8. Testing residuals

Thursday, January 20, 2022 7:58 PM

§1.6 in book

Our question: as we continue to process our data $\{x_t\}$, we get

$$x_t = m_t + y_t \leftarrow \text{residual}$$

m_t: some <u>model</u> (trend, seasonality, ARMA process, ...)

and we want to know if we've captured all the signal / structure

i.e., $\boxed{\text{is } y_t \text{ white noise ?}}$

"<u>noise</u>" is vague — could have "structure"

"<u>white noise</u>" means it has no easy to obtain structure

"<u>iid</u>" means nothing useful at all

<u>Warmup</u>

We're going to ask if $Y_t \sim \underbrace{WN(0, \sigma^2)}_{\text{pairs are uncorrelated}}$

Let's ask something you may have seen in another class:

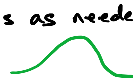
if I think $\{X_i\}_{i=1}^n$ are iid $N(\underbrace{\mu, \sigma^2}_{\text{sometimes known, sometimes not}})$, $\Big\}\ H_0$

how can I check ?

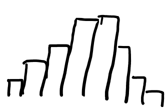
- χ^2 : sample variance $\hat{s}^2$ satisfies (under H_0)

  $$\frac{(n-1)}{\sigma^2}\hat{s}^2 \sim \chi^2_{n-1}$$ and use CDF of χ^2

- plot histogram (play w/ # bins as needed)

  Does it <u>look</u> <u>Gaussian</u> ?  $e^{-x^2/\sigma^2 \cdot 1/2}$

Don't need to know μ nor σ^2 Nice!



} plausible (how <u>stringent</u> we are depends on how large n is)

} seems inconsistent

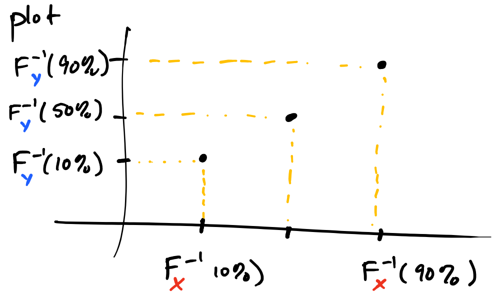
} seems inconsistent

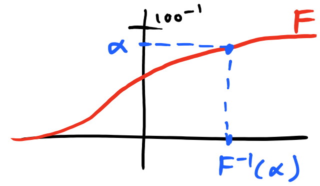
- QQ plot

Take 2 CDFS, F_X and F_Y

Pick a sequence of quantiles, say (10%, 50%, 90%)

plot





If $X \sim Y$ then data should lie on a straight line (since $F_X = F_Y$ so $F_X^{-1} = F_Y^{-1}$)

How to use it?

For Y use the empirical data } for length n, we have n quantiles to plot
(find F_Y and F_Y^{-1} via sorting)

For X use theoretical distribution
(by default, usually $X \sim \text{Normal}(0,1)$)

Advantage: for reference distr, say $X \sim N(\mu, \sigma^2)$, what if we don't know μ or σ^2?

Plot should still be a straight line ($y = mx + b$)
just maybe not $y = x$ line. No problem

Two ways to use: ① visually
② fit a line, run hypothesis test via R^2

= See R demos =

Note: No test is sufficient by itself.
The more you try, the better.

Back to testing if $Y_t \sim WN(0, \sigma^2)$

First, check if sample mean and sample variance are consistent w/ 0 and σ^2. "basic"

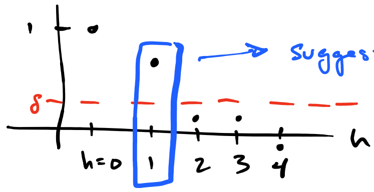
Now, we'll investigate if it has correlation
so primary tool is sample autocovariance $\hat{\rho}$

As $n \to \infty$, expect $\hat{\rho}(h) \sim N(0, 1/n)$ iid if it really was $WN(0,\sigma^2)$

So Not rigorous since asymptotic.

test 1

Check if $|\hat{\rho}(h)| < \delta$, eg. $\delta = \frac{1.96}{\sqrt{n}}$ for 95% Conf. Interval
for $h \neq 0$



suggests to reject $H_0: Y_t \sim WN(0, \sigma^2)$

Downsides: ① not rigorous since $\hat{\rho}$ not exactly iid $N(0, 1/n)$
(more of a problem if n is small)

② If we set δ = 95% C.I. threshold,
and check if $|\hat{\rho}(h)| > \delta$ for $h = 1, 2, \dots, 100$
then by pure luck alone even under H_0, we'd expect a violation

So could do a Bonferroni Correction but that's pessimistic and you lose power

test 2 portmanteau test

1) French for coatrack
2) merging words: "ginormous" = gigantic + enormous
3) H_0 specified, H_1 a bit vague

set $\hat{Q} = n \cdot \sum_{j=1}^{h} \hat{\rho}^2(j) \sim \chi^2_h$

if $\hat{\rho}(h)$ iid $N(0, 1/n)$

So check $\hat{Q} > \chi^2_{1-\alpha}(h)$

Turns out for small n, we can do slightly better than estimating $\hat{\rho}(h) \sim N(0, 1/n)$ iid   Ljung + Box '78

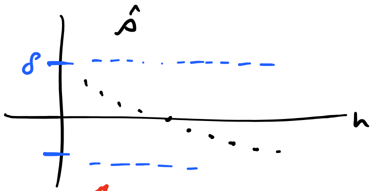
See book (p.31) for details

This complements some of the issues from the 1st test

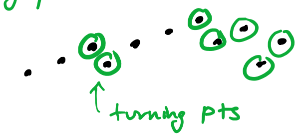
tests 3 and 4

turning point and difference-sign tests

See book for details. Idea:



turning point : local min or max



if $\{y_t\}$ is iid of length n, we expect a lot of turning pts : about $\frac{2}{3}n$

Each $|\hat\rho|$ is small so it might pass tests 1 and 2 but would fail turning pt. test

Differenced sign is similar, but counts how often $y_{t+1} > y_t$. Should happen $\frac{1}{2}$ the time

Related : if all $\hat\rho(h) > 0$ (or all $\hat\rho(h) < 0$) (and n is large), this is suspicious!

other tests : see book

special case : $H_0 : \{Y_t\} \sim N(0, \sigma^2)$ iid

then do normality tests : eg QQ plot