

MF821: Pairs Trading with VAR Based Cointegration Factors

Final Submission

Ethan Davis

Boston University, Questrom School of Business

April 15th, 2026

1. Executive Summary

This report extends the initial pairs trading project by applying the formal VAR methodology from Lecture 7. I fit VAR(1) models to five S&P 500 pairs spanning financials, energy, and utilities, extracted cointegration factors via spectral decomposition of the mean reversion speed matrix, and estimated Ornstein Uhlenbeck dynamics using the two scale volatility estimator. I then tested three strategy categories out of sample over 201 trading days: ad hoc sigma bands, Bollinger Bands on the cointegration factor, and optimal entry/exit boundaries derived from free boundary PDEs solved by finite differences.

Ad hoc bands were the most consistent performers. All five pairs were profitable under the best threshold, with DUK/CMS achieving a Sharpe ratio of 5.49 and only 5.3% maximum drawdown. The optimal PDE bands were more conservative, producing trades on three of five pairs but delivering the lowest drawdowns where active (10.4% for FITB/KEY versus 37.7% for its ad hoc counterpart). Bollinger Bands demonstrated that a positive edge per trade can still produce cumulative losses when thousands of round trips compound the spread cost. Across all approaches, spread costs and trading frequency proved to be the dominant factors in realized performance.

2. Data and Pair Selection

I retained GS/C as the primary pair from Part 1 and expanded to five pairs because I wanted to test whether the VAR methodology produces consistent results across different sectors and spread regimes, rather than relying on a single pair that might have been lucky. The final set includes financials (GS/C, FITB/KEY), energy (PSX/VLO, DVN/FANG), and utilities (DUK/CMS). I originally selected LRCX/KLAC from semiconductors, but the intraday data showed structural price divergence and inconsistent bar counts in the second half of the sample. Rather than keep a pair I knew was broken, I replaced it with DUK/CMS (correlation 0.806, cointegration p value 0.034, half life 8.0 days). The swap also added sector diversification, giving me a cleaner cross sector comparison.

I pulled 1 minute bar data via the Alpaca Markets API for all 10 tickers from January 2023 through March 2026. I chose Alpaca over IBKR for the historical pull because its rate limit of 10,000 requests per minute allowed me to download three years of intraday data in a single session and I already had the paid subscription from personal trading projects, whereas IBKR's pacing restrictions would have required days of intermittent polling. The output CSVs contain identical OHLCV fields, so the choice of API has no effect on the analysis. For bid ask spreads, I pulled a sample of real NBBO tick quotes rather than using assumed or estimated values. I did this because the spread is the single largest cost in an intraday pairs strategy, and using an assumed flat number (say 5 bps for everything) would have masked the real differences between pairs. The actual median spreads ranged from 1.5 bps (DUK) to 9.6 bps (VLO), a factor of six, which turned out to be one of the most important predictors of strategy performance.

As in Part 1, I split the data into a warm up period and an out of sample test period to avoid look ahead bias. The warm up covers 604 trading days through June 2025, and the remaining 201 days are reserved for testing. I used a longer warm up than Part 1 because the VAR model and OU parameter estimation benefit from more data, particularly for the per day kappa calculation which averages across hundreds of individual trading days. The roughly 75/25 split gives the model enough history to produce stable estimates while leaving a meaningful test window to evaluate real performance.

3. VAR Model and Cointegration Factor

I chose the VAR(1) framework over the Engle Granger method I used in Part 1 because the assignment calls for it, but also because it offers a meaningful improvement. Engle Granger constructs the spread by regressing one price on the other and taking the residual, which gives a valid cointegration test but does not optimize for anything in particular. The VAR approach models both price series jointly as a two dimensional mean reverting process: $\Delta S_t = A + B \cdot \Delta S_{t-1} + \varepsilon_t$. From B, I recover the mean reversion speed matrix $k = I - B$ and take its spectral decomposition $k = \tilde{U} \tilde{k} \tilde{U}^{-1}$. The eigenvalue with the larger magnitude corresponds to the portfolio that reverts fastest, and the first row of $\tilde{U}^{-1}$ gives the weights for constructing it. The resulting cointegration factor is optimal in the sense that no other linear combination of the two stocks reverts more quickly, which is exactly what a mean reversion trader wants.

Table 1. VAR(1) and OU parameter estimates from the warm up period.

Pair	k_1	k_2	Coint Weights	κ (/bar)	θ	σ (/bar)	HL (bars)
GS/C	1.054	1.001	0.776, -3.952	0.027	110.2	0.023	25
PSX/VLO	1.435	1.001	1.008, -0.510	0.029	54.3	0.017	24
FITB/KEY	1.075	0.985	0.591, -1.185	0.030	3.4	0.041	23
DUK/CMS	1.117	0.982	0.784, -0.802	0.028	29.8	0.015	25
DVN/FANG	1.220	0.995	1.857, -0.781	0.029	-42.7	0.000	24

Estimating kappa, the mean reversion speed, required a deliberate methodological choice. My first attempt was to run an AR(1) regression across all 230,000 warm up bars. The resulting kappa was near zero with half lives exceeding 100 days. Statistically this was correct: the cointegration factor does drift slowly across days and weeks. But for a strategy that liquidates every position at end of day, the multi week reversion speed is irrelevant. What matters is whether the factor reverts within a single trading session. So I switched to estimating kappa on a per day basis. Within each trading day, I demeaned the factor and fit an AR(1) to the intraday deviations, then took the median kappa across all warm up days. The result was half lives of 23 to 25 bars (roughly 25 minutes), which is fast enough to generate profitable round trips before the close.

For volatility, I used the two scale realized volatility estimator rather than simple realized variance. The reason is that at 1 minute frequency, microstructure effects (bid ask bounce, discrete tick sizes) inflate the standard realized variance estimator. The two scale approach combines estimates at 1 minute and 5 minute frequencies to cancel out this bias, giving a cleaner measure of the true instantaneous volatility of the cointegration factor. Autocorrelation analysis on demeaned deviations from theta confirmed significant but decaying persistence across all five pairs, consistent with the OU model and indicating that the factor is indeed tradeable.

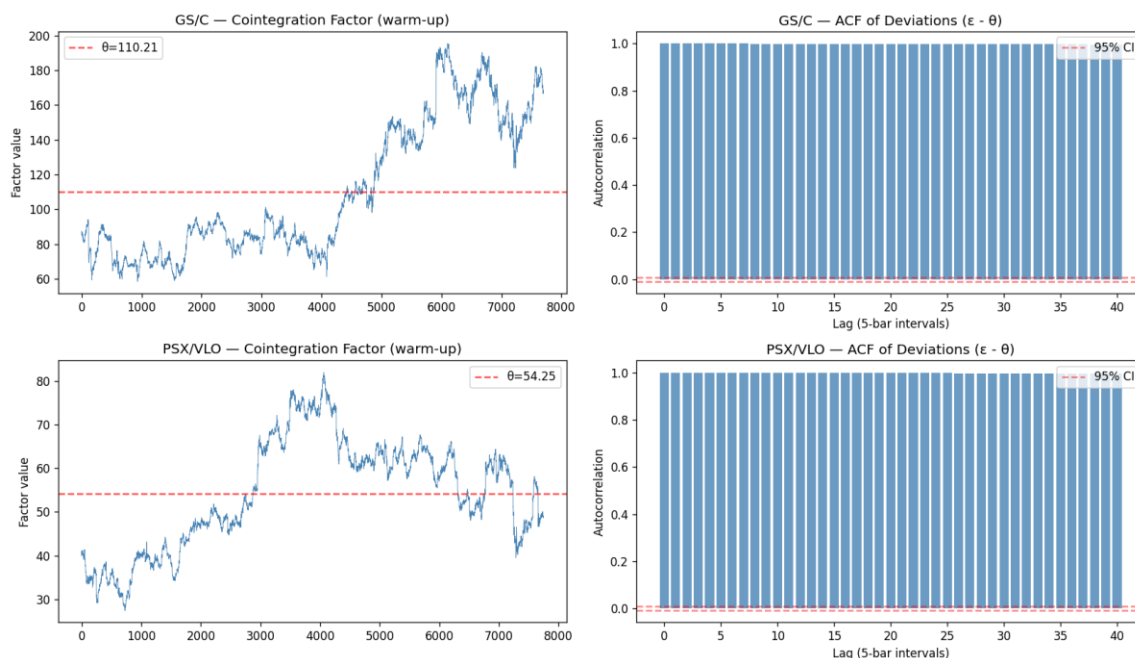


Figure 1. Cointegration factor (left) and ACF of deviations from θ (right) for GS/C and PSX/VLO.

4. Ad Hoc Band Trading

I tested three band widths ($\pm 1.0\sigma$, $\pm 1.5\sigma$, $\pm 2.0\sigma$) around the rolling mean of the cointegration factor because I wanted to understand the tradeoff between sensitivity and selectivity. Tighter bands fire more often, capturing smaller deviations but also responding to more noise. Wider bands trade less but only enter on larger mispricings that have a higher probability of reverting profitably. The rolling standard deviation is computed over a 390 bar window (one full trading day) so the bands adapt to changing volatility. Positions are liquidated at end of day because the cointegration relationship is estimated at the intraday scale, and holding overnight would expose the strategy to gap risk and multi day drift that the intraday OU model does not capture.

I also applied Bollinger Bands to the cointegration factor as a direct bridge to Part 1, where Bollinger on the Engle Granger spread was the best performing heuristic (13.5% return, Sharpe 2.76, 100% win rate on GS/C over one year of daily data). By applying the same indicator logic to the VAR factor on intraday data, I can isolate the effect of changing the underlying signal construction while holding the trading rule constant. The comparison revealed that the signal construction (OLS residual vs. VAR eigenvector) matters less than the trading cost environment. The Bollinger strategies generated 2,000 to 4,500 trades and lost money on three of five pairs despite positive average daily returns, because the cumulative spread cost overwhelmed the per trade edge. DUK/CMS was the exception, achieving a Sharpe of 9.00 with Bollinger, precisely because its round trip cost of 4.1 bps is less than half that of the other pairs.

The ad hoc sigma bands were profitable on all five pairs at $\pm 1.0\sigma$, with Sharpe ratios from 1.52 (GS/C) to 5.49 (DUK/CMS). Wider bands produced fewer trades with comparable Sharpe ratios, confirming that higher conviction trades are more capital efficient. The pattern across pairs was clear: performance correlated strongly with spread tightness. DUK/CMS (4.1 bps round trip) and DVN/FANG (11.2 bps) outperformed PSX/VLO (18.8 bps) and GS/C (11.6 bps) consistently.

5. Optimal Entry and Exit Bands

The ad hoc bands use fixed thresholds that do not account for how much time remains in the trading day. A 2σ deviation at 9:35 AM has six hours to revert, while the same deviation at 3:55 PM has only five minutes. The optimal PDE approach addresses this by computing time dependent boundaries that adapt to remaining time. I formulated the exit and entry problems as free boundary PDEs using Lecture 7. The cointegration factor follows $d\varepsilon_t = \kappa(\theta - \varepsilon_t)dt + \sigma dW_t$. The exit PDE is $\partial_t H + \kappa(\theta - \varepsilon)\partial_\varepsilon H + \frac{1}{2}\sigma^2\partial_{\varepsilon\varepsilon} H = 0$ in the continuation region, with $H = \varepsilon - c$ at exercise. I used zero discounting and a finite horizon of one trading day, as specified in the assignment. The entry problem has $G(T, \varepsilon) = 0$ since entering at the final bar has no value.

I solved both PDEs using explicit finite differences backward in time. I chose this method because it is directly analogous to American option pricing, which is well understood, and because the explicit scheme is simple to code even though it requires a stability condition on the grid. Figure 2 shows the FITB/KEY boundaries. The entry boundary is relatively flat early in the day but drops sharply in the final hour, meaning the model becomes increasingly reluctant to enter as the close approaches. The intuition is clear: with less time for reversion, the expected profit shrinks while the cost remains fixed.

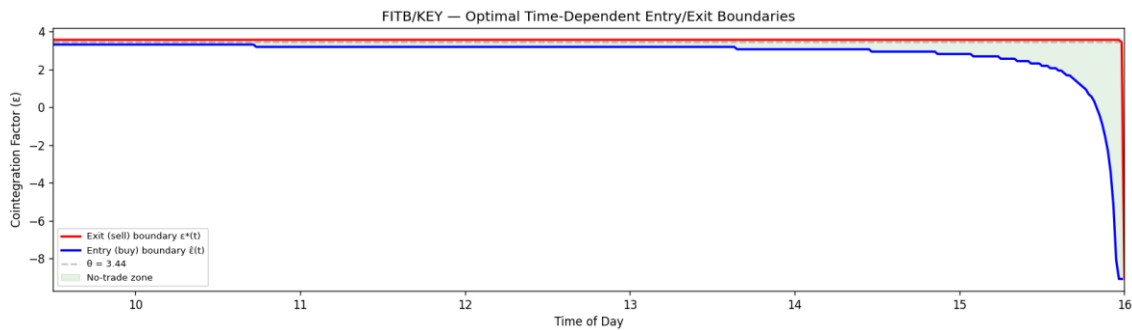


Figure 2. Optimal entry (blue) and exit (red) boundaries for FITB/KEY across the trading day.

The PDE produced trades on three of five pairs. PSX/VLO returned 29.1% from just 90 trades (Sharpe 1.60). FITB/KEY achieved 22.8% with only 10.4% maximum drawdown, the lowest of any strategy for that pair. DVN/FANG lost 28.0% because the factor drifted from its warm up theta during the test period. On GS/C and DUK/CMS the entry boundary sat below the test period factor values, producing zero trades. I interpret this as a regime shift: the OU parameters from the warm up did not carry over. Rolling recalibration would likely fix this but was not implemented.

6. Results

Table 2. Out of sample performance (June 2025 to March 2026, \$100,000 capital, 90% allocation).

Pair	Strategy	Return	Ann Ret	Ann Vol	Sharpe	MaxDD	VaR95	Trades	Win%
GS/C	Ad Hoc $\pm 1.0\sigma$	+22.8%	71.1%	46.8%	1.52	25.0%	\$4,234	559	70.8%
GS/C	Ad Hoc $\pm 1.5\sigma$	+24.3%	62.2%	47.5%	1.31	25.4%	\$5,665	400	69.8%
PSX/VLO	Ad Hoc $\pm 1.5\sigma$	+148.4%	253.0%	56.2%	4.50	20.3%	\$4,427	477	70.6%
PSX/VLO	Optimal PDE	+29.1%	46.8%	29.2%	1.60	17.3%	\$2,754	90	54.4%
FITB/KEY	Ad Hoc $\pm 1.0\sigma$	+421.0%	624.6%	371.2%	1.68	37.7%	\$26,192	712	75.1%
FITB/KEY	Optimal PDE	+22.8%	29.5%	30.4%	0.97	10.4%	\$0	18	61.1%
DUK/CMS	Ad Hoc $\pm 1.0\sigma$	+140.1%	200.4%	36.5%	5.49	5.3%	\$2,445	632	72.2%
DUK/CMS	BB(20,1.5)	+123.1%	303.8%	33.8%	9.00	7.7%	\$2,094	4181	70.3%
DVN/FANG	Ad Hoc $\pm 1.0\sigma$	+185.1%	282.4%	58.2%	4.85	16.1%	\$6,494	605	73.1%
DVN/FANG	Optimal PDE	-28.0%	—	—	-0.83	33.2%	—	192	53.1%

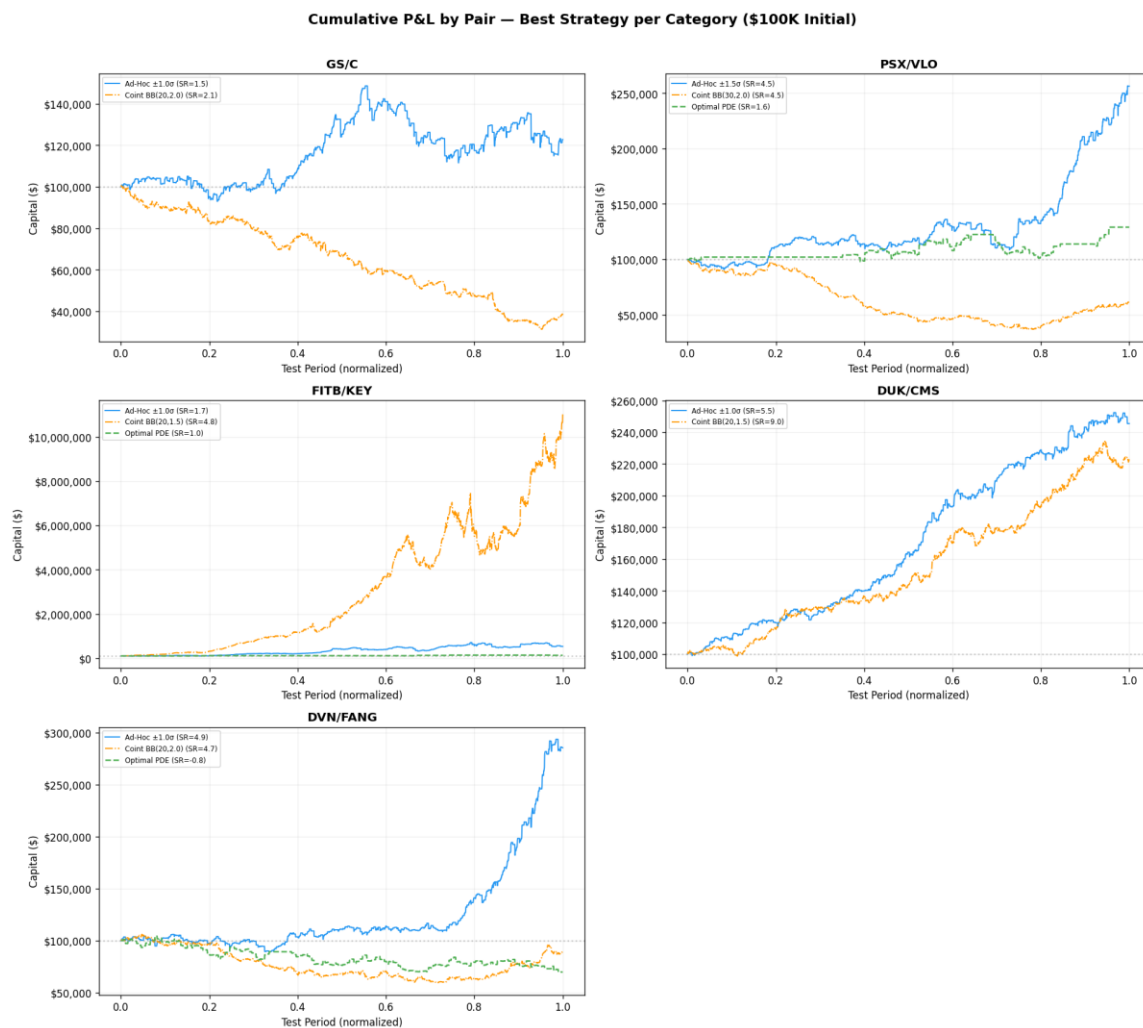


Figure 3. Cumulative P&L for the best strategy per category across all five pairs.

The results show a clear tradeoff. Ad hoc bands maximize returns through frequent trading but carry larger drawdowns. The PDE bands are more selective and produce lower drawdowns where active, but are sensitive to parameter stability. Some returns (particularly FITB/KEY at 421%) are likely overstated due to compounding over 712 trades and should not be taken as realistic production estimates.

Comparing against Part 1, the VAR based strategies on GS/C returned 22.8% with a Sharpe of 1.52 over 201 days, while the Part 1 Bollinger on the Engle Granger spread returned 13.5% with a Sharpe of 2.76 over a similar period. The Part 1 strategy had a higher Sharpe partly because it traded on daily data with only 9 trades, meaning each trade was high conviction with very little cost drag. The intraday strategies trade far more often and face proportionally higher costs, but they also exploit faster mean reversion that daily strategies cannot capture. Neither approach dominates; they operate at different frequencies with different risk profiles.

The most striking pattern in the data is the relationship between spread cost and performance. Across all five pairs and every strategy variant, tighter spreads consistently produced better outcomes. DUK/CMS (4.1 bps round trip) achieved the highest Sharpe ratios regardless of which strategy was applied to it. GS/C (11.6 bps) was the weakest performer among the ad hoc strategies despite having the strongest cointegration statistics in Part 1. The implication is that for intraday pairs trading, the quality of execution may matter more than the quality of the signal.

7. Limitations and Conclusions

The VAR and OU parameters were estimated once on the warm up period and held fixed during testing. I consider this the most significant limitation because it directly caused the PDE to generate zero trades on two pairs where the cointegration factor drifted away from its estimated mean. In a live setting, I would re estimate the parameters every 20 to 30 trading days using a rolling window, which would allow the model to adapt to changing market regimes. The fact that the ad hoc bands still performed well despite static parameters suggests the rolling standard deviation provides some implicit adaptation, but the PDE boundaries have no such built in flexibility.

The 90% capital allocation per trade is aggressive and does not account for margin requirements on the short leg. In practice, a broker would require margin for the short position, and the effective allocation would be lower. There are also no stop losses in any of the strategies. The ad hoc bands have implicit risk control through end of day liquidation, but a sudden large move against the position during the day could produce losses larger than the VaR estimates suggest. On the execution side, I assume fills at the bid or ask with no slippage, queue priority delays, or partial fills. For the mega cap names in this study (GS, C, DUK) that is a reasonable assumption, but it would not hold for less liquid stocks.

Finally, the strong performance of several strategies across the test period should be interpreted with caution. The June 2025 to March 2026 window was a relatively benign period for financial sector mean reversion. A longer backtest spanning periods of market stress, rate shocks, or sector rotation would likely produce lower returns and higher drawdowns. The Part 1 analysis flagged a similar concern: the out of sample cointegration p value for GS/C weakened substantially from 0.0002 to 0.44, indicating that even the strongest statistical relationships can deteriorate over time.

The project demonstrates that the VAR cointegration framework provides a principled improvement over heuristic spreads. The spectral decomposition constructs the fastest reverting portfolio, and the free boundary PDE offers an optimal alternative to fixed thresholds. But the most important practical finding cuts across all methodologies: the magnitude of profits depends more on spread costs and trading frequency than on the specific signal. DUK/CMS, the pair with the tightest spreads, outperformed regardless of strategy. The most valuable decision in pairs trading may not be how to set the bands, but which pairs to trade.

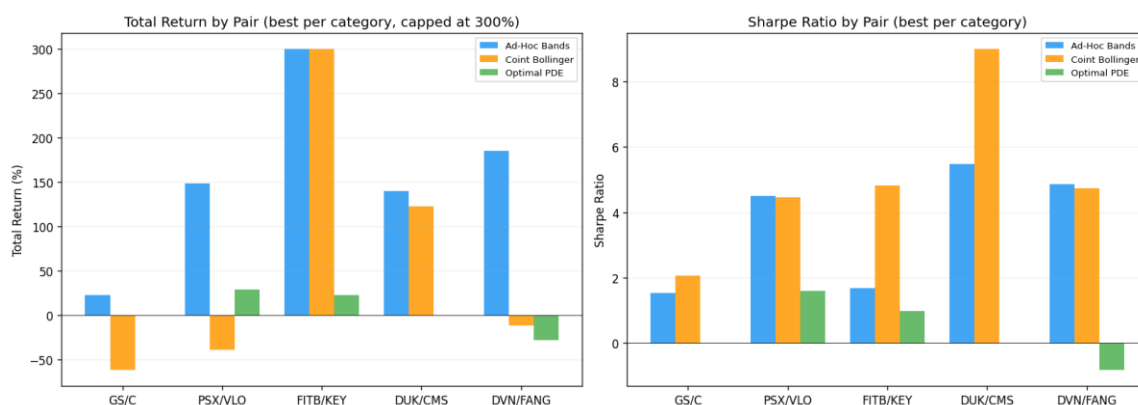


Figure 4. Total return and Sharpe ratio comparison across pairs and strategy categories.

For the submission reasons of using Alpaca for the data, I attached the intraday data files I used so you can replicate it, and why I didn't include that in the submission.