STA 542: Introduction to Time Series Analysis

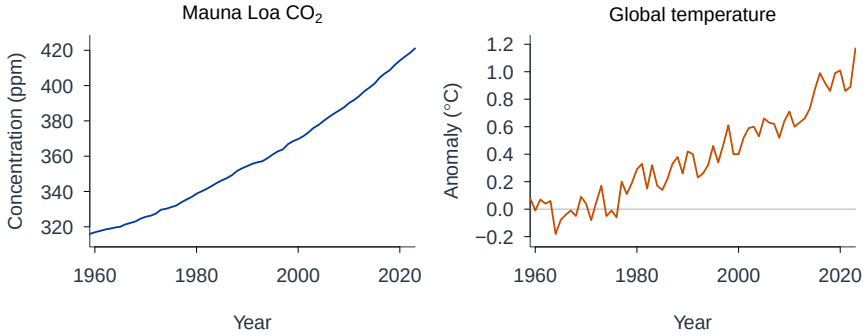
Lecture 13

Lasse Vuursteen

Duke
UNIVERSITY

Homework 6: carbon and temperature

Annual Mauna Loa CO_2 and global land-and-ocean temperature, 1959–2023.

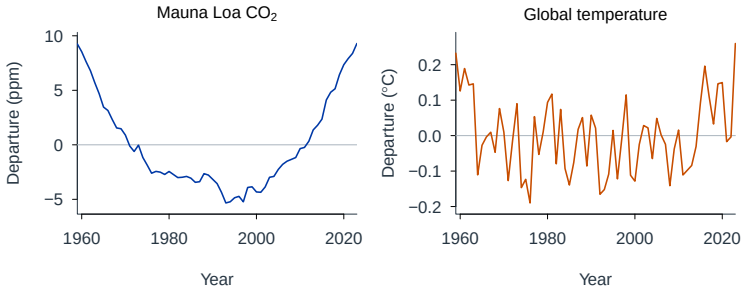


Does past CO_2 help predict temperature?

NOAA records from Homework 6. Temperature anomalies are relative to 1901–2000.

Linear detrending

$$y_t = x_t - (\widehat{\alpha} + \widehat{\beta} t)$$

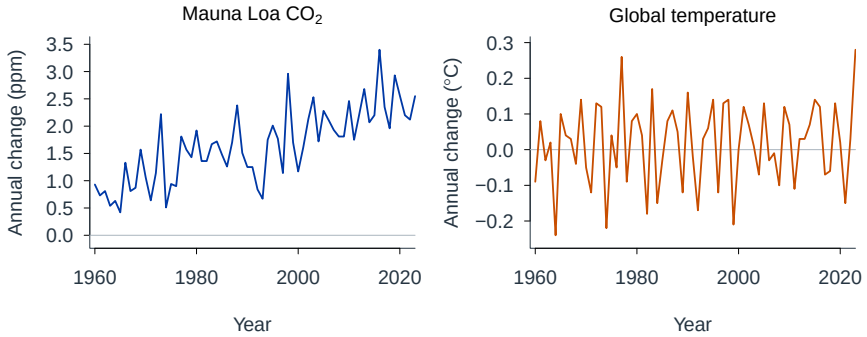


VAR(1) fit with intercept: $\rho(\widehat{A}) = 0.986$.

Is this persistent dependence, or trend we have not removed?

First differences

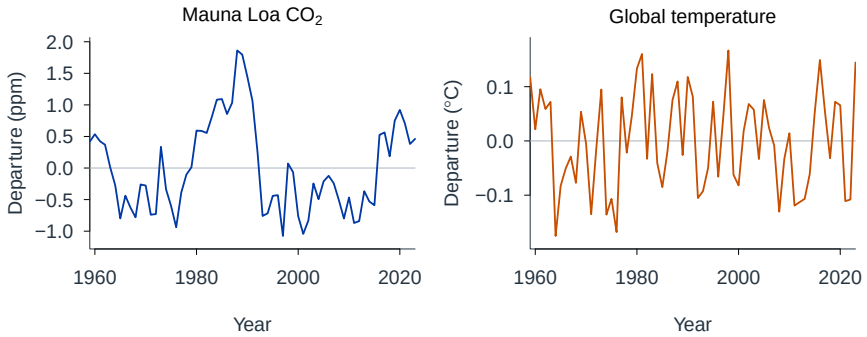
$$y_t = \Delta x_t = x_t - x_{t-1}$$



Is average CO_2 growth constant over this record?

Quadratic detrending

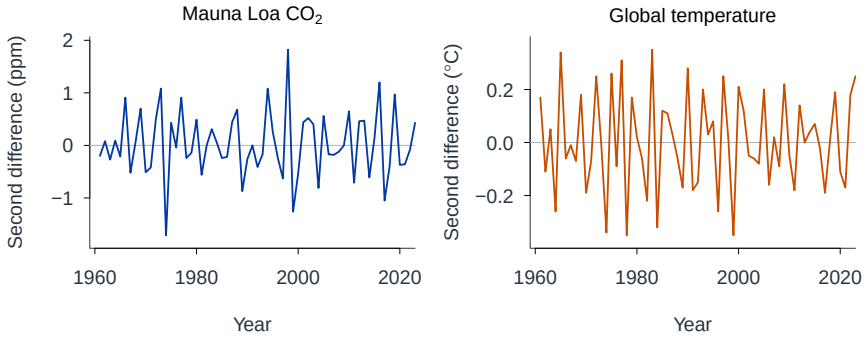
$$y_t = x_t - (\widehat{\alpha} + \widehat{\beta} t + \widehat{\gamma} t^2)$$



Each coordinate is a departure from its own fitted curve.

Second differences

$$y_t = \Delta^2 x_t = \Delta x_t - \Delta x_{t-1} = x_t - 2x_{t-1} + x_{t-2}$$



A positive value means the annual change increased.

Fit both equations

$Y_{C,t}$ and $Y_{T,t}$: transformed CO_2 and temperature.

$$Y_{C,t+1} = a_C + A_{CC} Y_{C,t} + A_{CT} Y_{T,t} + Z_{C,t+1},$$

$$Y_{T,t+1} = a_T + A_{TC} Y_{C,t} + A_{TT} Y_{T,t} + Z_{T,t+1}.$$

Two regressions, each with an intercept and both lagged coordinates.

Stationary causal VAR(1): $\rho(A) < 1$, $Z_t \overset{\text{iid}}{\sim} N_2(0, \Sigma_Z)$, $\Sigma_Z > 0$.

After quadratic detrending

Coordinate order: CO_2, temperature.

$$\widehat{y}_{t+1|t} = \underbrace{\begin{pmatrix} 0.0019 \\ -0.0014 \end{pmatrix}}_{\widehat{a}} + \underbrace{\begin{pmatrix} 0.7441 & 1.3612 \\ -0.0004 & 0.1795 \end{pmatrix}}_{\widehat{A}} \begin{pmatrix} y_{C,t} \\ y_{T,t} \end{pmatrix}.$$

Which entry predicts temperature from CO_2?

After quadratic detrending

Coordinate order: CO_2, temperature.

$$\widehat{y}_{t+1|t} = \underbrace{\begin{pmatrix} 0.0019 \\ -0.0014 \end{pmatrix}}_{\widehat{a}} + \underbrace{\begin{pmatrix} 0.7441 & 1.3612 \\ -0.0004 & 0.1795 \end{pmatrix}}_{\widehat{A}} \begin{pmatrix} y_{C,t} \\ y_{T,t} \end{pmatrix}.$$

Which entry predicts temperature from CO_2?

$\widehat{A}_{TC} = -0.0004$.

64 paired transitions; $\rho(\widehat{A}) = 0.743$.

After second differences

Coordinate order: CO_2, temperature.

$$\widehat{y}_{t+1|t} = \underbrace{\begin{pmatrix} 0.0367 \\ 0.0062 \end{pmatrix}}_{\widehat{a}} + \underbrace{\begin{pmatrix} -0.5677 & 1.8397 \\ -0.1191 & -0.3559 \end{pmatrix}}_{\widehat{A}} \begin{pmatrix} y_{C,t} \\ y_{T,t} \end{pmatrix}.$$

62 paired transitions; $\rho(\widehat{A}) = 0.649$.

Reading the CO_2 coefficient

Increase current transformed CO_2 by 1 ppm; hold current transformed temperature fixed.

Temperature target	Change in next year's prediction
Quadratic departure	$-0.0004\,^{\circ}\mathrm{C}$
Second difference	$-0.1191\,^{\circ}\mathrm{C}$

Does a negative coefficient mean that CO_2 causes cooling?

Reading the CO_2 coefficient

Increase current transformed CO_2 by 1 ppm; hold current transformed temperature fixed.

Temperature target	Change in next year's prediction
Quadratic departure	$-0.0004\,^{\circ}\mathrm{C}$
Second difference	$-0.1191\,^{\circ}\mathrm{C}$

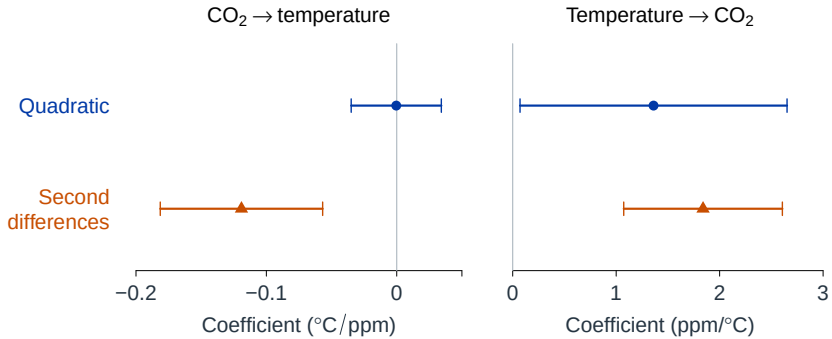
Does a negative coefficient mean that CO_2 causes cooling?

No. This is a prediction conditional on past observations.

Which intervals include zero?

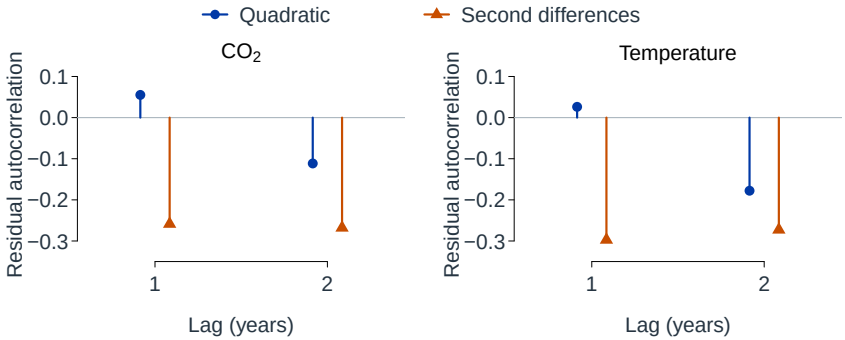
Granger non-causality in mean (Gaussian VAR, horizon one):

$$C \not\to T \iff A_{TC} = 0, \qquad T \not\to C \iff A_{CT} = 0.$$

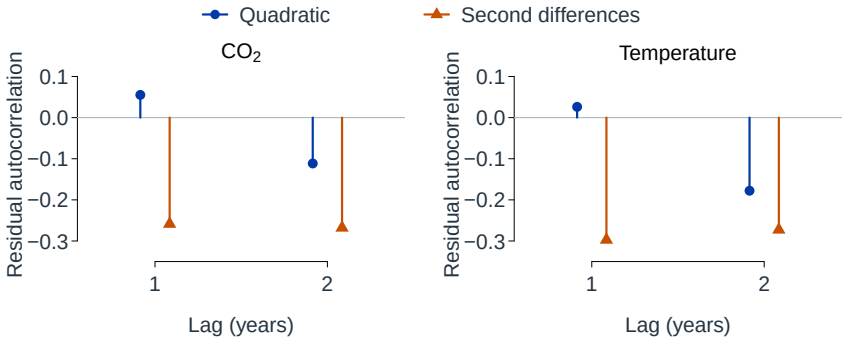


Individual nominal 95% rowwise regression intervals; fitted trends treated as fixed.

Which fit would we keep?



Which fit would we keep?



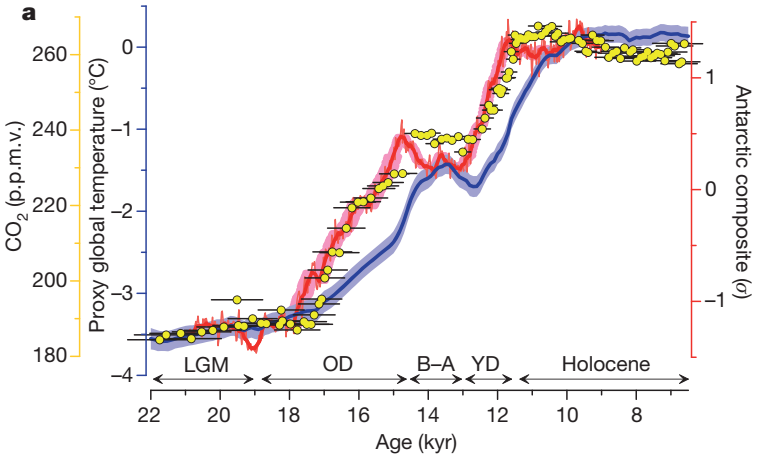
Quadratic detrending, provisionally: residual correlations closer to zero at both lags.

The interval calculations assume independent innovations.

What does the climate fit tell us?

- CO_2 drives long-term warming; temperature also affects carbon uptake and annual CO_2 growth.
- Our detrended VAR tests additional one-year predictive information, not the long-term warming effect of CO_2.

Warming at the end of the last ice age



Global temperature
Reconstruction from 80 proxy records

Antarctic temperature
Ice-core proxy

Atmospheric CO_2
Air trapped in ice cores

kyr = thousands of years ago

Shakun et al. (2012), *Nature* 484, 49–54, Figure 2a.

OLS with i.i.d. observations

Observe (X_i, Y_i), $i = 1, \ldots, n$, i.i.d., with $X_i \in \mathbb{R}^k$.

$$Y_i = X_i^\top \beta + Z_i \qquad \Longleftrightarrow \qquad Z_i = Y_i - X_i^\top \beta.$$

$$\widehat{\beta}_{\text{OLS}} = \underset{b \in \mathbb{R}^k}{\arg\min} \sum_{i=1}^{n} (y_i - x_i^\top b)^2.$$

The normal equations are

$$\sum_{i=1}^{n} x_i (y_i - x_i^\top \widehat{\beta}) = 0.$$

Assume $\sum_{i=1}^{n} x_i x_i^\top$ is invertible.

Method of moments

Assume finite second moments and $\mathbb{E}[X_i Z_i] = 0$. Substitute for Z_i:

$$0 = \mathbb{E}\Big[X_i(Y_i - X_i^\top \beta)\Big] = \mathbb{E}[X_i Y_i] - \mathbb{E}[X_i X_i^\top]\beta.$$

Replace expectations by sample averages:

$$\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i^\top\right)\widehat{\beta} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i.$$

These are the OLS normal equations. Gaussianity is unnecessary.

Gaussian maximum likelihood

Assume $Z_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$, independent of the predictors.

$$Z_i = Y_i - X_i^\top \beta \quad \Longrightarrow \quad Y_i \mid X_i = x_i \sim N(x_i^\top \beta, \sigma^2).$$

Conditional on the observed predictors,

$$\ell(\beta, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - x_i^\top \beta)^2.$$

Maximizing ℓ minimizes the residual sum of squares:

$$\widehat{\beta}_{\text{ML}} = \widehat{\beta}_{\text{OLS}} = \widehat{\beta}_{\text{MoM}}.$$

Sampling uncertainty

Under the Gaussian model, $\widehat{\beta}$ is exactly normal conditional on the predictors.

Without Gaussianity, use a large-sample result:

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) \xrightarrow{d} N_k(0, V).$$

A consistent estimate $\widehat{V}$ of $V > 0$ gives the approximate 95% confidence interval:

$$\widehat{\beta}_{n,j} \pm 1.96\sqrt{\widehat{V}_{jj}/n}.$$

A poor normal approximation or underestimated variance can cause undercoverage.

β_0: true coefficients. The limit requires sufficient moments and an identifiable regression.

Laplace errors

Keep $Z_i = Y_i - X_i^\top \beta$, independent of X_i. Use

$$Z_i \overset{\text{iid}}{\sim} \text{Laplace}(0, s), \qquad f_s(z) = \frac{1}{2s} e^{-|z|/s}, \qquad s > 0 \text{ fixed.}$$

Conditional on the predictors,

$$\ell(\beta) = -n \log(2s) - \frac{1}{s} \sum_{i=1}^{n} |y_i - x_i^\top \beta|.$$

Maximum likelihood gives least absolute deviations:

$$\widehat{\beta}_{\text{ML}} \in \underset{b \in \mathbb{R}^k}{\arg\min} \sum_{i=1}^{n} |y_i - x_i^\top b|.$$

OLS still solves the moment equations. It generally differs from this MLE.

Gaussian likelihood for a VAR(p)

For a mean-zero causal VAR with $X_t \in \mathbb{R}^d$,

$$X_t = \sum_{j=1}^{p} A_j X_{t-j} + Z_t, \qquad Z_t \overset{\text{iid}}{\sim} N_d(0, \Sigma_Z).$$

Rewrite the innovations:

$$Z_t = X_t - \sum_{j=1}^{p} A_j X_{t-j}.$$

Conditional on the starting observations $x_1, \ldots, x_p$,

$$L(A_1, \ldots, A_p, \Sigma_Z) = \prod_{t=p+1}^{T} \varphi_{\Sigma_Z}\left(x_t - \sum_{j=1}^{p} A_j x_{t-j}\right).$$

φ_{Σ_Z} is the $N_d(0, \Sigma_Z)$ density; $\Sigma_Z > 0$.

Fitting a VAR(p) by OLS

Every equation uses the same observed predictors $x_{t-1}, \ldots, x_{t-p}$.

$$(\widehat{A}_1, \ldots, \widehat{A}_p) = \underset{A_1,\ldots,A_p}{\arg\min} \sum_{t=p+1}^{T} \left\| x_t - \sum_{i=1}^{p} A_i x_{t-i} \right\|_2^2 .$$

For unrestricted coefficient matrices, Gaussian conditional MLE gives the same fit:

$$\widehat{A}_{j,\mathrm{ML}} = \widehat{A}_{j,\mathrm{OLS}}, \qquad j = 1, \ldots, p.$$

Normal confidence intervals require a time-series limit theorem.

Other innovation distributions

Keep the causal VAR. Let Z_t be i.i.d. with density f. Conditional on $x_1, \ldots, x_p$,

$$L(A_1, \ldots, A_p) = \prod_{t=p+1}^{T} f\left(x_t - \sum_{j=1}^{p} A_j x_{t-j} \right).$$

Independent Laplace$(0, s)$ coordinates, with fixed $s > 0$, give

$$(\widehat{A}_1, \ldots, \widehat{A}_p)_{\mathrm{ML}} \in \operatorname*{arg\,min}_{A_1, \ldots, A_p} \sum_{t=p+1}^{T} \left\| x_t - \sum_{j=1}^{p} A_j x_{t-j} \right\|_1.$$

$\|v\|_1 = \sum_{r=1}^{d} |v_r|$. Unrestricted coefficients; mean-zero innovations with finite covariance.

VARMA has an unobserved predictor

Mean-zero stationary causal VARMA(1,1):

$$X_t = A \underbrace{X_{t-1}}_{\text{observed lag}} + C \underbrace{Z_{t-1}}_{\text{unobserved lag}} + Z_t.$$

Recovering the missing innovation requires the unknown matrices:

$$Z_{t-1} = X_{t-1} - AX_{t-2} - CZ_{t-2}.$$

We first need estimates of the innovations to use as predictors.

$(Z_t) \sim \mathrm{WN}_d(0, \Sigma_Z)$, with Σ_Z positive definite.

Recall: invert the MA filter

$$\Phi(z) = I_d - Az, \qquad M(z) = I_d + Cz.$$

If $\det M(z) \neq 0$ for every $|z| \leq 1$,

$$\Pi(z) = M(z)^{-1}\Phi(z) = \sum_{k \geq 0} \Pi_k z^k,$$

$$Z_t = \sum_{k \geq 0} \Pi_k X_{t-k}, \qquad \Pi_0 = I_d.$$

The coefficients are absolutely summable.

Invertibility gives an infinite VAR

Move the lagged observations to the other side:

$$X_t = \sum_{k=1}^{\infty} F_k X_{t-k} + Z_t, \qquad F_k = -\Pi_k.$$

Keeping only m lags gives

$$X_t - \sum_{k=1}^{m} F_k X_{t-k} = Z_t + \underbrace{\sum_{k>m} F_k X_{t-k}}_{\text{omitted-lag term}}.$$

The omitted-lag term tends to zero in mean square as $m \to \infty$.

First stage: fit a long VAR

Fit by OLS:

$$x_t = F_1 x_{t-1} + \cdots + F_m x_{t-m} + \text{error}.$$

Keep the residuals:

$$\widetilde{z}_t = x_t - \sum_{k=1}^{m} \widehat{F}_k x_{t-k}, \qquad t = m+1, \ldots, T.$$

The workbook uses $m = 8$.

$\widetilde{z}_t$ approximates z_t: omitted lags and estimated coefficients add error.

Second stage: use its residuals as predictors

$$x_t = Ax_{t-1} + C\widetilde{z}_{t-1} + \text{error}.$$

Hold $\widetilde{z}_t$ fixed; fit A and C together:

$$(\widehat{A}, \widehat{C}) = \underset{A,C}{\arg\min} \sum_{t=m+2}^{T} \|x_t - Ax_{t-1} - C\widetilde{z}_{t-1}\|_2^2 .$$

One OLS regression per coordinate, with $2d$ predictors.

Gaussian fitting: invert the MA filter

For VARMA(p, q), write $A = (A_1, \ldots, A_p)$ and $C = (C_1, \ldots, C_q)$. If $\det M_C(z) \neq 0$ for $|z| \leq 1$,

$$Z_t = M_C(B)^{-1} \Phi_A(B) X_t.$$

For each proposed A, C, apply this filter to the observations:

$$e_t(A, C) = \left[M_C(B)^{-1} \Phi_A(B) x \right]_t, \qquad t = p + 1, \ldots, T.$$

Take $x_1, \ldots, x_p$ as given and evaluate the filter with preceding errors set to zero.

Gaussian fitting is nonlinear optimization

Now assume $Z_t \overset{\text{iid}}{\sim} N_d(0, \Sigma_Z)$, with $\Sigma_Z > 0$. The initialized Gaussian log likelihood, up to a constant, is

$$\ell_{\text{init}}(A, C, \Sigma_Z) = -\frac{T-p}{2} \log\det \Sigma_Z - \frac{1}{2} \sum_{t=p+1}^{T} \left\| \Sigma_Z^{-1/2} \left[M_C(B)^{-1} \Phi_A(B) x \right]_t \right\|_2^2.$$

For fixed Σ_Z, minimize these weighted squared errors.

The inverse filter depends on the unknown C_k: fitting all A_j, C_k together requires nonlinear optimization.

Estimate the innovation covariance

The model parameter is $\Sigma_Z = \mathrm{Cov}(Z_t)$. For each proposed A, C, its likelihood estimate is

$$\widehat{\Sigma}_Z(A, C) = \frac{1}{T-p} \sum_{t=p+1}^{T} e_t(A, C) e_t(A, C)^\top.$$

Substitute this estimate into the log likelihood:

$$\ell_{\mathrm{init}}\Big(A, C, \widehat{\Sigma}_Z(A, C)\Big) = \text{constant} - \frac{T-p}{2} \log\det \widehat{\Sigma}_Z(A, C).$$

Thus numerically minimize, jointly over A, C,

$$Q(A, C) = \log\det \widehat{\Sigma}_Z(A, C).$$

Assume $\widehat{\Sigma}_Z(A, C)$ is positive definite.

`MTS::VARMA` uses all three steps

1. Fit a preliminary VAR and retain its residuals $\widetilde{z}_t$.
2. Regress on observation and residual lags to obtain $A^{(0)}, C^{(0)}$.
3. Start there and jointly optimize the conditional Gaussian criterion, reconstructing errors at each proposed A, C.

The two OLS stages supply starting coefficients for the Gaussian fit.

The package chooses its own preliminary VAR order and initialization. `MTS::VARMA` source code.

An intercept is a constant in the model equation

For $V_t \in \mathbb{R}^d$, allow a constant vector $c \in \mathbb{R}^d$:

$$V_t = c + \sum_{i=1}^{p} A_i V_{t-i} + M_C(B) Z_t, \qquad \mathbb{E}[Z_t] = 0.$$

Fit the unrestricted c along with the other coefficients.

- Where OLS applies, add a column of ones.
- In Gaussian VARMA fitting, subtract c inside the inverse MA filter:

$$e_t(c, A, C) = \left[M_C(B)^{-1} \{ \Phi_A(B) v - c \} \right]_t.$$

The intercept generally differs from the mean

For a stationary model, write $\mu = \mathbb{E}[V_t]$. Taking expectations gives

$$\mu = c + \sum_{i=1}^{p} A_i \mu,$$

$$\underbrace{c}_{\text{equation intercept}} = \left(I_d - \sum_{i=1}^{p} A_i \right) \underbrace{\mu}_{\text{stationary mean}}.$$

The autoregressive terms also contribute to the mean.

Another option: Detrend first, then fit a stationary VARMA

Specify

$$V_t = \mu + \gamma t + X_t, \qquad \Phi_A(B) X_t = M_C(B) Z_t,$$

with X_t stationary and mean zero. Thus $\mathbb{E}[V_t] = \mu + \gamma t$.

1. Regress each coordinate of v_t on $1, t$, giving $\widehat{\mu}, \widehat{\gamma}$.
2. Form $r_t = v_t - \widehat{\mu} - \widehat{\gamma} t$.
3. Fit a mean-zero stationary VARMA to r_t.

With only a constant, the first step subtracts the sample mean: $r_t = v_t - \bar{v}$.

For stationary X_t with summable autocovariances, $\widehat{\mu}$ and $\widehat{\gamma}$ are consistent.

Joint Gaussian fitting also estimates the trend

For each proposed μ, γ, A, C, reconstruct the errors:

$$x_t(\mu, \gamma) = v_t - \mu - \gamma t,$$

$$e_t(\mu, \gamma, A, C) = \left[M_C(B)^{-1}\Phi_A(B)x(\mu, \gamma)\right]_t.$$

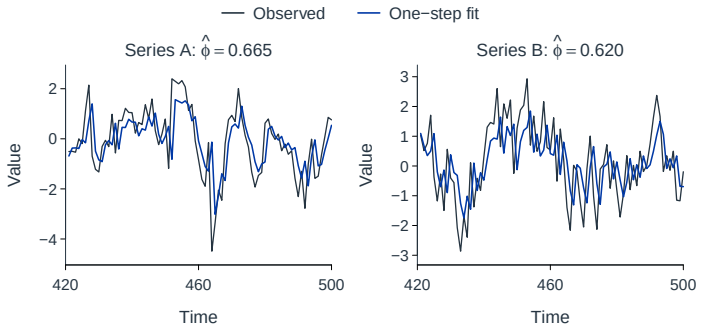
Use these errors in the same initialized Gaussian likelihood, optimizing μ, γ, A, C together.

This generally differs from fixing $\widehat{\mu}, \widehat{\gamma}$ by OLS first.

With $\gamma = 0$, this recovers the intercept model with $c = \Phi_A(1)\mu$.

Forecasting from an OLS fit

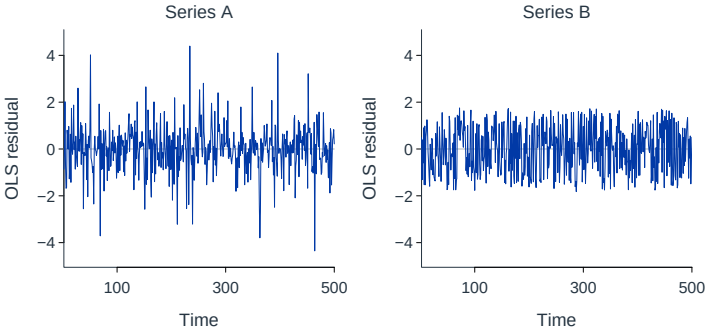
Simulated AR(1) examples: $X_t = 0.65X_{t-1} + Z_t$, with i.i.d. innovations of mean zero and variance one.



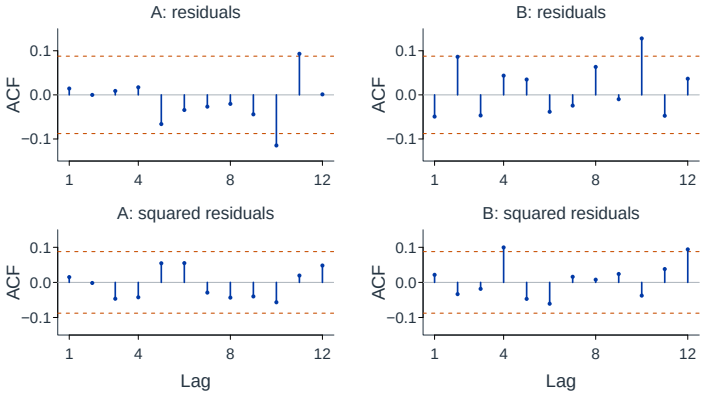
Fit an intercept and slope using the first 500 observations.

Do the residuals have a stable center and spread?

$$\widehat{z}_t = x_t - \widehat{a} - \widehat{\phi} x_{t-1}.$$

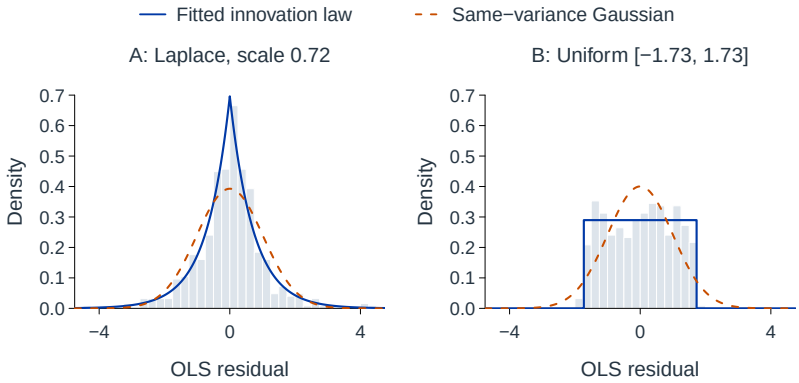


Is there dependence left in the residuals?



Simulated examples. Dashed bounds are approximate references; small ACFs do not establish independence.

Fit a distribution to the residuals

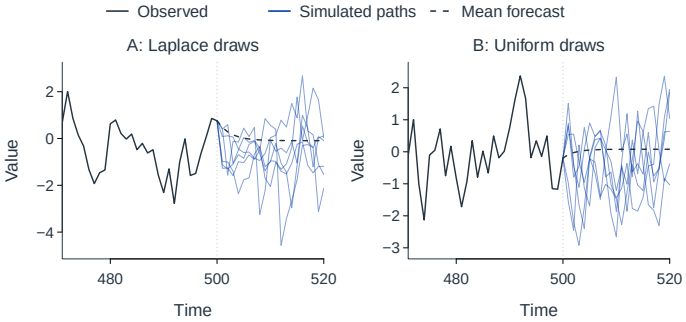


Retain the OLS coefficients; fit the innovation distribution to the residuals.

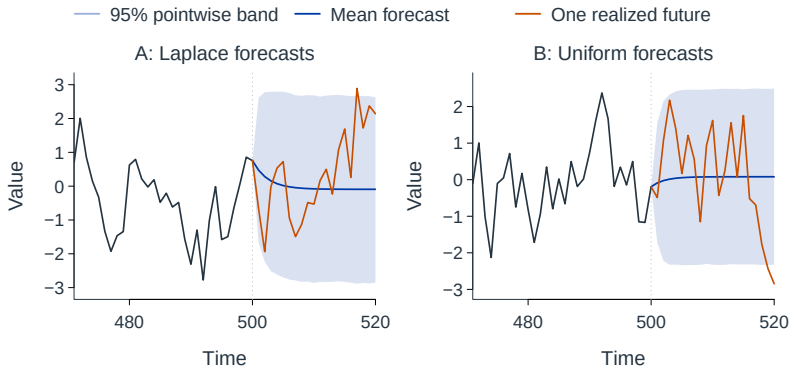
Draw new innovations and run the model forward

Start at $x^*_T = x_T$ and draw $z^*_{T+1}, z^*_{T+2}, \ldots$ independently from the fitted law:

$$x^*_{T+h} = \widehat{a} + \widehat{\phi} x^*_{T+h-1} + z^*_{T+h}.$$



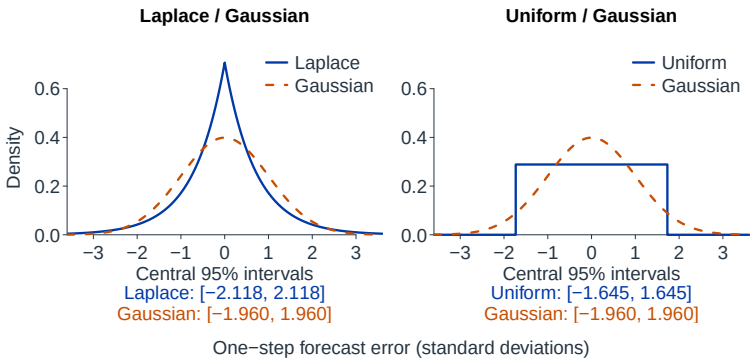
Read prediction intervals from the simulated paths



At each horizon: the 2.5% and 97.5% quantiles of 50,000 paths.

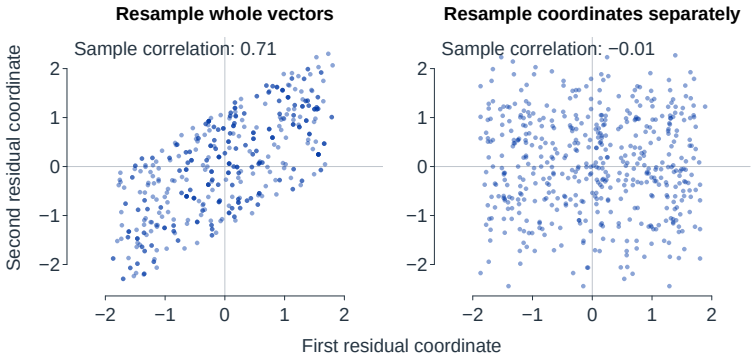
Pointwise 95% intervals under the fitted model; parameter estimates and the last observation are held fixed.

Same variance, different prediction intervals



For a fixed linear model, matching innovation means and covariances preserves forecast means and covariances.

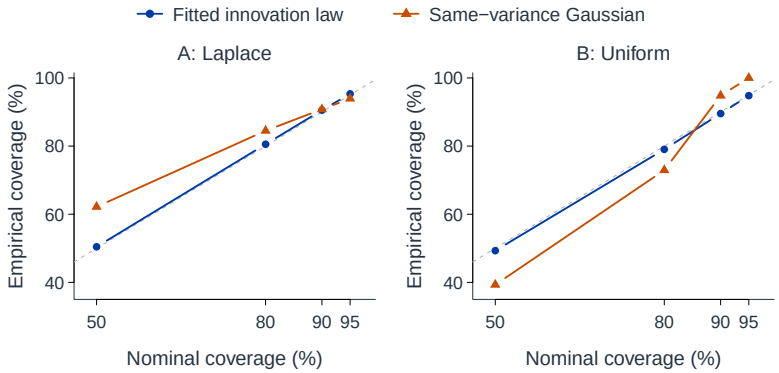
For a VAR, draw the residual vectors together



Fit a joint innovation law, or resample whole centered residual vectors.

A separate simulated bivariate VAR(1), with correlated innovation coordinates.

Check the forecasts on later observations



Compare nominal and observed coverage on 3,000 later observations.

One-step forecasts use the latest observed lag. All fitted parameters remain fixed.

What about combining VARMA with cointegration?

So far, our VARMA models assumed a stationary vector series.

Suppose instead the observed levels $V_t \in \mathbb{R}^d$ are $I(1)$ and cointegrated:

- the first differences ∇V_t are $I(0)$;
- some linear combinations $\beta^\top V_t$ of the levels are also stationary.

Fitting a VARMA to ∇V_t alone omits the level departures $\beta^\top V_{t-1}$. How can a VARMA use both?

Start with a VARMA model for the levels

Let $V_t \in \mathbb{R}^d$ collect the level series. Assume a VARMA model without an intercept:

$$V_t = \sum_{i=1}^{p} A_i V_{t-i} + M_C(B) Z_t, \qquad p \geq 1.$$

Here

$$M_C(B) = I_d + \sum_{j=1}^{q} C_j B^j, \qquad q \geq 0.$$

The levels may be nonstationary.

$(Z_t) \sim \mathrm{WN}_d(0, \Sigma_Z)$, with $\Sigma_Z > 0$.

Write earlier levels using changes

Each earlier level is V_{t-1} minus the changes since then:

$$V_{t-2} = V_{t-1} - \nabla V_{t-1},$$

$$V_{t-3} = V_{t-1} - \nabla V_{t-1} - \nabla V_{t-2},$$

$$V_{t-i} = V_{t-1} - \sum_{k=1}^{i-1} \nabla V_{t-k}, \qquad i = 1, \ldots, p.$$

Subtract V_{t-1} from the levels model:

$$\nabla V_t = (A_1 - I_d) V_{t-1} + \sum_{i=2}^{p} A_i V_{t-i} + M_C(B) Z_t.$$

Collect the coefficient of each change

Substituting the expressions for the earlier levels gives

$$\nabla V_t = \left(\sum_{i=1}^{p} A_i - I_d \right) V_{t-1} - \sum_{i=2}^{p} \sum_{k=1}^{i-1} A_i \nabla V_{t-k} + M_C(B) Z_t.$$

For a fixed k, the change ∇V_{t-k} appears for $i = k+1, \ldots, p$. Collecting these terms gives

$$\nabla V_t = \underbrace{\left(\sum_{i=1}^{p} A_i - I_d \right)}_{\Pi} V_{t-1} + \sum_{k=1}^{p-1} \underbrace{\left(- \sum_{i=k+1}^{p} A_i \right)}_{B_k} \nabla V_{t-k} + M_C(B) Z_t.$$

Cointegration restricts the level term

Now impose the $I(1)$, cointegrated structure. The differences and filtered innovations are jointly stationary, so the equation requires ΠV_{t-1} to be stationary too.

Let the columns of $\beta \in \mathbb{R}^{d \times r}$ span all coefficient vectors that give stationary linear combinations of the levels, with $0 < r < d$. Every row of Π is therefore a linear combination of those vectors. Consequently,

$$\Pi = \alpha\beta^{\top}, \qquad \alpha \in \mathbb{R}^{d \times r}.$$

Substitute the cointegrating relations

Substituting $\Pi = \alpha\beta^\top$ gives the error-correction form:

$$\nabla V_t - \alpha\beta^\top V_{t-1} - \sum_{k=1}^{p-1} B_k \nabla V_{t-k} = M_C(B) Z_t.$$

The stationary departures $\beta^\top V_{t-1}$ enter through the adjustment matrix α.

This is the assumed levels VARMA, rewritten in terms of changes and stationary level combinations.