Student: ____________________ **Date:** ____________________	**Instructor:** Richeng Piao **Course:** ECON 2560 - Applied Econometrics	**Assignment:** Practice Problem Set 14

1. Which of the following statements describe the possible problems that arise with big data? (*Check all that apply.*)

- ☐ **A.** Due to the presence of many variables, the number of predictors k is larger in comparison to the number of observations n.
- ☐ **B.** Making out of sample predictions is not possible as only cross-sectional data is included with big data while time series data is not.
- ☐ **C.** Big data sets can be nonstandard - for example, containing text or images and these nonstandard data needs to be converted into ı
- ☐ **D.** Testing multiple hypotheses can be difficult as there might be a potentially large set of coefficients representing different treatments.

Answer: A. Due to the presence of many variables, the number of predictors k is larger in comparison to the number of observations n.,
C.
Big data sets can be nonstandard - for example, containing text or images and these nonstandard data needs to be converted into numerical data.
, D.
Testing multiple hypotheses can be difficult as there might be a potentially large set of coefficients representing different treatments.

ID: Concept Exercise 14.1.1

---

2. Which of the following statements describes a problem with testing multiple hypotheses in the case of big data?

- ◯ **A.** For big data, the F-statistic is not well suited for the problem of testing many treatments to find out which of the treatments is effecti
- ◯ **B.** For big data, the F-statistic is not well suited for the problem of testing a single treatment as it is not possible to compute the statisti
- ◯ **C.** For big data, the F-statistic is not well suited for the problem of testing a single treatment to find out whether the treatment is effecti
- ◯ **D.** For big data, the F-statistic is not well suited for the problem of testing many treatments as it is not possible to compute the statistic

Answer: A.
For big data, the F-statistic is not well suited for the problem of testing many treatments to find out which of the treatments is effective.

ID: Concept Exercise 14.1.2

---

3. Using data from a random sample of elementary schools, a researcher regresses average test scores on the fraction of students who qualify for reduced-price meals. The regression indicates a negative coefficient that is highly statistically significant and yields a high R^2.

The given regression (1) ____________ useful in determining the causal effect of school meals on student test scores. This is because in the given scenario, a highly statistically significant coefficient (2) ____________ that OLS estimates are unbiased and consistent since (3) ____________ .

Is this regression useful for predicting test scores?

- **A.** This regression is useful for predicting test scores since a high value of R^2 indicates that the model fits the data well.
- **B.** It is not possible to say if this regression is useful for predicting test scores using information only on the value of R^2.
- **C.** This regression is not useful for predicting test scores since a high value of R^2 indicates regression standard errors are high.
- **D.** Whether this regression is useful for predicting the out-of-sample test scores depends on whether the model uses test samples fron

(1) is / is not

(2) indicates / does not indicate

(3)
- the problems that render the OLS estimates biased and inconsistent will not occur as the R^2 is high
- there may be other variables that affect average test scores that are not included in the regression

Answers (1) is not

(2) does not indicate

(3) there may be other variables that affect average test scores that are not included in the regression

B. It is not possible to say if this regression is useful for predicting test scores using information only on the value of R^2.

ID: Review Concept 14.1

4. Cross-validation uses in-sample observations.

Which of the following statements does not describe how it estimates the MSPE for out-of-sample observations? (*Check all that apply*.)

☐ **A.** The cross-validation procedure treats the data symmetrically by dividing the test sample into randomly chosen, approximately equa leaving out one of the subsamples. The cross-validation estimator of the MSPE is the average of the estimators of the MSPE calcul

☐ **B.** The cross-validation estimation of MSPE is given by $\widehat{\text{MSPE}}_{m-\textit{fold cross validation}} = \frac{1}{n_{test}} \sum_{\substack{\text{observations in} \\ \text{test samples}}} \left(Y_i - \widehat{Y}_i\right)^2$.

☐ **C.** The cross-validation estimation of MSPE is given by $\widehat{\text{MSPE}}_{m-\textit{fold cross validation}} = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{n_i}{n/m}\right) \widehat{\text{MSPE}}_i$.

☐ **D.** The cross-validation procedure treats the data asymmetrically by arbitrarily splitting the observations into two subsamples: an estim prediction errors from the test sample is the cross-validation estimator of MSPE.

Answer:
C. The cross-validation estimation of MSPE is given by $\widehat{\text{MSPE}}_{m-\textit{fold cross validation}} = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{n_i}{n/m}\right) \widehat{\text{MSPE}}_i$., D.

The cross-validation procedure treats the data asymmetrically by arbitrarily splitting the observations into two subsamples: an estimation subsample which is used to calculate $\widehat{Y}$, which is in turn used to calculate the prediction errors of the test subsample. The average of the squared prediction errors from the test sample is the cross-validation estimator of MSPE.

ID: Review Concept 14.2

---

5. Regression coefficients estimated using shrinkage estimators are biased.

Which of the following statements accurately describe why these biased estimators yield more accurate predictions than an unbiased estimator? (*Check all that apply*.)

☐ **A.** The shrinkage estimator reduces the variance of the estimator enough for it to more than compensate for the increase in the square

☐ **B.** The shrinkage estimator introduces bias by shrinking the OLS estimator toward a specific point thereby reducing the variance of the

☐ **C.** The shrinkage estimator introduces bias by shrinking the OLS estimator toward a specific point thereby reducing the mean of the es

☐ **D.** The shrinkage estimator reduces the mean of the estimator, but this reduction is not enough for it to compensate for the increase in

Answer: A.
The shrinkage estimator reduces the variance of the estimator enough for it to more than compensate for the increase in the squared bias, resulting in a lower MSPE than with OLS.
, B.
The shrinkage estimator introduces bias by shrinking the OLS estimator toward a specific point thereby reducing the variance of the estimator.

ID: Review Concept 14.3

---

6. A researcher is interested in predicting average test scores for elementary schools in Arizona. She collects data on three variables from 300 randomly chosen Arizona elementary schools: average test scores ($TestScore^*$) on a standardized test, the fraction of students who qualify for reduced-priced meals (RPM^*), and the average years of teaching experience for the school's teachers ($TEXP^*$). The table below shows the sample means and standard deviations from her sample.

Variable	Sample mean	Sample standard deviation
$TestScore^*$	745.12	66.50
RPM^*	0.58	0.27
$TEXP^*$	13.20	3.70

After standardizing RPM^* and $TEXP^*$ and subtracting the sample mean from $TestScore^*$, she estimates the following regression:

$$\widehat{TestScore} = -48.90 \times RPM + 8.80 \times TEXP, \; SER = 44.0.$$

You are interested in using the estimated regression to predict average test scores for an out-of-sample school with $RPM^* = 0.53$ and $TEXP^* = 10.90$.

The transformed (standardized) values, *RPM* and *TEXP* for this school, are ______ and ______, respectively.

(*Round your answers to two decimal places. Enter a minus sign if your answer is negative.*)

The predicted value of the average test score for this school is ______.

(*Round your answer to two decimal places.*)

If the actual average test score for the school is 777.40, then the error of the prediction is ______.

(*Round your answer to two decimal places.*)

Answers -0.19

-0.62

748.96

28.44

ID: Exercise 14.1

7. We know that the standard error of a regression, (SER) is given by:

$$SER = s_{\hat{u}} = \sqrt{s_{\hat{u}}^2} = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2} = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}\right)^2}{n-2}}.$$

And, the mean squared prediction error, (MSPE) is given by:

$$\widehat{\text{MSPE}} = \frac{1}{n_{test}} \sum_{\textit{observations in test sample}} \left(Y_i^{oos} - \hat{Y}_i\left(X_i^{oos}\right)\right)^2.$$

Which of the following statements correctly describes whether or not there exists a relationship between the standard error of a regression and the square root of the mean squared prediction error?

- **A.** A relationship exists because the estimate of the regressand, $\widehat{Y}_i$ required for calculating both SER and MSPE is calculated from sar
- **B.** A relationship does not exist because while SER measures the spread of the observations around the regression line in a given sar
- **C.** A relationship exists because both SER and the square root of MSPE are described as the square roots of the sum of squared erro
- **D.** A relationship does not exist because SER is calculated with $n-2$ degrees of freedom while MSPE is calculated with n degrees of f

Answer: B.

A relationship does not exist because while SER measures the spread of the observations around the regression line in a given sample, MSPE is calculated to make a prediction for an observation not in the data set.

ID: Exercise 14.4

---

8. Health insurance companies are generally faced with the problem of determining the appropriate premium. Generally, the premium charged by an insurance company (*P*, measured in hundred dollars) is decided on the basis of the age of the customer (*A*) and the duration for which the insurance is taken (*D*). A health insurance company, Healtek, collects data on 91 randomly selected customers. The estimated regression function is:

$$\hat{P} = \hat{\beta}_0 + \hat{\beta}_1 A + \hat{\beta}_2 D.$$

The company wants to estimate the *m*-fold cross validation MSPE using in-sample observations to predict the out-of-sample (oos). It divides the sample of 91 customers into three subsamples of 50, 25, and 16 customers each. The calculated values of $\sum\left(P_i^{oos} - \hat{P}_i\right)^2$ for the first, second, and third subsamples are 30.54, 23.47, and 16.71, respectively.

The *m*-fold cross validation MSPE calculated by Healtek is [ ].

(*Round your answer to four decimal places.*)

The *m*-fold cross validation MSPE estimated by another insurance company, Tensurance, after performing a similar experiment is 1.9563.

The predictive model used by (1) ____________ is better in terms of predictive accuracy.

(1) ○ Healtek
○ Tensurance

Answers 0.7771

(1) Healtek

ID: Concept Exercise 14.2.1

---

9. We have the following standardized predictive regression model:

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i.$$

Which of the following formulas is used to estimate the MSPE of OLS in the special case that the regression error u is homoskedastic?

- **A.** $\widehat{\text{MSPE}} = \sigma_u^2 + E\left[\left(\hat{\beta}_1 - \beta_1\right) X_1^{oos} + \ldots + \left(\hat{\beta}_k - \beta_k\right) X_k^{oos}\right]^2.$
- **B.** $\widehat{\text{MSPE}} = \frac{1}{n}\sum\left(Y_i - \hat{Y}_i\right)^2.$
- **C.** $\widehat{\text{MSPE}} \cong \left(1 + \frac{k}{n}\right)\sigma_u^2.$
- **D.** $\widehat{\text{MSPE}} = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{n_i}{n/m}\right)\widehat{\text{MSPE}}_i.$

The best possible prediction of the MSPE is called the (1) __________. From the perspective of minimizing the MSPE, the best possible prediction is given by the (2) __________.

(1)
- minimum variance prediction
- oracle prediction

(2)
- conditional mean
- conditional variance

Answers C. $\widehat{\text{MSPE}} \cong \left(1 + \frac{k}{n}\right)\sigma_u^2.$

(1) oracle prediction

(2) conditional mean

ID: Concept Exercise 14.2.2

---

10. The ridge regression estimator minimizes the sum of squared residuals, plus a penalty factor that increases with the sum of the squared coefficients. Suppose a data sample is such that the penalized sum of squared residuals, $S^{Ridge}\left(b; \lambda_{Ridge}\right) = 5{,}615$; the sum of squared residuals, $\sum_{i=1}^{n}\left(u_i\right)^2 = 200$, where $u_i = Y_i - b_1 X_{1i} - \ldots - b_k X_{ki}$; and, the sum of the squared coefficients or the penalty term, $\sum_{j=1}^{k} b_j^2 = 95.$

The ridge shrinkage parameter, λ_{Ridge} is __________.

(*Round your answer to four decimal places*.)

Suppose that the regressors are uncorrelated and the OLS estimator $\hat{\beta}_j$ is 1.97.

The ridge regression estimator, $\hat{\beta}_j^{Ridge}$, is __________.

(*Round your answer to four decimal places*.)

Answers 57

0.034

ID: Concept Exercise 14.3.1

11.
If the ridge shrinkage parameter $\lambda_{Ridge} \to \infty$, the ridge regression estimator equals (1) ____________.

Which of the following statements describes the best method of measuring the ridge estimator?

- **A.** The best ridge estimator is calculated by measuring λ_{Ridge} such that it minimizes $S^{Ridge}\left(b; \lambda_{Ridge}\right)$.
- **B.** The best ridge estimator is calculated by measuring λ_{Ridge} arbitrarily.
- **C.** The best ridge estimator is calculated by measuring λ_{Ridge} such that it minimizes the estimated MSPE.
- **D.** The best ridge estimator is calculated by measuring the OLS estimator as it is always the best ridge estimator.

(1)
- the OLS estimator
- 1
- 0

Answers (1) 0

C. The best ridge estimator is calculated by measuring λ_{Ridge} such that it minimizes the estimated MSPE.

ID: Concept Exercise 14.3.2

---

12. Ridge regression and Lasso are two regression estimators based on penalization.

Which of the following statements describe how they are similar and how they differ? (*Check all that apply*.)

- **A.** The ridge regression estimator minimizes the sum of squared residuals without the added penalty and the Lasso estimator minimi:
- **B.** The ridge regression penalty increases with the sum of the squared coefficients and the Lasso penalty increases with the sum of tl
- **C.** Both the ridge regression and the Lasso estimate many coefficients to be exactly 0, thereby excluding them from the prediction prı
- **D.** With ridge and Lasso, the regression fit and the estimated coefficients depend on the specific regressors chosen for the linear cor

Answer: B.

The ridge regression penalty increases with the sum of the squared coefficients and the Lasso penalty increases with the sum of the absolute values of the coefficients.

, D.

With ridge and Lasso, the regression fit and the estimated coefficients depend on the specific regressors chosen for the linear combinations.

ID: Review Concept 14.4

---

13. A regression model in which the coefficients are nonzero for only a small fraction of the predictors is called a sparse model.

The (1) ______________ estimator is designed for sparse models.

Let k denote the number of regression coefficients and n denote the sample size in a sparse model.

The Lasso can be applied when (2) ______________.

Which of the following statements are true about the Lasso regression? (*Check all that apply*.)

- ☐ **A.** The Lasso estimates many coefficients to be exactly 0, thereby excluding them from the prediction process.
- ☐ **B.** If the shrinkage parameter is zero, the Lasso estimate would just be OLS estimate.
- ☐ **C.** Unlike the ridge penalty, the Lasso penalty increases with the square of regression coefficients.
- ☐ **D.** The Lasso has a shrinkage parameter that can be estimated by minimizing the cross-validated MSPE.

(1) ○ ridge
○ Lasso

(2) ○ $k < n$
○ $k = n$
○ $k > n$

Answers (1) Lasso

(2) $k > n$

A. The Lasso estimates many coefficients to be exactly 0, thereby excluding them from the prediction process., B.
If the shrinkage parameter is zero, the Lasso estimate would just be OLS estimate., D.
The Lasso has a shrinkage parameter that can be estimated by minimizing the cross-validated MSPE.

ID: Concept Exercise 14.4.1

---

14. Which of the following statements about the Lasso are true with respect to the OLS estimates? (*Check all that apply*.)

- ☐ **A.** For a single regressor, when the OLS estimator is far from the true population parameter value, the Lasso estimator shrinks it towa
- ☐ **B.** For a single regressor, when the OLS estimator is far from zero, the Lasso estimator shrinks it toward 0; and, when the OLS estim
- ☐ **C.** When the OLS estimator is large, the Lasso shrinks it more than ridge, but when the OLS estimator is small, the Lasso shrinks it le
- ☐ **D.** When the OLS estimator is large, the Lasso shrinks it less than ridge, but when the OLS estimator is small, the Lasso shrinks it mo

Answer: B.
For a single regressor, when the OLS estimator is far from zero, the Lasso estimator shrinks it toward 0; and, when the OLS estimator is sufficiently small, the Lasso estimator becomes exactly 0.
, D.
When the OLS estimator is large, the Lasso shrinks it less than ridge, but when the OLS estimator is small, the Lasso shrinks it more than ridge.

ID: Concept Exercise 14.4.2

---

15. Suppose a data set with 10 variables produces a scree plot that is flat.

This indicates that there is a (1) ____________ correlation between the variables.

What does this suggest about the usefulness of using the first few principal components of these variables in a predictive regression?

- **A.** It suggests that one can replace the many predictors in the predictive model with far fewer principal components and use those pri
- **B.** It suggests that one can replace the few predictors in the predictive model with far more principal components and use the first few
- **C.** It suggests that one can replace the few predictors in the predictive model with far more principal components and use the first few
- **D.** It suggests that one can replace the many predictors in the predictive model with far fewer principal components and use those pri

(1) high
low

Answers (1) high

D.
It suggests that one can replace the many predictors in the predictive model with far fewer principal components and use those principal components as regressors and estimate the coefficients using OLS more reliably.

ID: Review Concept 14.5

---

16. A researcher is interested in estimating the relation between the total house rent a family pays (Y, measured in dollars) as a function of the size of the house (X_1, measured in square feet) and the number of rooms in the house (X_2) by surveying 200 prospective locations. Both X_1 and X_2 are standard normal variables. The estimated regression function is as follows:

$$Y = 11.25 + 1.12X_1 + 1.98X_2.$$

The correlation between X_1 and X_2 was calculated to be 0.74. Suppose the size of the house is 4.12 and the number of rooms in the house is 5.

Let PC_1 and PC_2 denote the first and the second principal components of the variables X_1 and X_2, respectively.

The variance of the first principal component, $\text{var}(PC_1)$, will be ____.

The variance of the second principal component, $\text{var}(PC_2)$, will be ____.

(*Round your answers to two decimal places.*)

The first principal component explains ____% of the variance of X_1 and X_2.

The second principal component explains ____% of the variance of X_1 and X_2.

(*Round your answers to whole numbers.*)

Answers 1.74

0.26

87

13

ID: Concept Exercise 14.5.1

17. Suppose a researcher is estimating a regression equation with two variables, X_1 and X_2, which are both standard normal variables.

Let PC_1 and PC_2 denote the first and the second principal components of X_1 and X_2, respectively.

The formulas for calculating the first and the second principal components are (1) __________ and (2) __________, respectively.

Which of the following statements regarding the variance of the first and the second principal components is true?

- **A.** When there are only two variables, the first principal component maximizes the variance of the linear combination of the variables,
- **B.** When there are only two variables, both the first and the second principal components maximize the variance of the linear combin
- **C.** When there are only two variables, both the first and the second principal components minimize the variance of the linear combina
- **D.** When there are only two variables, the first principal component minimizes the variance of the linear combination of the variables,

Suppose a researcher is estimating a regression equation with k variables, $X_1, X_2, \ldots, X_k$, where all the k variables are standard normal variables.

Let n denote the sample size.

Which of the following is not a property of the principal components of the k variables?

- **A.** The sum of the variances of the principal components equals the sum of the variances of the X's.
- **B.** The squared weights of the linear combinations sum to 1.
- **C.** The j^{th} principal component maximizes the variance of its linear combination, subject to it being uncorrelated with the first $(j-1)$ pr
- **D.** Assuming there is no perfect multicollinearity in X, the number of principal components is the maximum of n and k.

(1)		(2)	
	$\frac{(X_1+X_2)}{\sqrt{2}}$		$\frac{(X_1+X_2)}{\sqrt{2}}$
	$\frac{X_1+X_2}{2}$		$\frac{(X_1-X_2)}{\sqrt{2}}$
	$\frac{(X_1-X_2)}{\sqrt{2}}$		$\frac{(X_1+X_2)}{\sqrt{2}}$

Answers (1) $\frac{(X_1+X_2)}{\sqrt{2}}$

(2) $\frac{(X_1-X_2)}{\sqrt{2}}$

A.
When there are only two variables, the first principal component maximizes the variance of the linear combination of the variables, while the second principal component minimizes the variance of the linear combination of the variables.

D. Assuming there is no perfect multicollinearity in X, the number of principal components is the maximum of n and k.

ID: Concept Exercise 14.5.2

18. Which of the following statements is not true?

- **A.** It is the tendency of Lasso to shrink less than ridge for small coefficients but to shrink more than ridge for large ones.
- **B.** With many regressors, OLS tends to fit individual observations by large coefficients on specific variables.
- **C.** In a scatterplot of the actual values versus the predicted values, the tighter the spread of the scatter along the 45° line, the better t
- **D.** None of the coefficients on standardized regressors estimated either through OLS or many-predictor methods has a causal interpr

Answer: A. It is the tendency of Lasso to shrink less than ridge for small coefficients but to shrink more than ridge for large ones.

ID: Concept Exercise 14.6.1

---

19. Which of the following statements is not true in case of large data sets?

- **A.** The three many-predictor methods always perform equally well in these data.
- **B.** The many-predictor methods like ridge or Lasso succeed where OLS fails.
- **C.** The *m*-fold MSPE is close to the MSPE computed using the reserved test sample.
- **D.** The many-predictor methods allow the coefficient estimates to be biased in a way that reduces their variance by enough to compensate for the increased bias.

Answer: A. The three many-predictor methods always perform equally well in these data.

ID: Concept Exercise 14.6.2