SU Ekonometrics I Spring 2024

Problem Set 2: Marek Chadim[1]

---

1. (a)

$$\begin{aligned}
\epsilon'\epsilon &= (Y - X\beta)'(Y - X\beta) \\
\epsilon'\epsilon &= Y'Y - 2Y'X\beta + \beta'X'X\beta \\
\frac{\partial(\epsilon'\epsilon)}{\partial\beta} &= -2X'Y + 2X'X\beta = 0 \\
\hat{\beta}_{LS} &= (X'X)^{-1}X'Y
\end{aligned}$$

(b)

$$\begin{aligned}
\ln L(\beta; Y, X) &= -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i\beta)^2 \\
\frac{\partial \ln L(\beta; Y, X)}{\partial\beta} &= \frac{1}{\sigma^2}X'(Y - X\beta) = 0 \\
\hat{\beta}_{MLE} &= (X'X)^{-1}X'Y
\end{aligned}$$

(c)

$$\pi(\beta) = \frac{1}{\sqrt{2\pi\tau^2}}\exp\left(-\frac{(\beta-\theta)^2}{2\tau^2}\right), \; p(\beta|Y, X) \propto L(\beta; Y, X) \times \pi(\beta)$$

In general, the normal conjugate prior allows for expressing the posterior mean as $\frac{1}{\frac{1}{\sigma_0^2}+\frac{n}{\sigma^2}}\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^{n} x_i}{\sigma^2}\right)$. Specifically, since $Var(\hat{\beta}_{MLE}) = \sigma^2(X'X)^{-1}$, mean of the posterior distribution of β is a precision weighted linear combination of the form:

$$\hat{\beta}_{\text{post}} = \left(\frac{1}{\tau^2} + \frac{1}{\sigma^2(X'X)^{-1}}\right)^{-1}\left(\frac{1}{\tau^2}\theta + \frac{1}{\sigma^2(X'X)^{-1}}\hat{\beta}_{MLE}\right)$$

(d) As a $\to \infty$, the prior becomes non-informative. The posterior mode is $\hat{\beta}_{MLE}$ since the difference between the Bayesian posterior and the likelihood function is the prior.

(e)

$$\begin{aligned}
E(X'(Y - X\beta)) &= 0 \\
E(X'Y) &= E(X'X\beta) \\
X'Y &= \beta X'X \\
\hat{\beta}_{MM} &= (X'X)^{-1}X'Y
\end{aligned}$$

2. (a) i.

$$L(p) = \left[f_X(19,\lambda)\cdot\left((1-p) + \frac{p}{2}\right)\right]^{n_{19}} \cdot \left[f_X(20,\lambda) + f_X(19,\lambda)\cdot\frac{p}{2}\right]^{n_{20}}$$

where:

$f_X(x, \lambda)$ denotes the Poisson probability of scoring x before any re-review,
p is the probability that a test scored at 19 is sent for re-review,
n_x, n_{20} are the number of students observed with scores of 19 and 20, respectively.

---

[1] 42624@student.hhs.se

The corresponding log-likelihood function is:

$$\log L(p) = n_{19} \log\left(f_X(19,\lambda)\cdot\left((1-p)+\frac{p}{2}\right)\right) + n_{20}\log\left(f_X(20,\lambda)+f_X(19,\lambda)\cdot\frac{p}{2}\right)$$

ii.

$$\begin{aligned}\frac{d}{dp}\log L(p) &= n_{19}\cdot\frac{d}{dp}\left[\log\left((1-p)+\frac{p}{2}\right)\right] + n_{20}\cdot\frac{d}{dp}\left[\log\left(f_X(20,\lambda)+f_X(19,\lambda)\cdot\frac{p}{2}\right)\right]\\ &= n_{19}\cdot\frac{1}{p-2} + n_{20}\cdot\frac{f_X(19,\lambda)}{f_X(19,\lambda)p+2f_X(20,\lambda)}\end{aligned}$$

iii. Setting the score function to zero and solving for p

$$\hat{p} = \frac{2(n_{20}f_X(19) - n_{19}f_X(19))}{f_X(19,\lambda)(n_{20}+n_{19})}$$

iv.

$$H(\hat{p}) = -E\left[\frac{d^2}{dp^2}\log L(p) = -\frac{n_{19}}{(p-2)^2} - \frac{n_{20}\cdot[f_X(19,\lambda)]^2}{[f_X(19,\lambda)p+2f_X(20,\lambda)]^2}\right]$$

v. When $\lambda = 19$ the derivative $\frac{dH(\hat{p}}{d\lambda}$ is zero due to the factor coming from the exponent of $f_X(19;\lambda)$. The number of students scoring 19 is maximized, which in turn maximizes the number of draws the re-review process. This provides the most information about p, the propensity to inflate grades, making the test design most informative about teachers' grading behavior.

3. (a) Using the properties: summation of gamma random variables is also gamma distributed, inverse of a gamma random variable is inverse-gamma distributed:

$$\ln(L(\lambda,X)) = n\ln(\lambda) - \lambda\sum_{i=1}^{n}x_i$$

$$\frac{d}{d\lambda}\ln(L(\lambda,X)) = \frac{n}{\lambda} - \sum_{i=1}^{n}x_i = 0 \Rightarrow \hat{\lambda} = \frac{n}{\sum_{i=1}^{n}x_i}$$

$$B(\hat{\lambda}) = E\left(\frac{n}{\sum_{i=1}^{n}x_i}\right) - \lambda = \frac{\lambda n}{n-1} = \frac{1}{n-1}\lambda,\ \lim_{n\to\infty}B(\hat{\lambda}) = 0$$

(b)

$$\ln L(\mu,\sigma^2\,|\,x_1,x_2,\ldots,x_n) = \sum_{i=1}^{n}\left(-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

$$\frac{d}{d\mu}\ln L(\mu,\sigma^2\,|\,x_1,x_2,\ldots,x_n) = \sum_{i=1}^{n}\frac{1}{\sigma^2}(x_i-\mu)$$

$$\sum_{i=1}^{n}\frac{1}{\sigma^2}(x_i-\hat{\mu}) = 0 \Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}x_i$$

$$\text{WLLN: }\lim_{n\to\infty} = P(|\hat{\mu} - \frac{1}{\lambda}| > \epsilon) = 0,\ \forall\epsilon > 0$$

(c)

$$\frac{d}{d(\sigma^2)}\ln L(\mu,\sigma^2\,|\,x_1,x_2,\ldots,x_n) = \sum_{i=1}^{n}\left(-\frac{1}{2\sigma^2} + \frac{(x_i-\mu)^2}{2(\sigma^2)^2}\right)$$

$$-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2$$

$$E(\hat{\sigma}^2) = \sigma_x^2 - \left(\frac{1}{N}\right)^2 \text{Var}\left(\sum_{n=1}^{N} x_n\right) = \sigma_x^2 - \left(\frac{1}{N}\right)^2 N\sigma_x^2 = \frac{N-1}{N}\sigma_x^2$$

$$\lim_{n\to\infty} P\left(\left|\hat{\sigma}^2 - \frac{1}{\lambda^2}\right| > \epsilon\right) = 0, \quad \forall \epsilon > 0.$$

(d) Denote $I(x_i < \hat{\mu})$ an indicator function that equals 1 if $x_i < \hat{\mu}$ and 0 otherwise:

$$\hat{P}_n = \frac{1}{n}\sum_{i=1}^{n} I(x_i < \hat{\mu}), \quad \text{WLLN: } \hat{P}_n \xrightarrow{pr} E[I(X_i < \mu)] = P(X_i < \mu) = F(\mu)$$

An estimate significantly different from 0.5 would suggests that the distribution is not symmetric about the mean, indicating skewness or other forms of non-normality.

(e) The Exponential estimators generally have lower biases and MSEs compared to the Normal MLE estimator, especially for smaller values of $\tilde{x}$. The share method demonstrated low bias but higher variance. This approach, while non-parametric and straightforward, may be more sensitive to sample size and sample specifics.

4. (a)

$$c_t^{\gamma-1} = E_t\left[\beta c_{t+1}^{\gamma-1} r_{t+1}\right]$$

(b)

$$E\left[c_t^{\gamma-1} - \beta c_{t+1}^{\gamma-1} r_{t+1}\right] = 0$$

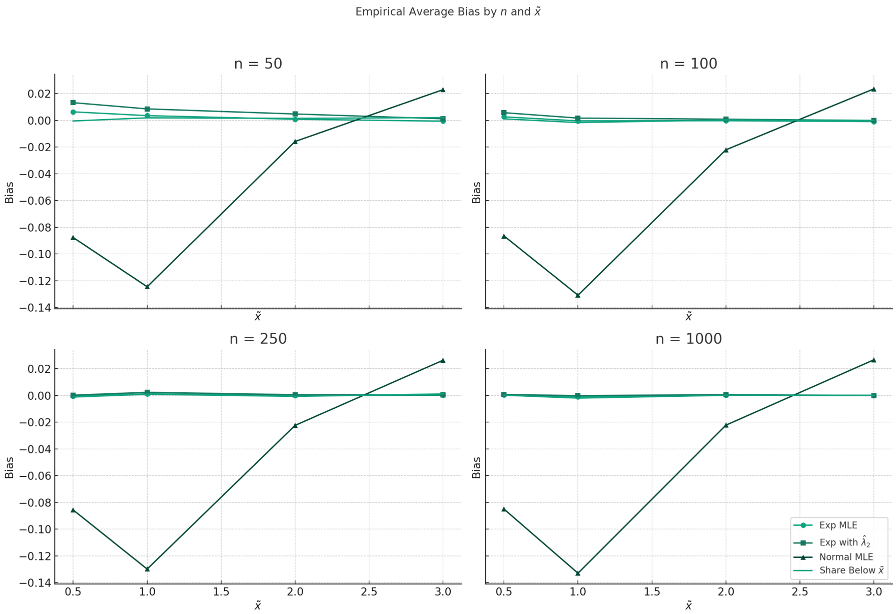
(c)

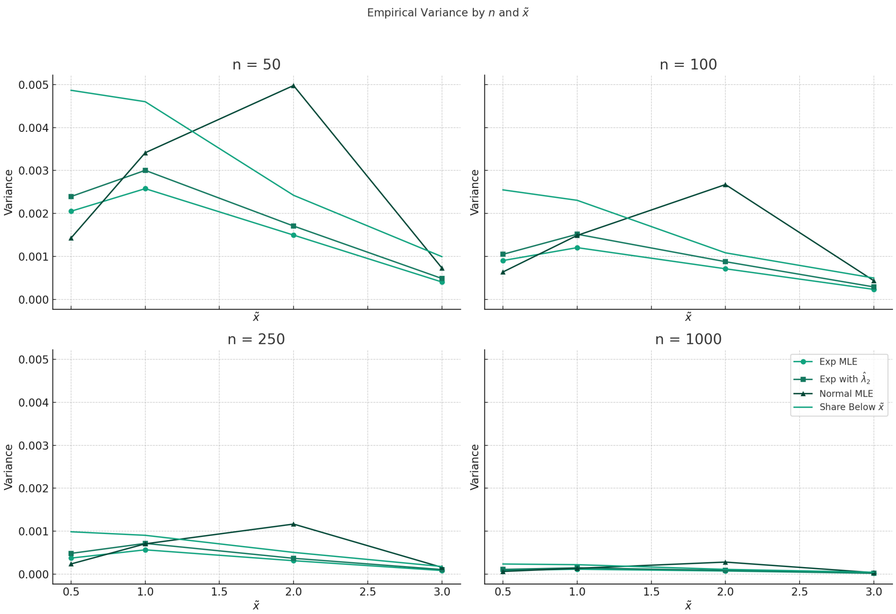
$$E_t\left[\left(c_t^{\gamma-1} - \beta c_{t+1}^{\gamma-1} r_{t+1}\right) r_{t+1}\right] = 0$$

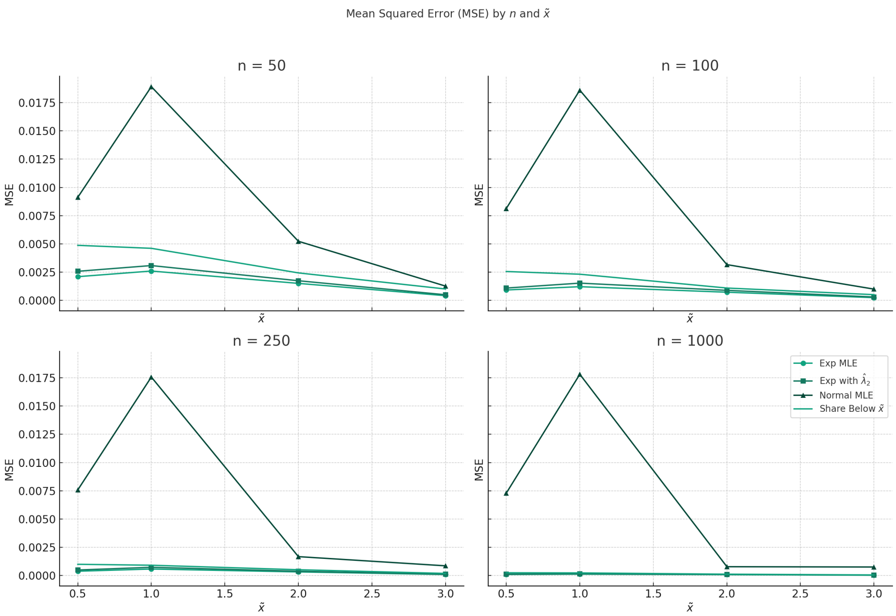
(d)

$$\begin{pmatrix} -c_{t+1}^{\gamma-1} r_{t+1} & c_t^{\gamma-1}\ln(c_t) - \beta c_{t+1}^{\gamma-1}\ln(c_{t+1}) r_{t+1} \\ -c_{t+1}^{\gamma-1} r_{t+1}^2 & (c_t^{\gamma-1}\ln(c_t) - \beta c_{t+1}^{\gamma-1}\ln(c_{t+1}) r_{t+1}) r_{t+1} \end{pmatrix}$$

(e) below


Empirical Average Bias by n and $\tilde{x}$
n = 50
n = 100
n = 250
n = 1000
Bias
$\tilde{x}$
Exp MLE
Exp with $\hat{\lambda}_2$
Normal MLE
Share Below $\tilde{x}$


Empirical Variance by n and $\tilde{x}$
n = 50
n = 100
n = 250
n = 1000
Variance
$\tilde{x}$
Exp MLE
Exp with $\hat{\lambda}_2$
Normal MLE
Share Below $\tilde{x}$


Mean Squared Error (MSE) by n and $\tilde{x}$
n = 50
n = 100
n = 250
n = 1000
MSE
$\tilde{x}$
Exp MLE
Exp with $\hat{\lambda}_2$
Normal MLE
Share Below $\tilde{x}$

```r
data <- read.csv("gmmdata.csv")
head(data)
```

```
##   time        c         r
## 1    1 1.000000 1.0058620
## 2    2 1.023381 1.0285820
## 3    3 1.052478 1.0235300
## 4    4 1.080480 1.0253520
## 5    5 1.137604 0.9997738
## 6    6 1.173770 1.0201980
```

```r
momentConditions <- function(theta, data) {
  beta <- theta[1]
  gamma <- theta[2]
  c_t <- data$c[1:(nrow(data) - 1)]  # Consumption at time t
  c_t1 <- data$c[2:nrow(data)]       # Consumption at time t+1
  r_t1 <- data$r[2:nrow(data)]       # Returns at time t+1
  g1 <- c_t^(gamma - 1) - beta * c_t1^(gamma - 1) * r_t1
  g2 <- r_t1 * (c_t^(gamma - 1) - beta * c_t1^(gamma - 1) * r_t1)
  g <- cbind(g1, g2)
  return(g)
}

initParams <- c(beta = 1, gamma = 3)

gmmResult <- gmm(g = momentConditions, x = data, t0 = initParams,
               vcov = 'iid', method = 'Nelder-Mead',
               control = list(reltol = 1e-25, maxit = 10000))

print(summary(gmmResult))
```

```
##
## Call:
## gmm(g = momentConditions, x = data, t0 = initParams, vcov = "iid",
##     method = "Nelder-Mead", control = list(reltol = 1e-25, maxit = 10000))
##
##
## Method:  twoStep
##
## Coefficients:
##        Estimate    Std. Error  t value     Pr(>|t|)
## beta   9.5000e-01  5.3459e-08  1.7771e+07  0.0000e+00
## gamma  2.0000e+00  2.1023e-06  9.5133e+05  0.0000e+00
##
## J-Test: degrees of freedom is 0
##                 J-test                P-value
## Test E(g)=0:    5.32190980560944e-24  *******
##
## #############
## Information related to the numerical optimization
## Convergence code =  0
## Function eval. =  253
## Gradian eval. =  NA
```

1