240D Cameron Fall 2012
Department of Economics, U.C.–Davis
Final Exam: March 24 2012

Compulsory. Closed book. 2 hours. Worth 50% of course grade.
Read question carefully so you answer the question.

Question scores (total 50 points and 50% of course grade)

Question	1	2	3	4
Points	12	12	14	12

1. Estimation.
Consider the random variable y with density

$$f(y) = \frac{y^{\alpha-1}\exp(-y/\lambda)}{\lambda^{\alpha}\Gamma(\alpha)}, \quad y > 0, \alpha > 0, \lambda > 0,$$

where $\Gamma(\alpha)$ is the gamma function, i.e. $\Gamma(\alpha) = \int_0^{\infty} e^{-t}t^{\alpha-1}dt$. The derivative of the gamma function is the digamma function $\psi(\alpha)$, i.e. $\Gamma'(\alpha) = \psi(\alpha)$, and its second derivative is the trigamma function, i.e. $\Gamma''(\alpha) = \psi'(\alpha)$. Unless α is an integer there are no closed form solutions for $\Gamma(\alpha)$ and $\psi(\alpha)$, which are evaluated by numerical methods.
The first two moments of y can be shown to be $\mathrm{E}[y] = \alpha\lambda$ and $\mathrm{V}[y] = \alpha\lambda^2$.
To form a regression model we suppose y_i given $\mathbf{x}_i$ has the density $f(y)$ given above, with

$$\begin{aligned} \lambda_i &= \exp(\mathbf{x}_i'\boldsymbol{\beta}) \\ \alpha &= \alpha. \end{aligned}$$

(a) Give the log-likelihood function for $\boldsymbol{\beta}$ and α.

(b) Give the first-order conditions for the MLE of $\boldsymbol{\beta}$ and α.

(c) Give the complete formula for an estimate of the variance matrix of the MLE of $\boldsymbol{\beta}$ and α that is consistent assuming that the density is correctly specified.
[Hint: Think carefully. There are several possible valid estimates. Give one that is easy to derive].

(d) Suppose the density is misspecified, but it is still the case that $\mathrm{E}[y_i|\mathbf{x}_i] = \alpha\lambda_i$ and $\mathrm{V}[y_i|\mathbf{x}_i] = \alpha\lambda_i^2$. Will the part (b) estimates of $\boldsymbol{\beta}$ and α still be consistent? Explain.

(e) Suppose α is known. Say $\alpha = 1$ for simplicity. Give a way to obtain a consistent estimator of $\boldsymbol{\beta}$ based on the conditional moment condition $\mathrm{E}[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta})$.

(f) Now suppose both α and $\boldsymbol{\beta}$ are unknown. Give the objective function for a consistent estimator of α and $\boldsymbol{\beta}$ based on the first two moments of the model of this question.

2. Binary, multinomial and truncation.
Consider a Poisson regression model where the nonnegative integer-valued random variable y^* has density

$$f^*(y^*) = e^{-\mu}\mu^{y^*}/y^*!, \qquad y^* = 0, 1, 2, ...$$

Given regressors $\mathbf{x}$, $\mu_i = \mathrm{E}[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta})$. Data are independent over i.

(a) Suppose we observe only

$$\begin{array}{ll} y_i = 1 & \text{if } y_i^* > 0 \\ y_i = 0 & \text{if } y_i^* = 0 \end{array}$$

Give with justification the objective function for a consistent estimator of $\boldsymbol{\beta}$.

(b) Suppose we observe only

$$\begin{array}{ll} y_i = 2 & \text{if } y_i^* \geq 2 \\ y_i = 1 & \text{if } y_i^* = 1 \\ y_i = 0 & \text{if } y_i^* = 0. \end{array}$$

Derive the log-likelihood function for $\boldsymbol{\beta}$.

(c) Suppose we observe only

$$y_i = y_i^* \quad \text{if } y_i^* \geq 1$$

(when $y^* = 0$ we do not observe anything).
Derive the log-density for $y_i|\mathbf{x}_i$.

(d) Show that for the model of part (c), $\mathrm{E}[y] = \mathrm{E}[y^*|y^* \geq 1] = \mu/(1 - e^{-\mu})$.
[There is some algebra here. Hint: $\sum_{j=0}^{\infty} jf(j) = \sum_{j=1}^{\infty} jf(j)$].

(e) Using the result in part (d), even if you could not derive it, propose a consistent estimator for $\boldsymbol{\beta}$ other than the MLE.

3.(a) Provide a formal definition of convergence in probability.

(b) Give an example of a central limit theorem in the scalar case, stating clearly the assumptions of the central limit theorem and the conclusion.

(c) Consider the following Stata code:

```
set obs 1000
set seed 10101
generate x = rnormal(0,1)
generate ystar = 1 + 2*x + rnormal(0,1*x)
generate y = ystar > 0
tobit y x
```

What is the data generating process for this model? Your answer should use $y, \mathbf{x}$ notation.

(d) Do you see any problems (inconsistency, wrong standard errors, inefficiency, ...) in the estimates that will be obtained from part (c)? Explain.

(e) Present the equations and stochastic assumptions for the Heckman two-step model. There is no need to present the estimator.

(f) For the Poisson regression model you wish to obtain the bootstrap estimate of the standard error of the average marginal effect of a change in a regression x_j on the conditional mean $\mathrm{E}[y|\mathbf{x}]$. Explain how you would do this.

(g) Suppose $\mathrm{E}[\mathbf{h}(\mathbf{w}_i, \boldsymbol{\theta})] = \mathbf{0}$ where the data vector $\mathbf{w}_i$ is $h \times 1$ and the parameter vector $\boldsymbol{\theta}$ is $q \times 1$, $q < h$. The data are independent over i. Give the objective function for the most efficient moment-based estimator in this case.

4. Keep answers brief. Note: Parts (f)-(h) are worth more than parts (a)-(e).
Consider the included table from regression of scalar y_{it} on an intercept and the single regressor x_{it} (coefficient $\boldsymbol{\beta}$) from a balanced panel with $T = 10$ and $N = 532$.

Table 1 Hours and wages: standard linear panel model estimators

	POLS	Between	Within	First Diff	RE-GLS	RE-MLE
α	7.44	7.48	7.22	.00	7.35	7.35
$\boldsymbol{\beta}$	.08	.07	.17	.11	.12	.12
Panel robust se	(.03)	(.02)	(.08)	(.08)	(.05)	(.05)
Default se	{.01}	{.02}	{.03}	{.03}	{.02}	{.01}

Note: Pooled OLS (POLS), Between, Within, First Differences, Random Effects (RE) GLS and ML linear panel regression of lnhrs on lnwg. Standard errors for the slope coefficients are panel robust in parentheses, and usual default estimates that assume iid errors in braces.

(a) Are you surprised by the difference between the panel robust and the default standard error of $\widehat{\beta}$ for the pooled OLS estimator? Give a brief explanation.

(b) Are you surprised by the difference between the panel robust and the default standard error of $\widehat{\beta}$ for the pooled RE-GLS estimator? Give a brief explanation.

(c) Provide the general method to obtain the within estimator.

(d) Give the specific Stata command to obtain the within estimator, with panel robust standard errors.

(e) State the assumptions under which the simple form for the Hausman test for the presence of fixed effects can be performed.

(f) Calculate this form of the Hausman test statistic (based on just β) given the data in the table. What do you conclude?

(g) Now suppose you are told that the assumptions for the simple form of the Hausman test do not hold, but you know $\mathrm{Cov}[\widehat{\beta}_{\mathrm{W}}, \widehat{\beta}_{\mathrm{RE\text{-}GLS}}] = 0.02^2$. Calculate the correct form of the Hausman test (based on just β) given the data in the table. What do you conclude?

(h) Consider pooled OLS estimation if y_{it} on $\mathbf{x}_{it}$ where for this part of the question the intercept is included in $\mathbf{x}_{it}$. Assume errors are independent over i but correlated over t for given i. Provide the formula for the cluster-robust estimate of the variance-covariance matrix of the OLS estimator. [Hint: Stack all observations over t for given i and then proceed.]