1 Handling Outliers with Dummy Variables

Suppose March's value, 35, is an outlier in an AR(2). The table shows part of a longer sample, with lags constructed using the original dates.

Month	y_t	y_{t-1}	y_{t-2}
Mar	**35**	9	8
Apr	10	**35**	9
May	11	10	**35**
Jun	12	11	10

Option 2a: exclude the outlier only as an outcome

Only March has 35 as the outcome. Add one dummy:

$$y_t = m + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \delta_{\text{Mar}} D_{\text{Mar},t} + \varepsilon_t.$$

Month	$D_{\text{Mar},t}$	Effect on AR estimation
Mar	1	Excluded
Apr	0	Kept: 35 is allowed as a lag
May	0	Kept: 35 is allowed as a lag
Jun	0	Kept

For March, the regression is

$$35 = m + 9\phi_1 + 8\phi_2 + \delta_{\text{Mar}} + \varepsilon_{\text{Mar}}.$$

For any given AR coefficients, choosing

$$\delta_{\text{Mar}} = 35 - m - 9\phi_1 - 8\phi_2$$

makes March's residual zero. **OLS chooses this adjustment because it minimizes March's squared residual without changing any other row's residual.**

Unlike the common intercept m, which shifts every month's fitted value, δ_{Mar} shifts only March's fitted value. OLS can therefore fit March exactly without worsening the fit for any other month.

For April, the dummy is zero:

$$10 = m + 35\phi_1 + 9\phi_2 + \varepsilon_{\text{Apr}}.$$

April's residual still affects the AR estimates. Thus, only March's row is effectively dropped.

Option 2b: exclude the outlier on either side

March, April, and May contain 35. Add three separate dummies:

$$\begin{aligned} y_t = {} & m + \phi_1 y_{t-1} + \phi_2 y_{t-2} \\ & + \delta_{\text{Mar}} D_{\text{Mar},t} + \delta_{\text{Apr}} D_{\text{Apr},t} + \delta_{\text{May}} D_{\text{May},t} + \varepsilon_t. \end{aligned}$$

Month	$D_{\text{Mar},t}$	$D_{\text{Apr},t}$	$D_{\text{May},t}$	Effect on AR estimation
Mar	1	0	0	Excluded
Apr	0	1	0	Excluded
May	0	0	1	Excluded
Jun	0	0	0	Kept

The three affected rows are

$$35 = m + 9\phi_1 + 8\phi_2 + \delta_{\text{Mar}} + \varepsilon_{\text{Mar}},$$
$$10 = m + 35\phi_1 + 9\phi_2 + \delta_{\text{Apr}} + \varepsilon_{\text{Apr}},$$
$$11 = m + 10\phi_1 + 35\phi_2 + \delta_{\text{May}} + \varepsilon_{\text{May}}.$$

Each row has its own adjustment, so OLS can minimize all three squared residuals to zero without affecting the other rows. The AR estimates are therefore the same as if these three rows were dropped.

For June, all three dummies are zero:

$$12 = m + 11\phi_1 + 10\phi_2 + \varepsilon_{\text{Jun}}.$$

June's residual still affects the AR estimates.

Keep the original lags: June still uses May's 11 and April's 10, even though their regression rows are excluded.

2 Forecasting with Trends and Seasonality

1. Forecast using future dates and quarters

Consider a model with a linear trend, quarterly seasonality, and an AR(1):

$$y_t = a + bt + \delta_2 D_{2t} + \delta_3 D_{3t} + \delta_4 D_{4t} + \phi y_{t-1} + \varepsilon_t.$$

Here, t counts quarters from the start of the sample, and $D_{qt} = 1$ in quarter q and 0 otherwise. The first quarter is the reference quarter.

Suppose estimation gives

$$\widehat{a} = 1, \quad \widehat{b} = 0.1, \quad \widehat{\delta}_2 = 2, \quad \widehat{\delta}_3 = -1, \quad \widehat{\delta}_4 = 3, \quad \widehat{\phi} = 0.5.$$

The sample starts in 2025.1, so 2026.4 is observation $T = 8$. The observed value in that quarter is $y_8 = 10$. We want to forecast the next two quarters:

Date	t	D_{2t}	D_{3t}	D_{4t}
2026.4 (last observed)	8	0	0	1
2027.1 (forecast)	9	0	0	0
2027.2 (forecast)	10	1	0	0

2027.1: all seasonal dummies are zero, and the lagged value is $y_8 = 10$:

$$y_{9|8} = \widehat{a} + \widehat{b}(9) + \widehat{\phi} y_8 = 1 + 0.1(9) + 0.5(10) = \boxed{6.9}.$$

2027.2: the second-quarter dummy is one. Since y_9 is not yet observed, use $y_{9|8} = 6.9$:

$$y_{10|8} = \widehat{a} + \widehat{b}(10) + \widehat{\delta}_2 + \widehat{\phi} y_{9|8} = 1 + 0.1(10) + 2 + 0.5(6.9) = \boxed{7.45}.$$

Both forecasts use only information available in 2026.4.

2. Convert growth forecasts to level forecasts

Let $x_t = \ln Y_t$ be the log level. A log change is the difference between two consecutive log levels:

$$\Delta \ln Y_t = \ln Y_t - \ln Y_{t-1}.$$

For example, a log change of 0.02 represents approximately 2% growth. If we report growth as $y_t = c\,\Delta \ln Y_t$, we must divide by c to recover the log change:

$$\ln Y_t - \ln Y_{t-1} = \frac{y_t}{c}.$$

Here, $c = 100$ for percentage log growth, $c = 400$ for quarterly annualized growth, and $c = 1200$ for monthly annualized growth.

Future log level = current log level + accumulated log changes.

For one period,

$$\ln Y_{T+1} = \ln Y_T + \frac{y_{T+1}}{c}.$$

For two periods, add the next change:

$$\ln Y_{T+2} = \ln Y_{T+1} + \frac{y_{T+2}}{c} = \ln Y_T + \frac{y_{T+1} + y_{T+2}}{c}.$$

Thus, replacing future growth with its forecasts,

$$\boxed{x_{T+h|T} = x_T + \frac{1}{c}\sum_{j=1}^{h} y_{T+j|T}}.$$

To convert the log forecast back to the original units, exponentiate:

$$Y_{T+h|T} = \exp\left(x_{T+h|T}\right).$$

3 Standard Errors and Hypothesis Tests

1. Where does the standard error come from?

Consider $y_t = \alpha + \beta x_t + u_t$. OLS chooses the intercept and slope that minimize the mean squared residual:

$$(\widehat{\alpha}, \widehat{\beta}) = \arg\min_{a,b} \frac{1}{T}\sum_{t=1}^{T}(y_t - a - bx_t)^2.$$

Let $\mathrm{SST}_x = \sum_{t=1}^{T}(x_t - \bar{x})^2$. The fitted slope is

$$\widehat{\beta} = \frac{\sum_{t=1}^{T}(x_t - \bar{x})y_t}{\mathrm{SST}_x} = \beta + \frac{\sum_{t=1}^{T}(x_t - \bar{x})u_t}{\mathrm{SST}_x}.$$

Hold the regressors X fixed and imagine drawing new errors. The y_t values change, so the fitted slope changes too. Its **standard error** estimates the standard deviation of these slope estimates across samples.

Under the classical assumptions, the errors have conditional mean zero, constant variance σ_u^2, and no serial correlation. Then

$$\mathrm{Var}(\widehat{\beta} \mid X) = \frac{\sigma_u^2}{\mathrm{SST}_x}, \qquad \widehat{\sigma}_u^2 = \frac{\sum_{t=1}^{T}\widehat{u}_t^2}{T-2}, \qquad SE_{\text{classical}}(\widehat{\beta}) = \sqrt{\frac{\widehat{\sigma}_u^2}{\mathrm{SST}_x}}.$$

The residuals $\widehat{u}_t = y_t - \widehat{\alpha} - \widehat{\beta}x_t$ estimate the noise around the line. Dividing the estimated error variance by SST_x gives the estimated variance of the slope.

2. Why does the standard error matter for a hypothesis test?

Suppose we estimate $\widehat{\beta} = 0.8$ and want to test $H_0 : \beta = 1$. Is a gap of 0.2 large? That depends on how much the estimated slope typically varies across samples. The t-statistic measures the gap in **standard errors**:

$$t = \frac{\widehat{\beta} - \beta_0}{\underbrace{\sqrt{\boxed{\widehat{\text{Var}}(\widehat{\beta})}}}_{SE(\widehat{\beta})}}.$$

The box contains the estimated variance of the slope. For the same estimate and the same null hypothesis, consider two possible standard errors:

$SE(\widehat{\beta})$	$t = (0.8 - 1)/SE(\widehat{\beta})$	Two-sided test at 5%
0.08	-2.50	Reject H_0
0.15	-1.33	Do not reject H_0

In the first row, the estimate is 2.50 standard errors below 1, beyond the large-sample cutoff of 1.96. In the second, the same gap is only 1.33 standard errors. To decide which standard error to use, we need to understand the regression errors.

3. Which standard error fits the regression?

Method	Error variance	Correlation across dates
Classical	Constant (homoskedastic)	Zero
Robust	May vary (heteroskedastic)	Zero
Newey–West	May vary	May be nonzero

Example. Why use Newey–West for a two-step direct regression? Write an AR(1) for two successive dates:

$$y_{t+1} = m + \phi y_t + \varepsilon_{t+1},$$
$$y_{t+2} = m + \phi y_{t+1} + \varepsilon_{t+2}.$$

Substitute the first equation into the second:

$$\begin{aligned} y_{t+2} &= m + \phi(m + \phi y_t + \varepsilon_{t+1}) + \varepsilon_{t+2} \\ &= m(1 + \phi) + \phi^2 y_t + \underbrace{\phi\varepsilon_{t+1} + \varepsilon_{t+2}}_{u_{t+2}}. \end{aligned}$$

With mean-zero future shocks conditional on information at t, u_{t+2} is the two-step forecast error using the true parameters. Moving the forecast origin forward one period gives

$$\text{Forecast made at } t: \quad u_{t+2} = \phi\varepsilon_{t+1} + \boxed{\varepsilon_{t+2}},$$
$$\text{Forecast made at } t+1: \quad u_{t+3} = \phi\boxed{\varepsilon_{t+2}} + \varepsilon_{t+3}.$$

The two errors share a shock. If the original shocks are uncorrelated with variance σ_ε^2, then $\text{Cov}(u_{t+2}, u_{t+3}) = \phi\sigma_\varepsilon^2$. When $\phi \neq 0$, the regression errors are serially correlated. Newey–West accounts for this correlation when estimating the standard error for a coefficient test.

4. How do we calculate the Newey–West standard error?

Step 1. Compute one value of z_t for each date. Use the same regression as in part 1. For each observation, calculate the OLS residual and multiply it by the centered regressor:

$$\widehat{u}_t = y_t - \widehat{\alpha} - \widehat{\beta}x_t, \qquad z_t = (x_t - \bar{x})\widehat{u}_t.$$

This gives a series $z_1, z_2, \ldots, z_T$.

Step 2. Pair the z values by lag. At lag 0, square each value. At lag 1, multiply neighbors. At lag 2, multiply values two dates apart. Divide each sum by T:

$$\begin{aligned}
\widehat{\gamma}_0 &= (z_1^2 + z_2^2 + \cdots + z_T^2)/T, \\
\widehat{\gamma}_1 &= (z_2 z_1 + z_3 z_2 + \cdots + z_T z_{T-1})/T, \\
\widehat{\gamma}_2 &= (z_3 z_1 + z_4 z_2 + \cdots + z_T z_{T-2})/T, \\
\widehat{\gamma}_3 &= (z_4 z_1 + z_5 z_2 + \cdots + z_T z_{T-3})/T.
\end{aligned}$$

If a question gives these autocovariances, start at Step 3.

Step 3. Apply the Newey–West weights.

Let $m-1$ be the largest lag included. For each lag $j = 1, \ldots, m-1$, multiply $\widehat{\gamma}_j$ by the weight $1 - j/m$. The estimated long-run variance is

$$\boxed{\widehat{V}_L = \widehat{\gamma}_0 + 2\sum_{j=1}^{m-1}\left(1 - \frac{j}{m}\right)\widehat{\gamma}_j}.$$

The weights decrease with the lag. The factor of 2 accounts for the equal autocovariances at lags $+j$ and $-j$. Lag zero appears only once.

Step 4. Convert $\widehat{V}_L$ to a slope standard error. First compute the sample variation in x:

$$\widehat{\sigma}_x^2 = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2.$$

Then divide $\widehat{V}_L$ by $T(\widehat{\sigma}_x^2)^2$ to get the estimated slope variance. Take its square root:

$$\boxed{SE_{\text{NW}}(\widehat{\beta}) = \sqrt{\frac{\widehat{V}_L}{T(\widehat{\sigma}_x^2)^2}}}.$$

Use this standard error in the denominator of the t-statistic.

Lecture notation. For the zero-mean model $y_t = \beta x_t + u_t$ fitted without an intercept, use $z_t = x_t\widehat{u}_t$ in Step 1 and $\widehat{\sigma}_x^2 = T^{-1}\sum_{t=1}^{T} x_t^2$ in Step 4.

4 Using Expert and Market Forecasts

Survey forecasts.

Let π_t be actual inflation and τ_t the survey's long-run inflation expectation. Their difference is the inflation gap:

$$y_t = \pi_t - \tau_t.$$

A positive gap means inflation is above the survey expectation; a negative gap means it is below.

Fit an AR model to the gap. For example,

$$y_t = m + \phi y_{t-1} + \varepsilon_t.$$

Use this model to forecast the future gap, $y_{T+h|T}$.

To recover an inflation forecast, start from the identity

$$\pi_{T+h} = \tau_{T+h} + y_{T+h}.$$

Future survey expectations are unknown, so assume they remain at today's value: $\tau_{T+h|T} = \tau_T$. Then

$$\pi_{T+h|T} = \tau_T + y_{T+h|T}.$$

Thus, forecast inflation equals today's long-run expectation plus the predicted gap above or below it.

Market forecasts.

Suppose today's forward rate is the expected future spot rate:

$$f_{t,t+h} = E(s_{t+h} \mid \Omega_t),$$

where Ω_t is the information available today. Define the forecast error as

$$y_{t+h} = s_{t+h} - f_{t,t+h}.$$

- **Regress the forecast error on information known today:**

  $$y_{t+h} = \beta_0 + \beta_1 z_t + u_{t+h},$$

  where z_t is known at time t.

  If the forecast already uses today's information correctly, z_t should not help predict its error. The slope hypothesis is

  $$H_0 : \beta_1 = 0.$$

  If $\beta_1 \neq 0$, the error has a predictable component that we could use to adjust the original forecast.

- **Regress the actual spot rate on the forward rate:**

  $$s_{t+h} = \beta_0 + \beta_1 f_{t,t+h} + u_{t+h}.$$

  Since

  $$s_{t+h} = f_{t,t+h} + y_{t+h},$$

  the benchmark coefficient on $f_{t,t+h}$ is one:

  $$H_0 : \beta_1 = 1.$$

  A one-unit increase in the forward rate predicts a one-unit increase in the future spot rate.

Under the exact conditional-mean benchmark above, the intercept β_0 is also zero in both regressions.

Exercises

Exercise 1: From growth forecasts to level forecasts

Let Y_t be a quarterly series and define annualized growth as

$$y_t = 400(\ln Y_t - \ln Y_{t-1}).$$

Let $x_t = \ln Y_t$. The latest observed level is $Y_T = 100$. The growth forecasts made at time T are

$$y_{T+1|T} = 4, \qquad y_{T+2|T} = 8.$$

1. Calculate the forecast change in $\ln Y$ for each quarter. Then calculate the total forecast change in $\ln Y$ from T to $T+2$.

2. Calculate the implied level forecast for Y_{T+2}. You may leave your answer in exponential form. A student uses $100e^{8/400}$. What is missing from this calculation?

Exercise 2: Newey–West standard errors

Consider the regression

$$y_t = \alpha + \beta x_t + u_t.$$

Suppose

$$\widehat{\beta} = 0.8, \qquad T = 100, \qquad \widehat{\sigma}_x^2 = 1,$$

where $\widehat{\sigma}_x^2$ is the estimated variance of x_t.

The estimated autocovariances of $w_t = (x_t - \bar{x})\widehat{u}_t$ are:

Lag j	0	1	2	3
$\widehat{\gamma}_j$	1.5	0.5	0	0

Use $m = 4$ and the formulas

$$\widehat{V}_L = \widehat{\gamma}_0 + 2\sum_{j=1}^{m-1}\left(1 - \frac{j}{m}\right)\widehat{\gamma}_j, \qquad SE_{\mathrm{NW}}(\widehat{\beta}) = \sqrt{\frac{\widehat{V}_L}{T(\widehat{\sigma}_x^2)^2}}.$$

1. Calculate $\widehat{V}_L$ and the Newey–West standard error of $\widehat{\beta}$. Show your calculations.

2. Test $H_0 : \beta = 1$ against $H_1 : \beta \neq 1$. Calculate the t statistic using your Newey–West standard error. Use critical values of -1.96 and 1.96 for a two-sided 5% test. Would you reject H_0?

Answers

Exercise 1: From growth forecasts to level forecasts

Let Y_t be a quarterly series and define annualized growth as

$$y_t = 400(\ln Y_t - \ln Y_{t-1}).$$

Let $x_t = \ln Y_t$. The latest observed level is $Y_T = 100$. The growth forecasts made at time T are

$$y_{T+1|T} = 4, \qquad y_{T+2|T} = 8.$$

1. Calculate the forecast change in $\ln Y$ for each quarter. Then calculate the total forecast change in $\ln Y$ from T to $T+2$.

   Since $y_t = 400\,\Delta \ln Y_t$, divide each growth forecast by 400:

   $$\text{First quarter:} \quad \frac{4}{400} = 0.01,$$
   $$\text{Second quarter:} \quad \frac{8}{400} = 0.02.$$

   Add the changes for both quarters:

   $$\text{Total forecast change in } \ln Y = 0.01 + 0.02 = \boxed{0.03}.$$

2. Calculate the implied level forecast for Y_{T+2}. You may leave your answer in exponential form. A student uses $100e^{8/400}$. What is missing from this calculation?

   The forecast log level is

   $$x_{T+2|T} = \ln(100) + 0.03.$$

   Exponentiating gives

   $$Y_{T+2|T} = e^{\ln(100)+0.03} = 100e^{0.03} \approx \boxed{103.05}.$$

   The student's calculation includes only the second quarter's growth. It omits the first quarter's forecast log change, $4/400 = 0.01$. To forecast the level two quarters ahead, we must include growth in both quarters.

Exercise 2: Newey–West standard errors

Consider the regression

$$y_t = \alpha + \beta x_t + u_t.$$

Suppose

$$\widehat{\beta} = 0.8, \qquad T = 100, \qquad \widehat{\sigma}_x^2 = 1,$$

where $\widehat{\sigma}_x^2$ is the estimated variance of x_t.

The estimated autocovariances of $w_t = (x_t - \bar{x})\widehat{u}_t$ are:

Lag j	0	1	2	3
$\widehat{\gamma}_j$	1.5	0.5	0	0

Use $m = 4$ and the formulas

$$\widehat{V}_L = \widehat{\gamma}_0 + 2\sum_{j=1}^{m-1}\left(1 - \frac{j}{m}\right)\widehat{\gamma}_j, \qquad SE_{\text{NW}}(\widehat{\beta}) = \sqrt{\frac{\widehat{V}_L}{T(\widehat{\sigma}_x^2)^2}}.$$

1. Calculate $\widehat{V}_L$ and the Newey–West standard error of $\widehat{\beta}$. Show your calculations.

   With $m = 4$, the weights for lags 1, 2, and 3 are

   $$1 - \frac{1}{4} = \frac{3}{4}, \qquad 1 - \frac{2}{4} = \frac{1}{2}, \qquad 1 - \frac{3}{4} = \frac{1}{4}.$$

   Apply these weights to the autocovariance estimates:

   $$\begin{aligned}\widehat{V}_L &= 1.5 + 2\left[\frac{3}{4}(0.5) + \frac{1}{2}(0) + \frac{1}{4}(0)\right] \\ &= 1.5 + 2(0.375) \\ &= \boxed{2.25}.\end{aligned}$$

   The estimated variance of the slope is

   $$\widehat{\text{Var}}(\widehat{\beta}) = \frac{2.25}{100(1)^2} = 0.0225.$$

   Take its square root:

   $$SE_{\text{NW}}(\widehat{\beta}) = \sqrt{0.0225} = \boxed{0.15}.$$

2. Test $H_0 : \beta = 1$ against $H_1 : \beta \neq 1$. Calculate the t statistic using your Newey–West standard error. Use critical values of -1.96 and 1.96 for a two-sided 5% test. Would you reject H_0?

   The hypothesized value is $\beta_0 = 1$, so

   $$t = \frac{\widehat{\beta} - \beta_0}{SE_{\text{NW}}(\widehat{\beta})} = \frac{0.8 - 1}{0.15} = \frac{-0.2}{0.15} \approx \boxed{-1.33}.$$

   Since -1.33 is inside $[-1.96, 1.96]$, we **do not reject** $H_0 : \beta = 1$ at the 5% level. There is insufficient evidence that β differs from 1.

5 Midterm 2 Review

Review Sheet

The following summarizes important formulas and concepts. It is intended for review and may not be comprehensive.

Do not bring this sheet to the exam. Enjoy! ♡

5.1 Lag Operators and ARMA Models

Lag operator:

$$Ly_t = y_{t-1}, \qquad L^j y_t = y_{t-j}.$$

First difference:

$$\Delta y_t = (1 - L)y_t = y_t - y_{t-1}.$$

Multiplying lag polynomials:

$$(1 - aL)(1 - bL) = 1 - (a + b)L + abL^2.$$

Geometric expansion:

$$\frac{1}{1 - \phi L} = 1 + \phi L + \phi^2 L^2 + \cdots, \qquad |\phi| < 1.$$

White noise:

$$E(\varepsilon_t) = 0, \qquad \mathrm{Var}(\varepsilon_t) = \sigma^2, \qquad \mathrm{Cov}(\varepsilon_t, \varepsilon_s) = 0 \quad (t \neq s).$$

AR(p):

$$y_t = m + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t.$$

MA(q):

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}.$$

ARMA(p, q):

$$y_t = m + \sum_{j=1}^{p} \phi_j y_{t-j} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}.$$

The MA terms above use the slides' minus-sign convention. The θ_j coefficients can be positive or negative. The order is the largest lag with a nonzero coefficient, not the number of nonzero coefficients.

Mean of a stationary AR(p) or ARMA(p, q):

$$\mu = E(y_t) = \frac{m}{1 - \phi_1 - \cdots - \phi_p}.$$

Lag-polynomial representation of an ARMA(p, q):
Let m be the intercept and $\mu = E(y_t)$ the mean of the stationary process.

Intercept form:

$\phi(L)y_t = m + \theta(L)\varepsilon_t.$

Mean-centered form:

$\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t.$

The two forms describe the same model, with

$$m = (1 - \phi_1 - \cdots - \phi_p)\mu, \qquad \mu = \frac{m}{1 - \phi_1 - \cdots - \phi_p}.$$

The lag polynomials are

$\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p,$

$\theta(L) = 1 - \theta_1 L - \cdots - \theta_q L^q.$

Special cases:

- AR(p): $\theta(L) = 1$.
- MA(q): $\phi(L) = 1$, and the intercept equals the mean: $m = \mu$.

Covariance stationarity:

$$E(y_t) = \mu, \qquad \mathrm{Var}(y_t) = \gamma_0 < \infty, \qquad \mathrm{Cov}(y_t, y_{t-k}) = \gamma_k.$$

The mean and variance do not depend on the date. The autocovariance depends only on the lag.

Stationarity condition for an AR(1):

$|\phi| < 1.$

Invertibility condition for an MA(1):

$$y_t - \mu = \varepsilon_t - \theta\varepsilon_{t-1}, \qquad |\theta| < 1.$$

Recovering shocks from an invertible MA(1):

$\varepsilon_t = (y_t - \mu) + \theta(y_{t-1} - \mu) + \theta^2(y_{t-2} - \mu) + \cdots .$

A finite MA process is stationary for any finite coefficients. Invertibility is a separate condition.

5.2 Wold Representation and Forecast Uncertainty

Wold representation for the stationary ARMA models:

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \qquad \psi_0 = 1, \qquad \sum_{j=0}^{\infty} \psi_j^2 < \infty.$$

AR(1) written as an infinite MA:

$$y_t - \mu = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \cdots, \qquad \psi_j = \phi^j.$$

Variance of the series:

$$\mathrm{Var}(y_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2.$$

Forecast using known shocks:

$$y_{T+h|T} = \mu + \sum_{j=h}^{\infty} \psi_j \varepsilon_{T+h-j}.$$

Treat parameters and past shocks as known. Future shocks have conditional mean zero given information at T, so their forecast contributions are zero.

Forecast error:

$$e_{T+h|T} = y_{T+h} - y_{T+h|T} = \sum_{j=0}^{h-1} \psi_j \varepsilon_{T+h-j}.$$

Forecast-error variance:

$$v_h^2 = \mathrm{Var}(e_{T+h|T}) = \sigma^2 \sum_{j=0}^{h-1} \psi_j^2.$$

Use the first h weights, starting with $\psi_0 = 1$. Negative weights still add variance because they are squared.

MA(2) forecasts:

$$y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}.$$

$$y_{T+1|T} = \mu - \theta_1\varepsilon_T - \theta_2\varepsilon_{T-1},$$
$$y_{T+2|T} = \mu - \theta_2\varepsilon_T,$$
$$y_{T+h|T} = \mu, \qquad h \geq 3.$$

MA(2) forecast-error variances:

$$v_1^2 = \sigma^2, \qquad v_2^2 = \sigma^2(1+\theta_1^2), \qquad v_h^2 = \sigma^2(1+\theta_1^2+\theta_2^2) \quad (h \geq 3).$$

For an MA(q), forecasts equal the mean when $h > q$, and the forecast-error variance no longer increases. In particular, the three-step and four-step intervals for an MA(2) have the same width.

Forecast intervals with normally distributed forecast errors:

$90\%:\quad y_{T+h|T} \pm 1.65 v_h.$

$95\%:\quad y_{T+h|T} \pm 1.96 v_h.$

Here $v_h = \sqrt{v_h^2}$ is the standard deviation of the forecast error, not the standard error of a regression coefficient.

5.3 Autocorrelations, Partial Autocorrelations, and Residual Checks

Autocorrelation at lag k: measures the linear relationship between observations k periods apart:

$$\rho_k = \frac{\text{Cov}(y_t, y_{t-k})}{\text{Var}(y_t)}.$$

Partial autocorrelation at lag k: measures the relationship between y_t and y_{t-k} after controlling for the intervening lags. It is the coefficient β_k in

$$y_t = m + \beta_1 y_{t-1} + \cdots + \beta_k y_{t-k} + u_t.$$

Model patterns

Model	Autocorrelations	Partial autocorrelations
AR(p)	Tail off	Cut off after lag p
MA(q)	Cut off after lag q	Tail off
ARMA(p, q), $p, q > 0$	Tail off	Tail off
White noise	Zero at nonzero lags	Zero at nonzero lags

Tail off: gradually approach zero.
Cut off: zero beyond a certain lag; sample values fluctuate around zero.

Reference bands and residual checks

Approximate 95% reference bands:

$$\pm \frac{1.96}{\sqrt{T}}, \qquad T = \text{sample size}.$$

Outside the bands suggests nonzero correlation. For residuals, this suggests remaining time dependence. Inside the bands means no clear evidence at that lag, not proof that the model is correct.

5.4 Autoregressive Forecasting and Lag Selection

Recursive forecasts (AR(2)):

$y_t = m + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t.$

Estimate one model and replace unknown future values with forecasts:

$y_{T+1|T} = \widehat{m} + \widehat{\phi}_1 y_T + \widehat{\phi}_2 y_{T-1}$,

$y_{T+2|T} = \widehat{m} + \widehat{\phi}_1 y_{T+1|T} + \widehat{\phi}_2 y_T$,

$y_{T+3|T} = \widehat{m} + \widehat{\phi}_1 y_{T+2|T} + \widehat{\phi}_2 y_{T+1|T}$.

Direct forecasts (AR(2)):

Estimate a separate regression for each horizon h:

$y_{t+h} = \alpha_h + \beta_h y_t + \gamma_h y_{t-1} + u_{t+h}$.

Use observed values directly, without intermediate forecasts:

$y_{T+h|T} = \widehat{\alpha}_h + \widehat{\beta}_h y_T + \widehat{\gamma}_h y_{T-1}$.

Only use estimation rows with $t + h \leq T$.

Neither method is always better. Compare out-of-sample accuracy at the same dates and horizons.

Lag selection: AIC and BIC

- Balance model fit against a penalty for more lags.
- Choose the **smallest value** of each criterion, using the same series and estimation sample.
- BIC usually penalizes extra lags more heavily, so it tends to select fewer lags than AIC.

5.5 Outliers, Date Dummies, and Missing Data

Date dummy:

$$D_{s,t} = \begin{cases} 1, & t = s, \\ 0, & t \neq s, \end{cases} \qquad y_t = m + \sum_{j=1}^{p} \phi_j y_{t-j} + \delta_s D_{s,t} + \varepsilon_t.$$

A separate date dummy gives the same AR estimates as excluding that regression row.

One outlier at date s in an AR(p):

Exclude the outlier as	Dates requiring separate dummies
An outcome only	s
An outcome or a regressor	$s, s+1, \ldots, s+p$

- Use one dummy per excluded row in the estimation sample.
- Keep the original calendar lags.
- Set past-date dummies to zero when forecasting future dates.

Interpolate one missing observation:

$$\widetilde{y}_t = \frac{y_{t-1} + y_{t+1}}{2}.$$

For logged data, interpolate the logs:

$$\widetilde{\ln Y_t} = \frac{\ln Y_{t-1} + \ln Y_{t+1}}{2}.$$

Both neighbors must be observed by the forecast date.

5.6 Trends, Seasonality, and Growth-to-Level Forecasts

Forecast components and add:

$$y_t = T_t + S_t + C_t, \qquad y_{T+h|T} = T_{T+h|T} + S_{T+h|T} + C_{T+h|T}.$$

Trend, seasonality, and AR(1):

$$y_t = m + bt + \sum_{q=2}^{4} \delta_q D_{qt} + \phi y_{t-1} + \varepsilon_t.$$

$D_{qt} = 1$ in quarter q and zero otherwise. With an intercept, use three quarterly dummies; quarter 1 is the reference.

$$y_{T+1|T} = \widehat{m} + \widehat{b}(T+1) + \sum_{q=2}^{4} \widehat{\delta}_q D_{q,T+1} + \widehat{\phi} y_T,$$
$$y_{T+2|T} = \widehat{m} + \widehat{b}(T+2) + \sum_{q=2}^{4} \widehat{\delta}_q D_{q,T+2} + \widehat{\phi} y_{T+1|T}.$$

Use the **future date and quarter**. Seasonal dummies repeat every year. Replace unknown lagged values with forecasts.

Growth to levels:

$$y_t = c(\ln Y_t - \ln Y_{t-1}).$$

Use $c = 100$ for percentage log growth, $c = 400$ for quarterly annualized growth, and $c = 1200$ for monthly annualized growth.

$$(\ln Y)_{T+h|T} = \ln Y_T + \frac{1}{c} \sum_{j=1}^{h} y_{T+j|T},$$
$$Y_{T+h|T} = Y_T \exp\left(\frac{1}{c} \sum_{j=1}^{h} y_{T+j|T} \right).$$

Sum growth over **all intervening periods** and divide by the correct scale.

5.7 Standard Errors and Hypothesis Tests

Properties of the regression errors:

- **Homoskedasticity:** constant error variance, $\mathrm{Var}(u_t \mid X) = \sigma_u^2$.
- **Heteroskedasticity:** error variance differs across observations, $\mathrm{Var}(u_t \mid X) = \sigma_t^2$.
- **Serial correlation:** errors at different dates are correlated, $\mathrm{Cov}(u_t, u_s \mid X) \neq 0$ for some $t \neq s$.

Here X denotes the regressors.

Choosing a standard error:

Method	Allows heteroskedasticity?	Allows serial correlation?
Classical	No	No
Robust	Yes	No
Newey–West	Yes	Yes

Test statistic for $H_0 : \beta = \beta_0$:

$$t = \frac{\widehat{\beta} - \beta_0}{SE(\widehat{\beta})}.$$

For a two-sided 5% test using the normal approximation, reject if $|t| > 1.96$.

95% confidence interval for β:

$$\widehat{\beta} \pm 1.96\, SE(\widehat{\beta}).$$

Reject H_0 if β_0 is outside this interval.

Changing the standard error method leaves $\widehat{\beta}$ unchanged, but can change the test statistic, interval, and conclusion.

Overlapping two-step errors from an AR(1):

$$u_{t+2} = \phi\varepsilon_{t+1} + \varepsilon_{t+2},$$
$$u_{t+3} = \phi\varepsilon_{t+2} + \varepsilon_{t+3}.$$

They share ε_{t+2}, so they are correlated when $\phi \neq 0$. Use Newey–West standard errors for coefficient tests.

Calculating the Newey–West standard error:

For the zero-mean regression $y_t = \beta x_t + u_t$, $\widehat{\gamma}_j$ estimates the lag-j autocovariance of $x_t u_t$.

$$\widehat{V}_L = \widehat{\gamma}_0 + 2\sum_{j=1}^{m-1}\left(1 - \frac{j}{m}\right)\widehat{\gamma}_j.$$

- $m - 1$ is the largest lag with a nonzero weight. At lag m, the weight $1 - m/m$ is zero.
- For example, $m = 4$ uses lags 1, 2, and 3, plus the lag-zero term.

$$SE_{\mathrm{NW}}(\widehat{\beta}) = \sqrt{\frac{\widehat{V}_L}{T(\widehat{\sigma}_x^2)^2}}.$$

Here T is the sample size and $\widehat{\sigma}_x^2$ estimates the variance of x_t. Use this standard error in the t statistic and confidence interval.

5.8 Survey and Financial-Market Forecasts

Forecast error: Let $y_{t+h|t}$ be the forecast made at time t.

$e_{t+h|t} = y_{t+h} - y_{t+h|t}$.

Positive error: the forecast was too low. Negative error: the forecast was too high.

Predictability of forecast errors:
For information x_t known at the forecast date, estimate

$e_{t+h|t} = \alpha + \beta x_t + u_{t+h}, \qquad H_0 : \beta = 0.$

- $\beta = 0$: x_t has no linear predictive power for the forecast error.
- Rejecting H_0 suggests that x_t contains information that could improve the forecast.
- Not rejecting H_0 does not prove the forecast is correct.

Standard errors: If overlapping horizons produce serially correlated errors, use Newey–West standard errors.