FINAL EXAM STAT 6869 SPRING 2000

ECONOMICS DEPARTMENT, UNIVERSITY OF UTAH

This is a closed book exam but you may bring one sheet with formulas with you; write your name on the formula sheet and submit it together with your exam.

Problem 23. *2 points*

The density function $f_y(y;\theta)$ which depends on the parameter θ belongs to the exponential family (in its canonical form) if it has the functional form

$$f_y(y;\theta) = \exp\bigl(y\theta - b(\theta) + c(y)\bigr) \tag{1}$$

in a certain region $U \subset \mathbb{R}$ and $= 0$ outside this region. Show that the Poisson distribution

$$\Pr[x=k] = \frac{\lambda^k}{k!}e^{-\lambda} \quad \textit{for } k = 0, 1, \ldots \tag{2}$$

Date of exam Thursday, May 4, 2000, 1:45–3:45 pm in the usual classroom.

is a member of the exponential family. Compute the canonical parameter θ and the function $b(\theta)$.

Answer. The probability mass function can be written as

$$\Pr[x{=}k] = \frac{e^{k \ln \lambda}}{k!} e^{-\lambda} = \exp(k \ln \lambda - \lambda - \ln k!) \quad \text{for } k = 0, 1, \dots \tag{3}$$

This is (1) for the Poisson distribution, where the values of the random variable are called k instead of x, and $\theta = \ln \lambda$. Substituting $\lambda = \exp(\theta)$ in (3) gives

$$\Pr[x{=}k] = \exp(k\theta - \exp(\theta) - \ln k!) \quad \text{for } k = 0, 1, \dots \tag{4}$$

from which one sees $b(\theta) = \exp(\theta)$.

□

Problem 24. *1 point In the case of the Poisson distribution (see Problem 23) compute $b'(\theta)$ and verify that it is the same as $\mathrm{E}[x]$, and compute $b''(\theta)$ and verify that it is the same as $\operatorname{var}[x]$. You are allowed, without proof, that a Poisson distribution with parameter λ has expected value λ and variance λ.*

Answer. $b(\theta) = \exp \theta$, therefore $b'(\theta) = b''(\theta) = \exp(\theta) = \lambda$. □

Problem 25. *Assume the density function of y depends on a parameter θ, write it $f_y(y; \theta)$, and θ° is the true value of θ. In this problem we will compare the expected value of y and of functions of y with what would be their expected value if the true parameter value were not θ° but would take some other value θ. If the random* *variable* t *is a function of* y*, we write* $\mathrm{E}_\theta[t]$ *for what would be the expected value of* t *if the true value of the parameter were* θ *instead of* θ°*. Occasionally, we will use the subscript* $\circ$ *as in* $\mathrm{E}_\circ$ *to indicate that we are dealing here with the usual case in which the expected value is taken with respect to the true parameter value* θ°*. Instead of* $\mathrm{E}_\circ$ *one usually simply writes* E*, since it is usually self-understood that one has to plug the right parameter values into the density function if one takes expected values. The subscript* $\circ$ *is necessary here only because in the present problem, we sometimes take expected values with respect to the "wrong" parameter values. The same notational convention also applies to variances, covariances, and the MSE.*

Throughout this problem we assume that the following regularity conditions hold: (a) the range of y *is independent of* θ*, and (b) the derivative of the density function with respect to* θ *is a continuous differentiable function of* θ*. These regularity conditions ensure that one can differentiate under the integral sign, i.e., for all function* $t(y)$ *follows*

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f_y(y;\theta)t(y)\,dy = \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f_y(y;\theta)t(y)\,dy = \frac{\partial}{\partial\theta}\,\mathrm{E}_\theta[t(y)] \tag{5}$$

$$\int_{-\infty}^{\infty} \frac{\partial^2}{(\partial\theta)^2} f_y(y;\theta)t(y)\,dy = \frac{\partial^2}{(\partial\theta)^2}\int_{-\infty}^{\infty} f_y(y;\theta)t(y)\,dy = \frac{\partial^2}{(\partial\theta)^2}\,\mathrm{E}_\theta[t(y)]. \tag{6}$$

All the following steps are fairly simple once you know the general procedures to tackle them. If you cannot do one of the steps, take it as given and do the next one.

• **a.** *1 point The* score *is defined as the random variable*

$$q(y;\theta) = \frac{\partial}{\partial\theta}\log f_y(y;\theta). \tag{7}$$

In other words, we do three things to the density function: take its logarithm, then take the derivative of this logarithm with respect to the parameter, and then plug the random variable into it. This gives us a random variable which also depends on the nonrandom parameter θ. Show that the score can also be written as

$$q(y;\theta) = \frac{1}{f_y(y;\theta)}\frac{\partial f_y(y;\theta)}{\partial\theta} \tag{8}$$

Answer. This is the chain rule for differentiation: for any differentiable function $g(\theta)$, $\frac{\partial}{\partial\theta}\log g(\theta) = \frac{1}{g(\theta)}\frac{\partial g(\theta)}{\partial\theta}$. □

• **b.** *1 point*

The density function $f_y(y;\theta,\phi)$ which depends on the two parameters θ and ϕ belongs to the exponential dispersion family if it has the functional form

$$f_y(y;\theta,\phi) = \exp\Big(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\Big). \tag{9}$$

in a certain region $U \subset \mathbb{R}$ and $= 0$ outside this region. If the density function is member of an exponential dispersion family (9), show that the score function has the form

$$q(y;\theta) = \frac{y - \frac{\partial b(\theta)}{\partial \theta}}{a(\psi)} \tag{10}$$

Answer. This is a simple substitution: if

$$f_y(y;\theta,\psi) = \exp\left(\frac{y\theta - b(\theta)}{a(\psi)} + c(y,\psi)\right), \tag{11}$$

then

$$\frac{\partial \log f_y(y;\theta,\psi)}{\partial \theta} = \frac{y - \frac{\partial b(\theta)}{\partial \theta}}{a(\psi)} \tag{12}$$

□

• **c.** *3 points If $f_y(y;\theta^\circ)$ is the true density function of y, then we know from a basic entropy inequality that $\mathrm{E}_\circ[\log f_y(y;\theta^\circ)] \geq \mathrm{E}_\circ[\log f(y;\theta)]$ for all θ. This explains why the score is so important: it is the derivative of that function whose expected value is maximized if the true parameter is plugged into the density function. The first-order conditions in this situation read: the expected value of this derivative must be zero for the true parameter value. This is the next thing you are asked to show: If θ° is the true parameter value, show that $\mathrm{E}_\circ[q(y;\theta^\circ)] = 0$.*

Answer. First write for general θ

$$\mathrm{E}_\circ[q(y;\theta)] = \int_{-\infty}^{\infty} q(y;\theta) f_y(y;\theta^\circ)\,dy = \int_{-\infty}^{\infty} \frac{1}{f_y(y;\theta)} \frac{\partial f_y(y;\theta)}{\partial\theta} f_y(y;\theta^\circ)\,dy. \tag{13}$$

For $\theta = \theta^\circ$ this simplifies:

$$\mathrm{E}_\circ[q(y;\theta^\circ)] = \int_{-\infty}^{\infty} \frac{\partial f_y(y;\theta)}{\partial\theta}\bigg|_{\theta=\theta^\circ} dy = \frac{\partial}{\partial\theta} \int_{-\infty}^{\infty} f_y(y;\theta)\,dy\bigg|_{\theta=\theta^\circ} = \frac{\partial}{\partial\theta} 1 = 0. \tag{14}$$

Here I am writing $\frac{\partial f_y(y;\theta)}{\partial\theta}\Big|_{\theta=\theta^\circ}$ instead of the simpler notation $\frac{\partial f_y(y;\theta^\circ)}{\partial\theta}$, in order to emphasize that one *first* has to take a derivative with respect to θ and *then* one plugs θ° into that derivative. □

• **d.** *Show that, in the case of the exponential dispersion family,*

$$\mathrm{E}_\circ[y] = \frac{\partial b(\theta)}{\partial\theta}\bigg|_{\theta=\theta^\circ} \tag{15}$$

Answer. Follows from the fact that the score function of the exponential family (10) has zero expected value. □

• **e.** *5 points If we differentiate the score, we obtain the* Hessian

$$h(\theta) = \frac{\partial^2}{(\partial\theta)^2} \log f_y(y;\theta). \tag{16}$$

From now on we will write the score function as $q(\theta)$ instead of $q(y;\theta)$; i.e., we will no longer make it explicit that q is a function of y but write it as a random variable which depends on the parameter θ. We also suppress the dependence of h on y; our notation $h(\theta)$ is short for $h(y;\theta)$. Since there is only one parameter in the density function, score and Hessian are scalars; but in the general case, the score is a vector and the Hessian a matrix. Show that, for the true parameter value θ°, the negative of the expected value of the Hessian equals the variance of the score, i.e., the expected value of the square of the score:

$$\mathrm{E}_\circ[h(\theta^\circ)] = -\,\mathrm{E}_\circ[q^2(\theta^\circ)]. \tag{17}$$

Answer. Start with the definition of the score

$$q(y;\theta) = \frac{\partial}{\partial\theta}\log f_y(y;\theta) = \frac{1}{f_y(y;\theta)}\frac{\partial}{\partial\theta}f_y(y;\theta), \tag{18}$$

and differentiate the rightmost expression one more time:

$$h(y;\theta) = \frac{\partial}{(\partial\theta)}q(y;\theta) = -\frac{1}{f_y^2(y;\theta)}\left(\frac{\partial}{\partial\theta}f_y(y;\theta)\right)^2 + \frac{1}{f_y(y;\theta)}\frac{\partial^2}{\partial\theta^2}f_y(y;\theta) \tag{19}$$

$$= -q^2(y;\theta) + \frac{1}{f_y(y;\theta)}\frac{\partial^2}{\partial\theta^2}f_y(y;\theta) \tag{20}$$

Taking expectations we get

$$\mathrm{E}_\circ[h(y;\theta)] = -\,\mathrm{E}_\circ[q^2(y;\theta)] + \int_{-\infty}^{+\infty}\frac{1}{f_y(y;\theta)}\left(\frac{\partial^2}{\partial\theta^2}f_y(y;\theta)\right)f_y(y;\theta^\circ)\,dy \tag{21}$$

Again, for $\theta = \theta^\circ$, we can simplify the integrand and differentiate under the integral sign:

$$\int_{-\infty}^{+\infty} \frac{\partial^2}{\partial\theta^2} f_y(y;\theta)\, dy = \frac{\partial^2}{\partial\theta^2} \int_{-\infty}^{+\infty} f_y(y;\theta)\, dy = \frac{\partial^2}{\partial\theta^2} 1 = 0. \tag{22}$$

□

• **f.** *Derive from* (17) *that, for the exponential dispersion family* (9),

$$\operatorname{var}_\circ[y] = \frac{\partial^2 b(\theta)}{\partial\theta^2} a(\phi)\Big|_{\theta=\theta^\circ} \tag{23}$$

Answer. Differentiation of (10) gives $h(\theta) = -\frac{\partial^2 b(\theta)}{\partial\theta^2}\frac{1}{a(\phi)}$. This is constant and therefore equal to its own expected value. (17) says therefore

$$\frac{\partial^2 b(\theta)}{\partial\theta^2}\Big|_{\theta=\theta^\circ} \frac{1}{a(\phi)} = \mathrm{E}_\circ[q^2(\theta^\circ)] = \frac{1}{\big(a(\phi)\big)^2} \operatorname{var}_\circ[y] \tag{24}$$

from which (23) follows. □

Problem 26. *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

$$\tilde{\mu} = \alpha_1 y_1 + \alpha_2 y_2 + \gamma \tag{25}$$

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

$$\mathrm{E}[\tilde{\mu} - \mu] = (\alpha_1 + \alpha_2 - 1)\,\mathrm{E}[\mu] + \gamma = 0 \tag{26}$$

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

$$\operatorname{var}[\tilde{\mu} - \mu] = \alpha^2\sigma^2 + (1-\alpha)^2\tau^2 \tag{27}$$

□

• **b.** *4 points Define ω^2 by*

$$\frac{1}{\omega^2} = \frac{1}{\sigma^2} + \frac{1}{\tau^2} \qquad \text{which can be solved to give} \qquad \omega^2 = \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}. \tag{28}$$

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

$$\hat{y} = \frac{\omega^2}{\sigma^2} y_1 + \frac{\omega^2}{\tau^2} y_2 \tag{29}$$

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 (27) takes its minimum value where its derivative with respect of α is zero, i.e., where

$$\frac{\partial}{\partial \alpha}\Bigl(\alpha^2\sigma^2 + (1-\alpha)^2\tau^2\Bigr) = 2\alpha\sigma^2 - 2(1-\alpha)\tau^2 = 0 \tag{30}$$

$$\alpha\sigma^2 = \tau^2 - \alpha\tau^2 \tag{31}$$

$$\alpha = \frac{\tau^2}{\sigma^2 + \tau^2} \tag{32}$$

In terms of ω one can write

$$\alpha = \frac{\tau^2}{\sigma^2+\tau^2} = \frac{\omega^2}{\sigma^2} \qquad \text{and} \qquad 1-\alpha = \frac{\sigma^2}{\sigma^2+\tau^2} = \frac{\omega^2}{\tau^2}. \tag{33}$$

□

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

$$\frac{1}{\omega^2} = \frac{1}{\sigma^2} + \frac{1}{\tau^2} \tag{34}$$

Answer. We already have introduced the notation ω^2 for the quantity defined by (34); therefore all we have to show is that the MSE or, equivalently, the variance of the estimation error is equal to this ω^2:

$$\operatorname{var}[\tilde{\mu} - \mu] = \Big(\frac{\omega^2}{\sigma^2}\Big)^2 \sigma^2 + \Big(\frac{\omega^2}{\tau^2}\Big)^2 \tau^2 = \omega^4 \Big(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\Big) = \omega^4 \frac{1}{\omega^2} = \omega^2 \tag{35}$$

□

Problem 27. *Prove the following facts about the diagonal elements of the so-called "hat matrix" $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$, which has its name because $\boldsymbol{H}\boldsymbol{y} = \hat{\boldsymbol{y}}$, i.e., it puts the hat on $\boldsymbol{y}$.*

- **a.** *1 point $\boldsymbol{H}$ is a projection matrix, i.e., it is symmetric and idempotent.*

Answer. Symmetry follows from the laws for the transposes of products: $\boldsymbol{H}^\top = (\boldsymbol{ABC})^\top = \boldsymbol{C}^\top \boldsymbol{B}^\top \boldsymbol{A}^\top = \boldsymbol{H}$ where $\boldsymbol{A} = \boldsymbol{X}$, $\boldsymbol{B} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}$ which is symmetric, and $\boldsymbol{C} = \boldsymbol{X}^\top$. Idempotency $\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$. □

- **b.** *1 point Prove that a symmetric idempotent matrix is nonnegative definite.*

Answer. If $\boldsymbol{H}$ is symmetric and idempotent, then for arbitrary $\boldsymbol{g}$, $\boldsymbol{g}^\top\boldsymbol{H}\boldsymbol{g} = \boldsymbol{g}^\top\boldsymbol{H}^\top\boldsymbol{H}\boldsymbol{g} = \|\boldsymbol{H}\boldsymbol{g}\|^2 \geq 0$. But $\boldsymbol{g}^\top\boldsymbol{H}\boldsymbol{g} \geq 0$ for all $\boldsymbol{g}$ is the criterion which makes $\boldsymbol{H}$ nonnegative definite. □

- **c.** *2 points Show that*

$$0 \leq h_{ii} \leq 1 \tag{36}$$

Answer. If $\boldsymbol{e}_i$ is the vector with a 1 on the ith place and zeros everywhere else, then $\boldsymbol{e}_i^\top \boldsymbol{H}\boldsymbol{e}_i = h_{ii}$. From $\boldsymbol{H}$ nonnegative definite follows therefore that $h_{ii} \geq 0$. $h_{ii} \leq 1$ follows because $\boldsymbol{I} - \boldsymbol{H}$ is symmetric and idempotent (and therefore nonnegative definite) as well: it is the projection on the orthogonal complement. □

• **d.** *2 points Show: the average value of the h_{ii} is $\sum h_{ii}/n = k/n$, where k is the number of columns of $\boldsymbol{X}$. (Hint: for this you must compute the trace $\operatorname{tr}\boldsymbol{H}$.)*

Answer. The average can be written as

$$\frac{1}{n}\operatorname{tr}(\boldsymbol{H}) = \frac{1}{n}\operatorname{tr}(\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top) = \frac{1}{n}\operatorname{tr}(\boldsymbol{X}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}) = \frac{1}{n}\operatorname{tr}(\boldsymbol{I}_k) = \frac{k}{n}.$$

Here we used $\operatorname{tr}\boldsymbol{B}\boldsymbol{C} = \operatorname{tr}\boldsymbol{C}\boldsymbol{B}$. □

• **e.** *1 point Show that $\frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}^\top$ is a projection matrix. Here $\boldsymbol{\iota}$ is the n-vector of ones.*

• **f.** *2 points Show: If the regression has a constant term, then $\boldsymbol{H} - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}^\top$ is a projection matrix.*

Answer. If $\boldsymbol{\iota}$, the vector of ones, is one of the columns of $\boldsymbol{X}$ (or a linear combination of these columns), this means there is a vector $\boldsymbol{a}$ with $\boldsymbol{\iota} = \boldsymbol{X}\boldsymbol{a}$. From this follows $\boldsymbol{H}\boldsymbol{\iota}\boldsymbol{\iota}^\top = \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{a}\boldsymbol{\iota}^\top$ $\boldsymbol{X}\boldsymbol{a}\boldsymbol{\iota}^\top = \boldsymbol{\iota}\boldsymbol{\iota}^\top$. One can use this to show that $\boldsymbol{H} - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}^\top$ is idempotent: $(\boldsymbol{H} - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}^\top)(\boldsymbol{H} - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}^\top) = \boldsymbol{H}\boldsymbol{H} - \boldsymbol{H}\frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}^\top - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}^\top\boldsymbol{H} + \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}^\top\frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}^\top = \boldsymbol{H} - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}^\top - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}^\top + \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}^\top = \boldsymbol{H} - \frac{1}{n}\boldsymbol{\iota}\boldsymbol{\iota}^\top$. □

• **g.** *1 point Show: If the regression has a constant term, then one can sharpen inequality (36) to $1/n \leq h_{ii} \leq 1$.*

Answer. $\boldsymbol{H} - \boldsymbol{\iota\iota}^{\top}/n$ is a projection matrix, therefore nonnegative definite, therefore its diagonal elements $h_{ii} - 1/n$ are nonnegative. □

Problem 28. *3 points What are the main concepts used in modern "Regression Diagnostics"? Can it be characterized to be a careful look at the residuals, or does it have elements which cannot be inferred from the residuals alone?*

Answer. Leverage (sometimes it is called "potential") is something which cannot be inferred from the residuals, it does not depend on $\boldsymbol{y}$ at all. □

Problem 29. *The model is $y = \alpha + x^{*}\beta + v$, but x^{*} is not observed; one can only observe $x = x^{*} + u$. The errors u and v have zero expected value and are independent of each other and of x^{*}. You have lots of data available, and for the sake of the argument we assume that the joint distribution of x and y is known precisely: it is*

$$\begin{bmatrix} y \\ x \end{bmatrix} \sim N\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 6 & -2 \\ -2 & 3 \end{bmatrix} \right). \tag{37}$$

The purpose of this question is to apply the three restrictions on the underlying parameters α and β which one can derive from the observations despite the lack of *identification in the errors-in-variables model:*

$$\mu_y = \alpha + \beta\mu_x \tag{38}$$

$$\beta \textit{ must have the same sign as } \sigma_{xy}. \tag{39}$$

$$\frac{|\sigma_{xy}|}{\sigma_x^2} \le |\beta| \le \frac{\sigma_y^2}{|\sigma_{xy}|}. \tag{40}$$

• **a.** *3 points What does the information about y and x given in equation (37) imply about α and β?*

Answer. (38) gives $\alpha - \beta = 1$, (39) gives $\beta \le 0$, and (40) $2/3 \le |\beta| \le 3$. □

• **b.** *3 points Give the* plims *of the OLS estimates of α and β in the regression of y on x.*

Answer. $\operatorname{plim}\hat\beta = \operatorname{cov}[x,y]/\operatorname{var}[x] = -\frac{2}{3}$, $\operatorname{plim}\hat\alpha = \mathrm{E}[y] - \mathrm{E}[x]\operatorname{plim}\hat\beta = \frac{1}{3}$. □

• **c.** *3 points Now assume it is known that $\alpha = 0$. What can you say now about β, σ_u^2, and σ_v^2? If β is identified, how would you estimate it?*

Answer. From $y = (x - u)\beta + v$ follows, by taking expectations, $\mathrm{E}[y] = \mathrm{E}[x]\beta$ (i.e., the true relationship still goes through the means), therefore $\beta = -1$, and a consistent estimate would be $\bar y/\bar x$. Now if one knows β one gets $\operatorname{var}[x^*]$ from $\operatorname{cov}[x,y] = \operatorname{cov}[x^*+u, \beta x^*+v] = \beta\operatorname{var}[x^*]$, i.e., $\operatorname{var}[x^*] = 2$. Then one can get $\operatorname{var}[u] = \operatorname{var}[x] - \operatorname{var}[x^*] = 3 - 2 = 1$, and $\operatorname{var}[v] = \operatorname{var}[y] - \operatorname{var}[y^*] = 6 - 2 = 4$.

Luckily, those variances came out to be positive; otherwise the restriction $\alpha = 0$ would not be compatible with (37). □

Problem 30. *Consider the "locally constant" dynamic linear model:*

$$y_t = \beta_t + \varepsilon_t \qquad \varepsilon_t \sim \text{NID}(0, \sigma^2) \tag{41}$$

$$\beta_t = \beta_{t-1} + \omega_t \qquad \omega_t \sim \text{NID}(0, \tau^2) \tag{42}$$

$$\beta_0 \sim (b_0, \tau^2\psi_0) \tag{43}$$

where all ω_t are independent of of all ε_t and of β_0. $\kappa^2 = \sigma^2/\tau^2$ is known but σ^2 and τ^2 separately are not. The objective of the estimation is to have the BLUE of the most recent level β_t based on the observations y_1 until y_t.

The best approach here is to compute the BLUE of β_{t+1} recursively with the help of the BLUE of β_t and the new observation y_{t+1}. The BLUE of β_{t+1} based on the observations $y_1, \dots, y_{t+1}$ is therefore the optimal combination of the following two unbiased estimators of β_{t+1}.

• **a.** *1 point The estimator is the BLUE of β_t before y_{t+1} was available; call this estimator b_t. For the purposes of this recursion b_t is known, it was computed in the previous iteration, and $\text{MSE}[b_t; \beta_t] = \tau^2\psi_t$ is known for the same reason. b_t is not only the BLUE of β_t based on the observations $y_1, \dots, y_t$, but it is also the BLUE of*

β_{t+1} based on the observations $y_1, \ldots, y_t$. Compute MSE$[b_t; \beta_{t+1}]$ *as a function of $\tau^2 \psi_t$.*

Answer. Since $\beta_{t+1} = \beta_t + \omega_{t+1}$ where $\omega_{t+1} \sim (0, \tau^2)$ is independent of b_t, b_t can also serve as a predictor of β_{t+1}, with MSE$[b_t; \beta_{t+1}] = \tau^2(\psi_t + 1)$. □

- **b.** *1 point The second unbiased estimator is the new observation y_{t+1}. What is its* MSE *as a estimator of β_{t+1}?*

Answer. Since $y_{t+1} = \beta_{t+1} + \varepsilon_{t+1}$ where $\varepsilon_{t+1} \sim (0, \sigma^2)$, clearly MSE$[y_{t+1}; \beta_{t+1}] = \sigma^2$. □

- **c.** *3 points The estimation errors of the two unbiased estimators, y_{t+1} and b_t are independent of each other. Therefore use problem 26 to compute the best linear combination of these estimators and the* MSE *of this combination?*

Answer. We have to take their weighted average, with weights proportional to the inverses of the MSE's.

$$b_{t+1} = \frac{\frac{1}{\tau^2(\psi_t+1)} b_t + \frac{1}{\sigma^2} y_{t+1}}{\frac{1}{\tau^2(\psi_t+1)} + \frac{1}{\sigma^2}} = \frac{\kappa^2 b_t + (\psi_t + 1) y_{t+1}}{\kappa^2 + \psi_t + 1} \tag{44}$$

(the second formula is obtained from the first by multiplying numerator and denominator by $\sigma^2(\psi_t + 1)$). The MSE of this pooled estimator is MSE$[b_{t+1}; \beta_{t+1}] = \tau^2 \psi_{t+1}$ where

$$\psi_{t+1} = \frac{\kappa^2(\psi_t + 1)}{\kappa^2 + \psi_t + 1}. \tag{45}$$

(44) and (45) are recursive formulas, which allow to compute ψ_{t+1} from ψ_t, and b_{t+1} from b_t and ψ_t.

-

□

Problem 31. *4 points We are working in a simple Keynesian income-expenditure model of a consumption function with investment i exogenous:*

$$c = \alpha + \beta y + \varepsilon \tag{46}$$

$$y = c + i \tag{47}$$

where the error term ε is independent of investment i. Show that OLS applied to equation (46) gives an estimate which is in the plim *larger than the true β.*

Answer. Plug $c = y - i$ into (46) to get

$$y - i = \alpha + \beta y + \varepsilon \tag{48}$$

$$\text{or} \qquad y(1-\beta) = \alpha + i + \varepsilon \tag{49}$$

$$\text{This gives the reduced form equations} \qquad y = \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} i + \frac{1}{1-\beta}\varepsilon \tag{50}$$

$$\text{and} \qquad c = y - i = \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta} i + \frac{1}{1-\beta}\varepsilon \tag{51}$$

From this one can see

$$\operatorname{cov}(y, \varepsilon) = 0 + 0 + \frac{1}{1-\beta} \operatorname{cov}(\varepsilon, \varepsilon) = \frac{\sigma^2}{1-\beta} \tag{52}$$

Therefore OLS applied to equation (46) gives inconsistent results.

$$\operatorname{plim}\hat{\beta} = \frac{\operatorname{cov}[y,c]}{\operatorname{var}[y]} = \frac{\frac{\beta}{(1-\beta)^2}\operatorname{var}[i] + \frac{1}{(1-\beta)^2}\operatorname{var}[\varepsilon]}{\frac{1}{(1-\beta)^2}\operatorname{var}[i] + \frac{1}{(1-\beta)^2}\operatorname{var}[\varepsilon]} = \tag{53}$$

$$= \frac{\beta\operatorname{var}[i] + \operatorname{var}[\varepsilon]}{\operatorname{var}[i] + \operatorname{var}[\varepsilon]} = \beta + \frac{(1-\beta)\operatorname{var}[\varepsilon]}{\operatorname{var}[i] + \operatorname{var}[\varepsilon]} > \beta$$

□

Problem 32. *It is the purpose of this question to show that the following two vector moving averages are empirically indistinguishable:*

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} = \begin{bmatrix} 1 & 1-L \\ 1 & \beta \end{bmatrix} \begin{bmatrix} \delta_t \\ \varepsilon_t \end{bmatrix} \tag{54}$$

and

$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} = \frac{1}{\sqrt{1+\beta^2}} \begin{bmatrix} 1-(1-\beta)L & 1+\beta-\beta L \\ 0 & 1+\beta^2 \end{bmatrix} \begin{bmatrix} \xi_t \\ \zeta_t \end{bmatrix} \tag{55}$$

where all error terms δ, ε, ξ, and ζ are independent with equal variances σ^2.

- **a.** *Show that in both situations*

$$\mathcal{V}\left[\begin{bmatrix} u_t \\ v_t \end{bmatrix}\right] = \sigma^2 \begin{bmatrix} 3 & 1+\beta \\ 1+\beta & 1+\beta^2 \end{bmatrix}, \qquad \mathcal{C}\left[\begin{bmatrix} u_t \\ v_t \end{bmatrix}, \begin{bmatrix} u_{t-1} \\ v_{t-1} \end{bmatrix}\right] = \sigma^2 \begin{bmatrix} -1 & -\beta \\ 0 & 0 \end{bmatrix} \tag{56}$$

and that the higher lags have zero covariances.

Answer. First scenario: $u_t = \delta_t + \varepsilon_t - \varepsilon_{t-1}$ and $v_t = \delta_t + \beta\varepsilon_t$. Therefore $\text{var}[u_t] = 3\sigma^2$; $\text{cov}[u_t, v_t] = \sigma^2 + \beta\sigma^2$, $\text{var}[v_t] = \sigma^2 + \beta^2\sigma^2$; $\text{cov}[u_t, u_{t-1}] = -\sigma^2$; $\text{cov}[u_t, v_{t-1}] = -\beta\sigma^2$, $\text{cov}[v_t, u_{t-1}] = \text{cov}[v_t, v_{t-1}] = 0$.

Second scenario: leaving out the factor $\frac{1}{\sqrt{1+\beta^2}}$, we have $u_t = \xi_t - (1-\beta)\xi_{t-1} + (1+\beta)\zeta_t - \beta\zeta_{t-1}$ and $v_t = (1+\beta^2)\zeta_t$ Therefore $\text{var}[u_t] = 3\sigma^2$; $\text{cov}[u_t, v_t] = \sigma^2 + \beta\sigma^2$, $\text{var}[v_t] = \sigma^2 + \beta^2\sigma^2$; $\text{cov}[u_t, u_{t-1}] = -\sigma^2$; $\text{cov}[u_t, v_{t-1}] = -\beta\sigma^2$, $\text{cov}[v_t, u_{t-1}] = \text{cov}[v_t, v_{t-1}] = 0$. □

• **b.** *Show also that the first representation has characteristic root $1-\beta$, and the second has characteristic root $\frac{1}{1-\beta}$. I.e., with $\beta < 1$, the first is not invertible but the second is.*

Answer. Replace the Lag operator L by the complex variable z, and compute the determinant:

$$\det \begin{bmatrix} 1 & 1-z \\ 1 & \beta \end{bmatrix} = \beta - (1-z) \tag{57}$$

setting this determinant zero gives $z = 1-\beta$, i.e., the first representation has a root within the unit circle, therefore it is not invertible. For the second representation we get

$$\det \begin{bmatrix} 1-(1-\beta)z & 1+\beta-\beta z \\ 0 & 1+\beta^2 \end{bmatrix} = (1-(1-\beta)z)(1+\beta^2) \tag{58}$$

Setting this zero gives $1-(1-\beta)z = 0$ or $z = \frac{1}{1-\beta}$, which is outside the unit circle. Therefore this representation is invertible. □

Problem 33. *Regression models incorporate seasonality often by the assumption that the intercept of the regression is different in every season, while the slopes remain the same. Assuming* $\boldsymbol{X}$ *contains quarterly data (but the constant term is not incorporated in* $\boldsymbol{X}$*), this can be achieved in several different ways: You may write your model as*

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{C}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \tag{59}$$

Alternatively, you may write your model in the form

$$
\boldsymbol{y} = \boldsymbol{\iota}\alpha + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{K}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\iota} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} \qquad \boldsymbol{K} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \end{bmatrix} \tag{60}
$$

In R *this is the default method to generate dummy variables from a seasonal factor variable. (*`Splus` *has a different default.) This is also the procedure shown in* [Gre97,

p. 383]. *But the following third alternative is often preferrable:*

$$(61) \qquad \boldsymbol{y} = \boldsymbol{\iota}\alpha + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{K}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\iota} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} \qquad \boldsymbol{K} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ \vdots & \vdots & \vdots \end{bmatrix}$$

In R *one gets these dummy variables from a seasonal factor variable if one specifies* `contrast="contr.sum"`.

3 points What is the meaning of the seasonal dummies δ_1, δ_2, δ_3, and of the constant term α or the fourth seasonal dummy δ_4, in models (59), (60), and (61)?

Answer. Clearly, in model (59), δ_i is the intercept in the ith season. For (60) and (61), it is best to write the regression equation for each season separately, filling in the values the dummies take for these seasons, in order to see the meaning of these dummies. Assuming $\boldsymbol{X}$ consists of one column only, (60) becomes

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} \alpha + \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ \vdots \end{bmatrix} \beta + \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \vdots \end{bmatrix}$$

or, written element by element

$$\begin{aligned} y_1 &= 1 \cdot \alpha + x_1 \cdot \beta + 0 \cdot \delta_1 + 0 \cdot \delta_2 + 0 \cdot \delta_3 + \varepsilon_1 && \text{winter} \\ y_2 &= 1 \cdot \alpha + x_2 \cdot \beta + 1 \cdot \delta_1 + 0 \cdot \delta_2 + 0 \cdot \delta_3 + \varepsilon_2 && \text{spring} \\ y_3 &= 1 \cdot \alpha + x_3 \cdot \beta + 0 \cdot \delta_1 + 1 \cdot \delta_2 + 0 \cdot \delta_3 + \varepsilon_3 && \text{summer} \\ y_4 &= 1 \cdot \alpha + x_4 \cdot \beta + 0 \cdot \delta_1 + 0 \cdot \delta_2 + 1 \cdot \delta_3 + \varepsilon_4 && \text{autumn} \end{aligned}$$

therefore the overall intercept α is the intercept of the first quarter (winter); δ_1 is the difference between the spring intercept and the winter intercept, etc.

(61) becomes

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ \vdots \end{bmatrix} \alpha + \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ \vdots \end{bmatrix} \beta + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \vdots \end{bmatrix}$$

or, written element by element

$$\begin{aligned} y_1 &= 1 \cdot \alpha + x_1 \cdot \beta + 1 \cdot \delta_1 + 0 \cdot \delta_2 + 0 \cdot \delta_3 + \varepsilon_1 && \text{winter} \\ y_2 &= 1 \cdot \alpha + x_2 \cdot \beta + 0 \cdot \delta_1 + 1 \cdot \delta_2 + 0 \cdot \delta_3 + \varepsilon_2 && \text{spring} \\ y_3 &= 1 \cdot \alpha + x_3 \cdot \beta + 0 \cdot \delta_1 + 0 \cdot \delta_2 + 1 \cdot \delta_3 + \varepsilon_3 && \text{summer} \\ y_4 &= 1 \cdot \alpha + x_4 \cdot \beta - 1 \cdot \delta_1 - 1 \cdot \delta_2 - 1 \cdot \delta_3 + \varepsilon_4 && \text{autumn} \end{aligned}$$

Here the winter intercept is $\alpha + \delta_1$, the spring intercept $\alpha + \delta_2$, summer $\alpha + \delta_3$, and autmumn $\alpha - \delta_1 - \delta_2 - \delta_3$. Summing this and dividing by 4 shows that the constant term α is the arithmetic mean of all intercepts, therefore δ_1 is the difference between the winter intercept and the arithmetic mean of all intercepts, etc. □

Maximum number of points: 55.

References

[Gre97] William H. Greene, *Econometric analysis*, third ed., Prentice Hall, Upper Saddle River, NJ, 1997. 22

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

E-mail address: ehrbar@econ.utah.edu

URL: http://www.econ.utah.edu/ehrbar