# 1 Outline.

1. Motivation.

2. Density Estimation.

3. Example- part 1.

4. Nonparametric Regression.

5. Semiparametric Regression.

6. Bootstrap.

7. Application Revisited.

# 2 Motivation.

- In this section, we consider the problem of estimating densities and running regressions without imposing strict functional form assumptions.

- The density estimation you have seen up to this point has assumed that distributions are parametric.

- However, in practice, theory provides very little guidance on the appropriate choice of a functional form.

- Misspecification of the functional form may bias our estimates.

- Discuss kernel density estimation which allows us to estimate distributions without imposing functional form restrictions.

- As an application of kernel density estimation we will consider Bajari and Hortacsu (2005, JPE).

- Recently, Guerre, Perrigne and Vuong (2000, Emet) have proposed methods for nonparametric estimation of structural auction models.

- These methods use a first order condition and a nonparametric bid distribution.

- Bajari and Hortacsu ask whether these nonparametric structural models generate the "right answer".

- Another application, Hendricks, Pinkse and Porter (2003, ReStud) use nonparametric auction models to assess whether bidders correct for the winner's curse in offshore oil auctions.

- A first (primitive) idea to estimate a density is to use the histogram.

- However, we might believe that the true distribution is "smoother" than the histogram.

- Gain efficiency by smoothing the histogram.

- There are drawbacks, however.

- Kernel density estimation may not work well for multidimensional distributions when the number of variables is large (curse of dimensionality).

- In a similar spirit, we will study various nonparametric regression techniques.

- In general, theory does not guarantee that there should be a strictly linear relationship between the exogeous and endogenous variables of our model.

- High order polynomials (as suggested by the text) as a method for flexibly approximating a function work very poorly in practice.

- The 4th, 5th and higher order coefficients may be quite unstable and the function can easily "blow up on the boundaries" where the higher order terms dominate the function.

- We discuss a few nonparametric alternatives to regression.

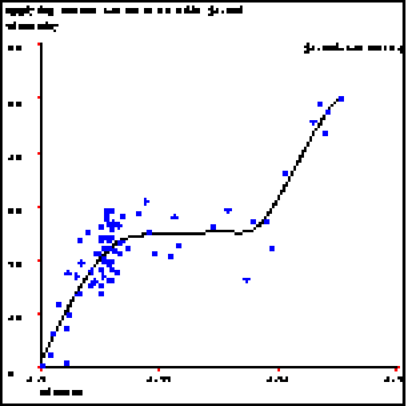
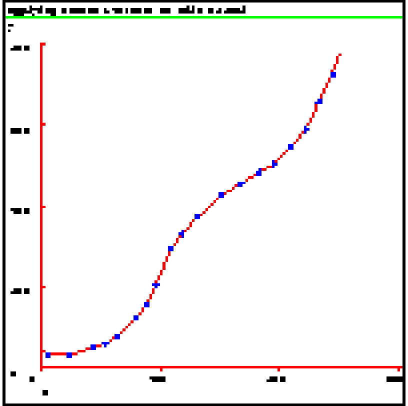
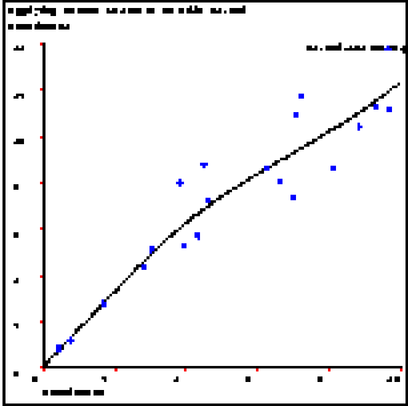
- Suppose we want to come up with a fitted value for $(y_0, x_0)$.

- The idea in our procedures will be to overweight points nearby to $(y_0, x_0)$.

- See figure 2.

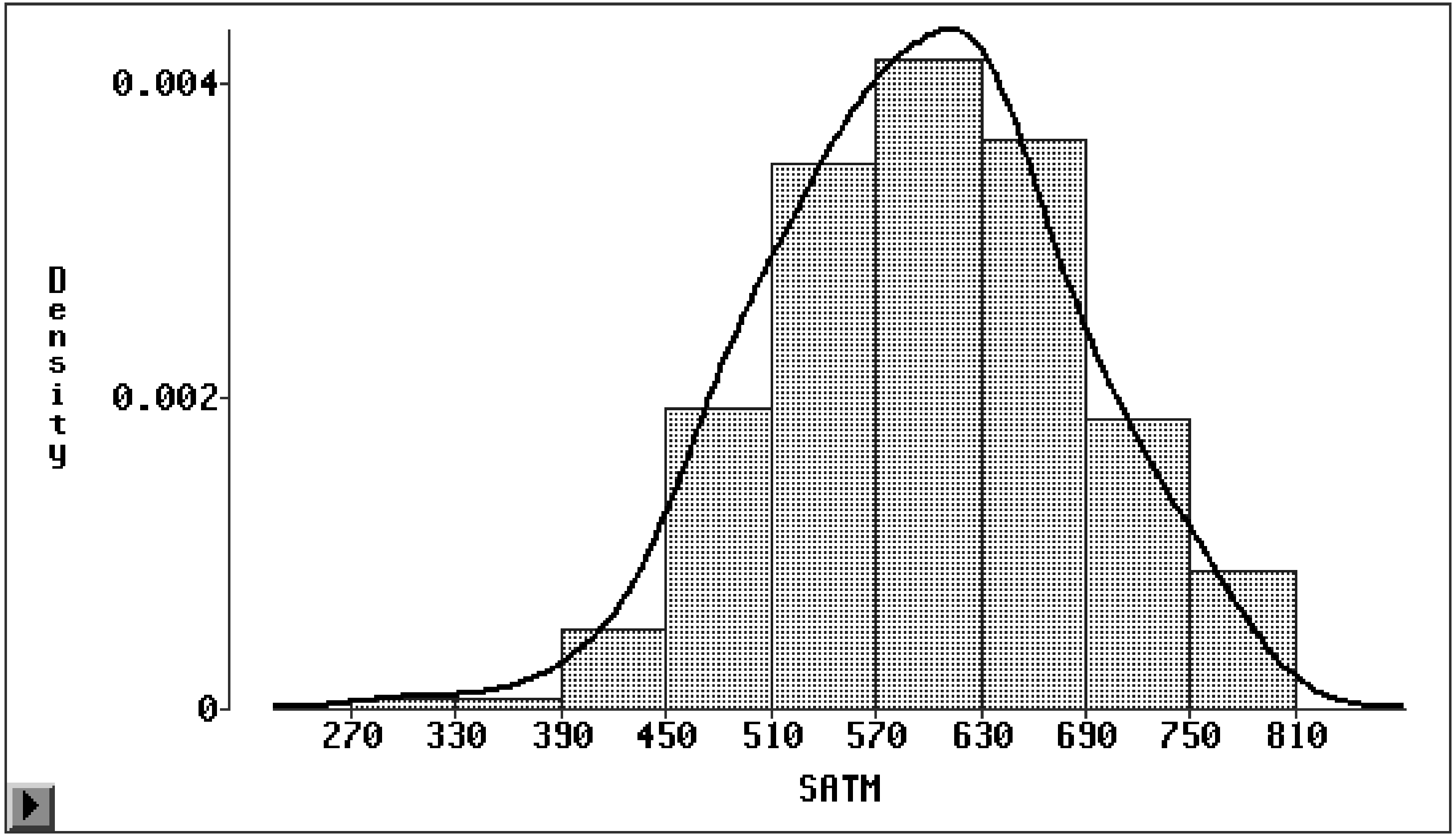
- We will focus on implimentation issues.

- Pagan and Ullah can be consulted for formal derivations of our claims.

SAS: Distribution SASUSER.GPA

File Edit Analyze Tables Graphs Curves Vars Help




| Kernel Density Estimation | | | | | | |
|---|---|---|---|---|---|---|
| Curve | Weight | Method | C Value | Bandwidth | Mode | AMISE (Normal) |
| ——— | Normal | AMISE | 0.7852 | 29.2634 | 612.5000 | 5.379E-05 |

# 3 Density Estimation

- Let $h$ denote the length of the cells in the histogram

- Let $f$ denote the density and $F$ the cdf, then:

$$f(x_0) = \lim_{h \to 0} \frac{F(x_0 + h) - F(x_0 - h)}{2h}$$

- A first (naive) estimator of a density would be to use the height of cells in a histogram.

$$\begin{aligned}
\widehat{f}_{HIST}(x_0) &= \frac{1}{N} \sum_{i=1}^{N} \frac{1(x_0 - h < x_i < x_0 + h)}{2h} \\
&= \frac{1}{Nh} \sum_{i=1}^{N} \frac{1}{2} 1 \left( \left| \frac{x_i - x_0}{h} \right| < 1 \right)
\end{aligned}$$

- This corresponds to the probability of falling into a bin of length 2h.

- In practice, note that this estimate of the density will be discontinuous.

- A more desirable (and efficient!) way to estimate the density would be to smooth out the discontinuities.

- A kernel density estimator generalizes our histogram estimator to:

$$\widehat{f}(x_0) = \frac{1}{Nh} \sum_{i=1}^{N} \frac{1}{2} K\left(\frac{x_i - x_0}{h}\right)$$

- where $K$ takes the place of the indicator function above.

- $K$ is called a kernel function and $h$ is smoothing parameter called a bandwidth.

- We will make the following assumptions about the kernel function:

**(i)** K(z) is symmetric around 0

**(ii)** $\int K(z)dz = 1$, $\int zK(z)dz = 0$, $\int |K(z)|\, dz < \infty$

**(iii)** (a) either $K(z) = 0$ for $|z| > z_0$ or (b) $|z|\, K(z) \to 0$ as $|z| \to \infty$

**(iv)** $\int z^2 K(z)dz = \kappa < \infty$

- We will commonly assume that $z \in [-1, 1]$ as a normalization on the domain in the case of (iii) a.

- Some commonly used kernels are:

$$\text{uniform } 1(|z| < 1)$$
$$\text{Epanechnikov } \frac{3}{4}(1 - z^2) \times 1(|z| < 1)$$
$$\text{normal } (2\pi)^{-1/2} \exp(-z^2/2)$$

- Note that as $h$ is larger, larger weights are given to observations further away from $x_0$.

- That is, larger values of $h$ smooth the observations more heavily.

- In an application, we will want $h \to 0$ as $N \to \infty$ so that in the limit (at an appropriate rate).

- Thus, we only include observations in an arbitrarily small neighborhood in our density estimate $\widehat{f}(x_0)$.

- In choosing the bandwidth, we will face a tradeoff between the bias of $\widehat{f}(x_0)$, denoted $b(x_0)$, and the variance of $\widehat{f}(x_0)$, denoted $\mathsf{V}[\widehat{f}(x_0)]$.

$$
\begin{aligned}
b(x_0) &= E[\widehat{f}(x_0)] - f(x_0) = \frac{1}{2}h^2 f''(x_0)\int z^2 K(z)dz \\
V[\widehat{f}(x_0)] &= \frac{1}{Nh} f(x_0)\int K(z)^2 dz + o(\frac{1}{Nh})
\end{aligned}
$$

- Note that a small $h$ decreases the bias but increases the variance.

- In the limit, we it is desirable to let $h \to 0$ and $Nh \to \infty$ so that both the bias and the variance eventually become zero.

- It can be shown that $\widehat{f}(x_0)$ is pointwise consistent if $h \to 0$ and $Nh \to \infty$

- Uniform consistency if $Nh/lnN \to \infty$ (this requires more smoothing).

- It can be shown that the kernel is (pointwise) asymptotically normal,

$$(Nh)^{1/2}\left(\widehat{f}(x_0) - f(x_0) - b(x_0)\right) \to^d N[0, f(x_0)\int K(z)^2 dz]$$

- This is potentially complicated object to compute.

- A practical alternative is to use a resampling procedure such as the bootstrap.

- Another important choice is the bandwidth.

- This can be found by minimizing the expected mean square error.

- There are also plug in estimates (such as Silvverman's plug in estimate).

# 4 Example-Part 1.

- Next, we consider the problem of the identification and estimation of auction models.

- In an auction, the economist sees the distribution of bids.

- The economist wishes to infer bidder's private information and utility functions.

- Key papers in the literature are Paarsch (1992), Elyakime, Laffont, Loisel and Vuong (1994) and Guerre, Perrigne and Vuong (2000).

# 5 First Price Auction Examples.

- Consider the first price auction with independent private values.

- In the model, there are $i = 1, ..., N$ symmetric bidders with valuation $v_i$ for a single and indivisible object.

- Valuations are iid with cdf $F(v)$ and pdf $f(v)$.

- In the auction, bidders simultaneously submit sealed bids $b_i$.

- Bidder $i$'s vNM utility is

$$u_i(b_1, ..., b_n, v_i) \equiv \begin{cases} v_i - b_i & \text{if } b_i > b_j \text{ for all } i \neq j \\ & 0 \text{ otherwise.} \end{cases} \tag{1}$$

- Let $\pi_i(b_i; v_i)$ denote the expected profit of bidder $i$ where $\phi$ is the inverse of the bid function:

$$\pi_i(b_i; v_i) \equiv (v_i - b_i) F(\phi(b))^{N-1}. \qquad (2)$$

- The first order condition for maximizing expected profits (2) implies that

$$v = b + \frac{F(\phi(b))}{f(\phi(b))\phi'(b)(N-1)} \ . \qquad (3)$$

- This looks hard to deal with.

- Guerre, Perrigne and Vuong (2000) propose an alternative approach.

- The econometrician observes $t = 1, ..., T$ independent replications of the auction described above.

- For each auction $t$, the econometrician observes all of the bids $b_{i,t}$.

- The object that GPV wish to estimate is $F(v)$.

- Let $G(b) = F(\phi(b_i))$ denote the equilibrium distribution of the bids.

- If we substitute $G(b)$ into (**??**) allows us to write expected utility as:

$$(v_i - b_i)G(b_i)^{N-1}.$$

The first order conditions can now be written as:

$$(v_i - b_i)\,(N-1)\,g(b_i) - G(b_i) = 0 \tag{4}$$

$$v_i = b_i + \frac{G(b_i)}{(N-1)g(b_i)} \tag{5}$$

- Let $\widehat{G}$ and $\widehat{g}$ denote estimates of $G$ and $g$

- we can form an estimate $\widehat{v}_{i,t}$ of bidder $i$'s private information $v_{i,t}$ in auction $t$ by substituting these terms into (5):

$$\widehat{v}_{i,t} = b_{i,t} + \frac{\widehat{G}(b_{i,t})}{(N-1)\widehat{g}(b_{i,t})} \tag{6}$$

To summarize, the estimator proposed by GPV:

1. Given bids $b_{i,t}$ for $i = 1, ..., N$ and $t = 1, ..., T$, estimate the distribution and density of bids $\widehat{G}(b)$ and $\widehat{g}(b)$.

2. Compute $\widehat{v}_{i,t}$ for $i = 1, ..., N$ and $t = 1, ..., T$ using equation (6). Use the empirical cdf of the $\widehat{v}_{i,t}$ to estimate $F$.

- This idea turns out to be quite general.

- The distribution of bids can be used to recover private information even in multiple unit auctions or auctions with dynamics.

- These estimators have been applied to offshore oil drilling, procurement, electronic commerce and treasury bill markets.

- There are still some interesting research questions left, however, particularly in the common values case.

## 5.1 The Risk-Averse Model.

- While the above estimator is quite flexible, a problem is that we may not be able to identify the model.

- The basic problem is that because there is only one equation for each $v_i$, we cannot distinguish $v_i$ from the vNM utility function.

- For example, suppose that we still have an IPV model, but that bidders are risk averse.

- CRRA utility function, $U(x) = x^\theta$. In this specification, $1 - \theta$ is the coefficient of relative risk aversion, with $\theta = 1$ corresponding to risk neutrality.

- In this model, the first order condition is:

$$v_i = b_i + \theta \cdot \frac{G(b_i)}{g(b_i)(N-1)}. \tag{7}$$

- Observe that when bidders are risk neutral, that is $\theta = 1$ then this is the risk neutral foc.

- For any fixed value of $\theta$, we could repeat the procedure above and recover the valuations conditional on $\theta$.

- For example, if $\theta = 2$, we could

1. Given bids $b_{i,t}$ for $i = 1, ..., N$ and $t = 1, ..., T$, estimate the distribution and density of bids $\widehat{G}(b)$ and $\widehat{g}(b)$.

2. Compute $\widehat{v}_{i,t}$ for $i = 1, ..., N$ and $t = 1, ..., T$ using equation

$$v_i = b_i + \theta \cdot \frac{\widehat{G}(b_i)}{\widehat{g}(b_i)(N-1)}.$$

Then use the empirical cdf of the $\widehat{v}_{i,t}$ to estimate $F$.

- Obviously, we could perfectly rationalize the bids when $\theta = 1$ or $\theta = 2$.

## 5.2 Structural Estimation

- In order to identify the model, something has to vary!

- One idea is the vary the bids so that the first order condition moves around.

- The logic of the estimator is similar to the previous section.

- Let $G(b; N)$ denote the distribution of bids with $N$ bidders.

- Let $N = 3$ or$N = 6$.

- Suppose that $F(v)$ is independent of $N$.

- Let $v_\alpha$ denote the $\alpha^{th}$ percentile of the distribution of valuations.

- Let $b_\alpha(3)$ denote the $\alpha^{th}$ percentile of $G(b; 3)$ and let $b_\alpha(6)$ denote the $\alpha^{th}$ percentile of $G(b; 6)$.

- By equation (7) it follows that

$$v_\alpha = b_\alpha(3) + \theta \cdot \frac{G(b_\alpha(3); 3)}{2g(b_\alpha(3); 3)} \tag{8}$$

$$v_\alpha = b_\alpha(6) + \theta \cdot \frac{G(b_\alpha(6); 6)}{5g(b_\alpha(6); 6)} \tag{9}$$

- By simple algebra, it follows from the equations (8) and (9) that:

$$b_\alpha(3) - b_\alpha(6) = \theta \cdot \left( \frac{G(b_\alpha(6); 6)}{5g(b_\alpha(6); 6)} - \frac{G(b_\alpha(3); 3)}{2g(b_\alpha(3); 3)} \right) \tag{10}$$

- Equation (10) suggests a simple way to estimate $\theta$.

- If we knew the distribution of bids in the 3 and 6 bidder experiments, given $\alpha$, all of the terms on the left and right hand in this equation would be directly observable except for $\theta$.

- By evaluating (10) at a large number of percentiles, we could then estimate $\theta$ using regression.

- Given an estimate $\widehat{\theta}$ of $\theta$, we can then estimate the valuations $v_i$ by evaluating the empirical analogue of equation (7) as in the previous section.

To summarize, we generate estimates $\widehat{v}_{i,t}$ of $v_{i,t}$ as follows:

1. Generate non-parametrically estimates $\widehat{G}(b; N)$ and $\widehat{g}(b; N)$ of $g(b; N, e)$ and $G(b; N, e)$.

2. Generate an estimate $\widehat{\theta}$ of $\theta$ by running the following regression, using a finite number of percentiles $\alpha$:

$$\widehat{b}_\alpha(3) - \widehat{b}_\alpha(6) = \theta \cdot \left( \frac{\widehat{G}(\widehat{b}_\alpha(6); 6)}{5\widehat{g}(\widehat{b}_\alpha(6); 6)} - \frac{\widehat{G}(\widehat{b}_\alpha(3); 3)}{2\widehat{g}(\widehat{b}_\alpha(3); 3)} \right) + \varepsilon_\alpha \tag{11}$$

3. Given $\widehat{\theta}, \widehat{G}(b; N)$ and $\widehat{g}(b; N)$ use the empirical analogue of (7) to generate an estimate $\widehat{v}_{i,t}$ of $v_{i,t}$.

$$\widehat{v}_{i,t} = b_{i,t} + \widehat{\theta} \cdot \frac{\widehat{G}(b_i)}{\widehat{g}(b_i)(N-1)} \tag{12}$$

- Bajari and Hortacsu (JPE, 2005) compare estimated valuations to experimental valuations for risk neutral, risk averse and two "behavioral models" where bidders may fail to correctly optimize.

- That is, let $v^e_{i,t}$ be the valuation assigned in the experiment to the bidder.

- We would like to examine the distance between the estimated and actual values in the $L^1$ and $L^2$ norms

$$
\begin{aligned}
L^1 &= \frac{1}{IT}\sum_{i,t}\left|v^e_{i,t} - \widehat{v}_{i,t}\right| \\
L^2 &= \frac{1}{IT}\sum_{i,t}\left(v^e_{i,t} - \widehat{v}_{i,t}\right)^2
\end{aligned}
$$

- Another measure of distance is the KS test statistic which measures the distance between the CDF of $v^e_{i,t}$ and $\widehat{v}_{i,t}$.

# 6 Nonparametric Regression.

- Suppose that the data generating process is:

$$\begin{aligned} y_i &= m(x_i) + \varepsilon_i \\ \varepsilon_i &\sim iid[0, \sigma_\varepsilon^2] \end{aligned}$$

- The most common case considered is where $x_i$ is a scalar or very low dimensional vector.

- In this model, we wish to avoid specifying the functional form for $m(x_i)$.

- Suppose that we wish to construct an estimate $m(x_0)$ of our regression function at $x_0$.

- Our strategy will be to overweight observations that are nearby to $x_0$.

- Let $w_{i0,h}(x_i, x_0, h)$ denote a set of weights that we assign to observation $x_0$ where $h$ is a smoothing parameter.

- The weights sum up to one.

- Our estimators will be of the form:

$$\widehat{m}(x_0) = \sum_i w_{i0,h}(x_i, x_0, h) y_i$$

- A first (naive) weighting scheme would be to put equal weight on the $k$ nearest neighbors.

- For example, if $k = 5$, we would put a weight of $1/5$th on the two observations to the left, $1/5$ on the two observations to the right and $1/5$ on $y_0$.

- That is, we rank the scalar $x$'s from highest to lowest and we average by putting equal weight on these nearest neighbors.

- This is obviously a somewhat crude scheme since it does not take account of the distance between $x_0$ and its neighbors, for instance.

- Also, as in density estimation, we might like to shrink the bandwidth in an appropriate manner as the sample size grows large.

- There is also a boundary problem when estimating this model near the endpoints.

- A better alternative might be to use kernel regression where $w_{i0,h}(x_i, x_0, h)$ is defined using kernel weights.

$$\widehat{m}(x_0) = \frac{\frac{1}{Nh}\sum_i K(\frac{x_i - x_0}{h}) y_i}{\frac{1}{Nh}\sum_i K(\frac{x_i - x_0}{h})}$$

- As with kernel density estimation, our estimator will be biased, but consistent if we let $h \to 0$ and $Nh \to \infty$.

- The estimator is pointwise asymptotically normal but with a slower convergence rate $(Nh)^{1/2}$

- Analytical forms of the variance may not be very useful and a resampling procedure may be needed to actually compute the standard errors.

- The choice of the bandwidth $h^*$ is more complicated and is often done using cross validation.

$$
\begin{aligned}
h^* &= \arg\min CV(h) = \sum_{i=1}^{N} (y_i - \widehat{m}_{-i}(x_0))^2 \pi(x_i) \\
\widehat{m}_{-i}(x_0) &= \sum_{j \neq i} w_{ji,h} y_j / \sum_{j \neq i} w_{ji,h}
\end{aligned}
$$

- In the above, $w_{ji,h}$ denote the kernel weights and $\pi(x_i)$ denotes weights for computing the mean square error.

- In practice, it is common to trim the tails to the 5th and 95th percentiles since they may include outliers and/or may influence the convergence of our estimator.

- In practice, researchers may choose a bandwidth by "eyeball" since estimated bandwidths may not always give a reasonable answer, especially in higher dimensions.

- That is, choose the bandwidth so the curve is not "too wiggly".

# 7 Other Approaches.

- Two other important approaches are local linear regression and series estimators.

- In a local linear regression, holding $x_0$ fixed, we wish to fit the model:

$$\sum_i K\left(\frac{x_i - x_0}{h}\right)(y_i - a_0 - b_0 x_i)^2$$

- There is a specific set of regression coefficients, $a_0$ and $b_0$ corresponding to each unique value of $x_0$.

- Unlike ols, we weight the observations using the kernel weights $K\left(\frac{x_i - x_0}{h}\right)$.

- Note that we overweight the observations close to $x_0$ and underweight other observations.

- Hence, the estimator corresponds to weighted least squares which can be expressed in closed form.

- Note that this estimator gives us an estimate of the marginal effect of increasing $x$ at every point, $x_i$.

- An attractive feature of this estimator is that if we oversmooth, we are biasing our results towards ols (hence we know the direction of the potential bias at least!).

- In series estimation, we start with a set of basis functions, $z_1(x), ..., z_K(x)$.

- In a simple (naive) example, these could be $z_0(x) = 1$, $z_1(x) = x$, $z_2(x) = x^2$, ..., $z_K(x) = x^K$

- More realistically, we would choose a set of flexible basis functions.

- These might include orthogonal polynomials, fourier series, etc...

- Our estimator is:

$$\widehat{m}(x) = \sum_{k=0}^{K} \widehat{\beta}_k z_k(x)$$

- $\widehat{\beta}_k$ are estimated by regression.

- In series estimation, we let $K \to \infty$ at an appropriate rate as $N \to \infty$.

- In general, as with other nonparametric techniques we expect the rate of convergence to be slow compared to parametric models.

- However, many linear functionals of $\widehat{m}(x)$ converge at a square root rate with normal asymptotics under the regularity conditions stated in Chen (2006, Handbook of Emet), Andrews (1994, Emet)

- Linear functional are mappings from $\widehat{m}(x)$ to the reals such as derivatives at a point, $x_0$, elasticites at $x_0$, $\int \widehat{m}(x)dx$ and so forth.

- In many cases, linear functionals are the main object of interest in terms of the results.

- This is an attractive feature of these estimators.

- A final useful approach for applied work is quantile regression.

- In this framework, the regression coefficients are functions of the quantile of the error term.

- This may be less flexible as the nonparametric regressions above.

- However, since the model is more parsimoniously specified, it may be more useful for certain applications.

- See in particular the Econometric Society Monograph Quantile Regression by Koenker.

# 8 Semiparametric Regression.

- In a semiparametric model, there is one parametric component and one nonparametric component.

- Some examples are:

$$\begin{aligned}\text{Partially linear } E[y|x,z] &= x'\beta + \lambda(z) \\ \text{Single Index } E[y|x,z] &= g(x'\beta) \\ \text{Generalized additive } E[y|x,z] &= c + \sum_{j=1}^{J} g_j(x)\end{aligned}$$

- Consider our example of testing rationality in an auction.

- That model has the form:

$$E\left(q_t^a - \lambda_i q_t^e + P(b_{i,t}) | I_t\right) = 0$$

- In the above, $P(b_{i,t})$ is the partial derivative of the "penalty function" associated with the bid $b_{it}$.
- The term $I_t$ is information at time $t$.
- This suggests that we could model the relationship between quantities (actual and estimated) and the penalty function as a partially linear model.
- There is a large (and sometimes complicated) literature on semiparametric estimation.
- The text discusses a couple of examples including the Robison difference estimator.

- Suppose:

$$y = x'\beta + \lambda(z) + u$$

- Note that $u = y - E[y|x, z]$

- By conditioning on $z$ note that:

$$E[y|z] = E[x|z]'\beta + \lambda(z)$$

- Substracting these two equations yields that:

$$y - E[y|z] = (x - E[x|z])'\beta + u$$

- Note that we have differenced out $\lambda(z)$.

- Let $\widehat{m}_{yi}$ and $\widehat{m}_{xi}$ be fitted values from a non-parametric regression of $y$ on $z$ and $x$ on $z$ respectively.

- Robinson proposed estimating $\beta$ by the following ols regression:

$$y - \widehat{m}_{yi} = (x - \widehat{m}_{xi})' \beta + v$$

- Under suitably regularity conditions, $N^{1/2}(\beta - \beta_0)$ is asymptotically normal.

- Hence, we can estimate the parametric part of our model at the "regular" rate even though the nonparametric first stage will converge at a much slower rate!

# 9 The bootstrap.

- Next we talk about the bootstrap.

- In many models, the formulas required to compute standard errors may be quite difficult to impliment empirically.

- We saw examples of this in our last section on semiparametrics.

- Also, standard errors and test statistics are typically based on first order asymptotic approximations.

- In finite samples, we might be concerned about the reliability of this approximation.

- The bootstrap and other resampling procedures are an important alternative.

- It is often quite easy to compute and in some cases has better small sample properties than alternatives based on asymptotic approximations.

- The bootstrap does require regularity conditions and cannot be used in all cases.

- Subsampling is an alternative when these regularity conditions cannot be met (as in the Bajari and Hortacsu application with the KS statistic).

## 9.1 Overview

- Suppose that we have an iid random sample $w_i = (y_i, x_i)$, $i = 1, ..., N$.

- Let $\widehat{\theta}$ be an estimator that is a function of the sample.

- Assume that $\widehat{\theta}$ is asymptotic normal at a rate of $N^{1/2}$

- We will be interested in the standard errors, $s_{\widehat{\theta}}$ and test statistics such as the t-stat, $t = (\widehat{\theta} - \theta_0)/s_{\widehat{\theta}}$ where $\theta_0$ is the null hypothesis value.

## 9.2 Simple example.

- Our asymptotic distributions were meant to approximate the sampling distribution of our test statistics.

- For example, a 95% confidence interval is constructed in such a way that 95 times out of 100, given our sample size of $N$, our true parameter will fall in this interval.

- Suppose that we could create $B$ random samples of size $N$.

- We could estimate $\theta$ using each of these samples and find an interval where the estimate falls each time.

- In the bootstrap, we do this by sampling from acutal data set without replacement.

- To take a simple example, suppose that we are interested in the sample mean:
$$\widehat{\mu} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

- Let us draw $B$ samples of size $N$ by sampling from $\{y_1, y_2, ..., y_N\}$ with replacement.

- Let $y_1^{(b)}, ..., y_N^{(b)}$ denote the sample $b = 1, ..., B$

- Let $\widehat{\mu}_b$ denote the mean from sample $b$ :

$$\widehat{\mu}_b = \frac{1}{N} \sum_{i=1}^{N} y_i^{(b)}$$

- We could then estimate the variance of the sample mean as:

$$\begin{aligned} Var[\widehat{\mu}] &= \frac{1}{B-1} \sum_{b=1}^{B} (\widehat{\mu}_b - \widetilde{\widehat{\mu}}) \\ \widetilde{\widehat{\mu}} &= \frac{1}{B} \sum_{b=1}^{B} \widehat{\mu}_b \end{aligned}$$

- More generally, the bootstrap is performed by drawing $B$ random samples, with replacement, of size N.

- Compute the distribution of the estimator using the empirical distribution of the statistic.

- An advantage of the bootstrap is that it can provide a better approximation.

- In our first order asymptotic approach, the approximation error was of order $O(N^{-1/2})$

- In the bootstrap, the approximation error in some cases (but not all!) is of order $O(N^{-1})$

- A sufficient condition for an asymptotic refinement is that the statistic is pivotal.

- We can use the bootstrap for hypothesis testing.

- Suppose that we wish to test the two-sided hypothesis that $\theta = \theta_0$ against the alternative that $\theta \neq \theta_0$ at the 1-$\alpha$ level of signifance.

- Using our bootstrap of the test statistic, we compute the $\alpha/2$ and the 1-$\alpha/2$ percentile of the boostrap.

- We reject the null if $\theta_0$ falls outside of this interval.

- Outside of estimators that are asympotically normal, you need to verify that the bootstrap works.

- Either consult the literature or do the proof.

- Subsampling works under more general conditions and is an increasingly popular alternative.

# 10 Application Revisited.

- There has been a recent growth in the use structural econometric modelling approach to analyze data on firm and consumer behavior.

- One of the most active research areas in this line of work has been on the analysis of auction data.

- Laffont and Vuong (1996) auction models appear especially well-suited "because of the availability of many data sets and the well-defined game forms associated with auctions."

- Examples: Paarsch (1992), Donald and Paarsch (1993, 1996), Elyakime, Laffont, Loisel and Vuong (1994), Flambard and Perrigne (2001), Campo (2001), Guerre, Perrigne and Vuong (2000) and Hendricks, Pinkse and Porter (2002).

- Many economists are uncomfortable with the rationality assumptions.

- This lack of comfort is not unwarranted.

- The equilibrium bid function in a first-price auction game is the solution to a complicated differential equation.

- Our goal: Which, if any, methods for estimating structural auction models yield reasonable estimates.

- Approach: Estimate the models using experimental data Dyer, Kagel, and Levin (1989).

- The four models we estimate are: 1) risk neutral Bayes-Nash 2) risk averse Bayes-Nash 3) Quantal Response Equilibrium (QRE) and 4) an adaptive model of learning.

# 11 A Simple Adaptive Model

- It is possible that the bidders "learn," rather than "know" $Q(b)$, the probability that a bid of $b$ will win the auction.

- Let $h_{it}$ denote the history of bids observed by the bidder $i$ who submits the bid $b_{(t)}$.

- Assume in this model that bidders estimate $G(b)$ using previously submitted bids.

- We denote this estimate as $\widehat{G}(b|h_t)$.

- Bidders choose their bids in order to maximize expected profit $\pi_i(b_i; v_i, \widehat{G}(b|h_t))$ which is equal to

$$\pi_i(b_i; v_i, \widehat{G}(b|h_t)) = (v_i - b_i)\widehat{G}(b|h_t)^{N-1}. \qquad (13)$$

- The first order condition for maximization in the learning model is then

$$\hat{v}_{it} = b_{it} + \frac{\hat{G}(b_{it}|h_{it})}{\hat{g}(b_{it}|h_{it})(N-1)} \; . \qquad (14)$$

where $\hat{v}_{it}$ is the valuation that rationalizes the $t^{th}$ bid.

# 12 The Logit Equilibrium Model

- In this logit equilibrium model, let $\widehat{\pi}(b_i; v_i)$ be the utility that the agent $i$ receives from bidding $b$ when she has a valuation $v_i$;

- This is a sum $\pi(b_i; v_i)$ and $\varepsilon(b_i, v_i)$ :

$$\widehat{\pi}(b; v_i) \equiv (v_i - b_i) * Q(b_i) + \varepsilon(b_i, v_i) = \pi(b_i; v_i) + \varepsilon(b_i, v_i) \ . \tag{15}$$

- The logit equilibrium model generalizes the Bayes-Nash model by including the term $\varepsilon(b_i, v_i)$ in an agent's payoffs.

- One can interpret $\varepsilon(b_i, v_i)$ as the agent's optimization error.

- Let $\sigma(b_i; v_i, \mathbf{B})$ be the probability that agent $i$ bids $b_i$ conditional on a value draw $v_i$ and that the $N-1$ other agents bid using the strategy $\mathbf{B}$.

- By well known properties of the extreme value distribution, it follows immediately that:

$$\sigma(b_i; v_i, \mathbf{B}) = \frac{\exp(\lambda\pi(b_i; v_i, \mathbf{B}))}{\sum_{b' \in \mathcal{B}} \exp(\lambda\pi(b'; v_i, \mathbf{B}))}. \tag{16}$$

## 12.1 Structural Estimation

- Let $p(b|\theta)$ denote the probability of the bid $b$ given $\theta$. Given $\widehat{Q}(b)$,

$$p(b|\theta, \lambda; \widehat{Q}) = \int_{\underline{v}}^{\bar{v}} \sigma(b_i|v) f(v|\theta) dv \tag{17}$$

$$= \int_{\underline{v}}^{\bar{v}} \frac{\exp(\lambda(v - b_i)\widehat{Q}(b_i))}{\sum_{b' \in \mathcal{B}} \exp(\lambda(v - b')\widehat{Q}(b'))} f(v|\theta) dv \tag{18}$$

- Our approach for estimating the logit equilibrium model can be summarized as follows:

1. Given a data set of $T$ bids, $b_{(1)}, ..., b_{(T)}$, form an estimate $\widehat{Q}(b)$ of $Q(b)$.

2. Estimate $\theta$ and $\lambda$ using maximum likelihood.

- Goeree, Holt, and Palfrey (2002) find that a QRE model with risk aversion seems to fit laboratory experiments in first-price sealed-bid auctions well.

- We attempted to estimate this model, a very poor fit resulted.

# 13 Results

- Data from IPV first-price auction experiments conducted by Dyer, Kagel, and Levin (1989).

- There were 3 experimental runs with 6 different subjects selected in each of them.

- In these experiments, bidders were assigned i.i.d. valuations drawn from a uniform distribution on $[\$0, \$30]$, and, in the event they won the auction, they were paid their assigned valuation minus their bid.

- Each subject participated in 28 auctions, during the course of a two hour experimental run.

- Following Kagel's suggestion, data from the first 5 runs of the experiments are excluded.

- This leaves us with 3 runs of 23 auctions.

- A novel feature of this experiment was that the bidders were faced with two possibilities as to how many competitors they were going to face: with probability $1/2$, the market they competed in contained $N = 3$ bidders and with probability $1/2$, $N = 6$ bidders.

- Bidders were asked to submit two "contingent" bids and one "non-contingent" bid.

- After the bidders submitted their three bids, a coin was tossed, first, to determine whether the "contingent" or "non-contingent" bids would be used in determining the winner.

- A second coin toss determined whether $N = 3$ or $N = 6$. If the "contingent" treatment was selected, the first "$N = 3$ contingent" bid was used if $N = 3$, and the "$N = 6$ contingent" bid was used if $N = 6$.

- After each auction, bids and corresponding private values were posted on a blackboard for the bidders to see.