Model specification

Consider the model

(1) $$\underset{\sim}{y} = X\underset{\sim}{\beta} + Z\underset{\sim}{\gamma} + \underset{\sim}{u} \qquad \underset{\sim}{u} \text{ iid } (\underset{\sim}{0}, \sigma^2 I_n)$$

y $n \times 1$ $\quad X_{n \times k_1}$ $\quad \beta_{k_1 \times 1}$

$\underset{\sim}{u}$ $n \times 1$ $\quad Z_{n \times k_2}$ $\quad \gamma_{k_2 \times 2}$

If the actual DGP was such that $\underset{\sim}{\gamma} = 0$ then this model is overspecified. The Actual DGP Being

(2) $$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{u}$$

OLS estimation of (1) when (2) is the true DGP is unbiased, But not efficient. unless X & Z are orthogonal.

Covariance of Overspec. Model.

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + Z\underset{\sim}{\gamma} + \underset{\sim}{u}$$

$$M_z = I - Z(Z^TZ)^{-1}Z^T$$
$$= I - P_z$$

FWL

$$M_z y = M_z X\beta + M_z Z\gamma + M_z \underset{\sim}{u}$$
$$= M_z X\beta + 0 + M_z \underset{\sim}{u}.$$

OLS yields

$$\hat{\beta} = (X^T M_z^T M_z X)^{-1} X^T M_z^T M_z \underset{\sim}{y}$$
$$= (X^T M_z X)^{-1} X^T M_z \underset{\sim}{y}$$

$$\mathrm{Cov}(\hat{\beta}) = \sigma^2 (X^T M_z X)^{-1}$$

$$b = (X^T X)^{-1} X^T y$$

correct ESTIMATOR for the overspec Model.

$$Cov(b) = \sigma^2 (X^T X)^{-1}$$

b is more efficient than $\tilde{\beta}$

$\frac{1}{4} - \frac{1}{8} = \frac{1}{8} > 0$ Provided $Cov(\tilde{\beta}) - Cov(b) = \Delta$ p.s.d.

$8 - 4 = 4 > 0$ This equivalent to

$$[Cov(b)]^{-1} - [Cov(\tilde{\beta})]^{-1} = \Delta' \text{ p.s.d.}$$

$$Cov(\tilde{\beta}) - Cov(b) = \sigma^2 (X^T M_z X)^{-1} - \sigma^2 (X^T X)^{-1}$$

Invert and change signs

$$\frac{(X^T X)^{+1}}{\sigma^2} - \frac{X^T M_z X}{\sigma^2}$$

$$= \frac{1}{\sigma^2}\left(X^T (I - M_z) X \right) = \frac{1}{\sigma^2}\left(X^T P_z X \right)$$

$$= \frac{1}{\sigma^2}\left(X^T P_z^T P_z X \right) = \frac{1}{\sigma^2} C^T C \quad \text{is p.s.d.}$$

A more serious error occurs if you estimate (2) when (1) is the true DGP. In this case you are omitting relevant variables and your model is <u>underspecified</u>.

LS applied to (1)

$$
\begin{aligned}
b &= (X^TX)^{-1}X^Ty \\
&= (X^TX)^{-1}X^T(X\beta + Z\gamma + u) \\
&= \beta + (X^TX)^{-1}X^TZ\gamma + (X^TX)^{-1}X^Tu
\end{aligned}
$$

$\cancel{E(b|X) = \beta +}$

$$
\begin{aligned}
E(b \mid X, Z) &= \beta + (X^TX)^{-1}X^TZ\gamma + E(X^TX)^{-1}X^Tu \\
&= \beta + (X^TX)^{-1}X^TZ\gamma \neq \beta
\end{aligned}
$$

unless $X^TZ = 0$ or $\gamma = 0$

True Model

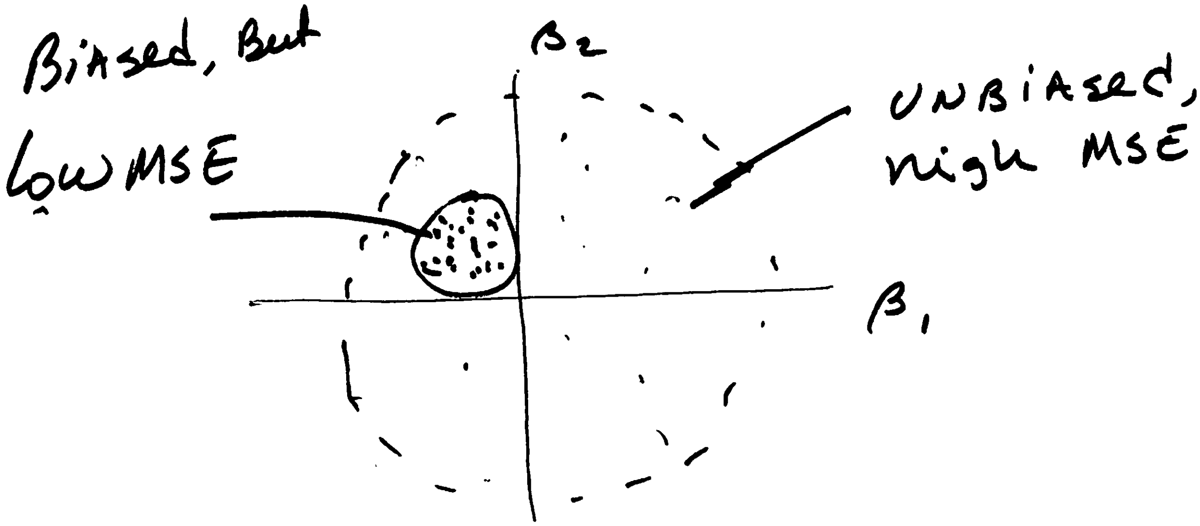
Est Model	$y = X\beta + u$	$y = X\beta + Z\gamma + u$
$y = X\beta + u$	OLS BLUE	OLS Biased, but more precise.
$y = X\beta + Z\gamma + u$	OLS unbiased, not efficient.	OLS BLUE

$$MSE :\equiv E\left[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)^T\right]$$

$$= \text{Bias} \cdot \text{Bias}^T + \text{Cov}(\tilde{\beta})$$

$$\text{Bias} = \cancel{E(\tilde{\beta} - \beta)} \quad E(\tilde{\underset{\sim}{\beta}}) - \underset{\sim}{\beta}$$

~~Cov~~ The increase in precision due to underspec. could outweigh the bias introduced.


Biased, But
Low MSE
β_2
Unbiased,
High MSE
β_1

In the underspecified Model

$\hat{\sigma}_1^2$ is Biased for σ^2

$$\hat{\sigma}_1^2 = \frac{\hat{u}_1^T \hat{u}_1}{n-k_1}$$

with $\hat{u}_1 = y - X b$

$b = (X^T X)^{-1} X^T y$

$$\begin{aligned} \hat{u}_1 &= y - X(X^T X)^{-1} X^T y = M_X y \\ &= ~~y~~ + \; M_X (X\beta + Z\gamma + u) \\ &= M_X(Z\gamma + u) = M_X Z\gamma + M_X u. \end{aligned}$$

$$E\left(\frac{\hat{u}_1^T \hat{u}_1}{n-k_1}\right) = E\left(\frac{(M_X Z\gamma + M_X u)^T (M_X Z\gamma + M_X u)}{n-k_1}\right)$$

$$= \frac{1}{n-k_1} E \Big(\gamma^T Z^T M_x^T M_x Z \gamma + u^T M_x^T M_x Z \gamma$$

$$+ u^T M_x^T M_x u + \gamma^T Z^T M_x^T M_x u \Big)$$

use tr & E( ) trick

$$= \frac{1}{n-k_1} E \left[\cancel{\mathrm{tr}} (\gamma^T Z^T M_x Z \gamma) + \mathrm{tr}(u^T M_x u) \right]$$

since $E(\gamma^T Z^T M_x u) = 0$

conditional on X & Z.

$$= \frac{1}{n-k_1} \cancel{\mathrm{tr}} \cancel{[} \cancel{\mathrm{tr}} \; \cancel{\mathrm{tr}} (\gamma^T Z^T M_x Z \gamma)$$

$$+ \frac{1}{n-k_1} \mathrm{tr}\, E (M_x u u^T)$$

$$= \frac{1}{n-k_1} \cancel{\mathrm{tr}} (\gamma^T Z^T M_x Z \gamma) + \frac{1}{n-k_1} \sigma^2 (n-k_1)$$

$$= \sigma^2 + \frac{\gamma^T Z^T M_x Z \gamma}{n-k_1} \geq \sigma^2$$

unless $M_x Z = 0$

But $M_x X = 0$ ∴

unless $X = Z$

Ad hoc Variable Selection

$$R_u^2 = 1 - \frac{\hat{u}^T\hat{u}}{y^T y}$$

or

$$R_c^2 = 1 - \frac{\hat{u}^T\hat{u}}{(\underset{\sim}{y} - \underset{\sim}{i}\bar{y})^T(\underset{\sim}{y} - \underset{\sim}{i}\bar{y})}$$

where $\hat{u} = y - X\hat{\beta}$ $\quad \hat{\beta} = (X^T X)^{-1} X^T y$

or $= M_X \underset{\sim}{y}$

One problem is that $\hat{u}^T\hat{u} = SSR$ can always be made smaller by adding add'l regressors.

$$\bar{R}_c^2 = 1 - \frac{n-1}{n-k}(1 - R_c^2)$$

$=$

Replacing $\hat{u}^T\hat{u}$ with an unbiased estimate $E(\hat{u}^T\hat{u}) = (n-k)\sigma^2$

$$\bar{R}^2 = 1 - \frac{\frac{1}{n-k}(\hat{u}^T\hat{u})}{\frac{1}{n-1}(\underset{\sim}{y} - \underset{\sim}{i}\bar{y})^T(\underset{\sim}{y} - \underset{\sim}{i}\bar{y})}$$

$$= 1 - \frac{n-1}{n-k} \cdot \frac{\hat{u}^T\hat{u}}{(\underset{\sim}{y} - \underset{\sim}{i}\bar{y})^T(\underset{\sim}{y} - \underset{\sim}{i}\bar{y})}$$

$0 \not\leq \bar{R}^2 \leq 1$ can be negative.

Others

Amemiya suggests

$$PC_j = \frac{\hat{u}_j^T \hat{u}_j}{n - k_j} \left(1 + \frac{k_j}{n} \right)$$

where $\hat{u}_j$ are residuals from j^{th} regression and k_j is # of regressors included. — minimize PC_j

$$AIC_j = \ln\left(\frac{\hat{u}_j^T \hat{u}_j}{n} \right) + \frac{2 k_j}{n}$$

minimize AIC_j

$$SC_j = \ln\left(\frac{\hat{u}_j^T \hat{u}_j}{n} \right) + k_j \frac{\ln(n)}{n}$$

AND dozens of others

NOTE SAS versions of AIC and SC are multiplied by n.

Take a R.S. of size n

$X = \{X_1, X_2, \ldots, X_n\}$ that has a joint p.d.f. $f(\underset{\sim}{x} \mid \underset{\sim}{\theta})$. The set of all possible values of X define the <u>sample space</u>. A test divides the sample space into two parts. One part contains a set of values that will lead you to believe $\theta \in \Omega_0$ H_0:
and the other " " $\theta \in \Omega_1$. H_1.

The subset that leads to the conclusion $\theta \in \Omega_1$ is called the <u>Rejection Region</u> (you are rejecting that $\theta \in \Omega_0$) of the test

In practice, A test statistic is computed using the sample. This stat maps the n dimensional sample into a single R.V. The test stat has a known dist when H_0: true. The rejection region (critical region) is the set of values that causes you to reject H_0:

Elements of A Test

(1) Null & Alternative hypotheses.

Null is maintained until enough evidence to its contrary is found.

· Alternative – what to conclude if Ho rejected or is NOT True.

(2) Test stat with known dist when Ho: True

(3) Choose an acceptable % of time that Ho can be rejected when it is act. true. "α"

(4) Decision rule.

Example:

$Y_1, Y_2, \ldots Y_n$ is a R.S. of size n

from ($Y \sim N(\beta, \sigma^2)$

$$H_0: \beta = 10$$
$$H_A\ \beta \neq 10$$

use sample mean to est β. (LS)

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \text{mean.}$$

$$\tilde{\beta} \overset{a}{\sim} N(\beta, \sigma^2/n) \qquad \text{its dist.}$$

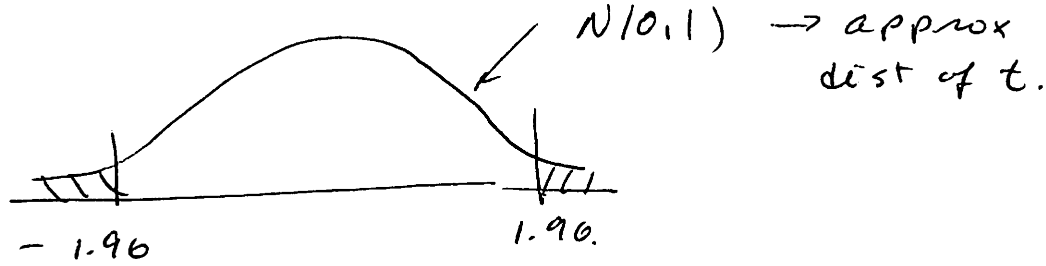
$$\frac{\tilde{\beta} - \beta}{\sqrt{\sigma^2/n}} \sim N(0,1) \qquad \text{standardize.}$$

let $s^2 \xrightarrow{P} \sigma^2$. Replacing σ^2 with s^2 leads to

$$t = \frac{\tilde{\beta} - \beta}{\sqrt{s^2/n}} \overset{a}{\sim} N(0,1)$$

If Ho True Then $t = \dfrac{\tilde{\beta} - 10}{S/\sqrt{10}} \overset{a}{\sim} N(0,1)$

let $\alpha = 5\%$



If $|t| > 1.96$ reject H_0.

NOTES: – α is the significance level of the test. Size of the rejection region. Prob of Type I error.

– Rejection of a True Null – Type I
– NOT reject when H_1 True – Type II
Failure to reject false hypoth.

– When the exact probability of Type I error is known, then we call the test exact.

– In our example, the exact dist of $\hat{\beta}$ was unknown – only its asy approx was available.

– Approximate a asympt. test.

In this case, α is called the <u>nominal</u> (significance) <u>level</u> of the test. The actual prob of type I error is likely different from this nominal level.

- <u>Power</u> of a test describes the prob of rejecting the null.

Consider a test

$X_1 \sim N(0,1)$ or $X_* \sim N(\lambda,1)$

$H_0: \lambda = 0$

$H_A: \lambda = \cancel{2}\ 1.96$

When $\lambda = 0$, $\text{Prob}(X \geq \overset{1.96}{2}) \doteq .025$

So, Prob of type I error $\doteq .025$

When $\lambda = 0$

When $\lambda = 0$, the Power of test is

Prob of Rejecting — also $\doteq .025$

When $\lambda = 0$, Prob of Type II error?

(only comes into play when H_1 true so ignore)

When $\lambda = \overset{1.96}{2}$, (unbeknownst to us).

$\text{Prob}(X \geq \overset{1.96}{2}) = .5$

Prob of Type I error $\doteq .025$

Prob of Type II error $\doteq .5$

Power of Test $= .5$ $= 1 - \text{Pr}(\text{Type II})$

As $\lambda \to \infty$ i) Prob type II error falls

ii) Power of Test rises

If truth is close to null, test has <u>low power</u>.

Also, note: As $\alpha \to 0$ Prob type II error $\to 1$

And $\alpha \to 1$ " " $\to 0$

Distributions Related to The Normal.

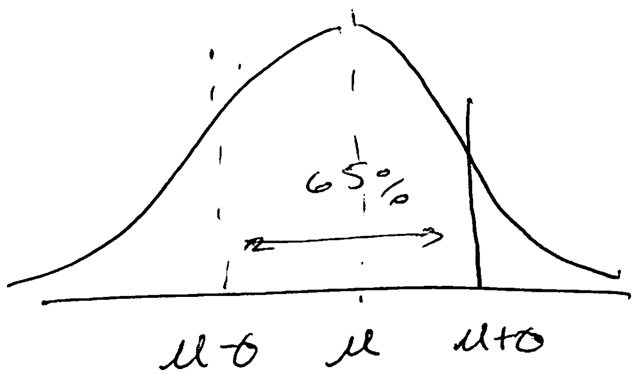
About 90% of hypoth testing & C.I. in econometrics is Based on Normal Dist Theory.
That leaves 6% for Simulation Based methods 4% other-
(I made these % up!)

A continuous r.v. has a normal p.d.f. with mean, μ, and var, σ^2,
denoted $X \sim N(\mu, \sigma^2)$ $\quad -\infty < \mu < \infty;\ \sigma^2 > 0$

if

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

$$-\infty < x < \infty$$



Mean μ.

Median μ.

Symmetric about μ.

Inflection at $\mu \pm \sigma$

When $\mu = 0,\ \sigma^2 = 1 \Rightarrow$ std normal

$X \sim N(0,1)$

$$f(x) \equiv \phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}x^2\right\}$$

<u>Property of the Normal</u> Let $X \sim N(\mu, \sigma^2)$ and a & b const

$Y = a + bX$

$Y \sim N(a + b\mu,\ b^2\sigma^2)$

Multivariate Normal

A continuous $n \times 1$ random vector $\underset{\sim}{X}$ has a MVN pdf with mean $\underset{\sim}{\mu}$ and $Var(X) = \Sigma$, denoted

$\underset{\sim}{X} \sim N(\underset{\sim}{\mu}, \Sigma)$ if X has

pdf.

$$g(\underset{\sim}{x} \mid \underset{\sim}{\mu}, \Sigma) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} (\underset{\sim}{x} - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{x} - \mu) \right\}$$

$\forall \; x \in R^n$

$$\underset{\sim}{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \underset{\sim}{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ & & & \\ \sigma_{n1} & \cdot & \cdot & \sigma_{nn} \end{pmatrix}$$

wh $\sigma_{ij} = Cov(X_i, X_j)$

$\sigma_{ii} = Var(X_i)$

<u>Property</u> "D Theorem" Linear trans of MVN

If $\underset{\sim}{X} \sim N(\underset{\sim}{\mu}, \Sigma)$ and D is $m \times n$ matrix of constants with $m \le n$

$$D\underset{\sim}{X} \sim N(D\underset{\sim}{\mu}, D\Sigma D^T)$$

<u>Other Properties</u>

Let $X_{n\times 1} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \begin{matrix} m\times 1 \\ n-m\times 1 \end{matrix}$ $\quad \underset{\sim}{\mu} = \begin{pmatrix} \underset{\sim}{\mu}_1 \\ \underset{\sim}{\mu}_2 \end{pmatrix}$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \qquad \begin{matrix} \Sigma_{11} \; m\times m \\ \Sigma_{22} \; (n-m)\times(n-m) \end{matrix}$$

Conditional dist of $X_1 | X_2$

$$\sim N\left(\underset{\sim}{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2-\mu_2), \; \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

The marginal dist of X_1 can be obtained using the "D" Thm.

$$D = (I_m \,\vdots\, 0)$$

$$DX \sim N(D\mu, D\Sigma D^T)$$

$$X_1 \sim N(\mu_1, \Sigma_{11})$$

NOTE:

Suppose $\Sigma_{12} = \Sigma_{21} = 0 \Rightarrow X_1$ & X_2 uncorr.

Note: $X_1 | X_2 \sim N(\mu_1 + 0, \; \Sigma_{11} + 0)$
$\sim N(\mu_1, \Sigma_{11}) =$

$X_1 | X_2$ has same dist as X_1

$\Rightarrow$ independent.

Chi-Square

If $\underset{\sim}{X}_{n \times 1} \sim N(\hat{\mu}, \Sigma)$ then

$$Y = (\underset{\sim}{X} - \underset{\sim}{\mu})^T \Sigma^{-1} (\underset{\sim}{X} - \underset{\sim}{\mu}) \sim \chi^2_n$$

univariate $\chi^2_1 = z^2$ where $z \sim N(0,1)$

$E(\chi^2_1) = 1$

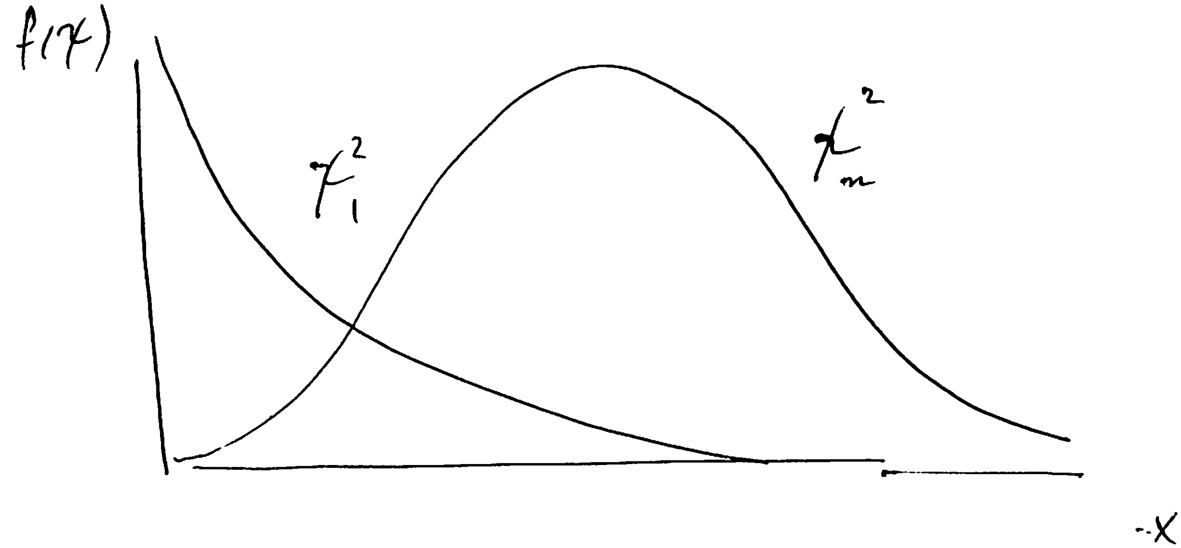
$Var(\chi^2_1) = 1$

$\chi^2_n + \chi^2_m \sim \chi^2_{n+m}$

i.e. $\chi^2_n = \sum_{i=1}^{n} \chi^2_1$ sum of indep χ^2_1

In general $E(\chi^2_n) = n$

$Var(\chi^2_n) = 2n.$



If M is a projection matrix with rank r and ~~$X \sim N(0, I)$~~ ~~$X \sim N(0, \Sigma)$~~) $X \sim N(0, \sigma^2 I_n)$ then

$$\frac{X^T M X}{\sigma^2} \sim \chi^2_r$$

t-distn

$$t = \frac{N(0,1)}{\sqrt{\chi^2_n / n}} \sim t_n$$

Provided the $N(0,1)$ and χ^2_n R.V.s are independent.

F-distn

$$F_{n,m} = \frac{\chi^2_n / n}{\chi^2_m / m}$$

where the 2 chi-square vars are indep.

$$n F_{n,m} \xrightarrow[m \to \infty]{} \chi^2_n$$

$$F_{n,m} \approx \chi^2_n / n$$