Exact finite-sample variance and BLUE weights for the weighted Hayashi–Yoshida covariance estimator

Dr Yang Azzollini
Oxford DPhil Statistics, 2016
yang.maths@gmail.com

April 2026

Setup

Let X, Y be a bivariate Brownian motion on $[0,1]$ with constant parameters σ_1^2, σ_2^2, σ_{12}. Observations $\{X_{t_i^X}\}_{i=0}^m$, $\{Y_{t_j^Y}\}_{j=0}^n$ arrive at fixed (but arbitrary) timestamps, independently of the price process. Let $R_i = X_{t_i^X} - X_{t_{i-1}^X}$, $S_j = Y_{t_j^Y} - Y_{t_{j-1}^Y}$ denote the returns, and $\tau_{ij} = |J_i^X \cap J_j^Y|$, $\tau_i^X = |J_i^X|$, $\tau_j^Y = |J_j^Y|$ the overlap structure. Write $\Theta = \sum_{ij} \tau_{ij}^2/(\tau_i^X \tau_j^Y)$.

The *extended weighted Hayashi–Yoshida class* of estimators of σ_{12} is $\{\hat\sigma_{12}(c) = \sum_{ij} c_{ij} R_i S_j : c_{ij} \in \mathbb{R}\}$, with $c_{ij} = 0$ whenever $\tau_{ij} = 0$. Unbiasedness requires $\sum_{ij} c_{ij}\tau_{ij} = 1$. The *normalised weighted ZHY estimator* is $\hat\sigma_{12}(w) = \sum_{ij} w_{ij} R_i S_j / \sum_{ij} w_{ij}\tau_{ij}$.

Results from the 2016 DPhil thesis (Chapter 5)

Theorem 1 (Exact finite-sample variance, thesis Theorem 5.1)**.** *For any weight matrix w,*

$$\mathrm{Var}(\hat\sigma_{12}(w)\,|\,\tau) \;=\; \sigma_1^2\sigma_2^2\, A_1(w) \;+\; \sigma_{12}^2\, A_2(w),$$

where $A_1(w)$ and $A_2(w)$ are explicit functions of $\{\tau_{ij}, \tau_i^X, \tau_j^Y\}$. The identity is exact, holds for any fixed timestamp structure, and requires no distributional assumption on the observation times beyond their independence from prices.

The proof of Theorem 5.1 on p. 108 of the thesis uses the following geometric fact, sometimes referred to separately as the *Interval-Structure Lemma*: for $i \neq k$ and $j \neq l$, the overlaps τ_{il} and τ_{kj} cannot both be strictly positive, because J_i^X, J_k^X are disjoint consecutive X-intervals and similarly for Y.

$\mathbf{A_1}$-optimal weights (thesis §5.4). Minimising $A_1(w)$ subject to the unbiasedness constraint gives $w_{ij}^* = \tau_{ij}/(\tau_i^X\tau_j^Y)$, and at this choice $A_1(w^*) = 1/\Theta$. Theorem 5.2 of the thesis extends these expressions to the microstructure-noise model.

New results (2026)

Theorem 2 (Full-variance optimum, general κ)**.** *Let $\kappa = \sigma_{12}^2/(\sigma_1^2\sigma_2^2) \in [0,1]$. Minimising $\sigma_1^2\sigma_2^2 A_1(w) + \sigma_{12}^2 A_2(w)$ under $\sum_{ij} w_{ij}\tau_{ij} = 1$ is a convex quadratic program whose KKT conditions reduce to an $(m+n+1)\times(m+n+1)$ linear system in the multipliers. The solution $w^*(\kappa)$ collapses to the thesis A_1-optimal weights at $\kappa = 0$. Numerical study gives a variance reduction of less than 1% over the A_1-optimal weights for all $\rho \in [0,1]$.*

Theorem 3 (Two-step plug-in). *The estimator* $\hat{\sigma}_{12}^{(2)} = \sum_{ij} w_{ij}^*(\hat{\kappa}) R_i S_j / \sum_{ij} w_{ij}^*(\hat{\kappa})\tau_{ij}$*, with* $\hat{\kappa}$ *constructed from a consistent first-step* $\hat{\sigma}_{12}^{(1)}$*, is unbiased and attains the fixed-*κ *minimum of Theorem 2 in the large-sample limit.*

Theorem 4 (Cramér–Rao match at $\rho = 0$). *At* $\rho = 0$*, the* A_1*-optimal estimator attains the Cramér–Rao lower bound:*

$$\mathrm{Var}\big(\hat{\sigma}_{12}(w^*) \mid \tau\big)\Big|_{\rho=0} = \frac{\sigma_1^2\sigma_2^2}{\Theta} = I(\sigma_{12} \mid \tau)^{-1}\big|_{\sigma_{12}=0},$$

where $I(\sigma_{12} \mid \tau)$ *is the Fisher information for* σ_{12} *conditional on the timestamp structure.*

Lean 4 formalisation (2026)

The 2016 thesis results on the extended-ZHY class and the new results above have been formalised in the Lean 4 proof assistant with Mathlib. The development consists of four modules and contains zero `sorry` statements. The probabilistic input is packaged as a single structure recording the Brownian covariance identities and the Gaussian (Isserlis) fourth-moment formula; everything downstream is algebraic.

- `ZHY_Core`: unbiasedness characterisation, A_1-optimality, Cramér–Rao equality at $\rho = 0$.
- `ZHY_BLUE`: $A_1(c) \geq 1/\Theta$ for every unbiased c (Cauchy–Schwarz); equality at w^*.
- `ZHY_Algebra`: combinatorial scaffolding (four `Finset.sum` identities) used by the variance decomposition.
- `ZHY_Variance`: exact finite-sample variance decomposition (Theorem 1 above), including the Interval-Structure Lemma as a hypothesis on τ.

Positioning with respect to the literature

To the author's best knowledge, no published work contains (i) an exact, finite-sample, timestamp-conditional variance formula for the weighted Hayashi–Yoshida class, (ii) a BLUE result for the class in finite samples, or (iii) a machine-checked proof of either. The existing literature — Hayashi–Yoshida (2005, 2008, 2011), Griffin–Oomen (2011), Bibinger (2014), Koike (2014–2026), Ogihara–Yoshida (2012) — works in the $\lambda \to \infty$ asymptotic regime or with renewal / kernel approximations.

Before submitting, the author would welcome a brief sanity check from an authority on the Hayashi–Yoshida estimator that these finite-sample exact-variance and BLUE results have not appeared elsewhere. Full proofs of Theorem 1 (from the 2016 thesis) and Theorems 2–4 are available on request, as is the Lean 4 source.

Repository (public): `github.com/ChubbyFaceEuler/Paper_2026`. DPhil thesis: *Correlation methods in the statistical analysis of financial trading data*, University of Oxford, 2016.