STA 542: Introduction to Time Series Analysis

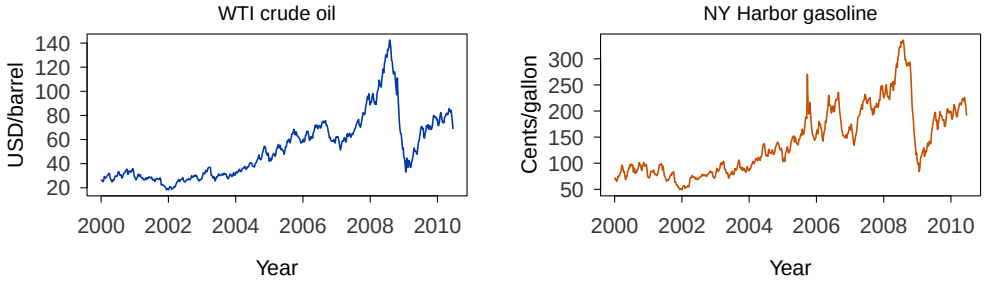
Lecture 10: Multivariate stationarity and cointegration

Lasse Vuursteen

Duke
UNIVERSITY

What do the oil and gasoline prices share?

Crude oil is an input to gasoline production. Consider their weekly spot prices.



Which movements occur together? What would separate models omit?

Real data: EIA via `astsa::oil` and `astsa::gas`; 545 weekly observations, 2000–mid-2010.

Covariance must compare coordinates as well as times

At time t, record $X_t = (X_{1t}, \ldots, X_{dt})^\top \in \mathbb{R}^d$. For random vectors U, V with finite second moments, define

$$\mathrm{Cov}(U, V) := \mathbb{E}[(U - \mathbb{E}[U])(V - \mathbb{E}[V])^\top].$$

Its (i, j) entry is $\mathrm{Cov}(U_i, V_j)$.

For paired prices, this records variation within each price and how the prices vary together.

Joint stationarity fixes every lagged covariance

Definition (Vector weak stationarity)

A process (X_t) in $\mathbb{R}^d$ is weakly stationary if it has finite second moments, a constant mean vector μ, and

$$\mathrm{Cov}(X_{t+h}, X_t) = \Gamma(h)$$

for every admissible t, h, with $\Gamma(h)$ depending only on the lag.

$\Gamma_{ij}(h)$ compares coordinate i at $t+h$ with coordinate j at t.

Stationary coordinates alone do not check the cross-covariances.

A lagged covariance matrix need not be symmetric

Let (Z_t) be white noise with variance one and set $X_t = (Z_t, Z_{t-1})^\top$. What is $\Gamma(1) = \text{Cov}(X_{t+1}, X_t)$?

A lagged covariance matrix need not be symmetric

Let (Z_t) be white noise with variance one and set $X_t = (Z_t, Z_{t-1})^\top$. What is $\Gamma(1) = \mathrm{Cov}(X_{t+1}, X_t)$?

$$\Gamma(1) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad \Gamma(-1) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

In general, $\Gamma(-h) = \Gamma(h)^\top$. At lag zero, $\Gamma(0)$ is symmetric and positive semidefinite.

We can compare times or combine coordinates

The vector backshift moves every coordinate together:

$$BX_t = X_{t-1} = (X_{1,t-1}, \ldots, X_{d,t-1})^\top.$$

- $(1 - B)X_t$ gives each coordinate's change since the previous time.
- $\beta^\top X_t$ combines coordinates observed at the same time.

Can combining coordinates remove variation that differencing would also remove?

Matrix coefficients combine both kinds of information

Definition (Finite matrix filter)

For fixed real $r \times d$ matrices $C_0, \ldots, C_m$, define

$$W_t := \sum_{j=0}^{m} C_j B^j X_t = \sum_{j=0}^{m} C_j X_{t-j} \in \mathbb{R}^r.$$

Differences use $C_0 = I_d$, $C_1 = -I_d$. A scalar combination uses $m = 0$ and $C_0 = \beta^\top$.

The output dimension can differ from the input dimension.

A finite filter preserves constant means

Suppose (X_t) is jointly weakly stationary, with mean μ. For $W_t = \sum_{j=0}^{m} C_j X_{t-j}$,

$$\mathbb{E}[W_t] = \sum_{j=0}^{m} C_j \mathbb{E}[X_{t-j}] = \sum_{j=0}^{m} C_j \mu.$$

This is constant in t. Each output coordinate also has a finite second moment, being a finite linear combination.

We still need to check the covariances.

The covariance calculation keeps only the lag

For $W_t = \sum_{j=0}^{m} C_j X_{t-j}$, the relevant input lag is $h - j + k$:

$$\begin{aligned}\operatorname{Cov}(W_{t+h}, W_t) &= \sum_{j=0}^{m} \sum_{k=0}^{m} C_j \operatorname{Cov}(X_{t+h-j}, X_{t-k}) C_k^\top \\ &= \sum_{j=0}^{m} \sum_{k=0}^{m} C_j \Gamma(h - j + k) C_k^\top.\end{aligned}$$

Thus the output is weakly stationary wherever all required inputs are defined.

In particular, $\gamma_{\beta^\top X}(h) = \beta^\top \Gamma(h) \beta$.

Which stationary vectors count as $I(0)$?

Definition (Vector $I(0)$)

(X_t) is $I(0)$ if it is jointly weakly stationary, $\sum_{h\in\mathbb{Z}} |\Gamma_{ij}(h)| < \infty$ for every i, j, and

$$\Omega := \sum_{h\in\mathbb{Z}} \Gamma(h)$$

has positive diagonal entries. The matrix Ω may be singular.

Each coordinate has positive long-run variance. A combination can have zero long-run variance and fall outside scalar $I(0)$.

An $I(1)$ vector accumulates mean-zero $I(0)$ increments

Definition (Vector $I(1)$)

For $t \geq 1$, write

$$X_t = X_0 + \sum_{j=1}^{t} V_j,$$

where (V_t) is a mean-zero $I(0)$ vector and X_0 has finite second moments.

Then $\nabla X_t = V_t$, and each coordinate satisfies $\mathrm{Var}(X_{it})/t \to \Omega_{V,ii} > 0$.

Constant shifts are allowed. Deterministic linear drift is outside this definition.

Could the levels themselves contain a stationary quantity?

Let $S_0 = 0$, $S_t = S_{t-1} + Z_t$. Suppose $(Z_t), (U_t), (V_t)$ are mutually independent $I(0)$ processes, with $\mathbb{E}[Z_t] = 0$.

$$X_t = S_t + U_t, \qquad Y_t = 2S_t + V_t.$$

Both coordinates accumulate the same shocks. Which fixed combination removes S_t?

Could the levels themselves contain a stationary quantity?

Let $S_0 = 0$, $S_t = S_{t-1} + Z_t$. Suppose $(Z_t), (U_t), (V_t)$ are mutually independent $I(0)$ processes, with $\mathbb{E}[Z_t] = 0$.

$$X_t = S_t + U_t, \qquad Y_t = 2S_t + V_t.$$

Both coordinates accumulate the same shocks. Which fixed combination removes S_t?

$Y_t - 2X_t = V_t - 2U_t$ is $I(0)$. The levels can wander while this gap remains stationary.

Cointegration is an $I(0)$ combination of $I(1)$ levels

Definition (Cointegration)

Suppose every coordinate of $X_t \in \mathbb{R}^d$ is $I(1)$. The coordinates are *cointegrated* if some fixed nonzero $\beta \in \mathbb{R}^d$ makes $\beta^\top X_t$ an $I(0)$ process.

The vector β is a *cointegrating vector*; its scalar combination is a *spread*.

Nonzero multiples describe the same direction. Subtracting a constant changes only the spread's mean.

The common-trend construction meets the definition

Return to $X_t = S_t + U_t$, $Y_t = 2S_t + V_t$. Write $\tau_Z^2, \tau_U^2, \tau_V^2 > 0$ for the long-run variances.

$$\begin{pmatrix} \nabla X_t \\ \nabla Y_t \end{pmatrix} = \begin{pmatrix} Z_t + \nabla U_t \\ 2Z_t + \nabla V_t \end{pmatrix}, \qquad \Omega_{\nabla(X,Y)} = \tau_Z^2 \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}.$$

The increments are jointly stationary with summable covariances. Their positive long-run variances make both levels $I(1)$.

The gap has long-run variance $\tau_V^2 + 4\tau_U^2 > 0$. Thus $(-2, 1)^\top$ is a cointegrating vector.

Differencing an $I(0)$ process gives zero long-run variance. The increment matrix here is singular, as allowed.

What changes if the common trend is deterministic?

Keep independent $I(0)$ processes U, V, but replace S_t by t:

$$X_t = t + U_t, \qquad Y_t = 2t + V_t, \qquad Y_t - 2X_t = V_t - 2U_t.$$

The gap is again $I(0)$. Are the coordinates cointegrated?

What changes if the common trend is deterministic?

Keep independent $I(0)$ processes U, V, but replace S_t by t:

$$X_t = t + U_t, \qquad Y_t = 2t + V_t, \qquad Y_t - 2X_t = V_t - 2U_t.$$

The gap is again $I(0)$. Are the coordinates cointegrated?

No. Their variances remain constant, so neither coordinate is $I(1)$. Their changing means make them nonstationary.

These independent processes can fit a line closely because their means rise together. Interpreting that fit as stochastic dependence would be misleading.

What if the coordinates are already stationary?

Let U_t, V_t be independent i.i.d. $N(0,1)$ sequences, and set

$$X_t = U_t, \qquad Y_t = 2U_t + V_t.$$

Again $Y_t - 2X_t$ is $I(0)$. Does this establish cointegration?

What if the coordinates are already stationary?

Let U_t, V_t be independent i.i.d. $N(0,1)$ sequences, and set

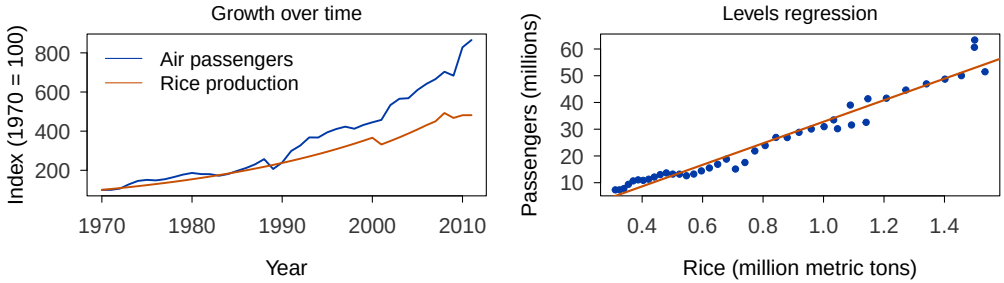
$$X_t = U_t, \qquad Y_t = 2U_t + V_t.$$

Again $Y_t - 2X_t$ is $I(0)$. Does this establish cointegration?

Both levels are $I(0)$, so the $I(1)$ premise fails. In fact, $Y_t - cX_t = (2-c)U_t + V_t$ is $I(0)$ for every fixed c.

A close regression fit can come from rising levels

Australian air passengers and rice production in Guinea, 1970–2011:



The levels regression has $R^2 = 0.958$. What does this establish about the processes?

Real data: 42 annual World Bank observations via Hyndman and Athanasopoulos (2021, Section 7.3).

A regression can overstate evidence of a relationship

Spurious regression is misleading regression evidence caused by ignoring trends or persistence.

High R^2 and apparently significant slopes can occur even for independent series. The usual regression reference distribution may fail.

Cointegration is a population property with $I(1)$ premises. The deterministic-trend example had an $I(0)$ gap and still allowed a misleading interpretation of the fit.

Independent random walks have no stationary direction

Let $A_0 = B_0 = 0$, with independent i.i.d. $N(0,1)$ increments in each walk. For fixed $(c_1, c_2) \neq (0,0)$,

$$\text{Var}(c_1 A_t + c_2 B_t) = t(c_1^2 + c_2^2).$$

Why does this rule out a stationary combination?

Independent random walks have no stationary direction

Let $A_0 = B_0 = 0$, with independent i.i.d. $N(0,1)$ increments in each walk. For fixed $(c_1, c_2) \neq (0,0)$,

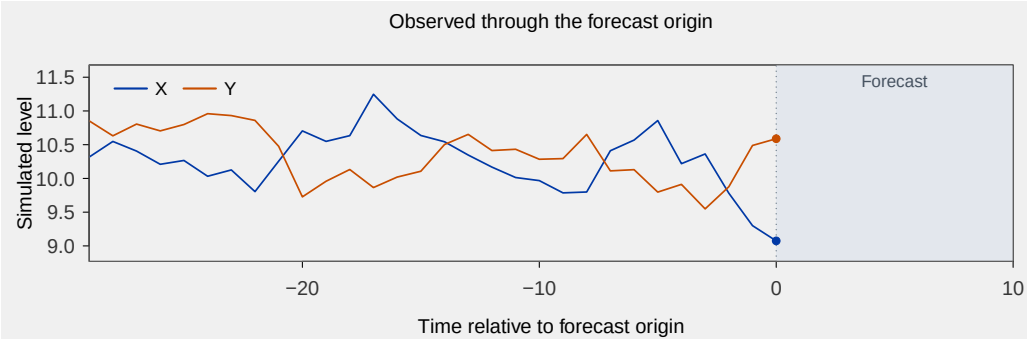
$$\mathrm{Var}(c_1 A_t + c_2 B_t) = t(c_1^2 + c_2^2).$$

Why does this rule out a stationary combination?

The variance grows with t for every nonzero direction. Yet finite paths can happen to move together and give a high regression R^2.

Both differences are $I(0)$; differencing and coordinate cancellation are different possibilities.

Should these two levels stay this far apart?



At $t = 261$, the observed levels are $x_t = 9.074$ and $y_t = 10.588$. Should their forecasts preserve this separation or move toward each other?

Simulated data: shared random walk and stationary AR(1), $\phi = .7$; innovation SDs .15, .3; seed 5421014.

A shared trend can coexist with mean reversion

The simulated levels share a random walk M and have opposite stationary deviations:

$$M_t = M_{t-1} + \eta_t, \quad U_t = .7U_{t-1} + Z_t,$$

$$X_t = M_t - U_t, \qquad Y_t = M_t + U_t.$$

The innovations are independent sequences of i.i.d. $N(0, .15^2)$ and $N(0, .3^2)$ variables, respectively. Take $M_0 = 10$ and the stationary causal solution for U.

Both levels are $I(1)$, while $Y_t - X_t = 2U_t$ is $I(0)$. Its conditional mean shrinks by 30% each step: the separation is mean-reverting. The common level M continues to receive new shocks.

Which way should each level move next?

Here $x_t = 9.074$ and $y_t = 10.588$, a separation of 1.514. The conditional mean of U_{t+1} is $.7U_t$. What does this imply for the two levels?

Which way should each level move next?

Here $x_t = 9.074$ and $y_t = 10.588$, a separation of 1.514. The conditional mean of U_{t+1} is $.7U_t$. What does this imply for the two levels?

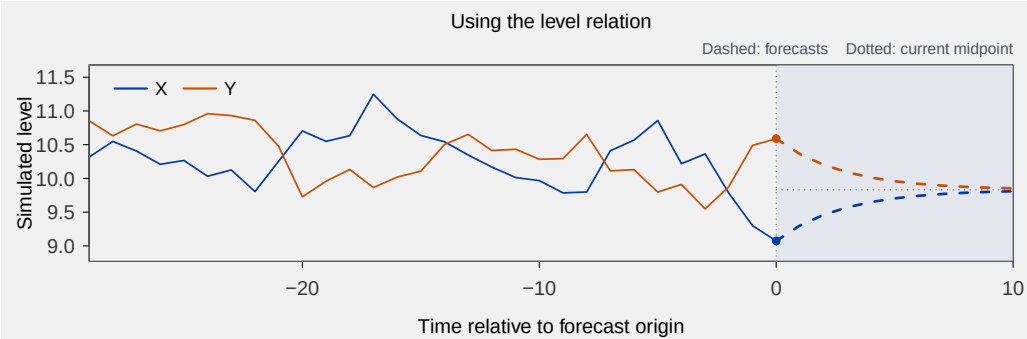
$$\widehat{x}_{t+1|t} = x_t + .15(y_t - x_t) \approx 9.30,$$

$$\widehat{y}_{t+1|t} = y_t - .15(y_t - x_t) \approx 10.36.$$

The forecast raises X and lowers Y. Both adjustments use their current separation.

Conditional mean forecasts under the stated model, with known parameters. Displayed observations and forecasts are rounded.

Mean reversion brings the level forecasts together



After h steps, $\widehat{y}_{t+h|t} - \widehat{x}_{t+h|t} = .7^h(y_t - x_t)$. Both forecasts approach the current midpoint, 9.831. The future common level remains uncertain.

Simulated data: shared random walk and stationary AR(1), $\phi = .7$; innovation SDs .15, .3; seed 5421014.

What if the adjustment uses only the latest changes?

The latest changes are $\nabla x_t = -0.226$ and $\nabla y_t = 0.100$. At each horizon, predict the cumulative change from this pair, then add it to today's level.

The best linear one-step adjustments under the same model are

What if the adjustment uses only the latest changes?

The latest changes are $\nabla x_t = -0.226$ and $\nabla y_t = 0.100$. At each horizon, predict the cumulative change from this pair, then add it to today's level.

The best linear one-step adjustments under the same model are

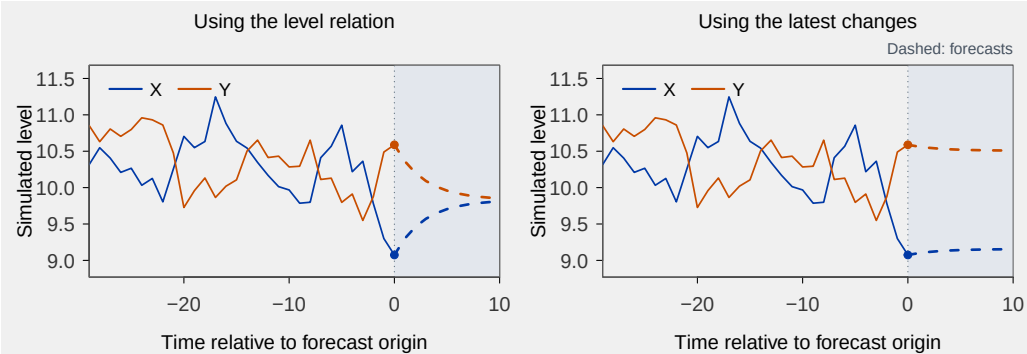
$$\widetilde{\nabla x}_{t+1|t} = .075(\nabla y_t - \nabla x_t) \approx +0.024,$$

$$\widetilde{\nabla y}_{t+1|t} = -.075(\nabla y_t - \nabla x_t) \approx -0.024.$$

These changes predict much smaller adjustments than the current separation.

The adjustment uses only the latest pair of differences. Both methods use known parameters and add their predicted changes to the current levels.

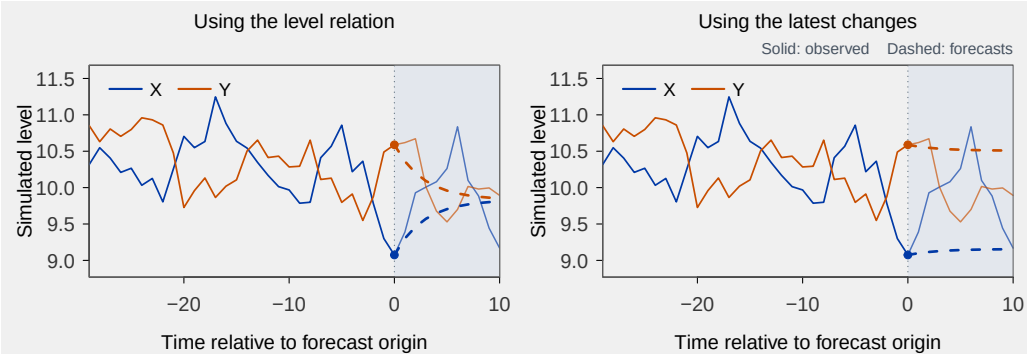
The current relation changes both level forecasts



The current separation produces much larger forecast adjustments than the latest changes. Both panels use the same observations, forecast origin, and known parameters.

Simulated data: shared random walk and stationary AR(1), $\phi = .7$; innovation SDs .15, .3; seed 5421014.

Realized paths still receive new shocks

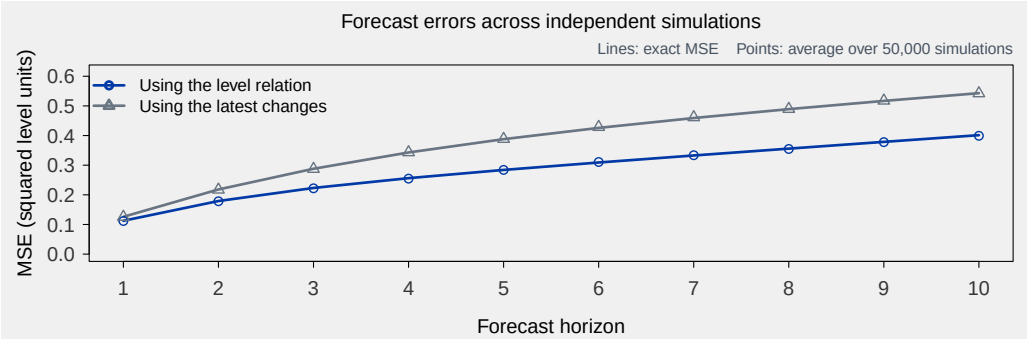


New shocks move the common level and can widen the separation. The shrinking separation is a statement about conditional means.

Simulated data: shared random walk and stationary AR(1), $\phi = .7$; innovation SDs .15, .3; seed 5421014.

The same future observations appear in both panels. They were not used to choose the forecast origin.

Using the relation reduces average forecast error



At five steps, level-forecast MSE falls from 0.388 to 0.284, a 26.8% reduction.

A complete difference history and the starting levels recover the current separation.

Exact MSE and 50,000 independent simulations, seed 5421015; errors averaged over X and Y, with all cases included.

Both observed levels can contain stationary fluctuations

Let $S_0 = 0$, $S_t = S_{t-1} + \eta_t$, and use stationary causal solutions

$$U_t = \phi_1 U_{t-1} + \varepsilon_t, \quad V_t = \phi_2 V_{t-1} + \xi_t,$$

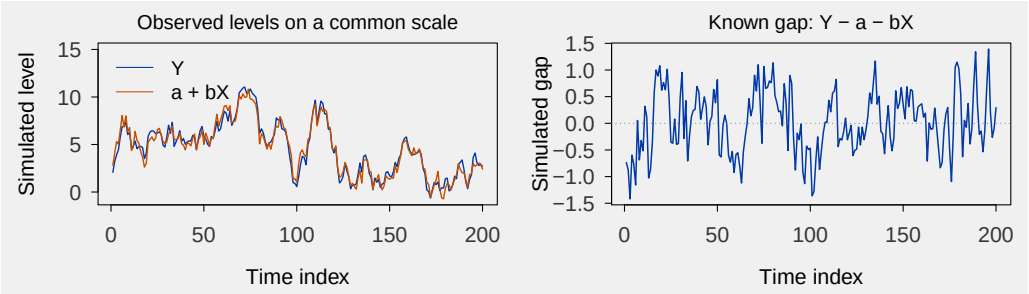
$$X_t = S_t + U_t, \qquad Y_t = a + bS_t + V_t.$$

$|\phi_1|, |\phi_2| < 1$, $\phi_1 \neq \phi_2$, and $b \neq 0$. The three innovation sequences are mutually independent, centered i.i.d., with positive finite variances $\sigma^2_\eta, \sigma^2_\varepsilon, \sigma^2_\xi$.

We observe X, Y. The components S, U, V are not separately observed. Take the parameters as known until we estimate the relation by OLS.

How does the gap behave compared with the levels?

With known a, b, compute $E_t := Y_t - a - bX_t$ from the observed pair.



Does the gap's variation grow with time in the same way as the levels' variation?

Simulated data: $T = 200$, $a = 2$, $b = 1.3$, $\phi_1 = .35$, $\phi_2 = .8$; Gaussian innovation SDs .7, .3, .3; seed 5421005. Independent stationary Gaussian initial values are used for U_0, V_0.

The component variances separate accumulation from fluctuation

The stationary component variances are

$$q_U := \frac{\sigma_\varepsilon^2}{1-\phi_1^2}, \qquad q_V := \frac{\sigma_\xi^2}{1-\phi_2^2}.$$

- $\mathrm{Var}(X_t) = t\sigma_\eta^2 + q_U$ and $\mathrm{Var}(Y_t) = b^2 t\sigma_\eta^2 + q_V$.
- The differences are $I(0)$ with long-run variances σ_η^2 and $b^2\sigma_\eta^2$.

Both observed levels are therefore $I(1)$. Their initial values U_0 and $a + V_0$ have finite second moments.

The observable gap combines the stationary components

The known coefficients give $E_t := Y_t - a - bX_t = V_t - bU_t$, with mean zero. Independence gives

$$\gamma_E(h) = q_V \phi_2^{|h|} + b^2 q_U \phi_1^{|h|},$$

$$\tau_E^2 = \frac{\sigma_\xi^2}{(1-\phi_2)^2} + \frac{b^2 \sigma_\varepsilon^2}{(1-\phi_1)^2} > 0.$$

The covariances are absolutely summable, so E is $I(0)$. The vector $(-b, 1)^\top$ cancels the common stochastic trend.

Only the cancelling slope removes accumulating error

For a fixed trial slope c, compare the residual variances:

$$\text{Independent walks:} \quad \operatorname{Var}(B_t - a - cA_t) = t(1 + c^2),$$

$$\text{Shared trend:} \quad \operatorname{Var}(Y_t - a - cX_t) = t(b - c)^2\sigma_\eta^2 + q_V + c^2 q_U.$$

Every slope fails for independent walks. In the shared-trend model, $c = b$ leaves the $I(0)$ gap.

A sample fit can be close in either setting.

Which stationary quantities should we retain?

For $X_t = S_t + U_t$ and $E_t = Y_t - a - bX_t$, both vectors below are jointly stationary:

$$\begin{pmatrix} \nabla X_t \\ \nabla Y_t \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} \nabla X_t \\ E_t \end{pmatrix}.$$

What does the current gap retain that a few recent differences may omit?

Which stationary quantities should we retain?

For $X_t = S_t + U_t$ and $E_t = Y_t - a - bX_t$, both vectors below are jointly stationary:

$$\begin{pmatrix} \nabla X_t \\ \nabla Y_t \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} \nabla X_t \\ E_t \end{pmatrix}.$$

What does the current gap retain that a few recent differences may omit?

It records the current distance from the level relation. All differences together with the initial prices can reconstruct it; a short window of differences generally cannot.

Differencing E_t also gives a stationary process, but its long-run variance is zero. It is outside our exact $I(0)$ class.

How do we estimate an unknown level relation?

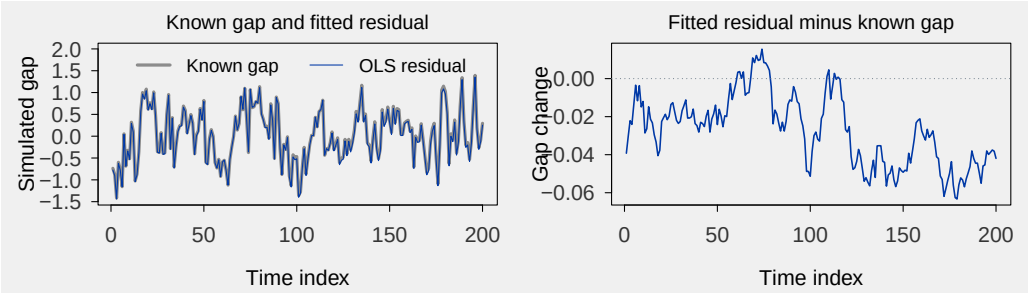
For observed pairs (x_t, y_t), OLS with an intercept chooses

$$(\widehat{a}, \widehat{b}) := \arg\min_{a,b} \sum_{t=1}^{T} (y_t - a - bx_t)^2, \qquad \widehat{u}_t := y_t - \widehat{a} - \widehat{b}x_t.$$

This fixes the coefficient of Y at one.

In our specified shared-trend model, OLS consistently estimates a, b. This does not justify conventional regression standard errors.

The estimated relation changes the observed gap



The fitted slope is 1.291, versus the known $b = 1.3$. The right panel isolates the change that is hard to see in the overlaid curves.

Fitting an intercept moves the sample mean from 0.027 to zero.

Simulated data: $T = 200$, $a = 2$, $b = 1.3$, $\phi_1 = .35$, $\phi_2 = .8$; Gaussian innovation SDs .7, .3, .3; seed 5421005.

Which spread does ordinary level KPSS assess?

A prespecified coefficient b_0 gives the observed scalar series $Y_t - b_0 X_t$. Level KPSS allows its unknown constant mean to be fitted.

An OLS slope is instead chosen using the same observations:

$$\widehat{u}_t = e_t - (\widehat{a} - a) - (\widehat{b} - b)x_t, \qquad e_t := y_t - a - bx_t.$$

Ordinary KPSS critical values do not account for estimating that slope.

Rejection for a fixed b_0 counts against that choice; it does not rule out every possible relation.

Could oil and gasoline have a stationary relative price?

Let O_t be WTI oil in dollars per barrel and G_t New York Harbor gasoline in cents per gallon. Specify the unit-slope log relation:

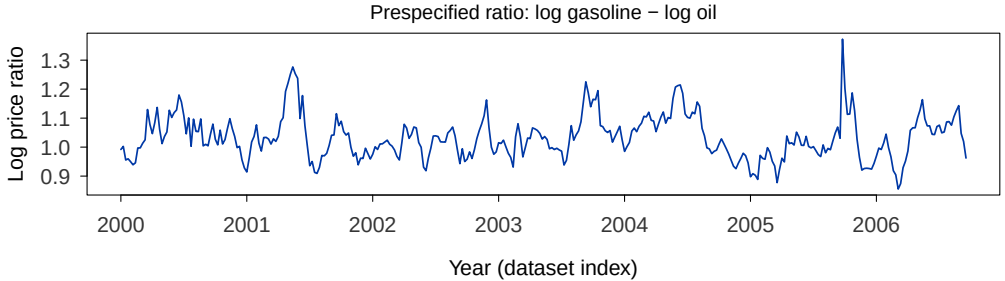
$$R_t := \log G_t - \log O_t.$$

Changing price units adds a constant to R_t, preserving level stationarity.

If both log prices are $I(1)$ and this ratio is $I(0)$, it gives a cointegrating relation. Use weeks 1–350 for assessment and fitting; retain weeks 351–545 for a later check.

Real data: EIA via `astsa::oil` and `astsa::gas`; 545 weekly observations, 2000–mid-2010.

Does the log ratio fluctuate around a constant mean?



The prespecified ratio is $R_t = \log G_t - \log O_t$. What features of these first 350 weeks count against a constant mean?

Real data: EIA via `astsa::oil` and `astsa::gas`; 545 weekly observations, 2000–mid-2010.

Assessment sample: weeks 1–350. No slope is estimated from these observations.

KPSS does not reject stationarity of the fixed ratio

Assess the first 350 weeks under the null that the prespecified ratio is $I(0)$ around an unknown constant mean.

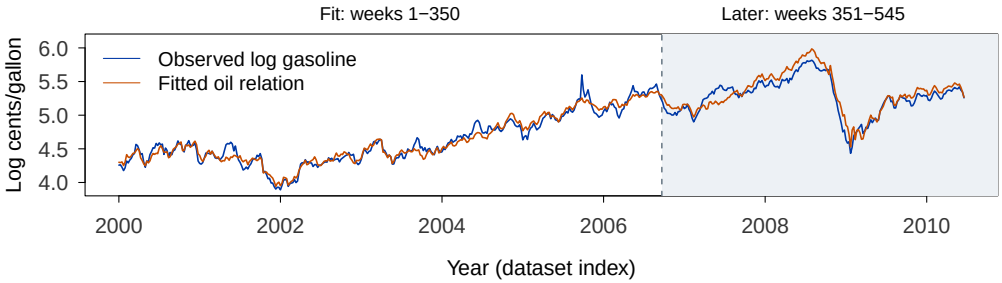
$$\underbrace{K_T = 0.085}_{\text{observed statistic}} \quad < \quad \underbrace{0.463}_{\text{5\% critical value}} \quad .$$

We do not reject the null at 5%. Nonrejection is compatible with the relation; it does not establish the null.

Level KPSS with Bartlett bandwidth 8. Calibration requires the functional-limit and bandwidth conditions; $I(0)$ alone does not supply them.

Fit a general relation on the first 350 weeks

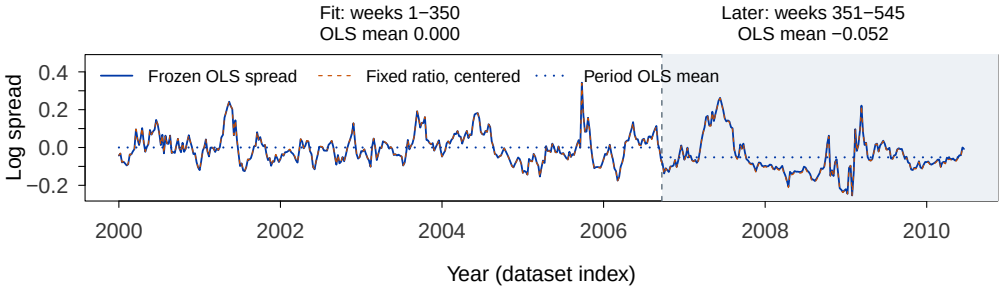
OLS gives $\widehat{a} = 1.051$ and $\widehat{b} = 0.995$ in $\log G_t = a + b \log O_t + U_t$.



Fit on weeks 1–350. Keep these coefficients fixed in the shaded later period.

Real data: EIA via `astsa::oil` and `astsa::gas`; 545 weekly observations, 2000–mid-2010.

How does the frozen relation behave later?



Later mean: -0.052. Later SD: 1.297 times the fitting-period SD.

The horizontal segments show the OLS spread's mean in each period.

Real data: EIA via `astsa::oil` and `astsa::gas`; 545 weekly observations, 2000–mid-2010.

The fixed ratio is centered with its first-350-week mean. Both curves use unchanged coefficients.

What did the oil and gasoline calculations establish?

- KPSS assessed the prespecified unit-slope log ratio on the first period.
- OLS estimated a more general candidate relation on that period.
- The later record showed how its frozen spread behaved beyond fitting.

The later shift and greater variability can reflect sampling variation, coefficient error, or a changing relation. These summaries do not separate them.

A forecast still needs a model for the gap's dynamics.

Relative-value trading aims to profit from mean reversion

A common hedge-fund strategy buys the relatively cheap side and shorts (sells borrowed units of) the relatively expensive side.

Consider an exchange-traded fund (ETF) holding a basket of stocks. A market maker quotes buy and sell prices and can hedge an ETF position with its basket.

Position	Price now	Price at closing	Price gain
Short one ETF share	\$102	\$101	+\$1
Buy its matching basket	\$100	\$101	+\$1

If the \$2 difference closes, the position gains \$2 before costs. If it widens, the position loses money.

Hypothetical prices; fixed basket, no distributions, financing charges, or trading costs.

Sources: AQR (2025), *Exploring Capital Efficiency*; BlackRock, *Authorised participants and market makers*.

An ETF and its basket form a vector price series

Let F_t and P_{it} be the ETF and constituent share prices, in dollars. For each ETF share, hold q_i shares of constituent i and cash c_0 fixed.

$$\boldsymbol{P}_t := \begin{pmatrix} F_t & P_{1t} & \cdots & P_{d-1,t} \end{pmatrix}^\top \in \mathbb{R}^d, \qquad \beta := \begin{pmatrix} 1 & -q_1 & \cdots & -q_{d-1} \end{pmatrix}^\top,$$

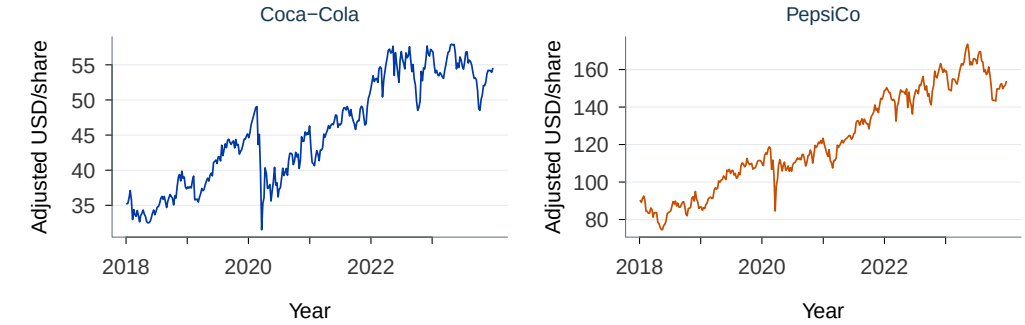
$$\underbrace{E_t}_{\text{ETF premium}} := F_t - \underbrace{\left(c_0 + \sum_{i=1}^{d-1} q_i P_{it} \right)}_{\text{basket value per ETF share}} = \beta^\top \boldsymbol{P}_t - c_0.$$

If $\boldsymbol{P}$ is $I(1)$ and E is $I(0)$, then β is a cointegrating vector.

Authorized participants can exchange large blocks of ETF shares for designated baskets and back. This mechanism encourages price alignment. A dynamic model describes the premium's mean reversion.

Stylized fixed-basket model. Source: SEC (2023), *Updated Investor Bulletin: Exchange-Traded Funds*.

Could the two beverage stocks share a long-run relation?

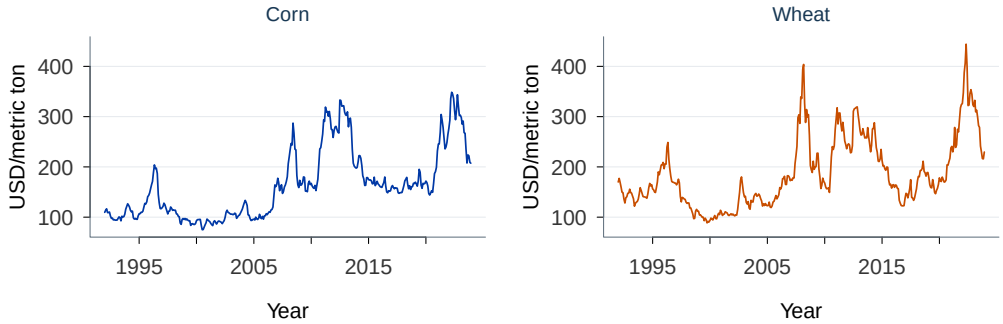


Both firms sell beverages; PepsiCo also sells food. What might affect their prices together?

Real data: Coca-Cola and PepsiCo, 313 weekly adjusted closes, 2018–2023. Source: Yahoo Finance (KO, PEP).

Homework 5, Problem 1. Original price scale, with split/dividend adjustments; separate vertical scales.

Could substitution link corn and wheat prices?

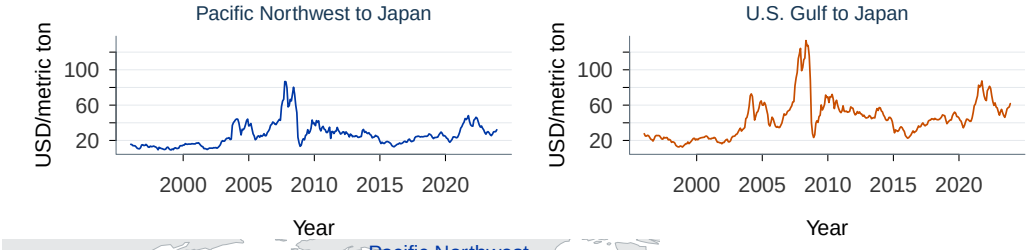


Wheat can substitute for corn in livestock feed. Harvest conditions and other uses can affect their prices differently.

Real data: 384 monthly global benchmark prices, 1992–2023. Source: IMF via FRED.

Homework 5, Problem 1. Original nominal prices; common vertical scale.

How much do grain freight rates depend on the route?



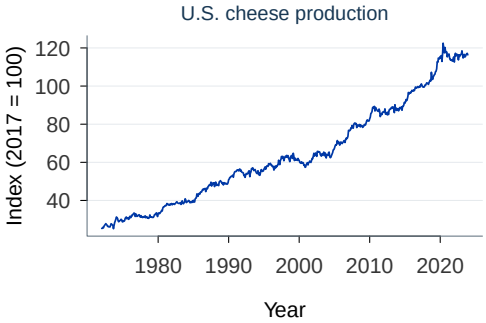
Bulk grain shipped to Japan from two U.S. regions.

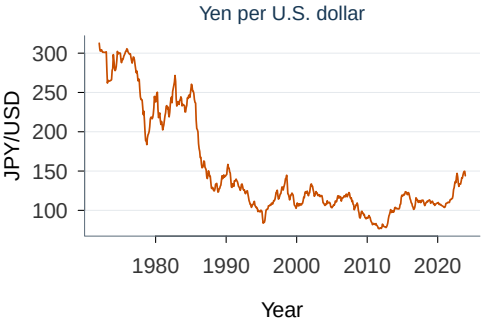
Schematic routes; Gulf route shown via Panama.

Real data: 336 monthly rates, 1996–2023. Source: USDA / O'Neil Commodity Consulting.

Homework 5, Problem 1. Original levels; common vertical scale. Map: Natural Earth.

Could cheese production help explain the yen?





The cheese series measures production volume. An increase in yen per dollar means that the yen buys fewer dollars.

Real data: 624 monthly observations, 1972–2023. Source: Federal Reserve via FRED (IPN311513S, DEXJPUS).

Homework 5, Problem 1. Original levels, different units; cheese production is seasonally adjusted.

Homework 5: a fund wants to trade relative prices

A hypothetical hedge fund pools investors' money for a market-neutral long–short strategy.



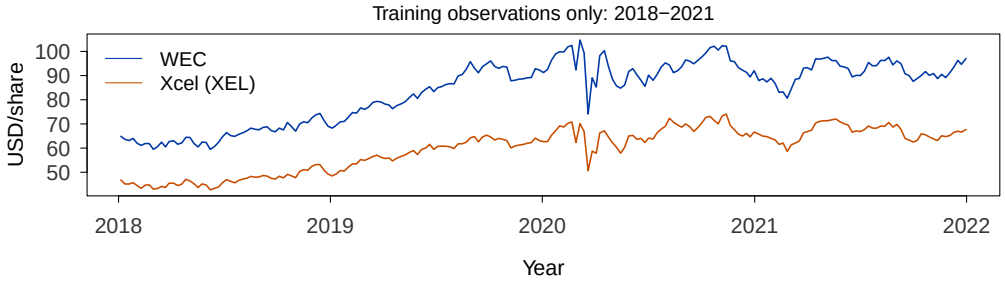
The aim is to reduce exposure to shared price movements.

WEC and Xcel both operate electric and natural-gas utilities in the Midwest. Common financing costs, demand, and regulation could affect both prices.

Hypothetical assignment: *Cointegration in Real Data*, Problem 2; company descriptions in the supplied workbook.

The supplied fit uses 2018–2021 prices

W_t denotes WEC and X_t Xcel, in dollars per share.



209 weekly closes, split adjusted but not adjusted for cash dividends. The assignment uses dollar prices directly.

Real data: supplied Homework 5 training CSV; Yahoo Finance WEC and XEL, last common trading session each week.

The assignment supplies an AR(1) model for the gap

OLS on the training prices, then an AR(1) fit to the gap, gives

$$W_t = a + bX_t + U_t, \quad U_{t+1} = \phi U_t + Z_{t+1},$$

$$a = 1.146277, \qquad b = 1.385665, \qquad \phi = 0.876663.$$

Treat these estimates as fixed parameters. Assume W, X share an integrated component and U is the stationary solution of the AR(1) equation.

For every t, Z_{t+1} has mean zero, positive finite variance, and is independent of $\mathcal{H}_t = ((W_s, X_s) : s \leq t)$.

The AR(1) assumption supplies a forecast law beyond cointegration.

Explain the trade before choosing the shares

Suppose the observed gap $u_t = w_t - a - bx_t$ is positive. Under the supplied AR model, its predicted change points toward zero.

Which stock would the fund buy, and which would it sell short? Why combine the positions instead of trading only one stock?

Explain the trade before choosing the shares

Suppose the observed gap $u_t = w_t - a - bx_t$ is positive. Under the supplied AR model, its predicted change points toward zero.

Which stock would the fund buy, and which would it sell short? Why combine the positions instead of trading only one stock?

A positive gap suggests shorting WEC and buying Xcel; a negative gap reverses them. Suitable holdings offset their shared price movement.

The assignment asks for the forecast derivation and the expected closure over 1, 4, and 13 weeks.

Choose holdings so the gain depends on the gap change

For signed share counts q_W, q_X held fixed from t to $t + h$,

$$G_{t,h} = q_W(W_{t+h} - W_t) + q_X(X_{t+h} - X_t).$$

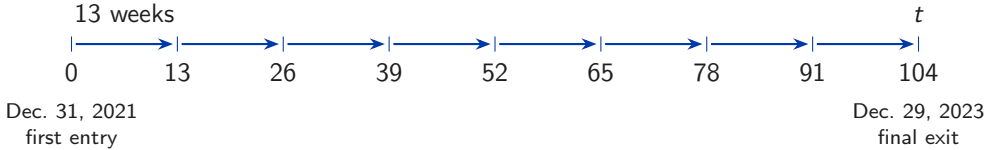
Substitute $W_t = a + bX_t + U_t$. Which coefficient must vanish to remove the unspecified movement in X?

- When $u_t \neq 0$: exactly 100 WEC shares in absolute value.
- Fractional Xcel shares and short positions are allowed.
- When $u_t = 0$: no position.

Derive the holdings and initial cash balance, with cash earning zero interest, before opening the evaluation file.

Each trade lasts 13 weeks

The evaluation file begins at $t = 0$, December 31, 2021, then records 104 weekly observations from 2022–2023.



Enter at $t = 0, 13, 26, 39, 52, 65, 78, 91$; close at $t + 13$. At a shared boundary, close the old position before opening the next.

Compare predicted and realized price gains

Keep the supplied coefficients and the holdings rule fixed. At entry and exit, assume immediate execution at the observed prices.

For each of the eight trades, report

- entry and exit dates;
- entry gap and both share positions;
- predicted and realized price gains.

Across the eight trades, report

- totals for both gains;
- number of realized losses.

Calculate price gains only, ignoring dividends, interest, and trading costs.

Continue with the workbook and Homework 5

- Use the workbook examples to distinguish a stationary combination from cointegration.
- In Homework 5, explain the proposed trade, derive its holdings, then evaluate the later observations.
- Explain why a positive expected gain can still produce a loss.