innovations \( e_t \). This assumption allows that \( e_t \) exhibits conditional heteroskedasticity, i.e. \( E(e_t^2|\mathcal{F}_{t-1}) \) is time-varying, but in a stationary sense, so that higher moments exist (fourth moments in RS) and \( e_t \) is unconditionally homoskedastic. The summability condition ensures that the dependence between distant error terms is limited. In our case, \( e_t \) is an innovation to volatility, not returns, so the lag polynomial \( C(L) \) generates serial correlation in the underlying series being analyzed, i.e., squared returns. Condition (c) allows time variation in conditional fourth moments of returns.

Empirical work typically uses volatilities or log-volatilities rather than the variances, and these alternative dispersion measures are covered by our theory, as well. Thus, we analyze the variance regression

\[
\sigma^2_{r,t}(J(t)) = \gamma_0 + \gamma_1 \sigma^2_{i,t}(J(t)) + \gamma_p \sigma^2_{p,t-H} + v_t, \tag{7}
\]

the volatility regression

\[
\sigma_{r,t}(J(t)) = \gamma^V_0 + \gamma^V_1 \sigma_{i,t}(J(t)) + \gamma^V_p \sigma_{p,t-H} + v^V_t, \tag{8}
\]
and the log-volatility regression

\[
\ln(\sigma_{r,t}(J(t))) = \gamma^L_0 + \gamma^L_1 \ln(\sigma_{i,t}(J(t))) + \gamma^L_p \ln(\sigma_{p,t-H}) + v^L_t. \tag{9}
\]

We develop the asymptotic theory for implied volatility regressions with a high degree of overlap. Accordingly, we have \( T \to \infty \) and \( K/T \to \delta \), where \( \delta \in (0,1) \) represents the fraction of the sample size corresponding to the maximal overlap length. For options data, the telescoping overlap dictated by the fixed expiration dates is unavoidable. Thus, we get a limiting function for \( J(t)/T \), which telescopes between \( \delta \) and 0 as \( J(t) \) telescopes between \( K \) and 1.

To formally develop this notion, we associate with each \( t = 0, \ldots, T \) a number \( \lambda \in [0,1] \) representing the fraction of the sample corresponding to the period from 0 to \( t \), i.e. \( \lambda = t/T \). Similarly, \( J(t)/T \) is the telescoping overlap stated as a fraction of sample size, and this may be expressed as a function of \( \lambda \) by writing \( t = \lambda T \) and dividing through by \( T \) in (1). As \( T \to \infty \) and \( K/T \to \delta \), the resulting function converges to the limiting function \( \psi(\lambda) \) defined as follows.

**Definition 1** The function \( \psi(\lambda) \): \([0,1] \to [0,\delta] \), defined as the limiting relative overlap \( J(t)/T \) as \( T \to \infty \), is given by

\[
\psi(\lambda) = \left(1 + \left\lfloor \frac{\lambda}{\delta} \right\rfloor \right) \delta - \lambda. \tag{10}
\]
Thus, $\psi(\lambda)$ indicates which fraction of the sample size is made up by the telescoping overlap length $J(t)$. In this way, $\lambda$, $\psi(\lambda)$, and $\delta$ are simply $t$, $J(t)$, and $K$ transformed into the unit interval, by dividing by $T$, for $T$ large.

### 2.1 Regression Coefficients

We now obtain the limiting distribution of the OLS slope coefficients of the regressions (7)-(9). Here, the assumption is that the sampling scheme includes a relation between the historical volatility window $H$ and the time $K$ between expiration dates, say, $H/K = \kappa$, e.g. if $H = 60$ and $K = 30$ then $\kappa = 2$. In Appendix B, the following theorem is proved.\(^5\)

**Theorem 1 (Limiting Distribution)** Under Assumption 1, the limiting distribution of the least squares estimator $\hat{\gamma} = \left[\hat{\gamma}_i \hat{\gamma}_p\right]'$ in (7) is a functional of Brownian motions $\gamma^* = \left[\gamma_i^* \gamma_p^*\right]'$ given by

\[
\begin{bmatrix}
\int_{\kappa \delta}^{1-\delta} \hat{I}_\psi(\lambda)(\lambda) d\lambda \\
\int_{\kappa \delta}^{1-\delta} \hat{I}_\psi(\lambda)(\lambda) P_{\kappa \delta}(\lambda) d\lambda
\end{bmatrix}
\begin{bmatrix}
\int_{\kappa \delta}^{1-\delta} \hat{\gamma}_i(\lambda) d\lambda \\
\int_{\kappa \delta}^{1-\delta} \hat{\gamma}_p(\lambda) d\lambda
\end{bmatrix}^{-1}
\begin{bmatrix}
\int_{\kappa \delta}^{1-\delta} \hat{I}_\psi(\lambda)(\lambda) \hat{R}_\psi(\lambda)(\lambda) d\lambda \\
\int_{\kappa \delta}^{1-\delta} \hat{P}_\lambda(\lambda) \hat{R}_\lambda(\lambda) d\lambda
\end{bmatrix}
\]

where $\hat{I}_\psi(\lambda)(\lambda)$, $\hat{R}_\psi(\lambda)(\lambda)$ and $\hat{P}_\lambda(\lambda)$ are stochastic processes defined in terms of independent standard Brownian motions $W^0$ and $W^1$ in the following manner:

\[
\begin{align*}
\hat{I}_\psi(\lambda)(\lambda) & \equiv I_\psi(\lambda)(\lambda) - \frac{1}{1 - \kappa(1 + \delta)} \int_{\kappa \delta}^{1-\delta} I_\psi(s)(s) ds, \\
\hat{R}_\psi(\lambda)(\lambda) & \equiv R_\psi(\lambda)(\lambda) - \frac{1}{1 - \kappa(1 + \delta)} \int_{\kappa \delta}^{1-\delta} R_\psi(s)(s) ds, \\
\hat{P}_\lambda(\lambda) & \equiv P_{\kappa \delta}(\lambda) - \frac{1}{1 - \kappa(1 + \delta)} \int_{\kappa \delta}^{1-\delta} P_{\kappa \delta}(s)(s) ds,
\end{align*}
\]

using the further definitions

\[
\begin{align*}
I_\psi(\lambda)(\lambda) & \equiv \frac{\sigma_\psi}{\psi(\lambda)} (W^0(\psi(\lambda) + \lambda) - W^0(\lambda)), \\
R_\psi(\lambda)(\lambda) & \equiv I_\psi(\lambda)(\lambda) + \frac{\sigma_\psi}{\psi(\lambda)} (W^1(\psi(\lambda) + \lambda) - W^1(\lambda)), \\
P_{\kappa \delta}(\lambda) & \equiv I_{\kappa \delta}(\lambda - \kappa \delta) + \frac{\sigma_\nu}{\kappa \delta} (W^1(\lambda) - W^1(\lambda - \kappa \delta)).
\end{align*}
\]

Furthermore, the limiting distribution of the least squares estimator $\hat{\gamma}^V = \left[\hat{\gamma}_i^V \hat{\gamma}_p^V\right]'$ in the volatility regression (8) is given by

\[
\gamma^V i = \left(\frac{\mu_i}{\mu}\right)^{1/2} \gamma_i^{*V}\gamma_p^{*V},
\]

\(^5\)The proof draws on the Functional Central Limit Theorem (FCLT) and the Continuous Mapping Theorem (CMT) (see e.g. Davidson, 1997).
and the limiting distribution of the least squares estimator \( \hat{\gamma}^L = [\hat{\gamma}_i^L \hat{\gamma}_p^L]' \) in the log-volatility regression (9) is given by

\[
\gamma^{L*} = \left[ \begin{pmatrix} \mu_i \\ \mu_p \end{pmatrix} \gamma^*_i \gamma^*_p \right]',
\]

(18)

where \( \mu_i \) and \( \mu_p \) are the unconditional means from (3) and (4).

The asymptotic distribution result is clearly very different from the consistency and asymptotic normality results from ordinary regression theory. The result generalizes the RS theorem in several ways. Historical volatility on the right hand side is not simply the lagged left hand side variable, but may typically cover a different interval. Our setting has multiple regressors. We allow for general driving processes (Assumption 1), and we account explicitly for the particular telescoping feature in realized and implied volatility resulting from the fixed, monthly option expiration dates, as reflected through the function \( \psi \) in the Theorem. In previous applications of the functional central limit theorem, the time index \( \lambda \) on the processes in the limiting stochastic integrals has only represented the fraction of sample size, and hence our deterministic time-change \( \psi(\lambda) \) designed to capture the effect of the telescoping feature is a novel aspect of our theory. Our theory allows accounting for the time varying degree of overlap in a precise manner.

Historical volatility is akin to a spurious regressor in the sense of Granger and Newbold (1974), in that it is strongly serially correlated, but may have little correlation with the left hand side variable. This may yield a spurious bias in the coefficients on both regressors, and our asymptotic theory allows calculating proper p-values in the presence of such effects. Corollary 1 gives the distributions of coefficients in the univariate regressions corresponding to the multiple regression analyzed in Theorem 1.

**Corollary 1 (Univariate Limiting Distribution)** Under Assumption 1, the limiting distribution of the least squares estimator \( \hat{\gamma}_{i,u} \) in the univariate regression on implied volatility, i.e. (7) with \( \gamma_p \equiv 0 \), is a functional of Brownian motions

\[
\gamma^*_{i,u} = \left[ \int_{\eta \delta}^{1-\delta} (\tilde{I}_{\psi(\lambda)}(\lambda))^2 d\lambda \right]^{-1} \left[ \int_{\eta \delta}^{1-\delta} \tilde{I}_{\psi(\lambda)}(\lambda) \tilde{R}_{\psi(\lambda)}(\lambda) d\lambda \right]
\]

(19)

and the corresponding distribution of \( \hat{\gamma}_{p,u} \) in the univariate regression on past realized volatility, (7) with \( \gamma_i \equiv 0 \), is a functional of Brownian motions given by

\[
\gamma^*_{p,u} = \left[ \int_{\eta \delta}^{1-\delta} (\tilde{P}_{\eta \delta}(\lambda))^2 d\lambda \right]^{-1} \left[ \int_{\eta \delta}^{1-\delta} \tilde{P}_{\eta \delta}(\lambda) \tilde{R}_{\psi(\lambda)}(\lambda) d\lambda \right]
\]

(20)

where \( \tilde{I}_{\psi(\lambda)}(\lambda), \tilde{R}_{\psi(\lambda)}(\lambda) \) and \( \tilde{P}_{\eta \delta}(\lambda) \) are defined in Theorem 1. Furthermore, the limiting distributions of the estimators \( \hat{\gamma}^V_{i,u} \) and \( \hat{\gamma}^V_{p,u} \) from the corresponding univariate volatility
regressions are given by

\[
\gamma_{V,i,u} = \left( \frac{\mu_i}{\mu} \right)^{1/2} \gamma_{i,u}^*, \quad \gamma_{V,p,u}^* = \gamma_{p,u}^*,
\]

(21)

and those of \( \hat{\gamma}_{i,u}^L \) and \( \hat{\gamma}_{p,u}^L \) from the univariate log-volatility regressions are given by

\[
\gamma_{L,i,u} = \left( \frac{\mu_i}{\mu} \right) \gamma_{i,u}^*, \quad \gamma_{L,p,u}^* = \gamma_{p,u}^*.
\]

(22)

From Theorem 1, the OLS coefficients on both regressors converge weakly to functionals of Brownian motions. Thus, the distributions of the coefficient on implied volatility and the coefficient on historical volatility are both dispersed, even in the limit. These inconsistencies are consequences of the use of overlapping data. Importantly, the inconsistency carries over to the univariate regressions, showing that the source of the problem is the overlapping data problem per se, not simply the presence of historical volatility on the right hand side. From (14)-(16), the two parameters \( \sigma_\omega \) and \( \sigma_\nu \) also enter in the limiting distribution of the OLS coefficients. Thus, the estimation of the limiting distribution involves an estimation of two nuisance parameters, namely the long-run variances of \( \omega_t \) and \( \nu_t \) (see Assumption 1).

The length of the overlap as a fraction of the sample size enters in the limiting distribution of the OLS coefficients. Here, it enters in two ways. First, via the standard Brownian motions, i.e. \( W^0(\psi(\lambda) + \lambda) - W^0(\lambda) \). Second, the limiting distributions of the mean-corrected variances include a multiplicative term that equals one divided by the overlap in fraction of the sample size. The term \( \sim P_{\kappa \delta}(\lambda) \) depends on \( \kappa \delta \), whereas the terms \( \sim I_{\psi(\lambda)}(\lambda) \) and \( \sim P_{\psi(\lambda)}(\lambda) \) depend on \( \psi(\lambda) \).

2.2 Test Statistics

In light of Theorem 1, statistical inferences may be seriously misleading if drawn using standard techniques without regard to the overlapping data problem. Commonly applied tests of the significance of the coefficients on implied and historical volatility are based on OLS or Newey and West (1987) standard errors and associated \( t \)-statistics. The overall fit of the regression is commonly judged by the standard coefficient of determination \( R^2 \), possibly adjusted for degrees of freedom. We now examine the asymptotic properties of these common practices.

Theorem 2 Let \( \hat{\Xi}(J(t)) \) denote the estimated OLS or Newey-West variance-covariance matrix of the regression coefficients, and \( DW \) the Durbin-Watson statistic. Under Assumption 1, the following results obtain for all the regressions considered in Theorem 1
(univariate and multivariate variance, volatility and log-volatility regressions) as $T \to \infty$:

\[
\hat{\Xi}(J(t)) \to 0 \quad (23)
\]

\[
t\text{-value} \to \infty \quad (24)
\]

\[
R^2 \Rightarrow R^{2*} \quad (25)
\]

\[
\text{adjusted } R^2 \Rightarrow R^{2*} \quad (26)
\]

\[
DW \to 0, \quad (27)
\]

where $R^{2*}$ is a functional of Brownian motions defined by

\[
R^{2*} \equiv \frac{\int_{\kappa\delta}^{1-\delta} \left( \gamma_i^* \tilde{I}_{\psi(\lambda)}(\lambda) + \gamma_p^* \tilde{P}_{\kappa\delta}^*(\lambda - \kappa\delta) \right)^2 d\lambda}{\int_{\kappa\delta}^{1-\delta} \left( \tilde{P}_{\psi(\lambda)}(\lambda) \right)^2 d\lambda}
\]

using the same notation as in Theorem 1.

The proofs are deferred to Appendix C. From the Theorem, standard inferences are highly misleading. Even though implied volatility is an efficient forecast under the null, implying that historical volatility should not enter the regression, all $t$-ratios diverge, including that for historical volatility. Using standard techniques, this variable may thus be considered significant when it is not. The true $p$-values are calculated using the new theory (Theorem 1). Similarly, the adjusted and unadjusted $R^2$ measures are of little value in judging goodness of fit, as they have dispersed distributions, even in the limit as $T \to \infty$. On the other hand, the Durbin-Watson statistic does contain potentially useful diagnostic information. In particular, from Theorem 2, a low $DW$ statistic may be a sign of an overlapping data problem, and should be interpreted as a call for concern.

We now consider an empirical application of our asymptotic theory to daily overlapping and monthly non-overlapping data on OEX implied volatility and S&P 100 realized return volatility. These samples serve to illustrate the general point: Conclusions based on ordinary inference procedures applied to daily data are highly sensitive to the sampling scheme chosen, and, in particular, to the telescoping overlap problem.

3 Data

We focus on at-the-money call options with between 7 and 35 calendar days to expiration. Clearly, the ratio of the maximal overlap to the total sampling interval ($\delta$ from the previous section) is higher for longer maturities, so any finding of adverse effects of the overlap for short term options is likely to be conservative. We sample the OEX prices from the Berkeley Option Data Base (BODB), which contains data on options traded at the Chicago Board Option Exchange (CBOE). The data cover the period from November 1983
through December 1995. OEX options by convention expire on the Saturday following
the third Friday of each month. For each trading date $t$, we record the bid-ask midpoint
for the option with between 7 and 35 calendar days to expiration that is closest to being
at-the-money in the last quote before 3PM (CST), and back out the associated implied
volatility $\sigma_{i,t}(J(t))$. This option has $J(t)$ days to expiration, and the next options in the
time series have $J(t + 1), J(t + 2), \ldots$ days to expiration. If $J(t) = k$, say, then these
following options have $k - 1, k - 2, \ldots$, 7, 35, 34, \ldots days to expiration, as captured by
the telescoping feature of the $J(\cdot)$ function.

We back out the implied volatility using a Cox, Ross and Rubinstein (1979) binomial
tree model with 30 time steps, where the dividend yield and possible early exercise are
accounted for. The daily dividend yield series is obtained from Datastream. As a proxy
for the risk-free interest rate we use the one month LIBOR (London Inter Bank Offer
Rate). For each implied volatility, we calculate the associated realized volatility $\sigma_{r,t}(J(t))$
over the remaining life of the relevant option, i.e., the sample standard deviation of the
$J(t)$ daily index returns over this period. In particular, there is no issue of maturity
mismatching. Finally, we calculate a 60-day historical volatility $\sigma_{p,t-H}$, as the sample
standard deviation of the daily index returns over the 60 trading days prior to the date
$t$ where the relevant option price is recorded. This procedure yields the three time series
for daily realized, implied, and past realized (historical) volatility.

Monthly non-overlapping series are obtained by selecting from the daily overlapping
series the implied volatility measured on the Wednesday immediately following the ex-
piration date of the given month. If this is not a trading day, we move to the following
Thursday, then the preceding Tuesday, and so on. The associated realized volatility is
the same as in the overlapping series, i.e. it covers the remaining life of the option. For
the non-overlapping series, past realized volatility is simply the (one month) lagged value
of realized volatility, i.e., the previous element in the time series.

Since index option markets have been found to behave differently before and after the
October 1987 stock market crash (Christensen and Prabhala, 1998), we consider separate
pre-crash and post-crash periods in the application. The pre-crash period is November
1983 through March 1987, well before the crash. Our post-crash period is February
1988 through December 1995. We start in February 1988 because volatilities are greatly
elevated in the first couple of months after the October 1987 crash. We end the sample
in December 1995 because BODB data are not available after this date.

We get similar results with 500 steps. The dividend and American feature are incorporated in our
study, but had little impact on our results.
3.1 Descriptive Statistics

Descriptive statistics for the volatilities and their logarithms appear in Table 1. Panel A shows statistics for the pre-crash period and Panel B for the post-crash period. As expected, daily and monthly measures of the same volatility (whether implied or realized) have about the same mean and standard deviation. Implied volatility is on average one to two percentage points higher than realized both before and after the crash, possibly due to imperfect correction for the American feature or excessive replication costs. Comparing across subperiods, average volatility has come down after the crash, whether judged by the implied or the realized measure. All variables are positively skewed and most are leptokurtic, but not alarmingly so, in a regression context, and the effects are virtually eliminated by taking logarithms.

Table 1 also shows the first, fifth and tenth autocorrelation coefficients. Here, it is seen that the daily overlapping data exhibit much larger persistence than the monthly non-overlapping data. Thus, historical volatility, the measure of past realized volatility in the daily data volatility regressions, has a first order autocorrelation of 0.99, and even on the tenth lag, the correlation is still 0.88 before the crash and 0.91 after the crash. Recall that two consecutive observations on historical volatility are based on 59 squared return observations that are common and two that differ. Clearly, the sampling methodology induces strong serial dependence in this measure, as reflected in Table 1, and the problem is not alleviated by taking logarithms. In the daily data, serial dependence is also strong in realized and implied volatility, but much less so in the monthly non-overlapping data.

To shed further light on the severity of the induced serial dependence, we apply an augmented Dickey-Fuller (DF) test with intercept, using four lags for the daily and two for the monthly data. The DF test appears in the last column in the table. From the results, serial dependence in historical volatility before the crash is so strong that the DF test fails to reject a unit root. After the crash, the test rejects strongly for all series. This suggests that the serial dependence introduced by the overlapping sampling method is a more serious problem in the shorter pre-crash period, where the 59-day overlap represents a greater fraction of the total sampling period, than in the longer post-crash period.

Non-overlapping monthly data have been considered by Christensen and Prabhala (1998) (henceforth CP), who found average volatility numbers within one percentage point of those for the non-overlapping series in Table 1, but also found that the standard deviation of realized volatility was relatively small before the crash - smaller than that of realized volatility after the crash, and even smaller than that of implied volatility before the crash. In Table 1, the standard deviation of realized volatility is similarly smaller before the crash than after, but not by much, and it exceeds the standard deviation of implied volatility before the crash. The latter makes good sense when thinking of implied volatility as a conditional expectation of realized volatility.
There are some differences in the post-crash periods here compared to in CP. The full pre-crash period was November 1983 through September 1987, which includes an additional six months just prior to the crash compared to ours. The post-crash period was December 1987 through May 1995, thus including two more months shortly after the crash, and excluding the last seven months of 1995. Another difference is that we use a binomial tree with adjustment for dividends and early exercise, whereas CP backed out simple Black-Scholes implied volatilities. Recalculating the standard deviations for our data, using the exact periods from the earlier study, we get 0.039 for realized volatility and 0.041 for implied before the crash, and 0.056 for realized after the crash. Here, the phenomenon that the standard deviation of pre-crash realized volatility is less than those of pre-crash implied and post-crash realized volatility reappears. This shows that the differences in the standard deviations are due to the period definitions, not the definition of implied volatility. Perhaps options market participants partly anticipated the crash, elevating the level and variance of implieds in the months before the crash (see Bates, 1991), but further analysis of this question is beyond the scope of the present paper. For the regression analysis below, we stay with the pre-crash period ending in March 1987, which seems most regular on these grounds. After the crash, we avoid the months closest to October 1987 and otherwise use all available data.

4 Empirical Results

4.1 Inference based on Standard Approach

Table 2 shows the regression results on the relation between implied and realized volatility for the pre-crash period November 1983-March 1987. Results are shown both for daily and monthly data.

For the daily overlapping data, the regression equation takes the form

$$\sigma_{r,t}(J(t)) = \gamma_0 \sigma_i + \gamma_i \sigma_{i,t}(J(t)) + \gamma_p \sigma_{p,t-60} + \epsilon_t, \quad t = 60, \ldots, T - K. \tag{28}$$

There are 784 observations in the sample period. The remaining option life $J(t)$ telescopes between 7 and 35 calendar days. Realized volatility $\sigma_{r,t}(J(t))$ associated with the option sampled at time $t$ is only observable at the end of this option’s life, i.e., at $t + J(t)$. In effect, $t$ is an index of option lives, and the daily time series are arranged so that $\sigma_{r,t}(J(t))$ and implied volatility $\sigma_{i,t}(J(t))$ are alternative (ex-post respectively ex-ante) measures of volatility over the interval $[t + 1, t + J(t)]$.

We next represent results from the corresponding non-overlapping regression. The regression can be written as
where we use a subset of the data with no overlap. In equation (29) \( T_i \) is the \( i \)'th expiration date in the sample. The exposition emphasizes that the option lives now are of equal length \( K \) (one month in the application), thus yielding monthly non-overlapping samples of both implied and realized volatility in the stylized sampling scheme from the previous section, where \( T_i - T_{i-1} = K \). In practice, we sample implied volatility on the Wednesday immediately following the option expiration date (the third Saturday) in the relevant month. There are \( M = 41 \) monthly observations in the pre-crash sample period. In (29), the past realized volatility forecast is simply the realized volatility for the previous month \( i - 1 \), i.e., the previous element in the monthly realized volatility series. Using instead the longer horizon measure, corresponding to the daily analysis, does not change any conclusions below.

Consider first the daily regression results in Table 2, based on specification (28). The first line shows the results when past realized volatility is excluded. Based on conventional analysis, the coefficient on implied volatility, 0.22, is significantly greater than zero (Newey and West (1987) standard errors in parentheses). This suggests that option prices do contain volatility information. However, the estimate is small in magnitude, and significantly less than one, suggesting that implied volatility is a downward biased forecast of future volatility. When past realized volatility is included as the only regressor, its coefficient is more than twice as high, at 0.49. The coefficient on past realized volatility remains almost as high in the encompassing regression including both volatility forecasts, but here, the coefficient on implied volatility halves to 0.10 and loses significance. Furthermore, the coefficient on past realized volatility is strongly significant, and that on implied volatility insignificant, at all conventional levels.

Panel B shows the daily and monthly results for the log-volatility series. The daily results are in the left hand side of the table. The slope coefficients are slightly larger, but the overall picture is the same. The univariate regressions yield a higher coefficient on past realized volatility than on implied. In the multivariate regression, the coefficient on implied volatility is small and insignificant, and the coefficient on past realized volatility is much larger and strongly significant. This conventional analysis of either volatilities or log-volatilities suggests that past realized volatility has incremental information over and beyond that in implied volatility, and, indeed, that past realized volatility subsumes the information content of implied volatility. A shared feature of the daily volatility and log-volatility regressions is that the Durbin-Watson statistics are low, at 0.23 –0.24, thus indicating strong residual autocorrelation. From Theorem 2, this is consistent with the presence of an overlapping data problem. Thus, the daily sampling scheme induces an
extreme degree of overlap.

The monthly results appear in the right hand side of Table 2. In the raw volatility regression where only implied volatility is included as regressor the estimated coefficient is 0.42, much higher than the daily estimate of 0.22. When only past realized volatility is included, the monthly estimate, 0.16, is less than half of the daily estimate, 0.49, and also less than half of the monthly estimate for implied volatility, 0.42, and the estimate is insignificantly greater than zero. In the encompassing regression, the coefficient on past realized volatility is even smaller, at 0.01. Implied volatility is strongly significant and subsumes the information content of past realized volatility, consistent with efficiency of the implied volatility forecast. Again, the picture is the same for log-volatilities (Panel B). Furthermore, the monthly Durbin-Watson statistics are close to 2.0 and show no sign of misspecification.

The daily and monthly regressions are quite different. The point estimates vary along with the conclusions about option market efficiency. What is the source of this discrepancy? Is it due to distortions induced by overlapping data? The low DW statistic certainly suggests that this is a distinct possibility.

4.2 Inference Based on New Asymptotic Theory

We now apply our new distribution theory to the empirical results. We show that the difference between the empirical results from overlapping and non-overlapping data is not particularly surprising, when using the proper asymptotics. Theorem 1 and Corollary 1 provide limiting representations of the least squares estimators as functionals of Brownian motions. These random functionals do not have a standard distribution, so the usual approach of looking up tabulated critical values is not feasible. The limiting representations provide a recipe for obtaining the asymptotic critical values by simulation. First, we estimate the long-run variances by the Andrews and Monahan (1992) method, obtaining the values $\hat{\sigma}_\omega = 0.341$ and $\hat{\sigma}_v = 0.762$. It is now straightforward to obtain the asymptotic critical values using Monte Carlo simulations (see e.g. Hamilton (1994) for simulation of functionals of Brownian motion). For this purpose we use 5,000 replications. This sample size yields invariant percentiles of the limiting distributions. The maximal overlap parameter $\delta$ in the integrals is 0.04, corresponding to a maximal term to maturity of 35 days, relative to 879 days in the sample. The telescoping overlap function $\psi$ runs from 0 to $\delta$.

Table 4 presents the $p$-values based on the new asymptotic theory. Consider the first numbers in the upper left portion of Table 4, for the pre-crash period. We first focus on the coefficient from the univariate regression of realized volatility on implied volatility.

---

7 The results turns out to be stable using 5,000 replications. Richardson and Smith (1991) also use 5000 replications in their Monte Carlo study.
Table 4 first recalls the empirical estimates from Table 2, i.e. 0.22 for the daily estimate and 0.42 for the monthly estimate. The issue is whether the asymptotic distribution from Corollary 1 implies so much noise in the estimates that the daily and monthly coefficients in fact are not statistically different. Accordingly, the Table next reports the asymptotic p-value based on Corollary 1 for an outcome where the daily coefficient falls short of the monthly coefficient by an even larger amount than $0.42 - 0.22 = 0.20$. From the Table, the asymptotic p-value is 0.25 and so clearly exceeds e.g. the conventional 1% and 5% levels. Thus, when properly accounting for the extreme degree and telescoping nature of the overlap in the daily regression in calculating the asymptotic distribution, we conclude that in fact there is no statistically significant difference between the results from the daily and the monthly methods. The daily overlapping data method produces an inconsistent estimate, and an unreliable standard error. From standard regression theory for non-overlapping data, the monthly estimate, in this case 0.42, is consistent, and so is the associated standard error, at 0.14 in the application.

Turning to the coefficient on past realized volatility before the crash, the daily and monthly coefficients are 0.49 and 0.16 in the univariate regressions. The question is whether the apparent difference is in fact significant, when accounting for the inconsistency of the daily results. The p-value is for an outcome where the daily estimate exceeds the monthly estimate by even more than $0.49 - 0.16 = 0.33$. At 0.21 the p-value exceeds the 1% and 5% levels, indicating that the difference in results is insignificant. The sampling distribution of the coefficients is so dispersed that the received difference is relatively unsurprising.

Next, in the multivariate regression before the crash, the result is similar for the coefficient on implied volatility. The p-value for the daily coefficient falling short of the monthly by more than in the empirical results is 0.15, and that for the daily coefficient on past realized to exceed the monthly by more than in the empirical results 0.12. Overall, there is little evidence of any statistically significant difference between the daily and monthly coefficients in univariate or multivariate regressions before the crash.

4.3 Post-Crash Subperiod

We also consider the regressions for the period February 1988 to December 1995, after the 1987 stock market crash. Here, the maximal overlap parameter $\delta$ is lower, at 0.02, corresponding to 35 days out of a total of 1,947 in the sampling window. The results appear in Table 3.

Consider first the daily frequency results in Table 3. In the univariate regressions,

---

8Both estimates may be subject to an error-in-variable bias towards zero, stemming from measurement error in implied volatility, but the reported p-values are for the differences between the estimates, not the bias in each.

17
the coefficient on implied volatility is now more than twice the value from the pre-crash period (Table 2), whether using raw volatilities or log-volatilities, and higher than the corresponding coefficient on past realized volatility. Based on standard inference, implied volatility remains strongly significant in the encompassing regression, and in the raw volatility specification even subsumes the information content of past realized volatility. This is in stark contrast to the pre-crash findings. However, the Durbin-Watson statistics remain low, so again, we also consider the corresponding monthly non-overlapping data regressions.

The monthly post-crash regression results appear in the right hand side of Table 3. The coefficients on implied volatility are larger and those on past realized volatility smaller than the associated daily estimates (left side of the Table). In the monthly encompassing regressions, implied volatility subsumes the information content of past realized volatility. The coefficient on implied volatility is insignificantly different from unity, both in univariate and multivariate regressions, and whether or not the log-transform is applied. Furthermore, in the raw volatility specification, the intercept is insignificant (it turns negative in the log-specification since realized volatility is more variable than implied volatility). The indication is that the implied volatility forecast is both unbiased and efficient. In these regressions, implied volatility explains about half of the future variation in volatility, judged by the adjusted $R^2$-statistic. Finally, the Durbin-Watson statistics show no sign of misspecification.

In the post-crash period, implied volatility is superior to historical volatility. On the methodological side, the important finding is that our results from overlapping and non-overlapping samples are closer. In the post-crash period the overlap ratio $\delta$ is lower, the post-crash value, 0.02, is only half of the pre-crash value. Invoking Theorems 1 and 2, the post-crash results (right hand side of Table 4) largely confirm the pre-crash pattern: The $p$-values exceed conventional levels, showing that when properly accounting for the overlapping data problem, there is no significant difference between results from overlapping and non-overlapping data analyses. The same conclusion emerges when (Table 5) we apply the new theory to the results for the log-volatility regressions. Overall, there is little evidence of any significant difference between the daily and monthly volatility and log-volatility regression results, once the adverse effects of the use of overlapping data in the daily approach are accounted for.

Figure 2 presents a graphical summary of the limiting distributions of the coefficients for implied and historical volatility based on the asymptotic theory (dashed lines) offered here, and provides a comparison with distributions from conventional asymptotics (solid lines). The true limiting distributions have thicker tails in the pre-crash period. Thus, taking the telescoping overlap into account leaves extra probability mass in the tails of the asymptotic distribution. In both periods, daily coefficients for historical volatility
are distributed to the left of the implied volatility distribution. The normal distribution comes closer to the data in the second period, which is what we expect because the overlap is a smaller fraction of the total sample size here compared to the pre-crash period.

5 Conclusion

Researchers using options data in their empirical analyses must confront a potentially severe overlap in the data. The overlap is induced by the fact that we have only a few option expirations each year but option prices are sampled at a much higher frequency. For instance, OEX option prices are usually sampled daily while there are only about 12 option maturities each year. This combination of fixed maturity dates and frequent sampling leads to a high degree of overlap that has a peculiar telescoping nature as the sample date approaches an option expiry date. We show that the telescoping overlap structure can induce potentially severe distortions in inferences. In the context of implied volatility regressions, we find that regression coefficients are inconsistent and remain disperse even in the limit. Standard inference procedures are similarly afflicted by this inconsistency; \(t\)-statistics diverge, Durbin-Watson statistic converge to zero, and \(R^2\) is inconsistent.

We apply the new asymptotic theory to reconcile seemingly contradictory overlapping and non-overlapping results in the liquid OEX options market. While non-overlapping samples suggest that option markets are informationally efficient, daily regressions suggest otherwise. With daily data, implied volatility appears to be subsumed by historical volatility in explaining future realized volatility. This contradiction is, however, resolved with the appropriate asymptotic theory that accounts for the telescoping overlap, which we develop here. With this theory, the message from overlapping data estimates is no longer different from non-overlapping estimates. We cannot reject the hypothesis that option markets aggregate volatility information efficiently.

An interesting avenue for further research suggested by our results is in the area of testing alternative option pricing models. There, researchers must also deal with problems of highly overlapping data with telescoping overlap, and also non-linear structural models for pricing options, possibly without closed-form solutions for option prices. We leave this topic for future research.
A  The GARCH Case

Return series are typically well described by GARCH processes. If returns follow a GARCH process then squared returns follow an ARMA model, and ARMA models are special cases of our framework for squared returns (see Stock (1994) p. 2746). Below we provide increasingly complex examples illustrating this point.

Example 1 (GARCH case)  Recall the model for the daily squared S&P 100 index returns from Eq. (2)

\[ r_i^2 = \mu + \omega_i + \nu_i. \]

Assume \( \omega_i \sim AR(1) \), with \( |\theta| < 1 \), which is equivalent to \( \omega_i \sim MA(\infty) \) and satisfies the first part of Assumptions 1. Assume furthermore \( \nu_i \sim iid(0, \sigma_{\nu}^2) \) i.e. white noise. Then \( r_i^2 \sim ARMA(1,1) \) and \( r_i \sim GARCH(1,1) \). To see this let

\[
\begin{align*}
\omega_i &= \theta \omega_{i-1} + \varepsilon_i, \quad |\theta| < 1 \\
r_i^2 &= \mu + \omega_i + \nu_i \\
&= \mu + \theta \omega_{i-1} + \varepsilon_i + \nu_i \\
&= \mu + \theta (r_{i-1}^2 - \mu - \nu_{i-1}) + \varepsilon_i + \nu_i \\
&= \underbrace{\mu(1 - \theta)}_{AR(1)} + \underbrace{\theta r_{i-1}^2}_{AR(1)} + \underbrace{\varepsilon_i + \nu_i - \theta \nu_{i-1}}_{MA(1)}. \tag{31}
\end{align*}
\]

The first part is the AR part and the last part is the MA part. This follows from Havey,1993 p. 31, since the term \( y_i = \varepsilon_i + \nu_i - \theta \nu_{i-1} \) has an autocorrelation function on the same form as an MA(1):

\[
\rho(\tau) = \begin{cases} 
-\frac{\theta \sigma_\nu^2}{\sigma_\varepsilon^2 (1 + \theta^2) + \sigma_\nu^2} & \tau = 1 \\
0 & \tau \geq 2 
\end{cases} \tag{32}
\]

This shows that Assumption 1 does allow for \( GARCH(1,1) \) processes driving the returns.

More generally, assume \( \omega_i \sim AR(p) \), with \( 0 < p < \infty \) and \( \nu_i \) is white noise. Then the argumentation as above shows that the squared returns follow an \( ARCH(p,p) \) and
thereby that the returns can be described by a GARCH\((p, p)\) process.

\[
\omega_i = \sum_{s=1}^{p} \theta_s \omega_{i-s} + \varepsilon_i, \quad \sum_{s=1}^{p} \theta_s < 1
\]

\[
r_i^2 = \mu + \omega_i + \nu_i
= \mu + \theta_1 \omega_{i-1} + \ldots + \theta_p \omega_{i-p} + \varepsilon_i + \nu_i
= \mu + \theta_1 (r_{i-1}^2 - \mu - \nu_{i-1}) + \ldots + \theta_p (r_{i-p}^2 - \mu - \nu_{i-p}) + \varepsilon_i + \nu_i
= \underbrace{\mu (1 - \theta_1 - \ldots - \theta_p) + \theta_1 r_{i-1}^2 + \ldots + \theta_p r_{i-p}^2}_{\text{AR}(p)} + \varepsilon_i + \nu_i - \theta_1 \nu_{i-1} - \ldots - \theta_p \nu_{i-p}.
\]

(33)

(34)

The autocorrelation function of the term \(y_i = \varepsilon_i + \nu_i - \theta_1 \nu_{i-1} - \ldots - \theta_p \nu_{i-p}\) reveals that \(y_i\) is a MA\((p)\), as

\[
\rho(\tau) = \begin{cases} 
\frac{-(\theta_s + \sum_{s=1}^{p} \theta_s \theta_{s-r}) \sigma_i^2}{\sigma_i^2 (1 + \sum_{s=1}^{p} \theta_s^2) + \sigma_N^2} & 1 \leq \tau \leq p \\
0 & \tau \geq p + 1
\end{cases}
\]

(36)

If we let \(\nu_i\) be a martingale difference sequence as in Assumption 1 (c) then (35) might exhibit an autocorrelation function similar to that of a MA\((q)\) with \(q \neq p\), and thereby allow the returns to be described by a general GARCH\((p, q)\).

It remains to be shown that the AR\((p)\) process describing the innovation series \(\omega_i\) satisfies the summability conditions in Assumption 1.

Let \(\Theta(L) = \sum_{s=1}^{p} \theta_s L^s\), where \(L\) is the lag operator, i.e. \(L^i \varepsilon_i = \varepsilon_{i-j}\). If \(\sum_{s=1}^{p} \theta_s < 1\), then \(\Theta(L)\) is invertible. To see that the summability condition in Assumption 1 (a) and (c) hold we consider \(C(L) = \Theta(L)^{-1}\), where \(C(L) = \sum_{j=0}^{\infty} c_j L^j\). Let

\[
d_j = \frac{\lambda_j^{p-1}}{\Pi_{s \neq j} (\lambda_j - \lambda_s)},
\]

where \(\lambda_j\) denotes the eigenvalues related to the polynomial. The eigenvalues \(|\lambda_j| < 1\) for all \(j = 1, \ldots, p\) since the \(p\)th order polynomial has roots outside the unit circle. From Hamilton (1994, p.35) we know that \(c_j = d_1 \lambda_1^j + d_2 \lambda_2^j + \ldots + d_p \lambda_p^j\). We can apply the ratio
\[
\frac{(j+1)^2c_{j+1}^2}{j^2c_j^2} = (1+2/j+1/j^2) \left( \frac{(d_1\lambda_1^{j+1} + d_2\lambda_2^{j+1} + \cdots + d_p\lambda_p^{j+1})^2}{(d_1\lambda_1^j + d_2\lambda_2^j + \cdots + d_p\lambda_p^j)^2} \right) \\
\leq (1+2/j+1/j^2) \left( \frac{(\max_{1\leq k\leq p} \lambda_k)^2}{(d_1\lambda_1^j + d_2\lambda_2^j + \cdots + d_p\lambda_p^j)^2} \right) \\
< \left( \max_{1\leq k\leq p} \lambda_k \right)^2 < 1 \text{ as } j \to \infty,
\]

which shows that the sum in condition in Assumption 1 (a) and (c) is finite. Hence, our Assumption 1 allow returns to be generated by a GARCH\((p,q)\) process.

Returns are widely conceived to be generated by ARCH or GARCH processes, i.e. the innovation term, \(\omega_t\), from Eq.(2) is generated by an AR\((p)\) process. The extended FCLT in Proposition 1 allow the innovation series to be independent but not identical distributed (i.n.i.d.). The only restriction is that the coefficients in the ARMA\((p,q)\) model satisfy the summability condition. It is sufficient that the AR polynomial is stationary. The limiting distribution of the least squares coefficients does not depend on the individual parameters in the generating ARCH or GARCH model. The long-run variance captures the features of the generating data process. In the regression with only one explanatory variable, this parameter cancels out in the limiting distribution.

**B Proof of Theorem 1**

The following proposition unifies several different versions of the functional central limit theorem (FCLT).

**Proposition 1 (FCLT)** Let \(\epsilon_j\) satisfy Assumption 1 with moving average representation \(C(L) = \sum_{j=0}^{\infty} \epsilon_j L^j\), then

\[
\frac{1}{\sqrt{T}} \sum_{i=1}^{[\lambda T]} \epsilon_i \Rightarrow C(1)\sigma W(\lambda)
\]

as \(T \to \infty\), where \(W\) is a Wiener process on \([0,1]\).

The FCLT based on Assumption (a) is the invariance principle of Phillips and Solo (1992) Theorem 3.4 (b) p.978, and the FCLT based on Assumption (c) is the invariance principle of Phillips and Solo (1992) Theorem 3.15 (b) p.983. The invariance principle in

---

9This ratio test is denoted the d ’Alembert ratio. It states if \(\lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right| = L < 1\), then the series \(\sum_{n=1}^{\infty} a_n\) is absolutely convergent (and therefore convergent).
Stock (1994) is includes as the special case where \( C(L) = 1 \). Note that the \( C(1)^2 \sigma^2_\epsilon \) is the long-run variance of \( \epsilon_t \).

The FCLT result for the partial sums of \( \{\epsilon_j\} \) gives

\[
\frac{1}{\sqrt{T}} \sum_{i=1}^{[\delta T]} \epsilon_{[\delta T]+i} = \frac{1}{\sqrt{T}} \sum_{i=1}^{[\delta+\lambda T]} \epsilon_i - \frac{1}{\sqrt{T}} \sum_{i=1}^{[\lambda T]} \epsilon_i \\
\Rightarrow \sigma P_\delta(\lambda) \equiv \sigma (W(\lambda + \delta) - W(\lambda)),
\]

as \( T \to \infty \) and \( K/T \to \delta \). Here, \( W \) is a Brownian motion restricted to the interval \([0, 1]\) and \( \sigma^2 \) is the long-run variance defined by

\[
\sigma^2 = C(1)^2 \sigma^2_\epsilon = \sigma^2_\epsilon \sum_{j=0}^{\infty} c_j = 2\pi f(0),
\]

where \( f(\omega) \) denotes the spectrum of \( \{\epsilon_j\} \).

The proof of Theorem 1 applies the FCLT and Continuous Mapping Theorem (CMT) to obtain the limiting distribution of the partial sum \( \frac{1}{\sqrt{T}} \sum_{i=1}^{J(t)} \omega_{t+i} \) multiplied by a deterministic function \( \frac{T}{J(t)} \to \frac{1}{\psi(\lambda)} \) as \( T \to \infty \).

**Proof of Theorem 1 (Limiting Distribution).** Consider the OLS estimate of the slope coefficients in the variance regression (28).

\[
\begin{bmatrix}
\hat{\gamma}_i \\
\hat{\gamma}_p
\end{bmatrix}
= \begin{bmatrix}
\frac{1}{T} \sum_{t=H}^{K} (\sigma^2_{\epsilon, t}(J(t)))^2 & \sum_{t=H}^{K} \sigma^2_{i, t}(J(t)) \sigma^2_{p, t-H} \\
\sum_{t=H}^{K} \sigma^2_{i, t}(J(t)) \sigma^2_{p, t-H} & \sum_{t=H}^{K} (\sigma^2_{p, t-H})^2
\end{bmatrix}
^{-1}
\begin{bmatrix}
\sum_{t=H}^{K} \bar{\sigma}^2_{i, t}(J(t)) \bar{\sigma}^2_{i, t}(J(t)) \\
\sum_{t=H}^{K} \bar{\sigma}^2_{p, t-H}(J(t)) \bar{\sigma}^2_{i, t}(J(t))
\end{bmatrix}
= \begin{bmatrix}
\frac{1}{T} \sum_{t=H}^{K} (\sqrt{T} \bar{\sigma}^2_{i, t}(J(t)))^2 & \frac{1}{T} \sum_{t=H}^{K} \sqrt{T} \bar{\sigma}^2_{i, t}(J(t)) \sqrt{T} \bar{\sigma}^2_{p, t-H} \\
\frac{1}{T} \sum_{t=H}^{K} \sqrt{T} \bar{\sigma}^2_{i, t}(J(t)) \sqrt{T} \bar{\sigma}^2_{p, t-H} & \frac{1}{T} \sum_{t=H}^{K} (\sqrt{T} \bar{\sigma}^2_{p, t-H})^2
\end{bmatrix}
^{-1}
\begin{bmatrix}
\frac{1}{T} \sum_{t=H}^{K} \sqrt{T} \bar{\sigma}^2_{i, t}(J(t)) \sqrt{T} \bar{\sigma}^2_{i, t}(J(t)) \\
\frac{1}{T} \sum_{t=H}^{K} \sqrt{T} \bar{\sigma}^2_{p, t-H} \sqrt{T} \bar{\sigma}^2_{i, t}(J(t))
\end{bmatrix},
\]
where
\[ \tilde{\sigma}^2_{r,t}(J(t)) \equiv \sigma^2_{r,t}(J(t)) - \frac{1}{T - (K + H) + 1} \sum_{k=H}^{T-K} \sigma^2_{r,k}(J(t)), \] (37)

\[ \tilde{\sigma}^2_{i,t}(J(t)) \equiv \sigma^2_{i,t}(J(t)) - \frac{1}{T - (K + H) + 1} \sum_{k=H}^{T-K} \sigma^2_{i,k}(J(t)), \] (38)

\[ \tilde{\sigma}^2_{p,t-H} \equiv \sigma^2_{p,t-H} - \frac{1}{T - (K + H) + 1} \sum_{k=H}^{T-K} \sigma^2_{p,k-H}. \] (39)

The historical variance includes the \( H \) days prior squared returns and the realized variance is an average of at most \( K \) days squared return. Accordingly, the regression is based on \( T - (K + H) + 1 \) daily observations. Note that the processes in (11)-(13) in Theorem 1 are obtained from those in (14)-(16) in a manner that corresponds to the mean adjustments (37)-(39). Furthermore, the processes (14)-(16) are the limits of the dispersion measures (37)-(39) less their means. The mean-corrected variables (37)-(39) consist of partial sums of the zero-mean variables.

Consider the rescaled mean corrected implied variance
\[ \sqrt{T} \tilde{\sigma}^2_{i,t}(J(t)) = \sqrt{T} \sigma^2_{i,t}(J(t)) - \frac{1}{T - (K + H) + 1} \sum_{k=H}^{T-K} \sqrt{T} \sigma^2_{i,k}(J(k)) \]

\[ = \frac{1}{\sqrt{T} J(t)} \sum_{j=1}^{J(t)} \omega_{t+j} \frac{1}{1 - (K + H)/T + 1/T} \sum_{k=H}^{T-K} \frac{1}{\sqrt{T} J(k)} \sum_{j=1}^{J(k)} \omega_{k+j} \frac{1}{T} \]

\[ = \frac{1}{\sqrt{T} J(t)} \sum_{j=1}^{J(t)} \omega_{\lambda T+j} \]

\[ - \frac{1}{1 - (K + H)/T + 1/T} \int_{H/T}^{1-K/T} \frac{1}{\sqrt{T} \psi(s)} \sum_{j=1}^{[\psi(s)T]} \omega_{[sT]+j} ds, \]

where the second sign of equation is due to the fact that the mean daily return \( \mu_t \) cancels out. The rescaled mean corrected realized variance and mean corrected historical variance can be rewritten similarly
\[ \sqrt{T} \tilde{\sigma}^2_{r,t}(J(t)) = \frac{1}{\sqrt{T} J(t)} \sum_{j=1}^{J(t)} (\omega_{[\lambda T]+j} + \nu_{[\lambda T]+j}) \]

\[ - \frac{1}{1 - (K + H)/T + 1/T} \int_{H/T}^{1-K/T} \frac{1}{\sqrt{T} \psi(s)} \sum_{j=1}^{[\psi(s)T]} (\omega_{[sT]+j} + \nu_{[sT]+j}) ds \]

The historical variance includes the \( H \) days prior squared returns and the realized variance is an average of at most \( K \) days squared return. Accordingly, the regression is based on \( T - (K + H) + 1 \) daily observations. Note that the processes in (11)-(13) in Theorem 1 are obtained from those in (14)-(16) in a manner that corresponds to the mean adjustments (37)-(39). Furthermore, the processes (14)-(16) are the limits of the dispersion measures (37)-(39) less their means. The mean-corrected variables (37)-(39) consist of partial sums of the zero-mean variables.

Consider the rescaled mean corrected implied variance
\[ \sqrt{T} \tilde{\sigma}^2_{i,t}(J(t)) = \sqrt{T} \sigma^2_{i,t}(J(t)) - \frac{1}{T - (K + H) + 1} \sum_{k=H}^{T-K} \sqrt{T} \sigma^2_{i,k}(J(k)) \]

\[ = \frac{1}{\sqrt{T} J(t)} \sum_{j=1}^{J(t)} \omega_{t+j} \frac{1}{1 - (K + H)/T + 1/T} \sum_{k=H}^{T-K} \frac{1}{\sqrt{T} J(k)} \sum_{j=1}^{J(k)} \omega_{k+j} \frac{1}{T} \]

\[ = \frac{1}{\sqrt{T} J(t)} \sum_{j=1}^{J(t)} \omega_{\lambda T+j} \]

\[ - \frac{1}{1 - (K + H)/T + 1/T} \int_{H/T}^{1-K/T} \frac{1}{\sqrt{T} \psi(s)} \sum_{j=1}^{[\psi(s)T]} \omega_{[sT]+j} ds, \]

where the second sign of equation is due to the fact that the mean daily return \( \mu_t \) cancels out. The rescaled mean corrected realized variance and mean corrected historical variance can be rewritten similarly
\[ \sqrt{T} \tilde{\sigma}^2_{r,t}(J(t)) = \frac{1}{\sqrt{T} J(t)} \sum_{j=1}^{J(t)} (\omega_{[\lambda T]+j} + \nu_{[\lambda T]+j}) \]

\[ - \frac{1}{1 - (K + H)/T + 1/T} \int_{H/T}^{1-K/T} \frac{1}{\sqrt{T} \psi(s)} \sum_{j=1}^{[\psi(s)T]} (\omega_{[sT]+j} + \nu_{[sT]+j}) ds \]

The historical variance includes the \( H \) days prior squared returns and the realized variance is an average of at most \( K \) days squared return. Accordingly, the regression is based on \( T - (K + H) + 1 \) daily observations. Note that the processes in (11)-(13) in Theorem 1 are obtained from those in (14)-(16) in a manner that corresponds to the mean adjustments (37)-(39). Furthermore, the processes (14)-(16) are the limits of the dispersion measures (37)-(39) less their means. The mean-corrected variables (37)-(39) consist of partial sums of the zero-mean variables.

Consider the rescaled mean corrected implied variance
\[ \sqrt{T} \tilde{\sigma}^2_{i,t}(J(t)) = \sqrt{T} \sigma^2_{i,t}(J(t)) - \frac{1}{T - (K + H) + 1} \sum_{k=H}^{T-K} \sqrt{T} \sigma^2_{i,k}(J(k)) \]

\[ = \frac{1}{\sqrt{T} J(t)} \sum_{j=1}^{J(t)} \omega_{t+j} \frac{1}{1 - (K + H)/T + 1/T} \sum_{k=H}^{T-K} \frac{1}{\sqrt{T} J(k)} \sum_{j=1}^{J(k)} \omega_{k+j} \frac{1}{T} \]

\[ = \frac{1}{\sqrt{T} J(t)} \sum_{j=1}^{J(t)} \omega_{\lambda T+j} \]

\[ - \frac{1}{1 - (K + H)/T + 1/T} \int_{H/T}^{1-K/T} \frac{1}{\sqrt{T} \psi(s)} \sum_{j=1}^{[\psi(s)T]} \omega_{[sT]+j} ds, \]

where the second sign of equation is due to the fact that the mean daily return \( \mu_t \) cancels out. The rescaled mean corrected realized variance and mean corrected historical variance can be rewritten similarly
\[ \sqrt{T} \tilde{\sigma}^2_{r,t}(J(t)) = \frac{1}{\sqrt{T} J(t)} \sum_{j=1}^{J(t)} (\omega_{[\lambda T]+j} + \nu_{[\lambda T]+j}) \]

\[ - \frac{1}{1 - (K + H)/T + 1/T} \int_{H/T}^{1-K/T} \frac{1}{\sqrt{T} \psi(s)} \sum_{j=1}^{[\psi(s)T]} (\omega_{[sT]+j} + \nu_{[sT]+j}) ds \]
\[
\sqrt{T} \sigma_{p,t}^2 = \frac{1}{\sqrt{T} H} \sum_{j=1}^{H} (\omega_{\lfloor \lambda T \rfloor + j} + \nu_{\lfloor \lambda T \rfloor + j}) 
- \frac{1}{1 - (K + H)/T + 1/T} \int_{H/T}^{1-K/T} \left( \frac{1}{\sqrt{T}} \sum_{j=1}^{\lfloor \kappa T \rfloor} (\omega_{\lfloor \lambda T \rfloor + j} + \nu_{\lfloor \lambda T \rfloor + j}) \right) ds.
\]

Therefore, we first need to find the convergence of the partial sums of the innovations. Consider the partial sum of the innovation anticipated by the option market, \(\omega_t\) which satisfies 1 with moving average representation \(C(L) = \sum_{\ell=0}^{\infty} e^{\ell L} J(t)\)

\[
\frac{T}{J(t) \sqrt{T}} \sum_{i=1}^{J(t)} \omega_{t+i} = \frac{T}{J(t)} \left( \frac{1}{\sqrt{T}} \sum_{i=1}^{t+J(t)} \omega_i - \frac{1}{\sqrt{T}} \sum_{i=1}^{t} \omega_i \right)
= \frac{T}{J(t)} \left( \frac{1}{\sqrt{T}} \sum_{i=1}^{\lfloor (\lambda + \psi(\lambda)) T \rfloor} \omega_i - \frac{1}{\sqrt{T}} \sum_{i=1}^{\lfloor \lambda T \rfloor} \omega_i \right)
\Rightarrow I_{\psi(\lambda)}(\lambda) \equiv \frac{\sigma_\omega}{\psi(\lambda)} (W^0(\psi(\lambda) + \lambda) - W^0(\lambda))
\]
as \(t/T \to \lambda, J(t)/T \to \psi(\lambda),\) as \(T \to \infty.\) Here \(\Rightarrow\) denotes convergence in distribution, \(\sigma_\omega^2 = C(1)^2 \sigma_\nu^2\) is the long-run variance of \(\omega_t\) as in Assumption 1, and \(W^0(\cdot)\) is a standard Brownian motion restricted to the \([0,1]\) interval. The term \(I_{\psi(\lambda)}(\lambda)\) is the limiting distribution of the innovations \(\omega_t\) multiplied by the fraction \(1/\psi(\lambda).\) Using this result with the CMT immediately gives us

\[
\sqrt{T} \sigma_{i,t}^2(J(t)) \Rightarrow \tilde{I}_{\psi(\lambda)}(\lambda) \equiv I_{\psi(\lambda)}(\lambda) - \frac{1}{1 - \delta(1 + \kappa)} \int_{\kappa \delta}^{1-\delta} I_{\psi(s)}(s) ds,
\]
as \(t/T \to \lambda, J(t)/T \to \psi(\lambda),\) as \(T \to \infty,\) and in particular, \(K/T \to \delta, H/T \to \kappa \delta.\)
Now consider the partial sum of both innovations $\omega_t$ and $\nu_t$
\[
\frac{1}{\sqrt{T}} \frac{T}{J(t)} \sum_{i=1}^{J(t)} (\omega_{t+i} + \nu_{t+i}) = \frac{T}{J(t)} \left( \frac{1}{\sqrt{T}} \sum_{i=1}^{t/J(t)} (\omega_i + \nu_i) - \frac{1}{\sqrt{T}} \sum_{i=1}^{t} (\omega_i + \nu_i) \right)
\]
\[
= \frac{T}{J(t)} \left( \frac{1}{\sqrt{T}} \sum_{i=1}^{[\lambda \psi(t)]} \omega_i - \frac{1}{\sqrt{T}} \sum_{i=1}^{[\lambda T]} \omega_i \right)
+ \frac{T}{J(t)} \left( \frac{1}{\sqrt{T}} \sum_{i=1}^{[\lambda \psi(t)]} \nu_i - \frac{1}{\sqrt{T}} \sum_{i=1}^{[\lambda T]} \nu_i \right)
\]
\[
\Rightarrow P_{\psi(t)}(\lambda)
= \frac{1}{\psi(t)} \left( \sigma_{\omega} W^0(\delta + \psi(t)) + \sigma_{\omega} W^0(\lambda) \right)
+ \frac{1}{\psi(t)} \left( \sigma_{\nu} W^1(\psi(t) + \lambda) - \sigma_{\nu} W^1(\lambda) \right),
\]

as $t/T \to \lambda$, $J(t)/T \to \psi(t)$, as $T \to \infty$, where $\sigma_{\omega}^2$ and $\sigma_{\nu}^2$ are the long-run variances of $\omega_t$ and $\nu_t$, respectively, $W^0$ and $W^1$ are Brownian motions restricted to $[0,1]$. Note that $P_{\psi(t)}(\lambda) = I_{\psi(t)}(\lambda) + R_{\psi(t)}(\lambda)$, where $I_{\psi(t)}(\lambda)$ is defined in (40) and $R_{\psi(t)}(\lambda) \equiv \sigma_{\psi(t)}(\lambda) (W^1(\psi(t) + \lambda) - W^1(\lambda))$.

Furthermore,
\[
\frac{1}{H} \frac{1}{\sqrt{T}} \sum_{i=1}^{H} (\omega_{t+i} + \nu_{t+i})
= \frac{T}{H} \left( \frac{1}{\sqrt{T}} \sum_{i=1}^{[\kappa \delta + \lambda]} \omega_i - \frac{1}{\sqrt{T}} \sum_{i=1}^{[\lambda]} \omega_i \right)
+ \frac{T}{H} \left( \frac{1}{\sqrt{T}} \sum_{i=1}^{[\kappa \delta + \lambda]} \nu_i - \frac{1}{\sqrt{T}} \sum_{i=1}^{[\lambda]} \nu_i \right)
\]
\[
\Rightarrow P_{\kappa \delta}^*(\lambda) \equiv \frac{\sigma_{\omega}}{\kappa \delta} (W^0(\kappa \delta + \lambda) - W^0(\lambda))
+ \frac{\sigma_{\nu}}{\kappa \delta} (W^1(\kappa \delta + \lambda) - W^1(\lambda)),
\]

as $t/T \to \lambda$, $J(t)/T \to \psi(t)$, $K/T \to \delta$, $H/T \to \kappa \delta$ as $T \to \infty$.

It now follows that the mean-corrected realized variance converges in distribution
\[
\sqrt{T} \sigma_{r,t}^2(J(t)) \Rightarrow \tilde{P}_{\psi(t)}(\lambda) \equiv P_{\psi(t)}(\lambda) - \frac{1}{1 - \delta(1 + \kappa)} \int_{\kappa \delta}^{1-\delta} P_{\psi(s)}(s) \, ds
\]  
(42)
and the mean-corrected historical variance covering a fixed period of $H$ days, converges in distribution as follows
\[
\sqrt{T} \sigma_{p,t-H}^2 \Rightarrow \tilde{P}_{\kappa \delta}^*(\lambda) \equiv P_{\kappa \delta}^*(\lambda) - \frac{1}{1 - \delta(1 + \kappa)} \int_{\kappa \delta}^{1-\delta} P_{\kappa \delta}^*(s - \kappa \delta) \, ds
\]  
(43)
as $J(t)/T \to \psi(t)$, $K/T \to \delta$, $H/T \to \kappa \delta$, $t/T \to \lambda$, as $T \to \infty$.  

26
By further use of the CMT we obtain the limiting distribution of the slope coefficients in the variance regression
\[
\begin{bmatrix}
\tilde{\gamma}_i \\
\tilde{\gamma}_p
\end{bmatrix} \sim \begin{bmatrix}
\int_{\kappa \delta}^{1-\delta} (I_{\psi(\lambda)}(\lambda))^2 d\lambda \\
\int_{\kappa \delta}^{1-\delta} T \tilde{\psi}_i(\lambda) \tilde{P}_{\nu \delta}(\lambda) d\lambda
\end{bmatrix}^{-1}
\begin{bmatrix}
\int_{\kappa \delta}^{1-\delta} \tilde{T} \tilde{I}_{\psi(\lambda)}(\lambda) \tilde{P}_{\nu \delta}(\lambda) d\lambda \\
\int_{\kappa \delta}^{1-\delta} \tilde{T} \tilde{P}_{\nu \delta}(\lambda) d\lambda
\end{bmatrix}
\]
as \(T \to \infty\) and \(J(t)/T \to \psi(\lambda), K/T \to \delta\) and \(H/T \to \kappa\delta\).

The limiting distributions the volatility regression (8) is derived as follows. We linearize the square root function using a first order Taylor expansion of \(\sqrt{\mu_i}\). Terms of order greater than one converge to zero as we normalize the expression by \(\sqrt{T}\),

\[
\sqrt{\mu_i} + \frac{1}{J(t)} \sum_{j=1}^{J(t)} \omega_{i,j} + \frac{1}{2\sqrt{\mu_i}} \frac{1}{J(t)} \sum_{j=1}^{J(t)} \omega_{i,j}
\]
for \(\frac{1}{J(t)} \sum_{j=1}^{J(t)} \omega_{i,j}\) close to 0, and the demeaned implied volatility is

\[
\tilde{\sigma}_{i,t}(J(t)) \approx \sqrt{\mu_i} + \frac{1}{2\sqrt{\mu_i}} \frac{1}{J(t)} \sum_{j=1}^{J(t)} \omega_{i,j} - \frac{1}{T - (K + H) + 1} \sum_{k=H}^{T-K} \left( \sqrt{\mu_i} + \frac{1}{2\sqrt{\mu_i}} \frac{1}{J(t)} \sum_{j=1}^{J(k)} \omega_{k,j} \right)
\]

\[
= \frac{1}{2\sqrt{\mu_i}} \left( \frac{1}{J(t)} \sum_{j=1}^{J(t)} \omega_{i,j} - \frac{1}{T - (K + H) + 1} \sum_{k=H}^{T-K} \left( \frac{1}{J(k)} \sum_{j=1}^{J(k)} \omega_{k,j} \right) \right).
\]

By FCLT and CMT we get

\[
\sqrt{T} \tilde{\sigma}_{i,t}(J(t)) \approx \frac{1}{2\sqrt{\mu_i}} \left( \frac{T}{J(t)} \frac{1}{\sqrt{T}} \sum_{j=1}^{J(t)} \omega_{i,j} - \frac{1}{1 - (K/T + H/T) + 1/T} \sum_{k=H}^{T-K} \left( \frac{T}{J(k)} \frac{1}{\sqrt{T}} \sum_{j=1}^{J(k)} \omega_{k,j} \right) \frac{1}{T} \right)
\]

\[
\Rightarrow \frac{1}{2\sqrt{\mu_i}} \left( I_{\psi(\lambda)}(\lambda) - \int_{\kappa \delta}^{1-\delta} I_{\psi(s)}(s) ds \right) \equiv \frac{1}{2\sqrt{\mu_i}} \tilde{I}_{\psi(\lambda)}(\lambda),
\]

In a similar way we can approximate the mean-corrected realized and historical volatilities by expansions around \(\mu_i\), and (17). The log-volatility regression (9) is handled similarly.
C Proof of Theorem 2

Proof of (23) OLS standard errors. The OLS variance-covariance matrix is given by

\[ \hat{\Sigma} = s^2 \left[ \sum_{t=H}^{T-K} \bar{\sigma}_t^2(J(t)) \right]^{-1}, \]

where

\[ s^2 = \frac{1}{T - (K + H) - 1} \sum_{t=H}^{T-K} \hat{v}_t(J(t))^2, \]

\[ \hat{v}_t(J(t)) = \bar{\sigma}_{r,t}^2(J(t)) - \bar{\sigma}_t^2(J(t)) \]

\[ \hat{v}_t(J(t)) = \hat{\gamma}_r^2(J(t)) - \hat{\gamma}_t^2(J(t)) = \hat{\gamma}_t^2(J(t)) - \hat{\gamma}_p^2(J(t)). \]

Here, \( \hat{v}_t(J(t)) \) denotes the residual from the regression (7). The regression includes two regressors and it is based on \( T - (K + H) + 1 \) observations, which leaves us with \( T - (K + H) - 1 \) degrees of freedom.

The limiting distribution results \( \hat{\gamma}' = [\hat{\gamma}, \hat{\gamma}_p] \Rightarrow \gamma^{\star}_\psi(\lambda) = [\gamma_{i,\psi}(\lambda), \gamma_{p,\psi}(\lambda)] \), from Theorem 1 and \( \sqrt{T} \bar{\sigma}_{r,t}^2(J) \Rightarrow \bar{I}_\psi(\lambda), \sqrt{T} \bar{\sigma}_{t}^2(J(t)) \Rightarrow \bar{P}_\psi(\lambda), \) and \( \sqrt{T} \bar{\sigma}_{r,t}^2(J(t)) \Rightarrow \bar{P}_{\psi}(\lambda - \kappa \delta), \) from equations (40), (42), and (44) imply that for \( \lambda \geq 2\delta, \)

\[ \sqrt{T} \hat{v}_{[\lambda T]}([\psi(\lambda)T]) \Rightarrow V_\psi(\lambda), \]

where

\[ V_\psi(\lambda) = \bar{P}_\psi(\lambda) - [\bar{I}_\psi(\lambda), \bar{P}_{\psi}(\lambda - \kappa \delta)] \gamma_\psi(\lambda), \] for \( \lambda \geq 2\delta. \]

This result in conjunction with the CMT yield the following weak convergence results

\[ \sum_{t=H}^{T-K} \bar{\sigma}_t^2(J(t)) \Rightarrow \Gamma_{\sigma}(0, \psi(\lambda)), \]

\[ \Gamma_{\sigma}(\rho, \psi(\lambda)) = \int_{\kappa \delta + |\rho|}^{1-\delta} \left[ \bar{I}_\psi(\lambda), \bar{P}_{\psi}(\lambda - \kappa \delta) \right] \left[ \bar{I}_{\psi}(\lambda - |\rho|), \bar{P}_{\psi}(\lambda - \kappa \delta - |\rho|) \right] d\lambda, \]

\[ T \cdot s^2 = \frac{T}{T - (K + H) - 1} \sum_{t=H}^{T-K} \hat{v}_t(J(t))^2 \Rightarrow \Gamma_{\nu}(0, \psi(\lambda)), \]

\[ \Gamma_{\nu}(\rho, \psi(\lambda)) = \frac{1}{1 - \delta(1 + \kappa)} \int_{\kappa \delta + |\rho|}^{1-\delta} V_\psi(\lambda) V_\psi(\lambda - |\rho|) d\lambda. \]

It now follows from further application of the CMT
Newey and West (1987) variance-covariance matrix of the OLS estimator that
and
The weight function is assumed to satisfy

\[
\hat{J}_v(t) \to \psi((\lambda)) \Rightarrow \Gamma_v(0, \psi(\lambda)) [\Gamma_\sigma(0, \psi(\lambda))]^{-1} = \Xi^*_\psi(\lambda)
\]  

(45)
as $J(t)/T \to \psi(\lambda)$, $K/T \to \delta$, $H/T \to \kappa \delta$, and $T \to \infty$. ■

**Proof of (23) Alternative standard errors.** The weighted Hansen (1982) or Newey and West (1987) variance-covariance matrix of the OLS estimator $\gamma$, allowing $K$ periods correlation, is given by

\[
\hat{\Psi}(J(t)) = \hat{\vartheta}(0)^{-1} \hat{\Psi}(J(t)) \hat{\vartheta}(0)^{-1}, \quad \text{where}
\]

\[
\hat{\Psi}(J(t)) = \frac{1}{T - (K + H) + 1} \sum_{j=-K}^{K} \hat{\Gamma}_T(j, J(t), K), \quad \text{and}
\]

\[
\hat{\Gamma}_T(j, J(t), K) = \sum_{t=H+j}^{T-K} k_T \hat{\psi}_t(J(t)) \hat{\psi}_{t-j}(J(t - j)) \hat{\sigma}^2_t(J(t)) \hat{\sigma}^2_{t-j}(J(t - j)),
\]

\[
k_T = k_T \left( \frac{J(t)}{T}, \frac{K}{T} \right) \text{ for simplification},
\]

\[
\hat{\vartheta}(j) = \sum_{t=H+j}^{T-K} \hat{\sigma}^2_t(J(t)) \hat{\sigma}^2_{t-j}(J(t - j)),
\]

where $\hat{\psi}_t(J(t)) = \hat{\sigma}^2_t(J(t)) - \hat{\sigma}^2_t(J(t)) \hat{\gamma}(J(t))$ again denotes the OLS regression residual.

The weight function is assumed to satisfy $k_T \left( \frac{J(t)}{T}, \frac{K}{T} \right) \to k(\rho, \psi(\lambda), \delta)$ as $i/T \to \rho$, $J(t)/T \to \psi(\lambda)$, $K/T \to \delta$, and $T \to \infty$. For $j/T \to \rho$, these results and CMT imply that

\[
T \left[ \hat{\sigma}^2_t(J(t)) \hat{\sigma}^2_{t-j}(J(t - j)) \right] \Rightarrow \Upsilon_\sigma(\rho, \psi(\lambda)), \quad \text{where}
\]

\[
\Upsilon_\sigma(\rho, \psi(\lambda)) = \begin{bmatrix}
\bar{I}_{\psi}(\lambda) \bar{I}_{\psi}(\lambda)(\lambda - |\rho|) & \bar{I}_{\psi}(\lambda) \bar{P}^*_\sigma(\lambda - \kappa \delta - |\rho|) \\
\bar{P}^*_\sigma(\lambda - \kappa \delta) & \bar{P}^*_\sigma(\lambda - \kappa \delta - |\rho|)
\end{bmatrix},
\]

and

\[
\hat{\vartheta}(j) \Rightarrow \vartheta(\rho, \psi(\lambda), \delta) = \int_{\kappa \delta + |\rho|}^{1-\delta} \Upsilon_\sigma(\rho, \psi(\lambda)) d\lambda,
\]

\[
T \cdot \hat{\Gamma}(J(t), J_{\max}) \Rightarrow \Gamma(\rho, \psi(\lambda), \kappa \delta),
\]

\[
\hat{\Gamma}(\rho, \psi(\lambda), \kappa \delta) = \int_{\kappa \delta + \rho}^{1-\delta} k(\rho, \psi(\lambda), \kappa \delta) \Upsilon_\psi(\rho, \psi(\lambda)) \Upsilon_\sigma(\rho, \psi(\lambda)) d\lambda.
\]
\[ T \cdot \hat{\Psi}(J(t)) \Rightarrow \frac{1}{1 - 2\delta} \int_{-\delta}^{\delta} \Gamma(\rho, \psi(\lambda), \kappa\delta)d\rho. \]

It follows from CMT that

\[ T \cdot \hat{\Xi}(J(t)) \Rightarrow \vartheta(0, \psi(\lambda), \delta)^{-1} \left[ \frac{1}{1 - 2\delta} \int_{-\delta}^{\delta} \Gamma(\rho, \psi(\lambda), \kappa\delta)d\rho \right] \vartheta(0, \psi(\lambda), \delta)^{-1} \equiv \Xi^*_{\psi(\lambda)}. \quad (47) \]

The weighted Hansen (1982) standard errors do not converge in probability to a constant but rather, after multiplying by the \( T \), converges weakly to a functional of Brownian motions. This result is similar to what we obtained for the OLS standard errors. Note that the limiting weight function, \( k \), is arbitrary.

**Proof of (24) t-values** Since the OLS and the alternative standard errors diverge as \( T \to \infty \), the \( t \)-ratios based on these standard errors diverge. Let \( \hat{\Xi}_i(J(t)) \) denote the variance estimator of \( \hat{\gamma}_i \), and \( \hat{\Xi}_p(J(t)) \) denotes the variance estimator of \( \hat{\gamma}_p \), with limiting distributions \( \Xi^*_{i,\psi(\lambda)} \) and \( \Xi^*_{p,\psi(\lambda)} \), respectively. Each \( t \)-ratio multiplied by \( T^{-1/2} \)

\[
T^{-1/2} t_i = \frac{\hat{\gamma}_i}{\sqrt{\hat{\Xi}_i(J(t))T}} \Rightarrow \gamma^*_{i,\psi(\lambda)} \Xi^{-1/2}_{i,\psi(\lambda)},
\]

\[
T^{-1/2} t_p = \frac{\hat{\gamma}_p}{\sqrt{\hat{\Xi}_p(J(t))T}} \Rightarrow \gamma^*_{p,\kappa\delta} \Xi^{-1/2}_{p,\psi(\lambda)},
\]

converges towards a functional of Brownian motions, using either OLS standard errors or weighted Hansen (1982) standard errors. Thus, both \( t \)-ratios diverge as \( T \to \infty \), a result similar to the spurious regression model of Phillips (1986).

**Proof of (25) and (26): \( R^2 \), Adjusted \( R^2 \).** The coefficient of determination

\[
R^2 = \frac{\sum_{t=J}^{T-K} (\hat{\gamma}_i^2 \hat{\sigma}_{i,t}^2 + \hat{\gamma}_p^2 \hat{\sigma}_{p,t-H}^2)^2}{\sum_{t=H}^{T-K} (\hat{\sigma}_{t}^2)^2} \Rightarrow R_{2*}^2 = \int_{\kappa\delta}^{1-\delta} \left( \gamma^*_i \tilde{T}(\psi(\lambda)) + \gamma^*_p \tilde{P}(\kappa\delta(\lambda - \kappa\delta)) \right)^2 d\lambda
\]

as \( J(t)/T \to \psi(\lambda) \), \( K/T \to \delta \), \( H/T \to \kappa\delta \), and \( T \to \infty \). The dependence of \( J(t) \) in the variances is suppressed to simplify the expression. In the empirical part of the paper we have reported the adjusted \( R^2 \). The two statistics are related as adjusted \( R^2 = 1 - \frac{T-(K+H)}{T-(K+H)-1}(1 - R^2) \).\(^{10}\) Consequently, the adjusted coefficient of determination

\(^{10}\)See Greene (2000, Chapter 6) for an overview of the coefficient of determination.
also converges in distribution to the stochastic variable \( R^2 \).

**Proof of (27) Durbin-Watson statistic.** Consider the Durbin-Watson statistic

\[
DW = \frac{\sum_{t=H+1}^{T-K} (\hat{v}_t - \hat{v}_{t-1})^2}{\sum_{t=H}^{T-K} \hat{v}_t^2},
\]

where \( \hat{v}_t = \hat{v}_t(J(t)) \) is the regression residual from Eq.(7). Let \( \hat{\sigma}^2 = \begin{bmatrix} \hat{\sigma}^2_{i,t}, \hat{\sigma}^2_{p,t-H} \end{bmatrix} \) and \( \hat{\gamma}' = [\hat{\gamma}_1, \hat{\gamma}_p] \), and \( z_t = [\hat{\sigma}^2_{r,t}, \hat{\sigma}^2_{i,t}, \hat{\sigma}^2_{p,t-H}] \) and \( j_t = [J(t), J(t), H] \) then \( z_t = \alpha z_{t-1} + \xi_t \) where \( \xi_t = [-(\omega_t + \nu_t), -\omega_t, -(\omega_t-H + \nu_t-H)] \) and \( \alpha = [J(t-1)/J(t), J(t-1)/J(t), 1] \). Since \( J(t-1) = J(t) + 1 \), we approximate \( \alpha \) by the vector \((1, 1, 1)\), which is a good approximation when \( J(t) \) is large. Now

\[
T^{-1} \sum_{t=H+1}^{T-K} (\hat{v}_t - \hat{v}_{t-1})^2 = T^{-1} \sum_{t=H+1}^{T-K} (\tilde{\sigma}^2_{r,t} - \tilde{\sigma}^2_{r,t-1} - (\hat{\sigma}^2_t - \tilde{\sigma}^2_{t-1})\hat{\gamma})^2
\]

\[
= \hat{\gamma}' T^{-1} \sum_{t=H+1}^{T-K} (\xi_t \xi_t') \hat{\gamma}.
\]

where \( \hat{\gamma}' = (1, -\hat{\gamma}) \). Furthermore, \( \omega_t \) and \( \nu_t \) satisfy Assumptions 1, therefore \( T^{-1} \sum_{t=H+1}^{T-K} (\xi_t \xi_t') \rightarrow a.s. \)

\[
T^{-1} \sum_{t=H}^{T-K} E(\xi_t \xi_t') = \Sigma_\xi \text{ as } T \rightarrow \infty.
\]

Finally, as \( J(t)/T \rightarrow \psi(\lambda), K/T \rightarrow \delta, \text{ and } T \rightarrow \infty \)

\[
\sum_{t=H}^{T-K} \hat{\sigma}^2_t = \sum_{t=H}^{T-K} (\tilde{\sigma}^2_{r,t} - (\hat{\gamma}_i \hat{\sigma}^2_{i,t} + \hat{\gamma}_p \hat{\sigma}^2_{p,t-H}))^2
\]

\[
\Rightarrow \int_{\kappa \delta}^{1-\delta} (\tilde{P}_\psi(\lambda)(\lambda))^2 - (\gamma_i^* \tilde{I}_\psi(\lambda)(\lambda) + \gamma_p^* \tilde{P}_\rho^*(\lambda - \kappa \delta))^2 d\lambda
\]

and

\[
T \cdot DW \Rightarrow \frac{\hat{\gamma}' \Sigma_\xi \hat{\gamma}^*}{\int_{\kappa \delta}^{1-\delta} (\tilde{P}_\psi(\lambda)(\lambda))^2 - (\gamma_i^* \tilde{I}_\psi(\lambda)(\lambda) + \gamma_p^* \tilde{P}_\rho^*(\lambda - \kappa \delta))^2 d\lambda},
\]

where \( \hat{\gamma}' = (1, -\hat{\gamma}) \), so the Durbin-Watson statistic converges to zero as \( T \rightarrow \infty \).
References


Table 1: Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Skewness</th>
<th>Kurto-</th>
<th>Auto Corr Coeffs</th>
<th>Unit Root</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>sisis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\rho_1$</td>
<td>$\rho_5$</td>
<td>$\rho_{10}$</td>
</tr>
</tbody>
</table>

**Panel A: November 1983 to March 1987**

**Daily**
- Realized: 0.1434, 0.0484, 1.06, 5.21, 0.90, 0.57, 0.32, -5.68
- Implied: 0.1558, 0.0432, 1.45, 11.69, 0.66, 0.48, 0.41, -5.60
- Historical: 0.1423, 0.0265, 0.24, 2.61, 0.99, 0.94, 0.88, -1.31
- Log-Realized: -1.985, 0.2934, -0.10, 3.62, 0.90, 0.60, 0.37, -5.43
- Log-Implied: -1.896, 0.2744, -0.31, 4.21, 0.78, 0.61, 0.53, -4.73
- Log-Historical: -1.968, 0.1887, -0.18, 2.48, 0.99, 0.95, 0.89, -1.39

**Monthly**
- Realized: 0.1452, 0.0407, 1.18, 5.49, 0.15, 0.22, -0.02, -2.95
- Implied: 0.1559, 0.0389, 0.21, 2.37, 0.52, -0.11, -0.32, -2.59
- Log-Realized: -2.173, 0.3857, 0.10, 3.50, 0.94, 0.40, 0.27, -3.30
- Log-Implied: -1.973, 0.2828, 0.33, 2.76, 0.94, 0.85, 0.79, -4.65
- Log-Historical: -2.118, 0.3273, 0.40, 2.97, 0.99, 0.97, 0.94, -3.06

**Panel B: February 1988 to December 1995**

**Daily**
- Realized: 0.1229, 0.0521, 1.87, 9.65, 0.95, 0.76, 0.58, -6.76
- Implied: 0.1448, 0.0434, 1.06, 4.10, 0.93, 0.84, 0.78, -4.80
- Historical: 0.1272, 0.0455, 1.46, 6.19, 0.99, 0.95, 0.91, -4.58
- Log-Realized: -2.173, 0.3857, 0.18, 3.50, 0.94, 0.78, 0.62, -6.41
- Log-Implied: -1.973, 0.2828, 0.33, 2.76, 0.94, 0.85, 0.79, -4.65
- Log-Historical: -2.118, 0.3273, 0.40, 2.97, 0.99, 0.97, 0.94, -3.06

**Monthly**
- Realized: 0.1224, 0.0461, 1.20, 4.80, 0.54, 0.35, 0.24, -4.26
- Implied: 0.1418, 0.0402, 0.87, 3.19, 0.76, 0.52, 0.41, -3.50
- Log-Realized: -2.165, 0.3584, 0.10, 3.01, 0.57, 0.40, 0.27, -3.30
- Log-Implied: -1.991, 0.2708, 0.31, 2.48, 0.76, 0.58, 0.47, -2.89

Descriptive statistic for the time series of realized, implied, and historical volatility, and their logarithms, reported for daily overlapping and monthly non-overlapping data. The implied volatility is computed each day using a binomial tree method for the at-the-money call OEX option with time to maturity between 7 and 35 calendar days. Realized volatility is the annualized ex-post daily return volatility (sample standard deviation) of the S&P 100 index returns over the remaining life of the option. The reported statistics are: mean, standard deviation, skewness, excess kurtosis, autocorrelations, and the augmented Dickey-Fuller unit root test (with an intercept, and 4 lags in the daily time series and 1 and 2 lags in the monthly time-series respectively in the pre and post-crash period), where $a$, $b$, and $c$ indicates that the parameter is significantly different from zero at a 1%, 5%, and 10% level of significance, respectively.

Data in the first subperiod consist of 784 daily and 41 monthly observations on each volatility series, and data in the second subperiod consist of 1,852 daily and 95 monthly observations on each volatility series.
Table 2: The relation between implied and realized volatility

November 1983 to March 1987

<table>
<thead>
<tr>
<th>Panel A: Volatilities</th>
<th>Daily overlapping data</th>
<th>Monthly non-overlapping data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intercept</td>
<td>Implied</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.109$^a$</td>
<td>0.223$^a$</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.064)</td>
</tr>
<tr>
<td></td>
<td>0.073$^a$</td>
<td>0.492$^a$</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.101)</td>
</tr>
<tr>
<td></td>
<td>0.069$^a$</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.066)</td>
</tr>
<tr>
<td>Panel B: Log-volatilities</td>
<td>-1.369$^a$</td>
<td>0.325$^a$</td>
</tr>
<tr>
<td></td>
<td>(0.126)</td>
<td>(0.066)</td>
</tr>
<tr>
<td></td>
<td>-0.883$^a$</td>
<td>0.560$^a$</td>
</tr>
<tr>
<td></td>
<td>(0.198)</td>
<td>(0.102)</td>
</tr>
<tr>
<td></td>
<td>-0.844$^a$</td>
<td>0.155</td>
</tr>
<tr>
<td></td>
<td>(0.192)</td>
<td>(0.081)</td>
</tr>
</tbody>
</table>

$a$ $p$-value $\leq 0.01$.

$b$ $p$-value $\in [0.01; 0.05]$.

This table reports the least squares estimates of the regressions

Panel A: $\sigma_{r,t} = \gamma_0 + \gamma_1 \sigma_{i,t} + \gamma_p \sigma_{p,t-60} + \epsilon_t$,

Panel B: $\ln(\sigma_{r,t}) = \gamma_0 + \gamma_1 \ln(\sigma_{i,t}) + \gamma_p \ln(\sigma_{p,t-60}) + \epsilon_t$,

where $\sigma_{r,t}$ denotes the realized volatility over the period from date $t$ to the nearest date $\tau$ on which the OEX option expires, $\sigma_{p,t-60}$ denotes the past realized 60-day volatility, and $\sigma_{i,t}$ denotes the binomial tree implied volatility of the at-the-money call option expiring at $\tau$. $\ln(\cdot)$ is the natural logarithm. For the monthly frequency, only $t$ corresponding to the Wednesday following the previous expiration date $\tau$ are included, and $\sigma_{p,t-60}$ is replaced by the lagged dependent variable. The numbers in brackets are the Newey-West heteroskedasticity consistent standard errors. DW denotes the Durbin-Watson statistic. Data consist of 784 daily and 41 monthly observations from November 1983 to March 1987.
### Table 3: Post-crash period regressions

February 1988 to December 1995

<table>
<thead>
<tr>
<th></th>
<th>Daily overlapping data</th>
<th>Monthly non-overlapping data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Past Adj.</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>Implied Realized R²</td>
<td>Intercept Implied Realized R²</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Intercepts</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Implied Realized R²</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel A: Volatilities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.023ᵃ</td>
<td>0.693ᵃ</td>
<td>33% 0.24</td>
</tr>
<tr>
<td>(0.008)</td>
<td>(0.055)</td>
<td>(0.012) (0.087)</td>
</tr>
<tr>
<td>0.047ᵃ</td>
<td>0.597ᵃ</td>
<td>27% 0.16</td>
</tr>
<tr>
<td>(0.009)</td>
<td>(0.075)</td>
<td>(0.012) (0.098)</td>
</tr>
<tr>
<td>0.023ᵃ</td>
<td>0.570ᵃ</td>
<td>34% 0.22</td>
</tr>
<tr>
<td>(0.008)</td>
<td>(0.089)</td>
<td>(0.012) (0.144)</td>
</tr>
</tbody>
</table>

Panel B: Log-volatilities

<table>
<thead>
<tr>
<th></th>
<th>Monthly non-overlapping data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Past Adj.</td>
</tr>
<tr>
<td></td>
<td>R²  DW</td>
</tr>
<tr>
<td>-0.550ᵃ</td>
<td>-0.304 0.934ᵃ</td>
</tr>
<tr>
<td>(0.107)</td>
<td>(0.163) (0.083)</td>
</tr>
<tr>
<td>-0.695ᵃ</td>
<td>-0.932 0.571ᵃ</td>
</tr>
<tr>
<td>(0.116)</td>
<td>(0.571) (0.078)</td>
</tr>
<tr>
<td>-0.483ᵃ</td>
<td>-0.303 0.886ᵃ</td>
</tr>
<tr>
<td>(0.117)</td>
<td>(0.162) (0.153)</td>
</tr>
</tbody>
</table>

³ p-value ≤ 0.01.

b p-value ∈[0.01;0.05].

This table reports the least squares estimates of the regressions

Panel A: \[ \sigma_{r,t} = \gamma_0 + \gamma_1 \sigma_{i,t} + \gamma_p \sigma_{p,t-60} + \epsilon_t, \]

Panel B: \[ \ln(\sigma_{r,t}) = \gamma_0 + \gamma_1 \ln(\sigma_{i,t}) + \gamma_p \ln(\sigma_{r,t-60}) + \epsilon_t, \]

where \( \sigma_{r,t} \) denotes the realized volatility over the period from date \( t \) to the nearest date \( \tau \) on which the OEX option expires, \( \sigma_{p,t-60} \) denotes the past realized 60-day volatility, and \( \sigma_{i,t} \) denotes the binomial tree implied volatility of the at-the-money call option expiring at \( \tau \). \( \ln(\cdot) \) is the natural logarithm. For the monthly frequency, only \( t \) corresponding to the Wednesday following the previous expiration date \( \tau \) are included, and \( \sigma_{p,t-60} \) is replaced by the lagged dependent variable. The numbers in brackets are the Newey-West heteroskedasticity consistent standard errors. DW denotes the Durbin-Watson statistic. Data consist of 1,852 daily and 95 monthly observations from February 1988 to December 1995.
Table 4: \( p \)-values accounting for the overlapping data in the volatility regression

<table>
<thead>
<tr>
<th></th>
<th>Before crash</th>
<th>After crash</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Implied</td>
<td>Past realized</td>
</tr>
<tr>
<td>Daily coeff.</td>
<td>0.223</td>
<td>0.693</td>
</tr>
<tr>
<td>Monthly coeff.</td>
<td>0.420</td>
<td>0.803</td>
</tr>
<tr>
<td>Asymp. ( p )-value</td>
<td>0.258</td>
<td>0.307</td>
</tr>
<tr>
<td>Daily coeff.</td>
<td>0.492</td>
<td>0.597</td>
</tr>
<tr>
<td>Monthly coeff.</td>
<td>0.164</td>
<td>0.531</td>
</tr>
<tr>
<td>Asymp. ( p )-value</td>
<td>0.201</td>
<td>0.386</td>
</tr>
<tr>
<td>Daily OLS</td>
<td>0.102</td>
<td>0.413</td>
</tr>
<tr>
<td>Monthly OLS</td>
<td>0.412</td>
<td>0.009</td>
</tr>
<tr>
<td>Asymp. ( p )-value</td>
<td>0.150</td>
<td>0.131</td>
</tr>
</tbody>
</table>

This table reports the asymptotic (Asymp.) \( p \)-values based on Theorem 1 that accounts for the overlapping data problem in the daily regression. Asymp. \( p \)-values are based on 5,000 replications and 5,000 observations for each simulation. Long-run variances are estimated using the Andrews and Monahan (1992) method. The table also repeats the relevant empirical daily and monthly coefficient estimates from Tables 2 and 3, and \( p \)-values indicate the probability of a more extreme outcome relative to the null that implied volatility is an efficient forecast, i.e. for implied volatility the probability that the daily coefficient falls short of the monthly coefficient by more than in the empirical results, and for the past realized volatility the probability that the daily coefficient exceeds the monthly coefficient by more than in the empirical results.
Table 5: \( p \)-values accounting for the overlapping data in the log-volatility regression

<table>
<thead>
<tr>
<th></th>
<th>Before crash</th>
<th>After crash</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Implied</td>
<td>Past realized</td>
</tr>
<tr>
<td>Daily coeff.</td>
<td>0.325</td>
<td>0.823</td>
</tr>
<tr>
<td>Monthly coeff.</td>
<td>0.476</td>
<td>0.934</td>
</tr>
<tr>
<td>Asymp. ( p )-value</td>
<td>0.313</td>
<td>0.327</td>
</tr>
<tr>
<td>Daily coeff.</td>
<td>0.560</td>
<td></td>
</tr>
<tr>
<td>Monthly coeff.</td>
<td>0.277</td>
<td></td>
</tr>
<tr>
<td>Asymp. ( p )-value</td>
<td>0.230</td>
<td></td>
</tr>
<tr>
<td>Daily OLS</td>
<td>0.155</td>
<td>0.431</td>
</tr>
<tr>
<td>Monthly OLS</td>
<td>0.443</td>
<td>0.063</td>
</tr>
<tr>
<td>Asymp. ( p )-value</td>
<td>0.185</td>
<td>0.150</td>
</tr>
</tbody>
</table>

This table reports the asymptotic (Asymp.) \( p \)-values based on Theorem 1 that accounts for the overlapping data problem in the daily regression. Asymp. \( p \)-values are based on 5,0000 replications and 5,000 observations for each replication. Long-run variances are estimated using the Andrews and Monahan (1992) method. The table also repeats the relevant empirical daily and monthly coefficient estimates from Tables 2 and 3, and \( p \)-values indicate the probability of a more extreme outcome relative to the null that implied volatility is an efficient forecast, i.e. for implied volatility the probability that the daily coefficient falls short of the monthly coefficient by more than in the empirical results, and for the past realized volatility the probability that the daily coefficient exceeds the monthly coefficient by more than in the empirical results.
This figure displays the telescoping overlap structure of the daily observations arising from the fixed maturity dates. The function $J(t)$ is defined in Eq. (1).
Figure 2: Limiting distributions

Before the Crash

![Graph showing implied and historical volatility distributions before the crash.]

After the Crash

![Graph showing implied and historical volatility distributions after the crash.]

The figures present the limiting distributions of the OLS coefficients for implied and historical volatilities from regression equations (28) and (29). Dashed lines are the limiting distributions using the new theory where the telescoping overlap is taken into account. Solid lines are the approximating normal distributions with means in the OLS coefficients and variances equal to the Newey-West estimates. The limiting distributions are based on 5,000 replications of 5,000 observations. The upper two figures are for the sample period before the crash and the lower two figures are for the sample period after the crash.