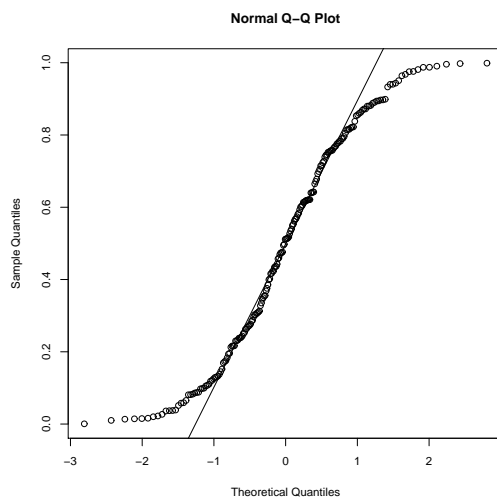


MIDTERM EXAM

- (1) *10 minutes.* Below is a qq plot against a normal cdf of a sample of 200 observations. Does the plot show the sample has approximately a normal distribution, a distribution with fatter tails than the normal, a distribution with thinner tails than the normal, or a distribution with a fatter tail than the normal on one side and a thinner tail on the other? Explain your answer.



The qq plot (quantile-quantile plot) plots the quantiles of one distribution against those of another. The p 'th quantile of X is the value q_p of the random variable X such that $P[X \leq q_p] = p$. The plot shows that, e.g. at the left end, the quantiles of the normal (the horizontal axis) are more widely spaced than those of the sample distribution, so the slope of the plot is low relative to the slope in the middle of the distribution. In one of the exercises, you computed a qq plot of a t -distributed sample against a normal, and that plot was steeper at the far left and far right than in the middle. A t distribution is fatter-tailed than a normal. This question's plot shows flatter slope at the far left and far right than in the middle. So it corresponds to a thinner-tailed distribution.

If you didn't remember the shape from the plot on the exercise, trying to figure out whether the plot is fat or thin tailed by thinking it through from first principles is a bit tricky. The qq plot plots y against x , where the relation between y and x is determined by

$$F_x(x) = F_y(y),$$

with F_x and F_y being the cdf's of x and y . The slope of the $y(x)$ function is then easily shown to be

$$y'(x) = \frac{f_x(x)}{f_y(y(x))}.$$

If you think that a relatively fat-tailed y distribution is one in which f_y/f_x is big in the tails (a natural first stab at a definition of "fat tail"), you might think that a low slope at the extremes of the distribution goes with a fat tail — when in fact the opposite is true. The reason is that we are not taking the ratio of f_x to f_y at $y = x$, but the ratio at $F_x(x) = F_y(y)$. One definition of relative tail fatness is that x has fatter tails than y if $f_x(x)/F_x(x)$ is smaller than $f_y(y)/F_y(y)$ if x and y are both close to the lower limit of the support of their respective distributions. (At the other extreme, the comparison is based on $f_x(x)/(1 - F_x(x))$.)

So if x is a t with one degree of freedom, this ratio behaves, as x approaches $-\infty$, like $-2x/(1+x^2)$ and thus approaches zero, whereas for a normal it behaves like $-x$, and thus approaches infinity. So the t has fatter tails, and for large negative x , at points where $F_x(x) = F_y(y)$ (with y normal) we will have $f_x/f_y = f_x F_y / (f_y F_x)$ very large. I.e., the slope will be very steep at the extremes in the qq plot when a $t(1)$ is plotted against a normal. The plot shown on the exam was of a sample from the uniform distribution against a normal.

- (2) 30 minutes. Here are the results from estimating a logit model from the Stock-Watson “Names” data set. These data resulted from sending out resumés to prospective employers with names that were likely to suggest the applicants were black, and/or female. The variable `call_back` is one for those resumés that generated a call back from the employer, 0 otherwise. `black` is 1 for resumés that were likely to be identified as black, zero otherwise, and `female` is one for resumés likely to be identified as female, 0 otherwise. There was also a dummy variable that was 1 for employers in Chicago. (There were a lot of other variables, but we are keeping this simple.) The logit regression has `call_back` as dependent variable, `black`, `female` and `chicago` as independent variables. Here is the summary output from the maximum likelihood fit:

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.2116      0.1246 -17.743  < 2e-16 ***
female        0.2745      0.1335   2.057   0.0397 *
black        -0.4431      0.1075  -4.120  3.78e-05 ***
chicago     -0.4599      0.1096  -4.198  2.70e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

- (a) Find an approximate 95% probability interval for the coefficient on `black` and use it to give a 95% probability interval for the difference in probability of a call back between a black male not in Chicago and a non-black male, also not in Chicago.

The 2.5% tails of the $N(0,1)$ distribution are at ± 1.96 . Thus the 95% interval on the `black` variable is

$$-.4431 \pm (1.96 \cdot .1075) = (-.2324, .3399).$$

For a male not in Chicago, the coefficients on `female` and `chicago` are not needed to compute the probabilities. The interval for the probabilities for a non-black and a black, respectively, are

$$\frac{e^{-2.2116}}{1 + e^{-2.2116}} = .0987, \quad \frac{e^{-2.2116 - .4431}}{1 + e^{-2.2116 - .4431}} = .0657.$$

- (b) The estimated covariance matrix of these coefficients is

	(Intercept)	female	black	chicago
(Intercept)	0.015537058	-0.012133751	-0.0046002636	-0.0027857895
female	-0.012133751	0.017814209	-0.0002528960	-0.0037471701
black	-0.004600264	-0.000252896	0.0115659895	0.0001393946
chicago	-0.002785789	-0.003747170	0.0001393946	0.0120026554

Use it to construct a chi-squared test statistic for the hypothesis that the `female` and `chicago` coefficients are both zero. Three significant figure accuracy is fine, and you will get full credit if you set up a correct numerical matrix expression for the statistic, even if you don't carry out all the matrix multiplications, inversions, etc. Explain how to determine whether the hypothesis is rejected at the 95% level by a frequentist test. Explain with a hand sketch what region of the space of coefficients on the two tested variables has probability $p_{\text{chisq}}(s, \text{df})$, where s is your statistic, df is the appropriate degrees of freedom for the chi-squared distribution, and

`pchisq()` is a function delivering the probability of a chi-squared variable being less than its first argument (`s` in the case at hand).

The test statistic is

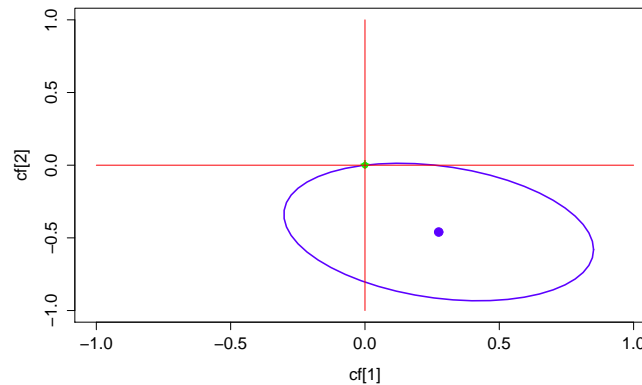
$$(0.2745, -0.4599) \begin{pmatrix} .01781 & -.003747 \\ -.003747 & .01200 \end{pmatrix}^{-1} \begin{pmatrix} 0.2745 \\ -0.4599 \end{pmatrix}.$$

The inverse of the matrix in the center is

$$\frac{1}{.01781 \cdot .012 - .003747^2} \begin{pmatrix} .012 & .003747 \\ .003747 & .01781 \end{pmatrix} = \begin{pmatrix} 60.09616 & 18.76502 \\ 18.76502 & 89.19271 \end{pmatrix}$$

And the test statistic evaluates to 18.655, which, being distributed as $\chi^2(2)$ under the null hypothesis, rejects the null at any reasonable significance level. The probability of a $\chi^2(2)$ variable exceeding this level is .0001.

The highest posterior density region of the parameter space for these two coefficients is an ellipse centered at the coefficient estimates and slightly negatively sloped (because the coefficients have negative covariance). The ellipse with probability .9999 corresponding to our test statistic passes through the null hypothesis point (0,0). A plot of it (more precise than you were required to produce) is below. Note that it includes some negative values of the coefficient on `female` (as we might have expected because of its smaller t statistic) but scarcely any positive values of the `chicago` coefficient.



(3) 20 minutes. We would like to estimate the coefficients in the equation

$$(1) \quad y_j = \beta_0 + \beta_1 x_j + \varepsilon_j.$$

The data are i.i.d. and ε_j has mean zero, but we do not actually have data on x . We have instead two error-ridden proxies for x

$$(2) \quad z_j = x_j + v_j$$

$$(3) \quad w_j = x_j + \zeta_j.$$

We are willing to assume that all the data are jointly i.i.d. across j , that v_j , ζ_j and ε_j all have zero mean conditional on x_j , and that v_j , ζ_j , and ε_j are uncorrelated with one another and have constant variance across j .

(a) Can we obtain consistent estimates of β_0 and β_1 by replacing x_j with z_j in the equation and then using w_j as an instrument for z_j ? Would the results be different if we did the reverse, using z_j as an instrument for w_j ?

The original equation implies that when we replace x_j by z_j we get

$$y_j = \beta_0 + \beta_1 z_j - \beta_1 v_j + \varepsilon_j = \beta_0 + \beta_1 z_j + \zeta_j.$$

The error term in this equation, $\zeta_j = -\beta_1 v_j + \varepsilon_j$, is clearly, under our assumptions, correlated with z_j , so OLS will not work. However, w_j is correlated with z_j because of their common dependence on x_j , and neither of w_j 's two components, x_j and ξ_j is correlated with ζ_j , so the model meets the basic requirements for validity of instrumental variables. The same arguments with the roles of z and w reversed imply that we could use z_j as an instrument for w_j instead. The two estimates would not be the same, though, since they are

$$(Z'W)^{-1}Z'Y \quad \text{and} \quad (W'Z)^{-1}W'Y,$$

where Z is the matrix consisting of the constant vector and the z_j 's and W is the matrix consisting of the constant vector and the w_j 's. The asymptotic covariance matrix would not even match between the two (though you were not asked to check that).

- (b) We could get a more accurate measure of x_j by taking an average of w_j and z_j . What about replacing x_j with that average and using z_j as an instrument for it? Would that give different results? Better results?

It would certainly give different results, and not better results. The residual once x_j was replaced by $(z_j + w_j)/2$ would depend on the noise terms in both z and w , so neither z nor w would be usable as an instrument.

- (4) 10 minutes. Suppose we have i.i.d. data on $\{y_j, x_j\}$ for $j = 1, \dots, n$ and wish to estimate the regression (with no constant term)

$$(4) \quad y_j = \alpha x_j + \varepsilon_j.$$

Suppose we know that $\varepsilon_j \mid x_j \sim N(0, \sigma^2 x_j^2)$, where σ^2 is an unknown parameter. Explain why ordinary least squares estimation is inefficient in this case and explain how to implement a more efficient estimator.

Ordinary least squares is maximum likelihood and unbiased when the errors are homoskedastic — i.e. $\varepsilon_j \sim N(0, \sigma^2)$ with the same σ^2 value for all j . Here the error variance varies, so GLS gives a lower variance. The weighted least squares version of GLS would tell us here to divide the data (y_j, x_j) for each j by x_j , then apply least squares to the transformed data. This results in the equation

$$\frac{y_j}{x_j} = \alpha + v_j,$$

where $v_j = \varepsilon_j/x_j$ and $v_j \sim N(0, \sigma^2)$. Applying least squares to this homoskedastic equation is just taking the sample mean of y_j/x_j .