ECONOMETRICS FIELD EXAM, SUMMER 2000, PART 1

ECONOMICS DEPARTMENT, UNIVERSITY OF UTAH

Part 1 of this exam (Hans Ehrbar) has three subparts a, b, and c. Part 2 will be provided by Peter Philips.

- For part 1a answer either Problem 1 or Problem 2 or Problem 3 (simple properties of conditional expectation).
- For part 1b you either have to answer Problem 4 or 5 (derivation of the Best Linear Unbiased estimator in certain simple situations).
- For part 1c you have to do either Problem 6 or Problem 7 (derivation of the formulas and properties of certain F-tests).

Date of exam August 2000.

Problem 1. *Let x and y be two jointly distributed variables. For every fixed value x, $\text{var}[y|x = x]$ is the variance of y under the conditional distribution, and $\text{var}[y|x]$ is this variance as a random variable, namely, as a function of x.*

• **a.** *1 point Prove that*

$$\text{var}[y|x] = \text{E}[y^2|x] - (\text{E}[y|x])^2. \tag{1}$$

This is a very simple proof. Explain exactly what, if anything, needs to be done to prove it.

Answer. For every *fixed* value x, it is an instance of the law

$$\text{var}[y] = \text{E}[y^2] - (\text{E}[y])^2 \tag{2}$$

applied to the conditional density given $x = x$. And since it is true for every fixed x, it is also true after plugging in the random variable x. □

• **b.** *3 points Prove that*

$$\text{var}[y] = \text{var}\big[\text{E}[y|x]\big] + \text{E}\big[\text{var}[y|x]\big], \tag{3}$$

i.e., the variance consists of two components: the variance of the conditional mean and the mean of the conditional variances. This decomposition of the variance is given e.g. in [Rao73, p. 97] *or* [Ame94, theorem 4.4.2 on p. 78].

Answer. The first term on the rhs is $\mathrm{E}[(\mathrm{E}[y|x])^2] - (\mathrm{E}[\mathrm{E}[y|x]])^2$, and the second term, due to (1), becomes $\mathrm{E}[\mathrm{E}[y^2|x]] - \mathrm{E}[(\mathrm{E}[y|x])^2]$. If one adds, the two $\mathrm{E}[(\mathrm{E}[y|x])^2]$ cancel out, and the other two terms can be simplified by the law of iterated expectations to give $\mathrm{E}[y^2] - (\mathrm{E}[y])^2$. □

• **c.** *2 points* [Coo98, p. 23] *The conditional expected value is sometimes called the population regression function. In graphical data analysis, the sample equivalent of the variance ratio*

$$\frac{\mathrm{E}\big[\mathrm{var}[y|x]\big]}{\mathrm{var}\big[\mathrm{E}[y|x]\big]} \tag{4}$$

can be used to determine whether the regression function $\mathrm{E}[y|x]$ *appears to be visually well-determined or not. Does a small or a big variance ratio indicate a well-determined regression function?*

Answer. For a well-determined regression function the variance ratio should be *small*. [Coo98, p. 23] writes: "This ratio is reminiscent of a one-way analysis of variance, with the numerator representing the average within group (slice) variance, and the denominator representing the varince between group (slice) means." □

Problem 2. *The figure on page 17 shows 250 independent observations of the random vector* $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$.

• **a.** *2 points Draw in by hand the approximate location of* $\mathcal{E}[\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]]$ *and the graph of* $\mathrm{E}[y|x]$. *Draw into the second diagram the approximate marginal density of* x.

• **b.** *2 points Is there a law that the graph of the conditional expectation* $\mathrm{E}[y|x]$ *always goes through the point* $\mathcal{E}[\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]]$*—for arbitrary probability distributions for which these expectations exist, or perhaps for an important special case? Indicate how this could be proved or otherwise give (maybe geometrically) a simple counterexample.*

Answer. This is *not* the law of iterated expectations. It is true for jointly normal variables, not in general. It is also true if $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent; then the graph of $\mathrm{E}[y|x]$ is a horizontal line at the height of the unconditional expectation $\mathrm{E}[y]$. A distribution with U-shaped unconditional distribution has the unconditional mean in the center of the U, i.e., here the unconditional mean does not lie on the curve drawn out by the conditional mean. □

• **c.** *2 points Do you have any ideas how the strange-looking cluster of points in the figure on page* *17* *was generated?*

Problem 3. *5 points Let* x *and* y *be jointly distributed scalar random variables, and assume conditional means exist. Define* $\varepsilon = y - \mathrm{E}[y|x]$*. Demonstrate the following* *equations:*

$$\mathrm{E}[\varepsilon|x] = 0 \tag{5}$$

$$\mathrm{E}[\varepsilon] = 0 \tag{6}$$

$$\mathrm{E}[x\varepsilon|x] = 0 \tag{7}$$

$$\mathrm{E}[x\varepsilon] = 0 \tag{8}$$

$$\mathrm{cov}[x, \varepsilon] = 0. \tag{9}$$

Say very explicitly which of the rules you are using in every step.

Answer. $\mathrm{E}[\varepsilon|x] = \mathrm{E}[y|x] - \mathrm{E}[\mathrm{E}[y|x]|x] = 0$ since $\mathrm{E}[y|x]$ is a function of x and therefore equal to its own expectation conditionally on x. The second statement follows from the first (the first statement is stronger than the second): if an expectation is zero conditionally on every possible outcome of x then it is zero altogether. Or one can also do it in one swoop: $\mathrm{E}[\varepsilon] = \mathrm{E}[y - \mathrm{E}[y|x]] = 0$ by the law of iterated expectations. $\mathrm{E}[x\varepsilon|x] = x\mathrm{E}[\varepsilon|x] = 0$; $\mathrm{E}[x\varepsilon] = \mathrm{E}[\mathrm{E}[x\varepsilon|x]] = \mathrm{E}[0] = 0$, and $\mathrm{cov}[x, \varepsilon] = \mathrm{E}[x\varepsilon] - \mathrm{E}[x]\,\mathrm{E}[\varepsilon] = 0 - \mathrm{E}[x] \cdot 0 = 0$. □

Problem 4. *5 points* [Lar82, example 5.4.1 on p 266] *Let y_1 and y_2 be two random variables with same mean μ and variance σ^2, but we do not assume that they are uncorrelated; their correlation coefficient is ρ, which can take any value $|\rho| \le 1$. Show that $\bar{y} = (y_1 + y_2)/2$ has lowest mean squared error among all linear unbiased estimators of μ, and compute its* MSE. *(An estimator $\tilde{\mu}$ of μ is linear iff it can be written in the form $\tilde{\mu} = \alpha_1 y_1 + \alpha_2 y_2$ with some constant numbers α_1 and α_2.)*

Answer.

$$\tilde{y} = \alpha_1 y_1 + \alpha_2 y_2 \tag{10}$$

$$\operatorname{var} \tilde{y} = \alpha_1^2 \operatorname{var}[y_1] + \alpha_2^2 \operatorname{var}[y_2] + 2\alpha_1\alpha_2 \operatorname{cov}[y_1, y_2] \tag{11}$$

$$= \sigma^2(\alpha_1^2 + \alpha_2^2 + 2\alpha_1\alpha_2\rho). \tag{12}$$

Here we used (**??**). Unbiasedness means $\alpha_2 = 1 - \alpha_1$, therefore we call $\alpha_1 = \alpha$ and $\alpha_2 = 1 - \alpha$:

$$\operatorname{var}[\tilde{y}]/\sigma^2 = \alpha^2 + (1-\alpha)^2 + 2\alpha(1-\alpha)\rho \tag{13}$$

Now sort by the powers of α:

$$= 2\alpha^2(1-\rho) - 2\alpha(1-\rho) + 1 \tag{14}$$

$$= 2(\alpha^2 - \alpha)(1-\rho) + 1. \tag{15}$$

This takes its minimum value where the derivative $\frac{\partial}{\partial\alpha}(\alpha^2 - \alpha) = 2\alpha - 1 = 0$. For the MSE plug $\alpha_1 = \alpha_2 - 1/2$ into (12) to get $\frac{\sigma^2}{2}\big(1+\rho\big)$. □

Problem 5. *You have two unbiased measurements with errors of the same quantity μ (which may or may not be random). The first measurement y_1 has mean squared error $\mathrm{E}[(y_1 - \mu)^2] = \sigma^2$, the other measurement y_2 has $\mathrm{E}[(y_1 - \mu)^2] = \tau^2$. The measurement errors $y_1 - \mu$ and $y_2 - \mu$ have zero expected values (i.e., the measurements are unbiased) and are independent of each other.*

• **a.** *2 points Show that the linear unbiased estimators of μ based on these two measurements are simply the weighted averages of these measurements, i.e., they can be written in the form $\tilde{\mu} = \alpha y_1 + (1-\alpha)y_2$, and that the MSE of such an estimator is $\alpha^2\sigma^2 + (1-\alpha)^2\tau^2$. Note: we are using the word "estimator" here even if μ is random. An estimator or predictor $\tilde{\mu}$ is unbiased if $\mathrm{E}[\tilde{\mu} - \mu] = 0$. Since we allow μ to be random, the proof in the class notes has to be modified.*

Answer. The estimator $\tilde{\mu}$ is linear (more precisely: affine) if it can written in the form

$$\tilde{\mu} = \alpha_1 y_1 + \alpha_2 y_2 + \gamma \tag{16}$$

The measurements themselves are unbiased, i.e., $\mathrm{E}[y_i - \mu] = 0$, therefore

$$\mathrm{E}[\tilde{\mu} - \mu] = (\alpha_1 + \alpha_2 - 1)\,\mathrm{E}[\mu] + \gamma = 0 \tag{17}$$

for all possible values of $\mathrm{E}[\mu]$; therefore $\gamma = 0$ and $\alpha_2 = 1 - \alpha_1$. To simplify notation, we will call from now on $\alpha_1 = \alpha$, $\alpha_2 = 1 - \alpha$. Due to unbiasedness, the MSE is the variance of the estimation error

$$\mathrm{var}[\tilde{\mu} - \mu] = \alpha^2\sigma^2 + (1-\alpha)^2\tau^2 \tag{18}$$

□

• **b.** *4 points Define ω^2 by*

$$\frac{1}{\omega^2} = \frac{1}{\sigma^2} + \frac{1}{\tau^2} \qquad \textit{which can be solved to give} \qquad \omega^2 = \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}. \tag{19}$$

Show that the Best (i.e., minimum MSE*) linear unbiased estimator (BLUE) of* μ *based on these two measurements is*

$$\hat{y} = \frac{\omega^2}{\sigma^2} y_1 + \frac{\omega^2}{\tau^2} y_2 \tag{20}$$

i.e., it is the weighted average of y_1 *and* y_2 *where the weights are proportional to the inverses of the variances.*

Answer. The variance (18) takes its minimum value where its derivative with respect of α is zero, i.e., where

$$\frac{\partial}{\partial \alpha}\Bigl(\alpha^2\sigma^2 + (1-\alpha)^2\tau^2\Bigr) = 2\alpha\sigma^2 - 2(1-\alpha)\tau^2 = 0 \tag{21}$$

$$\alpha\sigma^2 = \tau^2 - \alpha\tau^2 \tag{22}$$

$$\alpha = \frac{\tau^2}{\sigma^2 + \tau^2} \tag{23}$$

In terms of ω one can write

$$\alpha = \frac{\tau^2}{\sigma^2 + \tau^2} = \frac{\omega^2}{\sigma^2} \qquad \text{and} \qquad 1 - \alpha = \frac{\sigma^2}{\sigma^2 + \tau^2} = \frac{\omega^2}{\tau^2}. \tag{24}$$

□

- **c.** *2 points Show: the* MSE *of the BLUE* ω^2 *satisfies the following equation:*

$$\frac{1}{\omega^2} = \frac{1}{\sigma^2} + \frac{1}{\tau^2} \tag{25}$$

Answer. We already have introduced the notation ω^2 for the quantity defined by (25); therefore all we have to show is that the MSE or, equivalently, the variance of the estimation error is equal to this ω^2:

$$\operatorname{var}[\tilde{\mu}-\mu] = \Big(\frac{\omega^2}{\sigma^2}\Big)^2\sigma^2 + \Big(\frac{\omega^2}{\tau^2}\Big)^2\tau^2 = \omega^4\Big(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\Big) = \omega^4\frac{1}{\omega^2} = \omega^2 \tag{26}$$

□

Problem 6. [Seb77, exercise 4d-3] *Given $n+1$ observations y_j from a $N(\mu,\sigma^2)$. After the first n observations, it is suspected that a sudden change in the mean of the distribution occurred, i.e., that $y_{n+1} \sim N(\nu,\sigma^2)$ with $\nu \neq \mu$. We will use here three different approaches to derive the same test statistic for testing the hypothesis that the $n+1$st observation has the same population mean as the previous observations, i.e., that $\nu=\mu$, against the two-sided alternative. The formulas for this statistic should be given in terms of the observations y_i. It is recommended to use the notation $\bar{y}=\frac{1}{n}\sum_{i=1}^n y_i$ and $\bar{\bar{y}}=\frac{1}{n+1}\sum_{j=1}^{n+1} y_j$.*

- **a.** *3 points First you should derive this statistic by testing whether $\nu-\mu=0$ (the "Wald principle"). For this you must compute the BLUE of $\nu-\mu$ and its standard deviation and construct the t statistic from this.*

Answer. BLUE of μ is $\bar{y}=\frac{1}{n}\sum_{i=1}^n y_i$, and that of ν is y_{n+1}. BLUE of $\nu-\mu$ is $\bar{y}-y_{n+1}$. Because of independence $\operatorname{var}[\bar{y}-y_{n+1}] = \operatorname{var}[\bar{y}]+\operatorname{var}[y_{n+1}] = \sigma^2((1/n)+1) = \sigma^2(n+1)/n$. Standard deviation is $\sigma\sqrt{(n+1)/n}$.

For the denominator in the t-statistic you need the s^2 from the unconstrained regression, which is

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_j - \bar{y})^2 \tag{27}$$

What happened to the $(n+1)$st observation here? It always has a zero residual. And the factor $1/(n-1)$ should really be written $1/(n+1-2)$: there are $n+1$ observations and 2 parameters.

Divide $\bar{y} - y_{n+1}$ by its standard deviation and replace σ by s (the square root of s^2) to get the t statistic

$$\frac{\bar{y} - y_{n+1}}{s\sqrt{1+\frac{1}{n}}} \tag{28}$$

□

• **b.** *2 points One can interpret this same formula also differently (and this is why this test is sometimes called the "predictive" Chow test). Compute the Best Linear Unbiased Predictor of y_{n+1} on the basis of the first n observations, call it $\hat{\hat{y}}(n+1)_{n+1}$. Show that the predictive residual $y_{n+1} - \hat{\hat{y}}(n+1)_{n+1}$, divided by the square root of $\mathrm{MSE}[\hat{\hat{y}}(n+1)_{n+1}; y_{n+1}]$, with σ replaced by s (based on the first n observations only), is equal to the above t statistic.*

Answer. BLUP of y_{n+1} based on first n observations is $\bar{y}$ again. Since it is unbiased, $\mathrm{MSE}[\bar{y}; y_{n+1}] = \mathrm{var}[\bar{y} - y_{n+1}] = \sigma^2(n+1)/n$. From now on everything is as in part a. □

• **c.** *6 points Next you should show that the above two formulas are identical to the statistic based on comparing the SSEs of the constrained and unconstrained models (the likelihood ratio principle). Give a formula for the constrained SSE_r, the unconstrained SSE_u, and the F-statistic.*

Answer. According to the Likelihood Ratio principle, one has to compare the residual sums of squares in the regressions under the assumption that the mean did not change with that under the assumption that the mean changed. If the mean did not change (constrained model), then $\bar{\bar{y}}$ is the OLS of μ. In order to make it easier to derive the difference between constrained and unconstrained SSE, we will write the constrained SSE as follows:

$$SSE_r = \sum_{j=1}^{n+1}(y_j - \bar{\bar{y}})^2 = \sum_{j=1}^{n+1} y_j^2 - (n+1)\bar{\bar{y}}^2 = \sum_{j=1}^{n+1} y_j^2 - \frac{1}{n+1}(n\bar{y} + y_{n+1})^2$$

If one allows the mean to change (unconstrained model), then $\bar{y}$ is the BLUE of μ, and y_{n+1} is the BLUE of ν.

$$SSE_u = \sum_{j=1}^{n}(y_j - \bar{y})^2 + (y_{n+1} - y_{n+1})^2 = \sum_{j=1}^{n} y_j^2 - n\bar{y}^2.$$

Now subtract:

$$
\begin{aligned}
SSE_r - SSE_u &= y_{n+1}^2 + n\bar{y}^2 - \frac{1}{n+1}(n\bar{y} + y_{n+1})^2 \\
&= y_{n+1}^2 + n\bar{y}^2 - \frac{1}{n+1}(n^2\bar{y}^2 + 2n\bar{y}y_{n+1} + y_{n+1}^2) \\
&= (1 - \frac{1}{n+1})y_{n+1}^2 + (n - \frac{n^2}{n+1})\bar{y}^2 - \frac{n}{n+1}2\bar{y}y_{n+1} \\
&= \frac{n}{n+1}(y_{n+1} - \bar{y})^2.
\end{aligned}
$$

Interestingly, this depends on the first n observations only through $\bar{y}$.

Since the unconstrained model has $n+1$ observations and 2 parameters, the test statistic is

$$
(29) \qquad \frac{SSE_r - SSE_u}{SSE_u/(n+1-2)} = \frac{\frac{n}{n+1}(y_{n+1} - \bar{y})^2}{\sum_1^n (y_j - \bar{y})^2/(n-1)} = \frac{(y_{n+1} - \bar{y})^2\, n(n-1)}{\sum_1^n (y_j - \bar{y})^2\,(n+1)} \sim F_{1,n-1}
$$

This is the square of the t statistic (28). □

Problem 7. [Seb77, pp. 117–119] *Given a regression model with k independent variables. There are n observations of the vector of independent variables, and for each of these n values there is not one but $r > 1$ different replicated observations of the dependent variable. This model can be written*

$$
(30) \qquad y_{mq} = \sum_{j=1}^{k} x_{mj}\beta_j + \varepsilon_{mq} \qquad \textit{or} \qquad y_{mq} = \boldsymbol{x}_m^\top \boldsymbol{\beta} + \varepsilon_{mq},
$$

where $m = 1, \dots, n$, $j = 1, \dots, k$, $q = 1, \dots, r$, and $\boldsymbol{x}_m^\top$ is the mth row *of the $\boldsymbol{X}$-matrix. For simplicity we assume that r does not depend on m, each observation of the independent variables has the same number of repetitions. We also assume that the $n \times k$ matrix $\boldsymbol{X}$ has full column rank.*

• **a.** *2 points In this model it is possible to test whether the regression line is in fact a straight line. If it is not a straight line, then each observation of the dependent variables $\boldsymbol{x}_m$ has a different coefficient vector $\boldsymbol{\beta}_m$ associated with it, i.e., the model is*

$$y_{mq} = \sum_{j=1}^{k} x_{mj}\beta_{mj} + \varepsilon_{mq} \qquad \text{or} \qquad y_{mq} = \boldsymbol{x}_m^\top \boldsymbol{\beta}_m + \varepsilon_{mq}. \tag{31}$$

This unconstrained model does not have enough information to estimate any of the individual coefficients β_{mj}. Explain how it is nevertheless still possible to compute SSE_u.

Answer. Even though the individual coefficients β_{mj} are not identified, their linear combination $\eta_m = \boldsymbol{x}_m^\top \boldsymbol{\beta}_m = \sum_{j=1}^{k} x_{mj}\beta_{mj}$ is identified; one unbiased estimator, although by far not the best one, is any individual observation y_{mq}. This linear combination is all one needs to compute SSE_u, the sum of squared errors in the unconstrained model. □

• **b.** *2 points Writing your estimate of $\eta_m = \boldsymbol{x}_m^\top \boldsymbol{\beta}_m$ as $\tilde{\eta}_m$, give the formula of the sum of squared errors of this estimate, and by taking the first order conditions, show* *that the unconstrained least squares estimate of η_m is $\hat{\eta}_m = \bar{y}_{m\cdot}$ for $m = 1, \ldots, n$, where $\bar{y}_{m\cdot} = \frac{1}{r}\sum_{q=1}^{r} y_{mq}$ (i.e., the dot in the subscript indicates taking the mean).*

Answer. If we know the $\tilde{\eta}_m$ the sum of squared errors no longer depents on the independent observations $\boldsymbol{x}_m$ but is simply

$$SSE_u = \sum_{m,q}(y_{mq} - \tilde{\eta}_m)^2 \tag{32}$$

First order conditions are

$$\frac{\partial}{\partial \tilde{\eta}_h}\sum_{m,q}(y_{mq} - \tilde{\eta}_m)^2 = \frac{\partial}{\partial \tilde{\eta}_h}\sum_{q}(y_{hq} - \tilde{\eta}_h)^2 = -2\sum_{q}(y_{hq} - \tilde{\eta}_h) = 0 \tag{33}$$

□

• **c.** *1 point The sum of squared errors associated with this least squares estimate is the unconstrained sum of squared errors SSE_u. How would you set up a regression with dummy variables which would give you this SSE_u?*

Answer. The unconstrained model should be regressed in the form $y_{mq} = \eta_m + \varepsilon_{mq}$. I.e., string out the matrix $\boldsymbol{Y}$ as a vector and for each column of $\boldsymbol{Y}$ introduce a dummy variable which is $= 1$ if the given observation was originally in this colum. □

• **d.** *2 points Next turn to the constrained model (30). If $\boldsymbol{X}$ has full column rank, then it is fully identified. Writing $\tilde{\beta}_j$ for your estimates of β_j, give a formula for the sum of squared errors of this estimate. By taking the first order conditions,* *show that the estimate $\hat{\boldsymbol{\beta}}$ is the same as the estimate in the model without replicated observations*

$$z_m = \sum_{j=1}^{k} x_{mj}\beta_j + \varepsilon_m, \tag{34}$$

where $z_m = \bar{y}_{m\cdot}$ as defined above.

• **e.** *2 points If SSE_c is the SSE in the constrained model (30) and SSE_b the SSE in (34), show that $SSE_c = r \cdot SSE_b + SSE_u$.*

Answer. For every m we have $\sum_q (y_{mq} - \boldsymbol{x}_m^\top \hat{\boldsymbol{\beta}})^2 = \sum_q (y_{mq} - \bar{y}_{m\cdot})^2 + r(y_{m\cdot} - \boldsymbol{x}_m^\top \hat{\boldsymbol{\beta}})^2$; therefore $SSE_c = \sum_{m,q}(y_{mq} - \bar{y}_{m\cdot})^2 + r\sum_m (y_{m\cdot} - \boldsymbol{x}_m^\top \hat{\boldsymbol{\beta}})^2$; □

• **f.** *3 points Write down the formula of the F-test in terms of SSE_u and SSE_c with a correct accounting of the degrees of freedom, and give this formula also in terms of SSE_u and SSE_b.*

Answer. Unconstrained model has n parameters, and constrained model has k parameters; the number of additional "constraints" is therefore $n-k$. This gives the F-statistic

$$\frac{(SSE_c - SSE_u)/(n-k)}{SSE_u/n(r-1)} = \frac{rSSE_b/(n-k)}{SSE_u/n(r-1)} \tag{35}$$

□

Maximum number of points: 53.

References

[Ame94] Takeshi Amemiya, *Introduction to statistics and econometrics*, Harvard University Press, Cambridge, MA, 1994. 2

[Coo98] R. Dennis Cook, *Regression graphics: Ideas for studying regressions through graphics*, Series in Probability and Statistics, Wiley, New York, 1998. 3

[Lar82] Harold Larson, *Introduction to probability and statistical inference*, Wiley, 1982. 5

[Rao73] C. Radhakrishna Rao, *Linear statistical inference and its applications*, second ed., Wiley, New York, 1973. 2

[Seb77] G. A. F. Seber, *Linear regression analysis*, Wiley, New York, 1977. 9, 12

Economics Department, University of Utah, 1645 Campus Center Drive, Salt Lake City, UT 84112-9300, U.S.A