Asymptotic Inference about Predictive Accuracy using High Frequency Data

Jia Li and Andrew Patton

Department of Economics
Duke University

March 2014

Motivation

- Financial decision making relies on forecasts of risk and return
- "**Risk**" forecasting is difficult as it is generally **latent**
  - Variance, beta, correlation, idiosyncratic variance, jump risk, etc.
- Existing methods for forecast evaluation almost all rely on the target variable being **observable**
  - Diebold-Mariano (1995), West (1996), White (2000), Romano-Wolf (2005), Giacomini-White (2006), Hansen et al (2011)
- Recent work on the econometrics of **high frequency data** has provided estimators of many measures of risk, including those above
  - See Andersen, et al. (2006) for a review

Example: An infeasible test

- Example: forecasting the integrated variance $IV_{t+1} = \int_t^{t+1} \sigma_s^2 ds$.
- Compare forecast sequences $F_{1,t+1}$ and $F_{2,t+1}$ using loss function L
- Null hypothesis of interest:

$$H_0^\dagger : \mathbb{E}[L(F_{1,t+1}, IV_{t+1})] = \mathbb{E}[L(F_{2,t+1}, IV_{t+1})]$$

- The difficulty: IV_{t+1} is unobservable, even at $t+1$.
- If IV_{t+1} were observable, then this would be a standard problem. (Diebold-Mariano 95, West 96, Giacomini-White 06).

★ But unobservable target $\Rightarrow$ infeasible test

Example: A feasible test

- First, find a proxy for IV_t
  - eg: realized variance: $RV_t = \sum_i^{[1/\Delta]} |\Delta_{t,i}X|^2$
  - Δ is the sampling interval
  - $\Delta_{t,i}X$ is the ith return of the log-price X in day t
- Second, evaluate $F_{1,t+1}$ and $F_{2,t+1}$ w.r.t. the observable proxy
- Standard evaluation methods can be used, but only for the **proxy null hypothesis**:

$$H_0 : \mathbb{E}[L(F_{1,t+1}, RV_{t+1})] = \mathbb{E}[L(F_{2,t+1}, RV_{t+1})]$$

★ But "proxy null" $\neq$ "true null".

Example: A negligibility result

- Intuitively, the difference between the proxy null and the true null is "negligible" if:

  1. The proxy is "precise enough", and
  2. The loss function is "well behaved"

- This paper provides a general theory formalizing this intuition

  - allowing for almost all forecasting evaluation methods
  - allowing for almost all high-frequency based estimators.

Contributions of this paper

- Our primary contribution is methodological: we provide a general framework for tests of predictive ability with a latent target variable.
- The main result relies on two high level assumptions:
    1. General conditions on the inference method to be used when the target is observable
    2. Conditions on the accuracy of the (high frequency) proxy used for the target variable

    ★ Primitive conditions for these assumptions are provided
- Under these conditions we show that the asymptotic properties of standard predictive ability tests, implemented using high frequency proxies for the latent variable, are preserved

Contributions of this paper, cont'd

- En route to our main result, we make two additional contributions:

  1. We present a simple unifying framework for considering much of the extant literature on forecast evaluation: Diebold and Mariano (1995), West (1996), White (2000), Giacomini and White (2006) and McCracken (2007)

  2. We provide results on rates of convergence for a large collection of high frequency estimators: realized (co)variance, bipower (co)variation, truncated (co)variation, realized correlation, realized beta, realised skewness and kurtosis, jump power variation, realized semi-variance, realized Laplace transform.

A framework to bridge two literatures

Forecast evaluation *Methods*	**High freq econometrics** *Risk measures*
Diebold-Mariano (1995, JBES)	Realized variance (ABDL 2003 BNS 2004 ECTA)
West (1996, ECTA)	Bipower variation (BNS 2004 JFEC)
White (2000, ECTA)	Truncated variation (Mancini 2009 SJS)
Hansen (2005, JBES)	Avg sparse sampled RV (ZMA 2005 JASA)
Romano-Wolf (2005, ECTA)	Realized semi-variance (BNKS 2010 book)
Giacomini-White (2006, ECTA)	Realized Laplace transform (TT 2012 ECTA)
McCracken (2007, JoE)	General vol and jump f'nals (Jacod 08, Jacod-Protter 12)
HLN (2011, ECTA)	General jump-robust f'nals (Jacod-Rosenbaum 12 WP)
and others	and others

Existing work on forecasting latent target variables

- Andersen and Bollerslev (1998, IER), Meddahi (2002, JAE) and ABM (2004, IER) show that the apparent "poor" performance volatility forecasts is improved when a better volatility proxy is used
- Andersen, Bollerslev and Meddahi (2005, ECTA) show how to use asymptotic theory on the volatility proxy to better estimate the R^2 from predictive regressions for volatility
- Hansen and Lunde (2006, JoE) and Patton (2011, JoE) that standard forecast evaluation tests may be employed even when a proxy is used when certain conditions are satisfied:
  1. The proxy is (finite-sample) unbiased for the latent target variable
  2. The loss function satisfies a condition on its second derivative

Outline

1. Introduction and motivation
2. **Theory**
    1. General structure for existing forecast evaluation tests
    2. Negligibility result for applications with latent forecast target
    3. Primitive conditions for the negligibility result
3. Simulation study
4. Empirical application
5. Conclusion

Outline

1. Introduction and motivation
2. Theory
    1. **General structure for existing forecast evaluation tests**
    2. Negligibility result for applications with latent forecast target
    3. Primitive conditions for the negligibility result
3. Simulation study
4. Empirical application
5. Conclusion

Base case: Forecast evaluation with observable target

- Forecast models:
  - $F_{t+\tau}(\beta) = [F_{1,t+\tau}(\beta), ..., F_{K,t+\tau}(\beta)] \in \mathcal{F}_t$: K forecast models
  - $\hat{\beta}_t \in \mathcal{F}_t$: estimator of β
  - β^* : pseudo-true population parameter

- Example: volatility forecasting using a GARCH model:
  - $\sigma^2 = w + b\sigma^2_{t-1} + ar^2_{t-1}, \;\; \beta = [w, b, a]'$
  - *Rolling* window forecast: $\hat{\beta}_t$ uses data from $t - R + 1$ to t
  - *Expanding* window forecast: $\hat{\beta}_t$ uses data from 1 to t

- Sample size $T = R + P$ (Total = Regression + Prediction)

Base case: Forecast evaluation with observable target

Actual forecasts and forecast models

1. Inference concerning the actual forecast, based on $F_{t+\tau}(\hat{\beta}_t)$.
   - We can simply treat $F_{t+\tau}(\hat{\beta}_t)$ as an observable sequence $F_{t+\tau}$.
   - The form of $F_{t+\tau}(\cdot)$ and the structure of $\hat{\beta}_t$ are irrelevant.
   - Treated by Diebold-Mariano (1995) and Giacomini-White (2006).

2. Inference concerning the forecast model, based on $F_{t+\tau}(\beta^*)$.
   - $F_{t+\tau}(\beta^*)$ is the (unobservable) "oracle" forecast.
   - Need to know the form of $F_{t+\tau}(\cdot)$ and the structure of $\hat{\beta}_t$ to solve a "two-step" inference problem
   - Treated by West (1996), White (2000), McCracken (2007).

Base case: Forecast evaluation with observable target

Performance measures

- Relative performance: $f_{t+\tau} \equiv f\left(Y_{t+\tau}, F_{t+\tau}\left(\hat{\beta}_t\right), h_t\right)$
  - Example 1: Relative quadratic loss

$$f_{t+\tau} = \left(Y_{t+\tau} - F_{1,t+\tau}\left(\hat{\beta}_t\right)\right)^2 - \left(Y_{t+\tau} - F_{2,t+\tau}\left(\hat{\beta}_t\right)\right)^2$$

  - Example 2: General loss, compared with a common benchmark

$$f_{t+\tau} = L\left(Y_{t+\tau} - F_{j,t+\tau}\left(\hat{\beta}_t\right)\right) - L\left(Y_{t+\tau} - F_{0,t+\tau}\left(\hat{\beta}_t\right)\right)_{1\leq j\leq K}$$

  - Example 3: "instrumented" loss:

$$f_{t+\tau} = L\left(Y_{t+\tau} - F_{j,t+\tau}\left(\hat{\beta}_t\right)\right) - L\left(Y_{t+\tau} - F_{0,t+\tau}\left(\hat{\beta}_t\right)\right) h_t$$

Base case: Forecast evaluation with observable target

Null and alternative hypotheses

- Denote:
  - Relative performance: $f_{t+\tau} \equiv f(Y_{t+\tau}, F_{t+\tau}(\hat{\beta}_t), h_t)$
  - "Oracle" relative performance: $f^*_{t+\tau} \equiv f(Y_{t+\tau}, F_{t+\tau}(\beta^*), h_t)$
  - $\bar{f}_T \equiv P^{-1} \sum_{t=R}^{T} f_{t+\tau}$, and $\bar{f}^*_T \equiv P^{-1} \sum_{t=R}^{T} f^*_{t+\tau}$.
- **Proxy equal predictive ability** (DM95, West96, GW06, Mc07)

$$
\begin{aligned}
H_0 &: \mathbb{E}\left[f^*_{t+\tau}\right] = 0 \text{ for all } t \geq 1 \\
H_a &: \liminf_{T \to \infty} |\mathbb{E}[\bar{f}^*_{j,T}]| > 0, \quad \text{for some } j \in \{1, \ldots, dim(f)\}.
\end{aligned}
$$

- **Proxy superior predictive ability** (White00, Hansen05, RW05)

$$
\begin{aligned}
H_0 &: \mathbb{E}\left[f^*_{t+\tau}\right] \leq 0 \text{ for all } t \geq 1 \\
H_a &: \liminf_{T \to \infty} \mathbb{E}[\bar{f}^*_{j,T}] > 0, \quad \text{for some } j \in \{1, \ldots, dim(f)\}.
\end{aligned}
$$

Base case: Forecast evaluation with observable target

Assumptions

- **Assumption A1**: $\left(a_T\left(\bar{f}_T - \mathbb{E}[\bar{f}_T^*]\right), a'_T S_T\right) \xrightarrow{d} (\xi, S)$

- Comments:

  1. a_T and a'_T are normalizing sequences (eg, $a_T = \sqrt{P}$ and $a'_T = 1$)
  2. S_T is typically a HAC estimator, and S is the variance of ξ
  3. S_T may be inconsistent: nested models (McCracken 2007) or "fixed b" asymptotics (Kiefer-Vogelsang 2002)

Base case: Forecast evaluation with observable target I

Assumption A1: Limiting behavior of components of test statistic

- Assumption A1: $\left(a_T\left(\bar{f}_T - \mathbb{E}[\bar{f}_T^*]\right), a'_T S_T\right) \xrightarrow{d} (\xi, S)$.
- **Diebold-Mariano (1995) and Giacomini-White (2006):**
  1. $a_T = \sqrt{P}$, $a'_T = 1$
  2. $\sqrt{P}\left(\bar{f}_T - \mathbb{E}[\bar{f}_T^*]\right) \xrightarrow{d} \xi \sim N(0, \Sigma)$
  3. S_T=HAC estimator of Σ, such that $S_T \xrightarrow{p} \Sigma$

Base case: Forecast evaluation with observable target II

Assumption A1: Limiting behavior of components of test statistic

- Assumption A1: $\left(a_T\left(\bar{f}_T - \mathbb{E}[\bar{f}_T^*]\right), a_T' S_T\right) \xrightarrow{d} (\xi, S)$.

- **West (1996):**

  1. $a_T = \sqrt{P}$, $a_T' = 1$
  2. $\sqrt{P}\left(\bar{f}_T - \mathbb{E}[\bar{f}_T^*]\right) \xrightarrow{d} \xi \sim N(0, \Omega)$, where
     $\Omega = S_{ff} + \Pi\left(FBS_{fh}' + S_{fh}B'F'\right) + 2\Pi FV_\beta F'$
  3. S_T=HAC estimator of Ω, such that $S_T \xrightarrow{p} \Omega$

Base case: Forecast evaluation with observable target III

Assumption A1: Limiting behavior of components of test statistic

- Assumption A1: $\left(a_T\left(\bar{f}_T - \mathbb{E}[\bar{f}_T^*]\right), a_T' S_T\right) \xrightarrow{d} (\xi, S)$.

- **McCracken (2007) (OOS-t test):**

  1. $a_T = P$, $a_T' = 1$ (or $a_T = \sqrt{P}$, $a_T' = \sqrt{P}$)
  2. $S_T = P\hat{\Omega}_T$, where $\hat{\Omega}_T = \frac{1}{P}\sum_{t=R}^{T}\left(f_{t+\tau} - \bar{f}_T\right)^2$
  3. $\left(P\bar{f}_T, P\hat{\Omega}_T\right) \xrightarrow{d} (\xi, S)$, so $OOS\text{-}t \equiv \frac{\sqrt{P}\bar{f}_T}{\sqrt{\hat{\Omega}_T}} \xrightarrow{d} \frac{\Gamma_1 - \Gamma_2/2}{\Gamma_2^{1/2}}$, where

     Γ_1, Γ_2 are functionals of a vector Brownian motion

Base case: Forecast evaluation with observable target IV

Assumption A1: Limiting behavior of components of test statistic

- Assumption A1: $\left(a_T\left(\bar{f}_T - \mathbb{E}[\bar{f}_T^*]\right), a'_T S_T\right) \xrightarrow{d} (\xi, S)$.

- **Giacomini-White (2006) with Kiefer-Vogelsang (2005) std error**

  1. $a_T = \sqrt{P}$, $a'_T = 1$
  2. $\sqrt{P}\left(\bar{f}_T - \mathbb{E}[\bar{f}_T^*]\right) \xrightarrow{d} \xi \sim N(0, \Sigma)$
  3. S_T=*inconsistent* HAC estimator of Σ, where $S_T \xrightarrow{d} S$

- We consider a test statistic

$$\varphi_T = \varphi\left(a_T \bar{f}_T, a_T' S_T\right)$$

- **Assumption A2**: $\varphi(\cdot,\cdot)$ is continuous a.s. under the law of (ξ, S)
- This assumption is satisfied by all tests in the literature:
  - t-tests: $\varphi_{t\text{-stat}}(\xi, S) = \xi/\sqrt{S}$
  - F-tests: $\varphi_{F\text{-stat}}(\xi, S) = \xi' S^{-1} \xi$
  - Max tests: $\varphi_{Max}(\xi, S) = \max_i \xi_i$ or $\varphi_{St\text{-}Max}(\xi, S) = \max_i \xi_i / \sqrt{S_t}$

- **Assumption A3:** The distribution function of $\varphi(\xi, S)$ is continuous at its $1-\alpha$ quantile, $z_{1-\alpha}$. Moreover, the sequence of critical values $z_{T,1-\alpha} \xrightarrow{p} z_{1-\alpha}$.
- For many tests the limit distribution of φ_T under the null is *known*
  - DM, West and GW: $\varphi_T \xrightarrow{D} N(0,1)$ or χ_q^2 so critical values are known
  - McCracken (2007): Limit null distribution of φ_T is non-standard but known and critical values provided in that paper
- White (2000) and Hansen (2005): Critical values *not known*, but can be consistently estimated using the bootstrap

Base case: Forecast evaluation with observable target

Assumptions B1-B2: Behavior of test statistic (for alternative)

- **Assumption B1:** For any $\tilde{S} \in \mathcal{S}$
  1. $\varphi\left(u, \tilde{S}\right) \leq \varphi\left(u', \tilde{S}\right)$ whenever $u \leq u'$
  2. $\varphi\left(u, \tilde{S}\right) \rightarrow \infty$ whenever $u_j \rightarrow \infty$ for some $j \in \{1, ..., \dim(f)\}$
- **Assumption B2:** For any $\tilde{S} \in \mathcal{S}$, $\varphi\left(u, \tilde{S}\right) \rightarrow \infty$ whenever $|u_j| \rightarrow \infty$ for some $j \in \{1, ..., \dim(f)\}$
- So test statistic is increasing in $a_T \bar{f}_T$, and diverges if an element of the numerator diverges. (B1(b) corresponds to one-sided tests; B2 to two-sided tests)

- We consider non-randomized tests of the form

$$\phi_T = \mathbf{1}\left\{\varphi_T > z_{1-\alpha}\right\}$$

- **Proposition 1:** Assume A1–A3 and B1–B2.

  1. $E\left[\phi_T\right] \to \alpha$ under EPA proxy null, and
     $E\left[\phi_T\right] \to 1$ under proxy alternative.
  2. $\limsup_{T\to\infty} E\left[\phi_T\right] \leq \alpha$ under SPA proxy null, and
     $E\left[\phi_T\right] \to 1$ under proxy alternative.

Forecast evaluation with *unobservable* target

True hypotheses

- Now we consider the observable variable Y_t as a proxy for the true, latent, target variable $Y_t^\dagger$.
- Let $f_{t+\tau}^\dagger \equiv f(Y_t^\dagger, F_{t+\tau}(\beta^*), h_t)$ with sample mean $\bar{f}_T^\dagger$. Then
- **Equal predictive ability:**

$$\begin{aligned} H_0^\dagger &: \mathbb{E}\left[f_{t+\tau}^\dagger\right] = 0 \text{ for all } t \geq 1 \\ H_a^\dagger &: \liminf_{T\to\infty} |\mathbb{E}[\bar{f}_{j,T}^\dagger]| > 0, \quad \text{for some } j \in \{1, \ldots, dim(f)\}. \end{aligned}$$

- **Superior predictive ability:**

$$\begin{aligned} H_0^\dagger &: \mathbb{E}\left[f_{t+\tau}^\dagger\right] \leq 0 \text{ for all } t \geq 1 \\ H_a^\dagger &: \liminf_{T\to\infty} \mathbb{E}[\bar{f}_{j,T}^\dagger] > 0, \quad \text{for some } j \in \{1, \ldots, dim(f)\}. \end{aligned}$$

Forecast evaluation with *unobservable* target

Assumption C1: Approximation of hypothesis condition

- **Assumption C1:** $a_T \left(\mathbb{E}\left[f^*_{t+\tau}\right] - \mathbb{E}\left[f^\dagger_{t+\tau}\right] \right) \to 0$
- This high level assumption allows us to obtain our main result (two slides below)
- Under this condition, tests of the quantity of interest, $\mathbb{E}\left[f^\dagger_{t+\tau}\right]$, can be implemented on the proxy quantity, $\mathbb{E}\left[f^*_{t+\tau}\right]$.
- Note that the restrictiveness of this assumption depends on a_T, where a_T comes from $a_T \left(\bar{f}_T - \mathbb{E}[\bar{f}^*_T] \right) \xrightarrow{d} \xi$
    - So weaker for DM-West-GW tests, stronger for McC tests
- Assumption C1 can be shown to be implied by the following two assumptions:

- **Assumption C2:** There exist some bounded deterministic sequence $\{d_t\}_{t\geq 1}$ and constants $p \in [1,2)$, $\theta > 0$ and $C > 0$ s.t.

$$\left\| Y_{t+\tau} - Y^{\dagger}_{t+\tau} \right\|_p \leq C d_t^{\theta} \quad \text{for all} \quad t > R$$

- d_t is the (possibly irregular and time-varying) mesh of the high frequency grid
- This is a high level assumption on accuracy, which needs to be verified
    - We devote Section 3 of the paper to obtaining primitive conditions for the result to hold

- **Assumption C3:** For some sequence $m_{t+\tau}$,
  1. $\left\| f\left(Y_{t+\tau}, F_{t+\tau}, h_t\right) - f\left(Y^{\dagger}_{t+\tau}, F_{t+\tau}, h_t\right)\right\| \leq m_{t+\tau}\left\| Y_{t+\tau} - Y^{\dagger}_{t+\tau}\right\|$, and $\sup_t \left\| m_{t+\tau}\right\|_q \leq C$ for $q \geq p/\left(p-1\right)$, where p is from Assumption C2
  2. $a_T P^{-1} \sum_{t=R}^{T} d_t^{\theta} \to 0$, where a_T is from Assumption A1
- Part (a) is a Lipschitz-type condition, and in some cases $m_{t+\tau}$ is constant
  - Note that this places constraints on the behavior of f as a function of Y, not of β
- Part (b) requires d_t to be sufficiently small on average over the prediction sample

Forecast evaluation with *unobservable* target

Main result

- **Theorem 1:** Assume A1–A3, B1–B2 and C1–C2.
  1. $E[\phi_T] \to \alpha$ under EPA *true* null, and
     $E[\phi_T] \to 1$ under *true* alternative.
  2. $\limsup_{T\to\infty} E[\phi_T] \leq \alpha$ under SPA *true* null, and
     $E[\phi_T] \to 1$ under *true* alternative.
- That is, the asymptotic behavior of the test under the **true** null and alternative hypotheses is the same as under the **proxy** hypotheses.
  - Thus the error in the proxy Y_t for the latent target $Y_t^\dagger$ is **negligible**

Forecast evaluation with *unobservable* target

Discussion: Negligibility and West's "asymptotic irrelevance"

- Similar to our negligibility result, West (1996) defines in cases exhibiting "asymptotic irrelevance" as those where valid inference may be made while ignoring the presence of parameter estimation error
  - eg: $P/R \rightarrow 0$, or estimation and evaluation loss functions coincide
- Our negligibility result is quite different:
  1. Our latent quantity is a *process* $\left\{Y_t^{\dagger}\right\}_{t\geq 1}$ which grows in T, while in West it is a fixed and finite-dimensional vector, β^*
  2. West provides a correction to address the lack of asymptotic irrelevance where needed: the correction adjusts the variance. In our case no such correction is readily available: it requires dealing with a bias term (which appears infeasible to estimate consistently).

Forecast evaluation with *unobservable* target

Discussion: Weak versus Strong negligibility

- The above is a "**weak**" negligibility result: $E\left[\phi_T\right]$ under proxy hypothesis $\stackrel{a}{\approx} E\left[\phi_T\right]$ under true hypothesis
- A "**strong**" negligibility result would be $\Pr\left[\phi_T = \phi_T^{\dagger}\right] \longrightarrow 1$
  - ie: Rejection decision using a test based on the proxy is the same, asymptotically, as the decision made using a test based on the true target variable
- We argue that weak negligibility is all that is needed
  - Strong negligibility is too strict: requires feasible test to make same Type I errors as infeasible test
  - Obtaining this would require much stronger assumptions (and harder proofs).

Outline

1. Introduction and motivation
2. Theory
    1. General structure for existing forecast evaluation tests
    2. Negligibility result for applications with latent forecast target
    3. **Primitive conditions for the negligibility result**
3. Simulation study
4. Empirical application
5. Conclusion

Outline

1. Introduction and motivation
2. Theory
   1. General structure for existing forecast evaluation tests
   2. Negligibility result for applications with latent forecast target
   3. Primitive conditions for the negligibility result
3. **Simulation study**
4. Empirical application
5. Conclusion

Simulation study

- We consider three different DGPs for the simulation (described below). All three have:
- In-sample period: $R \in \{500, 1000\}$ observations
- Out-of-sample period: $P \in \{500, 1000, 2000\}$ obs
  - So $\pi \equiv P/R \in \{1/2, ..., 4\}$
- Euler discretization step (used to simulate the continuous-time processes) is $1/23400$
  - Corresponds to one-second sampling for a stock traded on the NYSE
- Two models for a latent target variable are compared using a quadratic loss function
- All simulations are repeated 250 times.

Simulation design A

- **DGP:** Diffusion as in Andersen-Bollerslev-Meddahi (05, ECTA):

$$\begin{aligned} dX_t &= 0.03d_t + \sigma_t\left(-0.58dW_{1t} + \sqrt{1\text{-}0.58^2}\,dW_{2t}\right) \\ d\log\sigma_t^2 &= -0.01\left(0.84 + \log\sigma_t^2\right)d_t + 0.11dW_{1t} \end{aligned}$$

- **Target variable:** quadratic variance of log-price process over one day
- **Proxies:** RV using 1-sec, 5-sec, 1-min, 5-min, 30-min returns
- **Model 1:** GARCH(1,1)
- **Model 2:** HAR model using five-min RV

Simulation design B

- **DGP:** Diffusion as in Andersen-Bollerslev-Meddahi (05, ECTA), with log-price jumps added following Tauchen and Zhou (2011, JoE):
  - Jump arrival is Poisson with $\lambda = 0.05$
  - Jump size distributed as $N\left(0.2, 1.4^2\right)$
- **Target variable:** integrated variance of log-price process over one day
- **Proxies:** BPV using 1-sec, 5-sec, 1-min, 5-min, 30-min returns
- **Model 1:** GARCH(1,1) model applied to sqr-root of one-min BPV
- **Model 2:** HAR model using one-min BPV

Simulation design C I

- **DGP:** Bivariate stochastic volatility model used in the simulation study of Barndorff-Nielsen and Shephard (2004, ECTA):

$$
\begin{aligned}
dX_t &= \sigma_t dW_t \\
\sigma_t \sigma_t' &= \begin{bmatrix} \sigma_{1t}^2 & \rho_t \sigma_{1t} \sigma_{2t} \\ \cdot & \sigma_{2t}^2 \end{bmatrix} \\
\text{where} \quad \sigma_{1t}^2 &= v_t + \tilde{v}_t \\
dv_t &= -0.04 \left(v_t - 0.11 \right) d_t + 0.28 \sqrt{v_t} dB_{1t} \\
d\tilde{v}_t &= -3.74 \left(\tilde{v}_t - 0.39 \right) d_t + 2.60 \sqrt{\tilde{v}_t} dB_{2t} \\
\\
\text{and} \quad d\sigma_{2t}^2 &= -0.035 \left(\sigma_{2t}^2 - 0.636 \right) d_t + 0.236 \sigma_{2t}^2 dB_{3t} \\
\\
\text{and} \quad \rho_t &= \left(e^{2x_t} - 1 \right) / \left(e^{2x_t} + 1 \right) \\
dx_t &= -0.03 \left(x_t - 0.64 \right) d_t + 0.118 x_t dB_{4t}
\end{aligned}
$$

Simulation design C II

- **Target variable:** integrated correlation over one day
- **Proxies:** Realized correlation using 1-sec, 5-sec, 1-min, 5-min, 30-min returns
- **Model 1:** DCC(1,1) model for daily returns
- **Model 2:** "DCC-R"(1,1) model, including lagged 30-min realized correlation in the model

Testing the null hypothesis

- We use Giacomini-White (2006) to compare the models in each simulation design, based on MSE:

$$H_0^\dagger : E\left[\left\{Y_t^\dagger - \hat{Y}_{1,t}\left(\hat{\beta}_{1,t,R}\right)\right\}^2 - \left\{Y_t^\dagger - \hat{Y}_{2,t}\left(\hat{\beta}_{2,t,R}\right)\right\}^2\right] = \chi$$

- In all 3 designs, the forecasts are *not* equally accurate, so $\chi \neq 0$.
  - Our theory does not require $\chi = 0$, so this is no problem
- We compute the population values of χ using simulations of length 500,000 obs, and using the **true latent** target variable, $Y_t^\dagger$.
- We implement two versions of the GW test:
  1. Using **Newey-West** std errors, with truncation $= 3P^{1/3}$
  2. Using **Kiefer-Vogelsang** std errs and crit vals, with truncation $= 0.5P$

Finite-sample rejection frequencies

Simulation design A: Forecasting log quadratic variation

	GW-NW			GW-KV		
$P=$	*500*	*1000*	*2000*	*500*	*1000*	*2000*
	$R = 500$					
True	0.04	0.11	0.06	0.01	0.02	0.02
$\Delta = 5$ sec	0.04	0.11	0.07	0.01	0.02	0.02
$\Delta = 1$ min	0.04	0.11	0.07	0.01	0.02	0.02
$\Delta = 5$ min	0.04	0.11	0.08	0.01	0.02	0.03
$\Delta = 30$ min	0.05	0.13	0.11	0.02	0.04	0.04
	$R = 1000$					
True	0.16	0.10	0.07	0.05	0.01	0.02
$\Delta = 5$ sec	0.16	0.10	0.07	0.05	0.01	0.02
$\Delta = 1$ min	0.16	0.10	0.07	0.06	0.01	0.02
$\Delta = 5$ min	0.16	0.09	0.08	0.06	0.01	0.02
$\Delta = 30$ min	0.21	0.13	0.11	0.09	0.02	0.02

Finite-sample rejection frequencies

Simulation design B: Forecasting log integrated variance

	GW-NW			GW-KV		
$P=$	*500*	*1000*	*2000*	*500*	*1000*	*2000*
	$R = 500$					
True	0.05	0.06	0.06	0.04	0.05	0.04
$\Delta = 5$ sec	0.06	0.06	0.06	0.04	0.04	0.05
$\Delta = 1$ min	0.07	0.08	0.07	0.05	0.06	0.06
$\Delta = 5$ min	0.03	0.05	0.04	0.02	0.06	0.05
$\Delta = 30$ min	0.03	0.02	0.00	0.03	0.03	0.01
	$R = 1000$					
True	0.03	0.04	0.04	0.03	0.04	0.05
$\Delta = 5$ sec	0.03	0.04	0.04	0.04	0.04	0.05
$\Delta = 1$ min	0.04	0.05	0.06	0.03	0.04	0.06
$\Delta = 5$ min	0.03	0.04	0.05	0.04	0.04	0.06
$\Delta = 30$ min	0.02	0.01	0.01	0.02	0.01	0.01

Finite-sample rejection frequencies

Simulation design C: Forecasting correlation

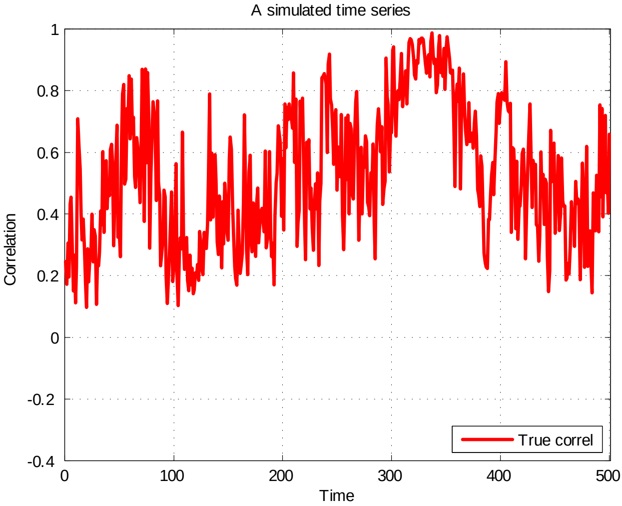
	GW-NW			GW-KV		
$P=$	*500*	*1000*	*2000*	*500*	*1000*	*2000*
	$R = 500$					
True	0.22	0.17	0.17	0.07	0.05	0.03
$\Delta = 5$ sec	0.22	0.17	0.17	0.07	0.05	0.03
$\Delta = 1$ min	0.22	0.17	0.18	0.07	0.05	0.02
$\Delta = 5$ min	0.22	0.17	0.18	0.08	0.05	0.02
$\Delta = 30$ min	0.20	0.16	0.17	0.07	0.04	0.03
	$R = 1000$					
True	0.24	0.22	0.20	0.13	0.08	0.05
$\Delta = 5$ sec	0.24	0.21	0.20	0.13	0.08	0.05
$\Delta = 1$ min	0.24	0.21	0.20	0.14	0.08	0.06
$\Delta = 5$ min	0.24	0.21	0.20	0.14	0.08	0.05
$\Delta = 30$ min	0.24	0.22	0.20	0.14	0.08	0.05

Conclusions from the simulation study

- Simulation results strongly support for the theory presented above
- Rejection rates from tests using high frequency proxies are comparable to infeasible tests based on the latent target variable
  - Some of these tests have a tendency to over-reject when sample size is small, and this tendency is “inherited” by tests that use a proxy
  - Use of a proxy does not lead to worse properties than tests based on the latent target variable
- Intuition for why this works:

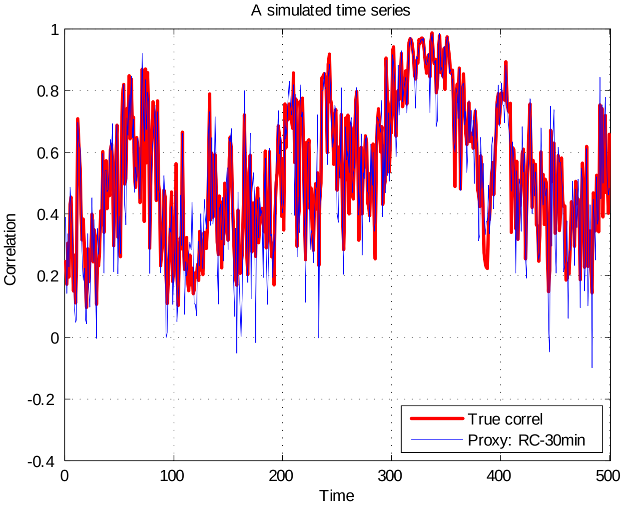
One simulation of a time series of integrated correlation

500 observations from the BNS specification



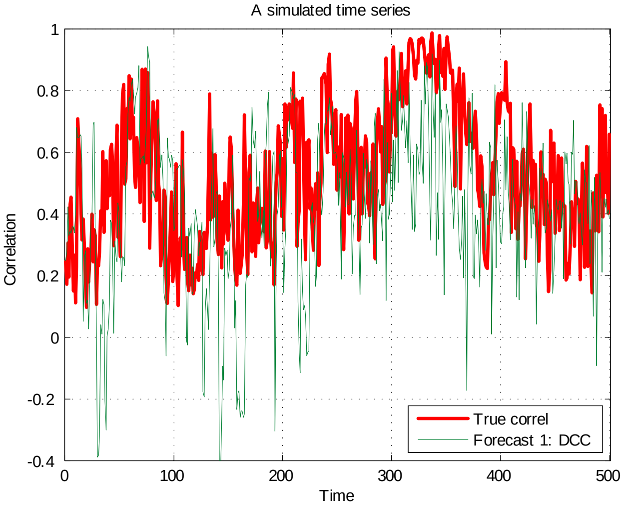
One simulation of a time series of integrated correlation

30-min realized correlation as a proxy for true integrated correlation



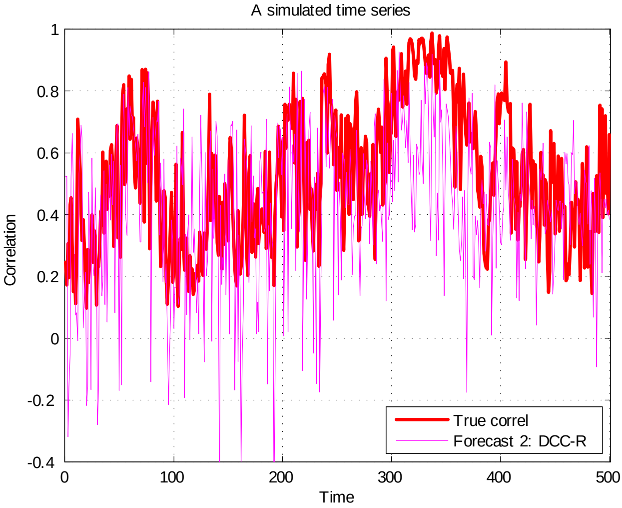
One simulation of a time series of integrated correlation

DCC forecasts of integrated correlation are noisier than the proxy (unsurprising)



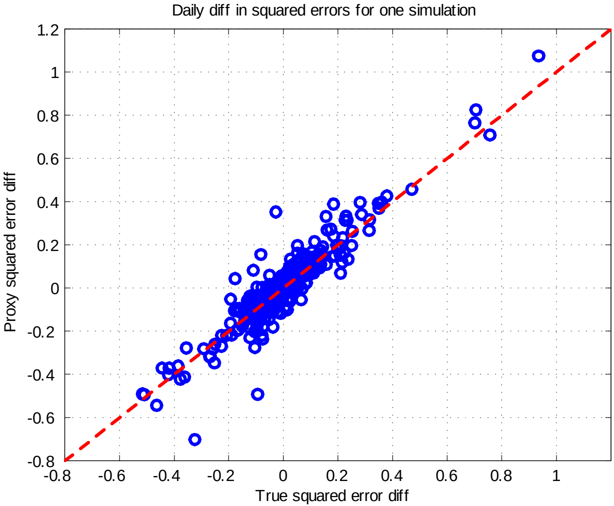
One simulation of a time series of integrated correlation

DCC-R forecasts of integrated correlation are noisier than the proxy (unsurprising)



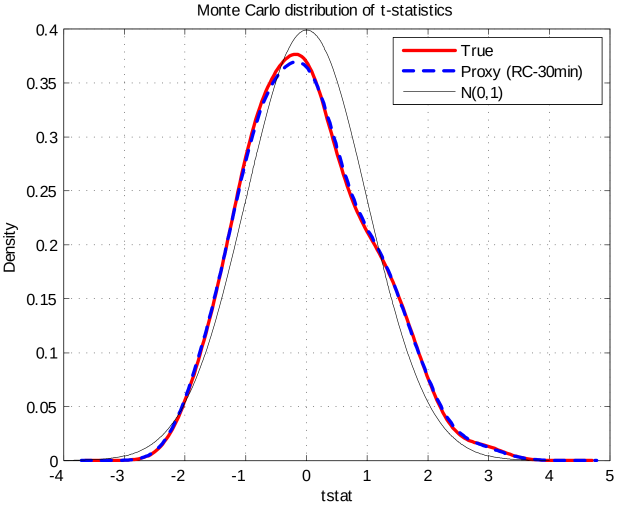
One simulation of a time series of integrated correlation

Squared errors using true correlation vs. squared errors using proxy



Monte Carlo distribution of test statistics

Proxy distribution very close to "true" distribution



Intuition for the (positive) simulation results

- Our "negligibility" condition is:

$$a_T \left(\mathbb{E}\left[f^*_{t+\tau}\right] - \mathbb{E}\left[f^\dagger_{t+\tau}\right] \right) \to 0$$

where a_T is such that

$$a_T \left(\bar{f}_T - \mathbb{E}[f^*_T] \right) \xrightarrow{d} \xi$$

- In the standard case we have $a_T = \sqrt{P}$, and $\xi \sim N(0, \Sigma)$.
- Strictly, the negligibility condition requires

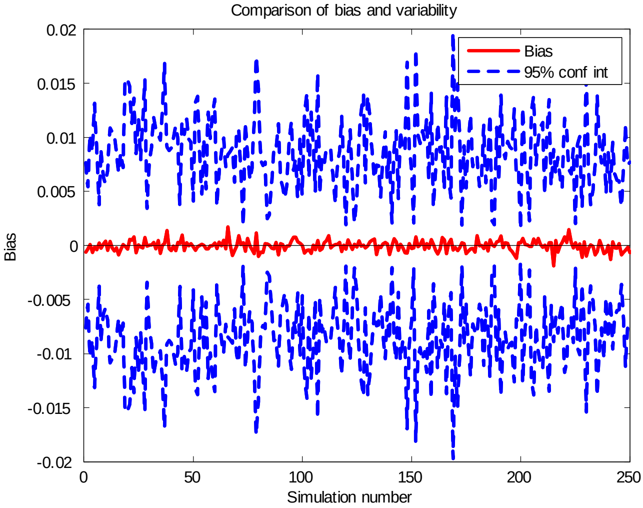
$$\mathbb{E}\left[f^*_{t+\tau}\right] - \mathbb{E}\left[f^\dagger_{t+\tau}\right] \ll 1/\sqrt{P}$$

but in practice we actually need

$$\mathbb{E}\left[f^*_{t+\tau}\right] - \mathbb{E}\left[f^\dagger_{t+\tau}\right] \ll \sqrt{\Sigma \; / \; P}$$

Comparison of bias and variability

The bias is very small relative to the variance of the test statistic, so negligibility holds



Outline

1. Introduction and motivation
2. Theory
    1. General structure for existing forecast evaluation tests
    2. Negligibility result for applications with latent forecast target
    3. Primitive conditions for the negligibility result
3. Simulation study
4. **Empirical application**
5. Conclusion

Empirical application

- We illustrate the theory proposed here with an application to correlation forecasting
  - **Economically interesting**: used for hedging, pairs trading, etc.
  - **Statistically difficult**: latent
- We will compare 4 different DCC-type models for predicting daily correlation (described in detail below)
  1. Baseline DCC model
  2. Extension to allow for a "leverage effect"
  3. Extension to use high frequency data to estimate current correlation
  4. Both extensions

Data

- We use daily asset returns on two pairs of stocks
  1. Proctor & Gamble and General Electric
  2. Microsoft and Apple

  - Extensions to include all four, or even more, stocks are possible

- Sample period: Jan 2000 – Dec 2010, so $T = 2733$ obs
  - In-sample: $R = 1500$, Out-of-sample: $P = 1233$

- To evaluate these forecasts we use high frequency data from TAQ, with realized correlation estimated using sampling frequency between 1 minute and 130 minutes.
  - We compute realized correlation using 10 equally-spaced subsamples.

Forecasting models I

- All models use the same mean and volatility specifications:

$$
\begin{aligned}
r_{it} &= \mu_i + \sigma_{it}\varepsilon_{it} \\
\sigma_{it}^2 &= \omega_i + \beta_i \sigma_{i,t-1}^2 + \alpha_i r_{i,t-1}^2 + \delta_i r_{i,t-1}^2 \mathbf{1}\{r_{i,t-1} \leq 0\} + \gamma_i RV_{i,t-1}^{1\,\mathrm{min}}
\end{aligned}
$$

1. **Baseline** DCC model (Engle, 2002, JBES):

$$
\begin{aligned}
Q_t &= \hat{R}(1-a-b) + bQ_{t-1} + a\varepsilon_{t-1}\varepsilon_{t-1}' \\
\rho_t &= \frac{Q_{12,t}}{\sqrt{Q_{11,t}Q_{22,t}}}
\end{aligned}
$$

Forecasting models II

2. DCC with an **asymmetric** term (Cappiello, et al., 2006, JFEC):

$$\begin{aligned} Q_t &= \hat{R}\left(1\text{-}a\text{-}b\text{-}d\right) + bQ_{t-1} + a\varepsilon_{t-1}\varepsilon'_{t-1} + d\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}'_{t-1} \\ \text{where} \quad \boldsymbol{\eta}_t &\equiv \varepsilon_t \odot \mathbf{1}\left\{\varepsilon_t \leq 0\right\} \end{aligned}$$

3. DCC with **realized** correlation:

$$Q_t = \hat{R}\left(1\text{-}a\text{-}b\text{-}g\right) + bQ_{t-1} + a\varepsilon_{t-1}\varepsilon'_{t-1} + g \cdot RCorr_{t-1}^{65\,\text{min}}$$

4. DCC with an **asymmetric** term **and realized** correlation:

$$Q_t = \hat{R}\left(1\text{-}a\text{-}b\text{-}d - g\right) + bQ_{t-1} + a\varepsilon_{t-1}\varepsilon'_{t-1} + d\boldsymbol{\eta}_{t-1}\boldsymbol{\eta}'_{t-1} + g \cdot RCorr_{t-1}^{65\,\text{mi}}$$

In-sample results

- Estimating these models in-sample period yields log-likelihoods below, along with *p*-values from LR tests against the DCC model:

	PG-GE	MS-AP
DCC	-2786.8	-4109.5
A-DCC	-2784.7	-4109.5
	(0.043)	(1.000)
R-DCC	-2781.8	-4094.3
	(0.002)	(0.000)
AR-DCC	-2781.4	-4094.3
	(0.005)	(0.000)

- So the realized term is significant in both pairs; asymmetric term only signif for PG-GE
- But do these additional terms help improve forecast performance?

Comparison of models for correlation forecasting: PG–GE

t-stats from GW test. Positive indicates model beats baseline DCC. KV crit val=2.77

	DCC vs *A*-DCC	DCC vs *R*-DCC	DCC vs *AR*-DCC
RRho1m	1.947	1.626	1.745
RRho5m	1.845	2.040	2.099
RRho15m	2.047	1.945	1.962
RRho30m	2.246	1.529	1.679
RRho65m	1.642	0.828	0.947
RRho130m	0.850	0.830	0.655

Comparison of models for correlation forecasting: MS–AP

t-stats from GW test. Positive indicates model beats baseline DCC. KV crit val=2.77

	DCC vs *A*-DCC	DCC vs *R*-DCC	DCC vs *AR*-DCC
RRho1m	-1.024	4.405*	3.712*
RRho5m	-1.156	4.357*	2.234
RRho15m	-1.195	4.279*	2.116
RRho30m	-1.055	3.948*	2.289
RRho65m	-1.168	3.506*	2.222
RRho130m	-1.243	3.342*	1.847

Summary and conclusion

- We present a general forecast evaluation framework allowing for a latent target variable
  - **Forecasting:** correctly specified or mis-specified, nested or non-nested, rolling or recursive, EPA or SPA, single or step-wise procedures
  - **High frequency:** general volatility and jump functionals, semi-variance, ratios (eg, beta and correlation)
- A set of realistic Monte Carlo studies provides support for the asymptotic theory
- Our empirical application shows evidence that the DCC model can be improved by including high frequency data

Appendix

Outline

1. Introduction and motivation
2. Theory
    1. General structure for existing forecast evaluation tests
    2. Negligibility result for applications with latent forecast target
    3. **Primitive conditions for the negligibility result**
3. Simulation study
4. Empirical application
5. Conclusion

Forecast evaluation with *unobservable* target I

Assumptions on continuous time DGP

- **Assumption HF:** Let $k \geq 2$ and $C > 0$ be constants. Then suppose:
  1. The process $(X_t)_{t \geq 0}$ is a d-dimensional Itô semimartingale with the following form

$$
\begin{aligned}
X_t &= X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + J_t, \\
J_t &= \int_0^t \int_{\mathbb{R}} \delta(s,z) 1_{\{\|\delta(s,z)\| \leq 1\}} \tilde{\mu}(ds, dz) \\
&\quad + \int_0^t \int_{\mathbb{R}} \delta(s,z) 1_{\{\|\delta(s,z)\| > 1\}} \mu(ds, dz),
\end{aligned}
$$

- b_s is a d-dimensional càdlàg adapted process
- W_s is a d' -dimensional standard Brownian motion
- σ_s is a $d \times d'$ càdlàg adapted process, and $c_s = \sigma_s \sigma_s^\intercal$
- $\delta(\cdot)$ is a d-dimensional predictable function
- μ is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ with compensator $\nu(ds, dz) = ds \otimes \lambda(dz)$ for some σ-finite measure λ and $\tilde{\mu} = \mu - \nu$.

Forecast evaluation with *unobservable* target II

Assumptions on continuous time DGP

2. The process σ_t is a $d \times d'$ Itô semimartingale with the form

$$\sigma_t = \sigma_0 + \int_0^t \tilde{b}_s ds + \int_0^t \tilde{\sigma}_s dW_s + \int_0^t \int_{\mathbb{R}} \tilde{\delta}\left(s, z\right) \tilde{\mu}\left(ds, dz\right),$$

where

- $\tilde{b}_s$ is a $d \times d'$ càdlàg adapted process
- $\tilde{\sigma}_s$ is a $d \times d' \times d'$ dimensional càdlàg adapted process
- and $\tilde{\delta}\left(\cdot\right)$ is a $d \times d'$ dimensional predictable function.

Forecast evaluation with *unobservable* target III

Assumptions on continuous time DGP

3. For some constant $r \in (0, 2]$, and nonnegative deterministic functions $\Gamma(\cdot)$ and $\tilde{\Gamma}(\cdot)$ on R, we have

$$\begin{aligned} \|\delta(s,z)\| &\leq \Gamma(z) \\ \text{and} \quad \|\tilde{\delta}(s,z)\| &\leq \tilde{\Gamma}(z) \quad \text{for all} \quad (s,z) \in \mathbb{R}_+ \times \mathbb{R} \end{aligned}$$

and

$$\begin{aligned} \int_{\mathbb{R}} (\Gamma(z)^r \wedge 1)\lambda(dz) + \int_{\mathbb{R}} \Gamma(z)^k 1_{\{\Gamma(z)>1\}} \lambda(dz) &< \infty, \\ \text{and} \quad \int_{\mathbb{R}} \tilde{\Gamma}(z)^k \lambda(dz) &< \infty \end{aligned}$$

Forecast evaluation with *unobservable* target IV

Assumptions on continuous time DGP

4. Let

$$
\begin{aligned}
b'_s &= b_s - \int_{\mathbb{R}} \delta(s,z)\, 1_{\{\|\delta(s,z)\| \leq 1\}} \lambda(ds) \quad \text{if} \quad r \in (0,1] \\
b'_s &= b_s \quad \text{if} \quad r \in (1,2]
\end{aligned}
$$

Then we have for all $s \geq 0$,

$$
\mathbb{E}\|b_s\|^k + \mathbb{E}\|b'_s\|^k + \mathbb{E}\|\sigma_s\|^k + \mathbb{E}\|\tilde{b}_s\|^k + \mathbb{E}\|\tilde{\sigma}_s\|^k \leq C.
$$

5. For each day t, the process X is sampled at deterministic discrete times $t-1=\tau(t,0)<\cdots<\tau\left(t,n_t\right)=t$, where n_t is the number of intraday returns. Moreover, with

$$d_{t,i}=\tau(t,i)-\tau(t,i-1)$$

we have

$$\begin{aligned} d_t &= \sup_{1\le i\le n_t} d_{t,i}\to 0 \\ \text{and}\quad n_t &= O(d_t^{-1}) \quad \text{as} \quad t\to\infty \end{aligned}$$

Example: Volatility functions for continuous case

- Notation: $\Delta_{t,i}X$ is the ith return on day t, with sampling interval $d_{t,i}$.

- Proxy: $Y_t \equiv \widehat{\mathcal{I}}_t\left(g\right) = \sum_{i=1}^{n_t} g(\Delta_{t,i}X / d_{t,i}^{1/2}) d_{t,i}$.
- The latent target:

$$Y_t^{\dagger} \equiv \mathcal{I}_t(g) \equiv \int_{t-1}^{t} \rho\left(c_s; g\right) ds$$

where c_s is the spot variance-covariance matrix of X_t and $\rho\left(c_s; g\right) = E\left[g\left(U\right)|c_s\right]$ for $U \sim N\left(0, c_s\right)$.

- Examples:
  - $g(x) = xx^{\intercal}$: $I_t(g) = \int_{t-1}^{t} c_s\, ds$
  - $g(x) = x^4/3$: $I_t(g) = \int_{t-1}^{t} \sigma_s^4\, ds$
  - $g(x) = \cos(\sqrt{2u}x)$: $I_t(g) = \int_{t-1}^{t} \exp(-u\sigma_s^2)\, ds$.

Example: Volatility functions for continuous case I

- **Proposition:** Let $p \in [1,2)$. Suppose for some constant $C > 0$ the following conditions hold:

  1. X_t is continuous
  2. $g(\cdot)$ and $I(\cdot; g)$ are differentiable and for some $k_1 \geq 0$, $\|\partial g(x)\| \leq C(1 + \|x\|^{k_1})$ and $\|\partial \rho(A; g)\| \leq C(1 + \|A\|^{k_1/2})$
  3. Assumption HF with some $k \geq \max\{2k_1 p/(2-p), 2\}$
  4. $E[|\rho(c_s; g^2)| + |\rho(c_s; g)|^p] \leq C$ for all $s \geq 0$.

  Then

$$\|\widehat{\mathcal{I}}_t(g) - \mathcal{I}_t(g)\|_p \leq K d_t^{1/2}.$$

Example: Volatility functions for continuous case II

- This establishes the applicability of our "negligibility" result to (no-jump) high frequency applications under quite weak conditions.
- The paper presents similar results for jump functionals, jump-robust volatility functionals, and some additional special cases:
  - Realized variance and covariance (in presence of jumps)
  - Bi-power variation
  - Realized semi-variance
  - Realized correlation and beta

Comparison of models for correlation forecasting: PG–GE

t-stats from GW test. Positive indicates second model beats first model. KV crit val=2.77

	A-DCC vs *R*-DCC	*A*-DCC vs *AR*-DCC	*R*-DCC vs *AR*-DCC
RRho1m	1.231	1.426	0.762
RRho5m	1.627	1.819	0.517
RRho15m	1.470	1.703	1.000
RRho30m	0.881	1.271	0.486
RRho65m	-0.153	0.413	0.973
RRho130m	0.688	0.516	-0.031

Comparison of models for correlation forecasting: MS–AP

t-stats from GW test. Positive indicates second model beats first model. KV crit val=2.77

	A-DCC vs *R*-DCC	*A*-DCC vs *AR*-DCC	*R*-DCC vs *AR*-DCC
RRho1m	3.134*	3.657*	-1.580
RRho5m	4.506*	6.323*	-1.586
RRho15m	4.044*	5.449*	-1.441
RRho30m	4.635*	7.284*	-0.882
RRho65m	6.059*	7.868*	-0.635
RRho130m	3.392*	5.061*	-1.582