Index Arbitrage and Refresh Time Bias in Covariance Estimation

Dale W.R. Rosenthal Jin Zhang

University of Illinois at Chicago

10 May 2011

Variance and Covariance Estimation

Classical problem with many financial applications:

- Risk management: VaR, ES, risk budgeting;
- Portfolio optimization and asset allocation;
- Option valuation and hedging;
- Market making: inventory risk;
- Pairs trading/relative value strategies;
- Forecasting, rate of information flow, “liquidity.”
- Estimate corporate bond variances, default probabilities.
- Covariance risk: time-varying betas, Sharpe ratios.

High-Frequency Data

Estimation increasingly done with high-frequency data. Allows:

- Study intraday pattern of volatility/covariance;
- Improve volatility/covariance forecasts;
- Portfolio performance gain worth 50–200 bp annually[1].
- Time-varying correlations, variances, betas, Sharpe ratios;
- High-frequency estimates crucial for market making, HFT;
- Post IPO/merger: quicker estimation allows more investment.

[1] Fleming, Kirby, Ostdiek (2003)

Variance, Covariance Literature

Much work on high-frequency variance and covariance estimation:

- Handling microstructure noise/asynchronicity
  1. Kernel-based approach:
     Barndorff-Nielsen, Hansen, Lunde, Shephard (2008, 2010)
  2. Pre-averaging: Podolskij, Vetter (2009);
     Christensen, Kinnebrock, Podolskij (2010)
  3. Two-scales Realized Variance, Covariance:
     Zhang, Mykland, Aït-Sahalia (2005); Zhang (2010)
- Handling jumps
  1. Bipower Variation, Covariation:
     Barndorff-Nielsen, Shephard (2004a, 2004b)
  2. Median Realized Volatility:
     Andersen, Dobrev, Schaumburg (2008)

Index Arbitrage

- *Index Arbitrage*: Trade index members vs. futures/ETF.
  - Simple application of APT; has been done for decades;
  - Increasing automation greatly eases trading.
- US indexes: Dow 30, Nasdaq 100, S&P 500, Russell 2000 (!).
- Myth: Too expensive/fussy to trade all those stocks. (Why?)
- *Spread*: $\delta_t = \underbrace{\sum_{i=1}^{N} w_i S_{it}}_{\text{Index}} - \underbrace{F_t}_{\text{Futures}}$
- Strategy: Trade stocks vs. futures/ETF when $|\delta|$ "large."

Index Arbitrage Bias

- Index arb pushes index, futures, ETF toward each other.
- Worse: trades determine (mostly) contemporaneous returns.
    - Index arbitrage creates simultaneous index members trades.
    - Index arbitrage often create trades when $|\delta_t|$ large.
- Thus price co-movement is due to *two* DGPs:
    - Similarity of economic fundamentals (Σ);
    - Reversion (O-U?) of index-ETF-futures prices (δ).
- Spread δ_t biases estimates of variance, covariance.
- We suspect the bias is larger for illiquid stocks.

Classical Model for Financial Data

Often assume geometric Brownian motion:
Let X_t be a vector of log-stock prices $\log(S_t)$ at time t;

$$dX_t = \underbrace{\mu dt}_{\text{drifts}} + \underbrace{\Sigma dW_t}_{\text{diffusions}} . \quad (1)$$

We also can augment this to address inadequacies:

- Stochastic volatility;
- Leverage effects;
- Account for microstructure noise; and,
- Incorporate jumps.

Classical Model with Index Arbitrage

Index arbitrage adds an O-U term to the standard drift+diffusion:

$$dX_t = \underbrace{\mu dt}_{\text{drifts}} + \underbrace{\Sigma dW_t}_{\text{diffusions}} - \underbrace{\gamma \delta_t / S_t}_{\substack{\text{index arb}\\ \text{effects}}} \tag{2}$$

$$d\delta_t = \lambda(\delta_t^* - \delta_t) + \sigma_\delta dZ_t \tag{3}$$

where

γ = price sensitivities to spread δ_t, $\gamma > 0$, $\gamma \overset{?}{\propto} w/2$; and,

λ = speed of mean reversion.

Index Arbitrage and Variance Estimation

- Spread δ_t biases estimates of variances σ_i^2.
- Continuous-time bias is easy to determine:

$$\mathrm{Var}(dX_{it}) = \sigma_i^2 + \gamma_i^2 \frac{\sigma_\delta^2}{2\lambda S_{it}^2}; \quad (4)$$

$$E(\hat{\sigma}_i^2) > \sigma_i^2. \quad (5)$$

- We often "sample" by computing returns between trades.
- But index arb causes trades $\Rightarrow$ not sampling at random.
- More trades if $|\delta_t|$ large $\Rightarrow$ larger observed effect.
  - Endogeneity in trade times *and* spread return sizes[2].

[2] Thus Li, Mykland, Renault, Zhang, Zheng (2011 WP) does not hold.

Index Arbitrage and Covariance Estimation

- Spread δ_t also biases estimated covariances $\hat{\Sigma}_{ij}$.
- Continuous-time bias is easy to determine:

$$\mathrm{Cov}(dX_{it}, dX_{jt}) = \Sigma_{ij} + \gamma_i \gamma_j \frac{\sigma_\delta^2}{2\lambda S_{it} S_{jt}}; \tag{6}$$

$$E(\hat{\Sigma}_{ij}) > \Sigma_{ij}. \tag{7}$$

- Recall: index arb causes trades $\Rightarrow$ not random sampling.
- However, bias may be worse for covariance estimation.
  - Covariance estimation must handle asynchronous trading.
  - Most covariance estimates use "refresh times;" but,
  - Refresh times amplify over-sampling of index arb comovement.

Refresh Times and Asynchronous Trading

We use refresh times to handle asynchrony of trading. However:

- Many non-index-arb trades do not create refresh times.
- Index arbitrage trades create refresh times.
  - ⇒ Refresh times discard few/no index arb trades.
- Thus over-sampling (and bias) likely worse than for variance.

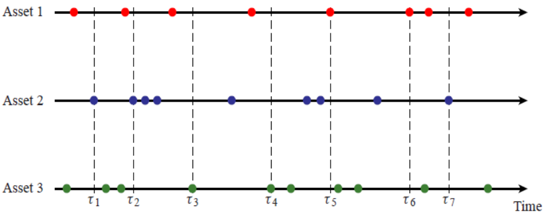


Figure 1: Example refresh times. Source: Barndorff-Nielsen *et al.* (2010)

Data

- Look at some data to see if we find these effects.
- Index: Dow Jones Industrial Average (DJIA)
- ETF: S&P depository receipt, (DIA) $=$ DJIA/100 $\pm c$
- Sample period: 1–31 October 2008.
- Data source: NYSE Trade and Quote (TAQ) Database.
- Data cleaning as in Barndorff-Nielsen *et al.* (2009).

Index Arbitrage Spread Construction

- Construct tick-by-tick DJIA bid and ask prices.
- To avoid exchange clock differences, compare to DIA ETF.
- *Spread*: $\delta_t = \underbrace{\sum_{i=1}^{N} w_i S_{it}}_{\text{Index}} - \underbrace{S_{DIA,t} \times 100}_{\text{ETF}}$
- Note more trading when spread is large: over-sampling.

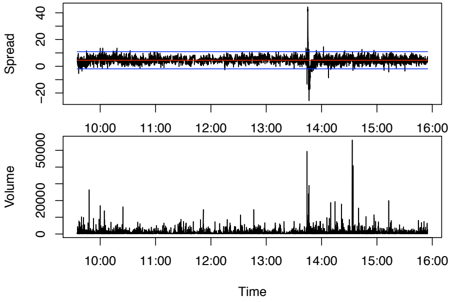


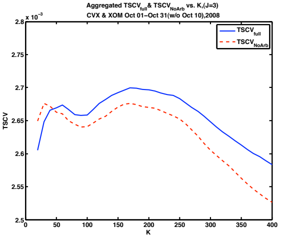
Figure 2: Dow 30 index-ETF spread on 1 Oct 2008

Index Arb Covariance Bias: Cleaning

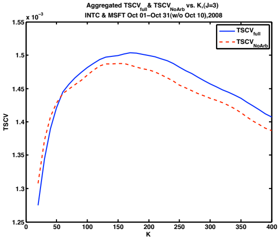
- To see if index arb refresh times have an effect, remove them.
- Flag trades when spread $\delta_t > 2$ s.d.s from daily mean.
- Compute TSCV with and without flagged trades.
- Allow slow time scale to vary to see limiting behavior.
- Same-sector pairs all show overestimation of covariance.

Index Arb Covariance Bias: Plots

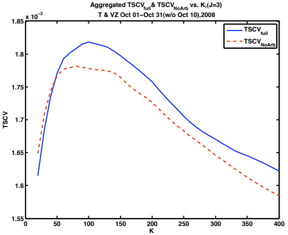
Same-sector pairs all show overestimation of covariance.



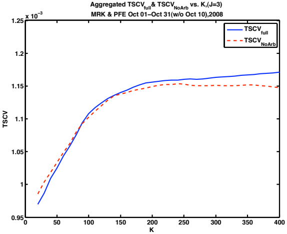
ChevronTexaco vs. ExxonMobil



Intel vs. Microsoft



T-Mobile vs. Verizon



Merck vs. Pfizer

Conclusion

- Shown index arbitrage biases variance, covariance estimates.
- Biases all high-frequency variance, covariance estimation.
- Reasons why refresh times can exacerbate this problem.
- Data analysis: some covariances overestimated by about 3%.
- Overestimated covariances may cause over-diversification.
  - Seems innocuous, but this can raise investors costs.
- Combined overestimates reduce allocations to risky assets.
- Data analysis remains to be done for variance bias.
- Suggests more careful data cleaning needed.