Bootstrapping Realized Volatility

Sílvia Gonçalves

Université de Montréal, CIREQ and CIRANO

and

Nour Meddahi

Université de Montréal, CIREQ and CIRANO

Financial Econometrics Conference

May 7-8, 2004

Preliminary and Incomplete
Introduction

- Realized volatility (defined as the sum of squared high frequency returns) is a popular measure of volatility in empirical finance (Andersen, Bollerslev and Diebold 2004, Barndorff-Nielsen and Shephard 2004).

- Under the assumption of no microstructure noise, realized volatility consistently estimates integrated volatility.

- For inference on integrated volatility, the key ingredient is the sampling distribution of realized volatility.

- Two possible approaches:
  - Asymptotic theory.
  - The bootstrap.
Motivation

• Barndorff-Nielsen and Shephard (BN-S) establish a CLT for realized volatility over, e.g., a day, as the number of intra-day returns increases to infinity.

• Monte Carlo simulations show that the quality of the asymptotic approximation may be poor if the sampling frequency is small.

• Sampling too frequently is not done in practice to avoid potential microstructure biases.

• The bootstrap often provides a more accurate approximation to the finite-sample distribution of an estimator than its asymptotic distribution.

• So, the bootstrap may be an important method of inference in this context.
Contributions of this Paper

• We propose bootstrap methods useful for inference on realized volatility.

• We consider two bootstrap methods:
  1. i.i.d. bootstrap.
  2. wild bootstrap (WB).

• We prove the asymptotic validity of these methods for general stochastic volatility models, under conditions similar to those used by BN-S.

• Our Monte Carlo simulations show that bootstrap percentile-$t$ intervals based on the i.i.d. bootstrap outperform the asymptotic theory-based intervals. The wild bootstrap is somewhat conservative when the sampling frequency is small.
The Problem

- Stochastic volatility (SV) model:
  \[ d\log S_t = \sigma_t dW_t. \]

- Integrated volatility for day \( t = 1 \):
  \[ IV = \int_0^1 \sigma_i^2 du. \]

- Intra-day \( h \)-period returns:
  \[ r_{ih,h} = \log S_{ih} - \log S_{(i-1)h}, \quad \text{for } i = 1, \ldots, \frac{1}{h}. \]

- Realized volatility:
  \[ RV_h = \sum_{i=1}^{1/h} r_{ih,h}^2. \]

- Goal: form a 95% confidence interval for \( IV \).
Asymptotic Theory of BN-S

• CLT for \( RV_h \): as \( h \to 0 \),
\[
\sqrt{h^{-1}} (RV_h - IV) \xrightarrow{d} N(0, 1).
\]

\( IQ = \int_0^1 \sigma_u^4 du \) is the integrated quarticity.

• Realized quarticity:
\[
RQ_h = \frac{1}{3} h^{-1} \sum_{i=1}^{1/h} r_{ih,h}^4 \xrightarrow{P} IQ.
\]

• Feasible CLT for \( RV_h \): as \( h \to 0 \),
\[
\sqrt{h^{-1}} (RV_h - IV) \xrightarrow{d} N(0, 1).
\]

• A feasible asymptotic 95% confidence interval for \( IV \) is:
\[
CI_{BN-S} = RV_h \pm 1.96 \sqrt{2RQ_h}.
\]
The Bootstrap Approach

• Let \( \{ r_{ih}^*: i = 1, \ldots, \frac{1}{h} \} \) be a bootstrap sample from the original set of intra-day returns.

• Let

\[
RV_h^* = \sum_{i=1}^{1/h} r_{ih}^* \cdot
\]

• Bootstrap percentile interval:

\[
CI_{perc} = RV_h \pm Q_{0.95}^* \cdot
\]

with \( Q_{0.95}^* \) the 95% percentile of the distribution of an unstudentized bootstrap version of \( \sqrt{h^{-1}} (RV_h - IV) \).

• Bootstrap percentile-t interval:

\[
CI_{perc-t} = RV_h \pm q_{0.95}^* \frac{1}{\sqrt{h^{-1}}} \sqrt{2RQ_h} \cdot
\]

with \( q_{0.95}^* \) the 95% percentile of the distribution of a studentized bootstrap version of \( \sqrt{h^{-1}} (RV_h - IV) \).
Which bootstrap should we use?

- Under SV, conditional on the volatility path, intra-day returns are independent (normally distributed) but *heteroskedastic.*
- This suggests applying a *wild bootstrap* (Gonçalves and Kilian 2004):
  \[ r_{ih}^{\circ} = r_{ih} \eta_i, \]
  \[ \eta_i \sim i.i.d. N(0, 1). \]

- Under constant volatility, intra-day returns are i.i.d, which suggests applying an i.i.d. bootstrap:
  \[ r_{ih,h}^* \sim i.i.d. \text{ from } \{r_{ih,h}\}. \]
- Empirically, volatility is very persistent and close to being constant over a day.
- Even under SV, the i.i.d. bootstrap is theoretically valid if we apply it to an appropriately studentized statistic (Gonçalves and Vogelsang 2004).
Mean and Variance of the Wild bootstrap RV

Let

\[ r_{ih}^\diamond = r_{ih} \eta_i, \quad \eta_i \sim i.i.d. \ N(0,1). \]

Then:

(i) \( E^\diamond (RV_h^\diamond - RV_h) = 0 \) for any \( h \).

(ii) \( v_{h}^\diamond = Var^\diamond \left( \sqrt{h^{-1}} RV_h^\diamond \right) = 2h^{-1} \sum_{i=1}^{1/h} r_{ih,h}^4 \) is such that

\[
v_{h}^\diamond^2 - v^\diamond^2 \xrightarrow{P} 0 \quad \text{as} \quad h \to 0,
\]

where

\[
v^\diamond^2 = 6 \int_0^1 \sigma_u^4 du > 0.
\]

• Under SV, the WB variance of \( \sqrt{h^{-1}} RV_h^\diamond \) does not converge to \( 2IQ \).

• Valid inference based on the WB requires some correction.
Wild bootstrap CLT

Under conditions similar to those of BN-S, as $h \to 0$,

$$\sup_{x \in \mathbb{R}} \left| P^\circ \left( \frac{\sqrt{h^{-1}} (RV_h^\circ - RV_h)}{v_h^\circ} \leq x \right) - \Phi(x) \right| \overset{P}{\to} 0,$$

where $\Phi(x) = P(Z \leq x)$, with $Z \sim N(0, 1)$.

Implications for percentile bootstrap confidence intervals:

- Since $v_h^\circ$ does not converge to $2IQ$, a bootstrap percentile interval based on the quantiles of the distribution of

$$\sqrt{h^{-1}} (RV_h^\circ - RV_h)$$

is not valid.

- Thus, a naive application of the WB percentile should not work.

- However, a bootstrap percentile interval based on the quantiles of the distribution of

$$\frac{\sqrt{h^{-1}} (RV_h^\circ - RV_h)}{\sqrt{3}}$$

is valid.
Implications for percentile-t bootstrap confidence intervals:

- A bootstrap percentile-t interval based on the quantiles of the distribution of

\[
\frac{\sqrt{h^{-1}} (RV_h^\diamond - RV_h)}{\hat{v}_h^\diamond}
\]

is valid, if

\[
\hat{v}_h^{\diamond 2} \rightarrow v_h^{\diamond 2}.
\]

- Here we propose:

\[
\hat{v}_h^{\diamond 2} = \frac{2}{3} h^{-1} \sum_{i=1}^{1/h} r_{ih,h}^{\diamond 4} = 2RQ_h^\diamond.
\]

- Note that \( E^\diamond (\hat{v}_h^{\diamond 2}) = 2h^{-1} \sum_{i=1}^{1/h} r_{ih,h}^{4} = v_h^{\diamond 2} \).
Mean and Variance of the I.I.D. bootstrap RV

Let

\[ r_{ih,h}^* \sim \text{i.i.d. from } \{r_{ih,h}\}. \]

Then:

(i) \( E^* (RV_h^* - RV_h) = 0 \) for any \( h \).

(ii) \( v_{h}^2 = Var^* \left( \sqrt{h^{-1}} RV_h^* \right) = h^{-1} \sum_{i=1}^{1/h} r_{ih,h}^4 - \left( \sum_{i=1}^{1/h} r_{ih,h}^2 \right)^2 \)

is such that

\[ v_{h}^2 - v^2 \xrightarrow{P} 0 \text{ as } h \to 0, \]

where

\[ v^2 = 3 \int_0^1 \sigma_u^4 du - \left( \int_0^1 \sigma_u^2 du \right)^2 \neq 2 \int_0^1 \sigma_u^4 du. \]

- Under SV, the i.i.d. bootstrap second moment is incorrect.

- Under constant volatility,

\[ v^2 = 2\sigma^4 = 2 \int_0^1 \sigma_u^4 du, \]

so the i.i.d. bootstrap variance is correct.
I.I.D. bootstrap CLT

Under conditions similar to those of BN-S, as $h \to 0$,

$$\sup_{x \in \mathbb{R}} \left| P^* \left( \frac{\sqrt{h^{-1}} (RV^*_h - RV_h)}{v^*_h} \leq x \right) - \Phi(x) \right| \xrightarrow{P} 0.$$

Implications for percentile bootstrap confidence intervals:

- Since $v^*_h$ does not converge to $2IQ$, a bootstrap percentile interval based on the quantiles of the distribution of

$$\sqrt{h^{-1}} (RV^*_h - RV_h)$$

is not valid.

Implications for percentile-t bootstrap confidence intervals:

- A bootstrap percentile-t interval based on the quantiles of the distribution of

$$\frac{\sqrt{h^{-1}} (RV^*_h - RV_h)}{\hat{v}^*_h},$$

$$\hat{v}^*_h = h^{-1} \sum_{i=1}^{1/h} r_{ih,h}^* - \left( \sum_{i=1}^{1/h} r_{ih,h}^{*2} \right)^2,$$

is valid if $\hat{v}^*_h \to v^*_h$. 
Monte Carlo Analysis: Models

We consider four models without drift and no leverage effect:

\[ d \log S_t = \sigma_t dW_{1t} \]

**Log-Normal diffusion** (Andersen, Benzoni and Lund 2002):

\[ d \log (\sigma_t^2) = -0.0136 \left[0.8382 + \log (\sigma_t^2)\right] dt + 0.1148dW_{2t} \]

**GARCH(1,1) diffusion** (Andersen and Bollerslev 1998):

\[ d\sigma_t^2 = 0.035 (0.636 - \sigma_t^2) \, dt + 0.144\sigma_t^2 dW_{2t} \]

**Two-Factors Affine diffusion** (Bollerslev and Zhou 2002, ABM 2004):

\[
\begin{align*}
\sigma_t^2 &= \sigma_{1,t}^2 + \sigma_{2,t}^2 \\
d\sigma_{1,t}^2 &= 0.5708 (0.3257 - \sigma_{1,t}^2) \, dt + 0.2286\sigma_{1,t} dW_{2t} \\
d\sigma_{2,t}^2 &= 0.0757 (0.1786 - \sigma_{2,t}^2) \, dt + 0.1096\sigma_{2,t} dW_{3t}
\end{align*}
\]

**Two-Factors diffusion** (Chernov et al. 2003, Huang and Tauchen 2003):

\[
\begin{align*}
\sigma_t &= s\text{-exp} \left(-1.2 + 0.04\sigma_{1,t}^2 + 1.5\sigma_{2,t}^2\right) \\
d\sigma_{1,t}^2 &= -0.00137\sigma_{1,t}^2 dt + dW_{2t} \\
d\sigma_{2,t}^2 &= -1.386\sigma_{2,t}^2 dt + (1 + 0.25\sigma_{2,t}^2) \, dW_{3t}
\end{align*}
\]
Methods

• Asymptotic theory of BN-S:
  1. Raw: $RV_h \pm 1.96 \frac{1}{\sqrt{h-1}} \sqrt{2RQ_h}$
  2. Log: $\log(RV_h) \pm 1.96 \frac{1}{\sqrt{h-1}} \sqrt{2 \frac{RQ_h}{(RV)^2}}$

• i.i.d. and wild bootstrap:
  1. perc-n: naive percentile bootstrap, based on the quantiles of
     $\sqrt{h^{-1}} (RV_h^* - RV_h)$
     for the i.i.d. bootstrap, and
     $\sqrt{h^{-1}} (RV_h^\diamond - RV_h)$,
     for the WB.
  2. perc: percentile bootstrap, corrected according to our theory. For the WB, we use the quantiles of
     $\sqrt{h^{-1}} (RV_h^\diamond - RV_h) / \sqrt{3}$.
  3. perc-t: percentile-t bootstrap.
     – Raw: based on the quantiles of
       $\sqrt{h^{-1}} (RV_h^* - RV_h) / \hat{v}_h^*$
       and
       $\sqrt{h^{-1}} (RV_h^\diamond - RV_h) / \hat{v}_h^\diamond$. 
– *Log*: we use the quantiles of

\[
\sqrt{h^{-1}} (\log (RV^*_h) - \log (RV_h))/\sqrt{\hat{\nu}_h^* / RV^*_h},
\]

and similarly for the WB.
Conclusions

• We prove the asymptotic validity of the i.i.d. and wild bootstrap for realized volatility.

• Our simulations show that the bootstrap improves upon the asymptotic theory of BN-S in finite samples.

Extensions

• More general models, with drift and leverage.

• Applications to jump tests (BN-S 2004, Huang and Tauchen 2003, ABD 2004).

• Other bootstrap methods: block bootstrap; subsampling (Mykland, Zhang, Aït-Sahalia 2003).
Coverage Rates of Nominal 95% Intervals for RV

Log-normal diffusion

\[ d \log (\sigma_t^2) = -0.0136 \left[ 0.8382 + \log (\sigma_t^2) \right] dt + 0.1148dW_{2t} \]

<table>
<thead>
<tr>
<th>$h$</th>
<th>BN-S</th>
<th>$\text{perc-n}$</th>
<th>$\text{perc-t}$</th>
<th>$\text{perc-n}$</th>
<th>$\text{perc}$</th>
<th>$\text{perc-t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/12</td>
<td>86.02</td>
<td>90.76</td>
<td>83.28</td>
<td>93.86</td>
<td>96.10</td>
<td>94.46</td>
</tr>
<tr>
<td>1/48</td>
<td>92.44</td>
<td>93.82</td>
<td>91.56</td>
<td>94.90</td>
<td>95.44</td>
<td>98.32</td>
</tr>
<tr>
<td>1/96</td>
<td>93.14</td>
<td>93.98</td>
<td>92.76</td>
<td>94.88</td>
<td>95.46</td>
<td>99.26</td>
</tr>
<tr>
<td>1/288</td>
<td>94.26</td>
<td>94.36</td>
<td>94.00</td>
<td>94.70</td>
<td>94.92</td>
<td>99.80</td>
</tr>
<tr>
<td>1/576</td>
<td>94.22</td>
<td>94.22</td>
<td>93.98</td>
<td>94.48</td>
<td>94.46</td>
<td>99.86</td>
</tr>
<tr>
<td>1/1152</td>
<td>94.68</td>
<td>94.70</td>
<td>94.56</td>
<td>94.82</td>
<td>94.76</td>
<td>99.88</td>
</tr>
</tbody>
</table>

- Note: 5000 Monte Carlo replications, with 999 bootstrap replications each, in every table.
Coverage Rates of Nominal 95% Intervals for RV

GARCH(1,1) diffusion

\[ d\sigma_t^2 = 0.035 \left(0.636 - \sigma_t^2\right) dt + 0.144\sigma_t^2 dW_{2t} \]

<table>
<thead>
<tr>
<th></th>
<th>iid Bootstrap</th>
<th></th>
<th>Wild Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BN-S</td>
<td>perc-n</td>
<td>perc-t</td>
</tr>
<tr>
<td></td>
<td>raw</td>
<td>log</td>
<td>raw</td>
</tr>
<tr>
<td>(h)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/12</td>
<td>86.00</td>
<td>90.76</td>
<td>83.32</td>
</tr>
<tr>
<td>1/48</td>
<td>92.40</td>
<td>93.90</td>
<td>91.48</td>
</tr>
<tr>
<td>1/96</td>
<td>93.14</td>
<td>93.94</td>
<td>92.76</td>
</tr>
<tr>
<td>1/288</td>
<td>94.28</td>
<td>94.28</td>
<td>93.88</td>
</tr>
<tr>
<td>1/576</td>
<td>94.10</td>
<td>94.22</td>
<td>94.08</td>
</tr>
<tr>
<td>1/1152</td>
<td>94.66</td>
<td>94.64</td>
<td>94.62</td>
</tr>
</tbody>
</table>
Coverage Rates of Nominal 95% Intervals for RV

Two-factors affine diffusion

\[
\begin{align*}
\sigma_t^2 &= \sigma_{1,t}^2 + \sigma_{2,t}^2 \\
\frac{d\sigma_{1,t}^2}{\sigma_{1,t}^2} &= 0.5708 \left(0.3257 - \sigma_{1,t}^2\right) dt + 0.2286 \sigma_{1,t} dW_{2t} \\
\frac{d\sigma_{2,t}^2}{\sigma_{2,t}^2} &= 0.0757 \left(0.1786 - \sigma_{2,t}^2\right) dt + 0.1096 \sigma_{2,t} dW_{3t}
\end{align*}
\]

<table>
<thead>
<tr>
<th></th>
<th>iid Bootstrap</th>
<th>Wild Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BN-S</td>
<td>perc-n</td>
</tr>
<tr>
<td>$h$</td>
<td>raw</td>
<td>log</td>
</tr>
<tr>
<td>1/12</td>
<td>84.80</td>
<td>89.58</td>
</tr>
<tr>
<td>1/48</td>
<td>92.32</td>
<td>93.94</td>
</tr>
<tr>
<td>1/96</td>
<td>93.16</td>
<td>93.84</td>
</tr>
<tr>
<td>1/288</td>
<td>94.22</td>
<td>94.76</td>
</tr>
<tr>
<td>1/576</td>
<td>94.06</td>
<td>94.60</td>
</tr>
<tr>
<td>1/1152</td>
<td>94.96</td>
<td>95.02</td>
</tr>
</tbody>
</table>
Coverage Rates of Nominal 95% Intervals for RV

Two-factors diffusion of Chernov et al. (2003)

\[
\sigma_t = \text{s-exp}(-1.2 + 0.04\sigma^2_{1,t} + 1.5\sigma^2_{2,t})
\]

\[
d\sigma^2_{1,t} = -0.00137\sigma^2_{1,t}dt + dW_{2t}
\]

\[
d\sigma^2_{2,t} = -1.386\sigma^2_{2,t}dt + (1 + 0.25\sigma^2_{2,t})dW_{3t}
\]

<table>
<thead>
<tr>
<th></th>
<th>iid Bootstrap</th>
<th></th>
<th>Wild Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BN-S</td>
<td>perc-n</td>
<td>perc-t</td>
</tr>
<tr>
<td>h</td>
<td>raw</td>
<td>log</td>
<td>raw</td>
</tr>
<tr>
<td>1/12</td>
<td>78.42</td>
<td>85.56</td>
<td>77.16</td>
</tr>
<tr>
<td>1/48</td>
<td>88.10</td>
<td>90.74</td>
<td>88.68</td>
</tr>
<tr>
<td>1/96</td>
<td>90.48</td>
<td>92.56</td>
<td>91.42</td>
</tr>
<tr>
<td>1/288</td>
<td>93.04</td>
<td>93.82</td>
<td>94.88</td>
</tr>
<tr>
<td>1/576</td>
<td>94.34</td>
<td>94.54</td>
<td>96.06</td>
</tr>
<tr>
<td>1/1152</td>
<td>94.88</td>
<td>94.90</td>
<td>96.38</td>
</tr>
</tbody>
</table>