1 Old problem sets

These are not assigned, but they will give you an example of what to expect on this year's problem sets.

1.1 Problem set 1: Notation and asymptotics

To do some of these problems you might need to consult some of the references.

1. For p a $n \times 1$ vector with typical element $p_t = [1 + \exp(-\mathbf{x}_t'\theta)]^{-1}$,where $\mathbf{x}_t$ and θ are $K \times 1$ vectors, and for s a $n \times 1$ "sum vector" that has each element equal to 1, write
$$\frac{\partial}{\partial\theta} s' \ln(p)$$
in matrix notation.

2. Show that $\stackrel{a.s.}{\to} \Rightarrow \stackrel{p}{\to}$.

3. Show that $\stackrel{d}{\to}$ does not imply $\stackrel{p}{\to}$.

4. Given $\hat{\beta}_n = (X'X)^{-1}(X'Y)$, what are the orders of the terms in parentheses?

   (a) Re-write the expression, using powers of n and any appropriate substitutions, so that you obtain an expression that contains individual terms that are $O_p(1)$ and $o_p(1)$.

5. Can a random variable that is $O_p(2)$ be asymptotically normally distributed? Why or why not?

6. The maximum likelihood estimator $\hat{\theta} = \arg\max s_n(\theta) = \frac{1}{n} \ln \mathcal{L}(\theta)$. Let $g_n(\theta) = \frac{\partial}{\partial\theta} s_n(\theta)$ and let $H_n(\theta) = \frac{\partial^2}{\partial\theta\partial\theta'}$ be the gradient and the Hessian, respectively. Using an exact second order Taylor's expansion, and noting that $g_n\left(\hat{\theta}\right) \equiv 0$, one can write
$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = -H_n(\theta^*)^{-1}\sqrt{n} g_n(\theta_0)$$
where $\theta^* = \lambda\hat{\theta} + (1-\lambda)\theta^0$, $0 < \lambda < 1$. We can write this in turn as
$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = -\left[H_n(\theta^0) + A_n\right]^{-1} \sqrt{n} g_n(\theta_0).$$
What are the orders of $H_n(\theta^0)$, A_n and $\sqrt{n} g_n(\theta_0)$? Explain how you determine this. Assume the usual regularity conditions.

7. Give an example of a function that converges pointwise to a limiting function, but which does not converge uniformly.

1.2 Problem Set 2: Extremum estimators

1. Suppose that $x_i \sim$ uniform(0,1), and $y_i = 1 - x_i^2 + \varepsilon_i$, where ε_i is iid$(0,\sigma^2)$. Suppose we estimate the misspecified model $y_i = \alpha + \beta x_i + \eta_i$ by OLS. Find the numeric values of α^0 and β^0 that are the probability limits of $\hat{\alpha}$ and $\hat{\beta}$

2. Suppose that y_t conditional on $\mathbf{x}_t$ is independently distributed Poisson. A Poisson random variable is a *count data* variable, which means it can take the values $\{0,1,2,...\}$. The Poisson density is

$$f(y_t) = \frac{\exp(-\lambda_t)\lambda_t^{y_t}}{y_t!}, y_t \in \{0,1,2,...\}.$$

The mean of y_t is λ_t, as is the variance. Note that λ_t must be positive. Suppose that the true mean is

$$\lambda_t^0 = \exp(\mathbf{x}_t'\beta^0),$$

which enforces the positivity of λ_t. Suppose we estimate β^0 by nonlinear least squares:

$$\hat{\beta} = \arg\min s_n(\beta) = \frac{1}{T}\sum_{t=1}^{n}\left(y_t - \exp(\mathbf{x}_t'\beta)\right)^2$$

   (a) Show that $\hat{\beta}$ is consistent. You may assume without verification that a LLN for independent r.v.'s may be applied as needed.

   (b) Determine the limiting distribution of $\sqrt{n}\left(\hat{\beta} - \beta^0\right)$. This means finding the the specific forms of $\frac{\partial^2}{\partial\beta\partial\beta'}s_n(\beta)$, $\mathcal{J}(\beta^0)$, $\left.\frac{\partial s_n(\beta)}{\partial\beta}\right|$, and $\mathcal{I}(\beta^0)$. Again, use a CLT as needed, no need to verify that it can be applied.

3. Use the asymptotic normality theorem to find the asymptotic distribution of the ML estimator of β^0 for the model $y = x\beta^0 + \varepsilon$, where $\varepsilon \sim N(0,1)$ and is independent of x. This means finding $\frac{\partial^2}{\partial\beta\partial\beta'}s_n(\beta)$, $\mathcal{J}(\beta^0)$, $\left.\frac{\partial s_n(\beta)}{\partial\beta}\right|$, and $\mathcal{I}(\beta^0)$.

4. To answer this question you should use Mathematica, or some similar program. Suppose that data is generated according to the probit model, $Pr(y = 1|x) = \Phi(-0.5 + x)$ where $\Phi(\cdot)$ is the N(0,1) distribution function, x is IID

uniform[0,1]. Mistakenly, we apply a logit model $Pr(y = 1|x) = 1/[1 + exp(-\alpha - \beta x)]$.

(a) find the probability limits of α, β.

(b) find $limVar\sqrt{n}(\beta - \beta^0)$

(c) Now repeat the exercise (several times) using different true values for the probit parameters. Is there an approximate linear relationship between the true values and the logit probability limit?

1.3 Problem set 3: Numeric methods

1. Work through the second computer lab on the web page. This illustrates various methods to

2. Run the mathematica notebook judge.nb, using VNC on Babial, to see how to calculate derivatives. This would not be hard to do by hand, but for more complicated problems it is sometimes nice to get derivatives with the help of the computer.

   - elements of the gradient
     $\frac{\partial}{\partial\alpha}\sum_{t=1}^{n}\left(y_t - \alpha - \beta x_t - \beta^2 z_t\right)^2 = \sum_{t=1}^{n}(-2y_t) + 2n\alpha + \sum_{t=1}^{n} 2\beta x_t + \sum_{t=1}^{n} 2\beta^2 z_t$
     $\frac{\partial}{\partial\beta}\sum_{t=1}^{n}\left(y_t - \alpha - \beta x_t - \beta^2 z_t\right)^2 = \sum_{t=1}^{n}(-2y_t x_t) + \sum_{t=1}^{n}(-4y_t\beta z_t) + \sum_{t=1}^{n} 2\alpha x_t + \sum_{t=1}^{n} 4\alpha\beta z_t + \sum_{t=1}^{n} 2\beta x_t^2 + \sum_{t=1}^{n} 6\beta^2 x_t z_t + \sum_{t=1}^{n} 4\beta^3 z_t^2$
   - elements of the Hessian
     $\frac{\partial^2}{\partial\alpha^2}\sum_{t=1}^{n}\left(y_t - \alpha - \beta x_t - \beta^2 z_t\right)^2 = 2n$
     $\frac{\partial}{\partial\alpha}\frac{\partial}{\partial\beta}\sum_{t=1}^{n}\left(y_t - \alpha - \beta x_t - \beta^2 z_t\right)^2 = \sum_{t=1}^{n} 2x_t + \sum_{t=1}^{n} 4\beta z_t$
     $\frac{\partial^2}{\partial\beta^2}\sum_{t=1}^{n}\left(y_t - \alpha - \beta x_t - \beta^2 z_t\right)^2 = \sum_{t=1}^{n}(-4y_t z_t) + \sum_{t=1}^{n} 4\alpha z_t + \sum_{t=1}^{n} 2x_t^2 + \sum_{t=1}^{n} 12\beta x_t z_t + \sum_{t=1}^{n} 12\beta^2 z_t^2$

3. Now run judge2.ox to see how one can use analytic gradients to speed up convergence.

4. Estimate logit and probit models using the data provided on the web page, using your favorite estimation software. The probit model specifies the bid acceptance probability as

$$\Pr(y = 1) = \Phi(\alpha + \beta A).$$

where $\Phi(.)$ is the standard normal distribution function. Calculate estimated compensating variation for the two models. One means of doing this would be to modify the judge1.ox program to read in the proper data and to provide the appropriate likelihood functions to maximize.

1.4 Problem set 4: GMM

1. Using Ox or some other econometrics package, generate data from the logit dgp (for example using the logit_dgp.ox routine provided on the class web page). Recall that $E(y_t|\mathbf{x}_t) = \mathbf{p}(\mathbf{x}_t, \theta) = [1 + \exp(-\mathbf{x}_t\prime\theta)]^{-1}$. Consider the moment condtions (exactly identified):

$$m_t(\theta) = [y_t - p(\mathbf{x}_t, \theta)]\mathbf{x}_t.$$

   (a) Estimate by GMM. Estimate by MLE. The two estimators should coincide (why?).

2. For the logit model,

   (a) prove that the MLE first order conditions reduce to

$$\frac{1}{n}\sum_{t=1}^{n}[y_t - p(\mathbf{x}_t, \hat{\theta})]\mathbf{x}_t \equiv 0$$

   (b) Show that $\mathbf{x}_t$ are the optimal instruments when the conditional moment is chosen to be $E[(y_t - p(\mathbf{x}_t, \theta)|\mathbf{x}_t] = 0$.

3. Suppose one wasn't aware of the fact that $\mathbf{x}_t$ are optimal instruments, and instead one estimates by GMM using as instruments $\mathbf{x}_t$ and all of the variables in $\mathbf{x}_t$ (except the column of ones) squared.

   (a) Use the specification test to see if the moment conditions appear to be well-specified.

   (b) Using a Monte Carlo experiment, compare the variances of the GMM estimators based upon the two sets of moment conditions. In the course of the Monte Carlo experiment, compare the nominal and actual sizes of the specification test.

4. What is the asymptotic distribution of $n \cdot m(\theta^0)'\Omega^{-1}m(\theta^0)$, supposing the number of moment conditions g is greater than the number of parameters (*i.e.*, we have overidentification.)?

5. Verify the missing steps needed to show that $n \cdot m(\hat{\theta})'\hat{\Omega}^{-1}m(\hat{\theta})$ has a $\chi^2(g-K)$ distribution. That is, show that the monster matrix is idempotent and has trace equal to $g-K$.

2 Old exams

Second year econometrics, final exam

(IDEA, UAB, Fall 1994, Prof. M. Creel)

Answer question 1 and three questions from 2-6. Total time 3 hours.

1. Consider the nonlinear model $y_t = f(x_t, \theta^0) + \varepsilon_t$, where $\varepsilon_t \sim N(0, \sigma^2), \forall t$, but $E(x_t\varepsilon_t) \neq 0$.

   (a) Show that the NLS estimator $\hat{\theta}$ will not be consistent.

   (b) Suggest a method of estimating θ^0 consistently. Give an brief, informal argument explaining why the estimator is consistent.

   (c) Find the asymptotic variance-covariance matrix of the proposed estimator and explain how to estimate it consistently.

2. Funtional forms:

   (a) Define Diewert flexibility.

   (b) Define Sobolev flexibility.

   (c) Explain why models based upon a Diewert flexible functional forms do not in general lead to consistent estimates.

   (d) Explain why models based upon a Sobolev flexible functional form do in general lead to consistent estimates.

3. Explain a method of finding the θ that minimizes $s(\theta) = \| y - f(x,\theta) \|^2$ where θ is a $K \times 1$ vector.

4. Given the models $M_0 : y = x\beta + \varepsilon$ and $M_1 : y = z\gamma + \eta$, supposing M_1 is the dgp, show that the P-test statistic will reject model M_0 with probability tending to 1 as the sample size tends to infinite.

5. Given the model

$$
\begin{aligned}
y_{it} = \alpha_i + x_{it}\beta + \varepsilon_{it} \\
E(\varepsilon_{it}\varepsilon_{js}) &= \sigma_{ij}, t = s \\
&= 0, t \neq s \\
E(x_{it}\varepsilon_{js}) = 0, \forall i, t, j, s \\
t, s \in \{1, 2, ..., T\} \\
i, j \in \{1, 2, ...G\} \\
\varepsilon_{it} \sim N(0, \sigma_{it}), \forall i, t
\end{aligned}
$$

   (a) a) Write the model in matrix form $Y = X\theta + V$, noting the structure of each matrix (vector).
   (b) b) Indicate the structure of Σ, the variance-covariance matrix of V.
   (c) c) Indicate how to consistently estimate every element of Σ.
   (d) d) Explain how to estimate θ asymptotically efficiently, informally noting why the estimator has this property.

6. Given a model $y = \alpha + \beta x + \varepsilon$, x a scalar random variable, explain how to construct an asymptotic 95% confidence interval for the elasticity of y with respect to x.

Second year econometrics, final exam

(IDEA, UAB, Fall 1995, Prof. M. Creel)

Answer two questions from {1,2,3}, and two questions from {4,5,6}.

1. Consider the logit model for a random variable y that can take on the values 0 or 1 :
$$\Pr(y_t = 1) = [1 + \exp(-x_t\beta)]^{-1}$$
Suppose we have a sample of size T.
   (a) Write the log-likelihood function.
   (b) Define moment conditions $g(\beta) = \frac{1}{T}\sum_{t=1}^{T} h_t(\beta)$ which are equivalent to MLE. The individual terms $h_t(\beta)$ should be defined explicitly.
   (c) Find the efficient weighting matrix for GMM estimation.

2. The Poisson distribution for count data (data where $y \in \{0,1,2,...\}$) is
$$f_Y(y) = \frac{\exp(-\lambda)\lambda^y}{y!}, \lambda > 0.$$
The distribution has mean λ and variance λ. To explain y as a function of X, the parameterization $\lambda = \exp(X\beta)$ has been used by many authors. This ensures that the mean is positive. Explain how one could construct an *overidentified* GMM estimator.

3. Consider the model
$$y = X\beta + \varepsilon$$
$$\varepsilon \sim N(0,\Sigma)$$
where Σ is of unknown form. Explain how to test the hypothesis $R\beta = r$ in a way such that the test statistic has a known asymptotic distribution under the null hypothesis. Explain why the test has this property.

4. Explain either the Gauss-Newton or Newton-Raphson methods of optimizing a function. Your explanation should include:
   (a) A full description of the algorithm

(b) An explanation of why the algorithm might converge to the optimal point

(c) A description of circumstances under which the algorithm might not work

5. Discuss the Nadaraya-Watson kernel regression estimator. What is the effect of the choice of the window width parameter on the estimator?

6. Suppose $y_t = X_t\beta + \varepsilon_t$, and ε_t is an MA(3) process. Discuss how to estimate β asymptotically efficiently, making clear the formulas for the estimators of all parameters. Explain briefly and informally why the estimators proposed are consistent.

Second year econometrics, final exam

(IDEA, UAB, Fall 1996, Prof. M. Creel)

Answer 5 of the following questions.

1. The method of simulated moments estimator based upon H simulation draws has an asymptotic variance that is $\left(1+\frac{1}{H}\right)$ times that of the equivalent method of moments estimator (supposing it were feasible). Explain why this is the case.

2. Suppose the true conditional density for y given $\mathbf{x}$ is $p(y|\mathbf{x},\theta^0)$. Consider a misspecified density $f(y|\mathbf{x},\phi)$. Suppose we have random sampling (that is, independently and identically distributed observations).

   (a) Define the quasi-maximum likelihood estimator, $\hat{\phi}$.

   (b) To what does $\hat{\phi}$ converge? Is $\hat{\phi}$ consistent for θ^0? Why or why not?

   (c) Give an example of a QML estimator that *is* consistent for at least some elements of the true parameter vector.

3. Explain why the covariance matrix of the QML estimator may not be consistently estimable. (That is, discuss circumstances under which we will have difficulites defining a consistent estimator).

4. Suppose that data is generated by random sampling of

$$y_t = f(\mathbf{x}_t,\theta^0)+\varepsilon_t$$

   but that

$$\mathcal{E} f(\mathbf{x},\theta)\varepsilon = B(\theta)$$

   and

$$\arg\min_{\Theta} B(\theta) = \theta^* \neq \theta^0$$

   (a) Is the NLS estimator consistent? Explain your answer.

   (b) Suppose that there exists a vector of variables $\mathbf{z}_t$ such that

$$\begin{aligned}\mathcal{E}(\mathbf{z}_t\varepsilon_t) &= 0\\ \mathcal{E}\left(\varepsilon_t^2|\mathbf{z}_t\right) &= \sigma_0^2\end{aligned}$$

Propose a GMM estimator that is consistent, give its asymptotic distribution, and indicate how to estimate the asymptotic covariance matrix consistently.

5. Suppose we wish to estimate an MA(2) process

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

where the ε_t are iid $N(0, \sigma^2)$ random variables. Although estimation is possible using ML, explain how you could estimate the parameters of the model using indirect inference. Also indicate how the model's specification can be tested using indirect inference.

6. Suppose that you have binary response data generated by random sampling of

$$\begin{aligned} y_i^* &= g(\mathbf{x}_i, \theta^0) - \varepsilon_i \\ y_i &= 1(y_i^* > 0) \\ \Pr(\varepsilon_i &< z) = F(z) \end{aligned}$$

Explain how one could estimate $\Pr(y_i = 1|\mathbf{x}_i)$ consistently *without* knowledge of the form of the functions $g(\mathbf{x}_i, \theta^0)$ and $F(z)$. Explain informally why the proposed estimator has this property.

7. Define stochastic equicontinuity and explain its usefulness in showing the consistency of extremum estimators.

Second year econometrics, final exam
(IDEA, UAB, Fall 1997, Prof. M. Creel)
Answer 6 of the following 7 questions.

1. Discuss the following propositions.

   (a) Every GMM estimator can be thought of as an extremum estimator.

   (b) Every extremum estimator can be thought of as a GMM estimator.

2. *OLS as a GMM estimator.* Assume that the model

$$y_t = \mathbf{x}_t'\beta^0 + \varepsilon_t$$

   satisfies the classical assumptions so that the OLS estimator is consistent and fully asymptotically efficient. Present a GMM estimator that is equivalent to the OLS estimator. Show that the GMM estimator is consistent and derive its asymptotic distribution.

3. Suppose we estimate using GMM with conditional moment restrictions

$$m_t(\theta) = \mathbf{z}_t'h(y_{t,}\mathbf{x}_t,\theta)$$

   where

$$\begin{aligned} \mathcal{E}_{\theta^0} m_t(\theta) &= 0, \theta = \theta^0 \\ \mathcal{E}_{\theta^0} m_t(\theta) &\neq 0, \theta \neq \theta^0. \end{aligned}$$

   (a) Explain how one can estimate the asymptotic variance of $\hat{\theta}$ assuming that the $\mathbf{z}_t$ are not optimal instruments, that $\dim(\mathbf{z}_t) > \dim(\theta)$, and that the optimal weighting matrix is used.

   (b) Proposition: the choice of the weighting matrix W is unimportant if optimal instruments are used. True or false? Provide discussion.

4. *Logit model with unobserved heterogeneity*. Suppose we have i.i.d. observations on $(y_t, \mathbf{x}_t)$ where y_t is a binary 0/1 variable and $\mathbf{x}_t$ is a vector of exogenous variables. Suppose that

$$\Pr(y_t = 1) = \Lambda(\mathbf{x}_t'\beta^0 + \sigma^0\varepsilon_t)$$

where ε_t is an unobserved variable that is distributed i.i.d. as

$$\varepsilon_t \sim N(0,1).$$

Discuss **two** different means of estimating the parameters (β^0, σ^0) consistently. Note, the function

$$\Lambda(z) = (1 + \exp(-z))^{-1}.$$

5. Discuss the following two propositions, which may be true or false.

   (a) The GMM estimator is less susceptible to problems of misspecification than is the maximum likelihood estimator, since the GMM estimator is based upon fewer distributional assumptions.

   (b) The above proposition applies to the simulated versions of GMM and MLE.

6. The GMM estimator, which minimizes $m_n(\theta)' W_n m_n(\theta)$, is most efficient, asymptotically, if W_n is put equal to a consistent estimator of $\Omega_\infty = \lim_{n \to \infty} Var \sqrt{n} m_n(\theta^0)$. If one is interested in the small sample performance of the estimator, should one always use a consistent estimator of Ω_∞ as the weighting matrix?

7. The asymptotic efficiency of the GMM estimator is never reduced if one adds additional moment conditions. If one is interested in the small sample performance of the estimator, should one always use additional moment restrictions if available?

Econometrics Exam, 9 Jan. 2001

Answer 4 out of the following 5 questions.

1. The asymptotic covariance matrix of the method of simulated moments, based on H simulations, is $1+1/H$ times that of the GMM estimator that uses the same moments, but without simulation (assume that this is feasible). Explain why this is the case.

2. Assume that y conditional on $\mathbf{x}$ follows the Poisson density

$$\begin{aligned} f_{y|\mathbf{x}}(y|\mathbf{x},\beta) &= \frac{exp(-\lambda)\lambda^y}{y!} \\ \lambda &= exp(\mathbf{x}\beta) \end{aligned}$$

   (a) Explain what are the moment conditions that lead to fully asymptotically efficient estimation of β.

   (b) What is the efficient weighting matrix corresponding to the efficient moments?

3. Assume the model

$$\begin{aligned} y &= \alpha+\beta x+\varepsilon \\ \varepsilon &\sim IIN(0,1) \\ x &\sim U(0,10) \end{aligned}$$

   is correctly specified. Assume that data are selected such that the sample only contains obervations such that $x>5$. Is the OLS estimator using this selected sample consistent? Explain.

4. Explain how to use cross validation to choose the window width for a kernel regression estimator.

5. Assume that y conditional on $\mathbf{x}$ follows the Poisson density

$$\begin{aligned} f_{y|\mathbf{x}}(y|\mathbf{x},\beta) &= \frac{exp(-\lambda)\lambda^y}{y!} \\ \lambda &= exp(\mathbf{x}\beta) \end{aligned}$$

   The conditional mean of y is $E(y|\mathbf{x}) = exp(\mathbf{x}\beta)$. As such we can write

$$y = exp(\mathbf{x}\beta)+\varepsilon$$

where $E(\varepsilon|\mathbf{x}) = 0$. Obviously, ε is not normally distributed. Is the quasi-ML estimator of the model $y = exp(\mathbf{x}\beta) + \varepsilon$ treating ε as normally distributed consistent for β? Explain.

Econometrics Exam, 21 Jan. 2002

Answer question 1 and any 3 additional questions.

1. The idea of *modeling* is to abstract from reality, to capture the essential points while supressing unimportant details. On the other hand, the ML estimator is based upon a *complete* specification of the density of the endogenous variables, conditional on the (weakly) exogenous variables. Does it make sense to speak of ML estimation of a *model*? How should the ML estimator be interpreted in this context? Do you have any opinion regarding how the asymptotic covariance matrix of the "ML" estimator should be estimated, in light of the preceeding?

2. Suppose y_t, conditional on $\mathbf{x}_t$, has the same first two moments as a negative binomial-I random variable. That is, $E(y_t|\mathbf{x}_t) = exp(\mathbf{x}_t\prime\beta)$ and $V(y_t|\mathbf{x}_t) = exp(\mathbf{x}_t\prime\beta)(1+\alpha)$. Suppose that the observations on $(y_t, \mathbf{x}_t)$ are independently and identically ditributed. Explain how we could estimate α, β using the GMM estimator.

3. Explain how to test the hypothesis $R\beta = r$ if β is estimated using GMM. Suppose that R is a $q \times k$ matrix of given constants and that r is a $q \times 1$ vector of given constants. You may directly write down the asymptotic distribution of the GMM estimator assuming the optimal weight matrix and work from there.

4. Suppose we have i.i.d. observations for $(y_t, \mathbf{x}_t)$ where y is a binary 0/1 variable such that

$$\begin{aligned} Pr(y_t = 1|\mathbf{x}_t) &= p(\mathbf{x}_t, \theta) \\ Pr(y_t = 0|\mathbf{x}_t) &= 1 - p(\mathbf{x}_t, \theta) \end{aligned}$$

   Find the limiting objective function for the ML estimator. To do this, work with the average log-likelihood function.

5. Suppose we have a model with a latent variable η_t. We have

$$\begin{aligned} y_t &= exp(\alpha\eta_t + \beta x_t) + \varepsilon_t \\ Pr(\eta_t = 1) &= p \\ Pr(\eta_t = 0) &= 1 - p, \end{aligned}$$

where $0 < p < 1$. We have n i.i.d. observations on (y_t, x_t), while η_t is not observed. Both η_t and ε_t are i.i.d. and independent x_t. Explain how α, β can be estimated consistently.

6. In the context of GMM estimation, define $D_n = D_\theta \left[m_n(\theta)' \right]$, so that it is a $k \times g$ matrix, where k is the number of parameters and g is the number of moment conditions. Let D_∞ be the almost sure limit of D_n. Explain the consequences of D_∞ not being of full row rank, for both for consistency and for asymptotic normality.

7. Discuss how kernel regression could be used to informally check the specification of a parametric model for the conditional mean of a dependent variable.