Intro Comments:

The purpose of the lecture today is to talk a little about quantile regression.

Quantiles

[This is closely based on Koenker and Hallock, Quantile Regression, Journal of Economic Perspectives, 15(4), 143-156, 2001.]

What is the point of quantile regression? - Boardwork

A little formalism:

OLS can be derived as the solution to the following fitting criteria:

$$\min_{\mu} \sum_{i=1}^{n} (y_i - \mu)^2$$

This gives the sample mean μ (ie. OLS with just a constant)

The median γ can be found by solving

$$\min_{\mu} \sum_{i=1}^{n} |y_i - \gamma|$$

That is minimizing the sum of absolute residuals (think through the geometry)

Similarly, the τth quantile can be found by minimising

$$\min_{\mu} \sum_{i=1}^{n} \tau \left(y_i - \gamma\right)^{+} + \left(\tau - 1\right)\left(y_i - \gamma\right)^{-}$$

If we want to get conditional medians we solve

$$\min_{\mu} \sum_{i=1}^{n} |y_i - \gamma(x, \beta)|$$

An analogously for other quantiles

Inference, works in standard ways. Turns out you can set this up as a GMM estimator if you want etc etc.

Computationally its a bit more of a pain that OLS, but canned programs like STATA do it pretty well. Solving for coefficients is a linear programming problem.

The assumptions we can get away with quantile regressions are slightly different than with expected value regressions (such as OLS).

For instance say I am estimating the model:

$$y_{it} = \phi(x_{it}\beta + \epsilon_{it}) \tag{1}$$

which is a non-additive measurement error model.

If we use a usual OLS regression, we will have a misspecification error for the estimates of $\widehat{\phi}$ due to the fact that $\phi(x + E(\epsilon)) \neq E(\phi(x + \epsilon))$ by Jensen's inequality. But this is not the case with median regression.

Example

```
. reg tfp ldinv ldnpt

      Source |       SS       df       MS              Number of obs =
-------------+------------------------------           F(  2,  2968) =
       Model |  11.1709857     2  5.58549283           Prob > F      =
    Residual |  915.972334  2968  .308616015           R-squared     =
-------------+------------------------------           Adj R-squared =
       Total |  927.143319  2970  .312169468           Root MSE      =

------------------------------------------------------------------------
         tfp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Int
-------------+----------------------------------------------------------
       ldinv |   .0862107   .0143293     6.02   0.000     .0581143     .
       ldnpt |  -.0797567   .0140316    -5.68   0.000    -.1072693    -.
       _cons |   .1258334    .031044     4.05   0.000     .0649634    .1
------------------------------------------------------------------------

. qreg tfp ldinv ldnpt, quantile(50)

Median regression                                     Number of obs =
  Raw sum of deviations 1098.531 (about .03077334)
  Min sum of deviations 1064.351                      Pseudo R2     =

------------------------------------------------------------------------
         tfp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Int
-------------+----------------------------------------------------------
       ldinv |   .1581972   .0141362    11.19   0.000     .1304794     .
       ldnpt |  -.1513372   .0138433   -10.93   0.000    -.1784806    -.1
       _cons |   .2948238   .0306275     9.63   0.000     .2347705    .3
------------------------------------------------------------------------
```

Next I want to show you the usefulness of Quantiles.

The next table compares OLS to quantile approaches to understanding the determinants of low birthweight in newborns.

The estimated model is birthweight on stuff. The sample size is just under 200k babies.

Birthweight is a strong predictor of health problems for kids.

Figure 4

Ordinary Least Squares and Quantile Regression Estimates for Birthweight Model

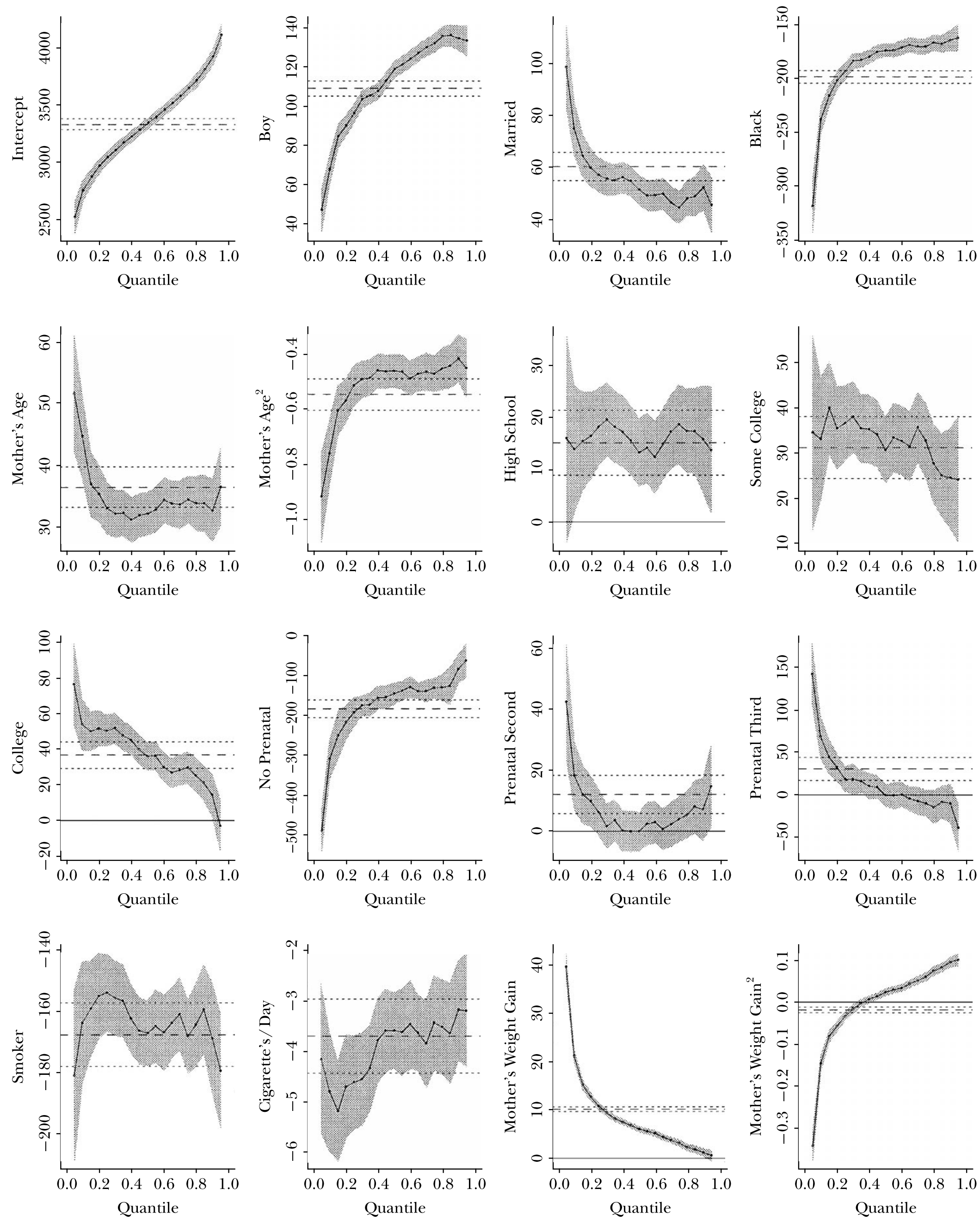

Figure P2: Table from Koenker and Hallock

What I want to do now is to talk through: Goldberg, Dealer Price Discrimination in New Car Purchases: Evidence form the Consumer Expenditure Survey, JPE, 1996.

This paper is a great example of how powerful quantile regression can be in analyzing market type data.

Background: Ayres (1991) on Dealer Price Discrimination from a field experiment, and Search.

Search

- Buyers have a willingness to pay of v.
- There is a distribution of prices given by $f(p)$: the distribution of prices that consumers expect to encounter.
- I can search for another price, or take the current price offered to me.
- Can search for a new price at a cost of c.
- Have a **Reservation Price** r, which is a price below which I accept, above which I keep searching.
- It is given by:

  $$r = V - c$$

  where the value of searching V is:

  $$V = \int_p \text{max}\{r - p, V - c\} df(p)$$

  where $f(p)$ is the distribution of prices I expect to receive.

- In other words the reservation price makes me indifferent between taking my current offer and searching.

Notice a few things:

- The more dispersed the distribution of prices, the more I search and the lower my reservation price.

- The higher the cost of searching, the higher my reservation price.

- Helps explain the “carpet store” puzzle: why are many carpet stores located next to each other: in order to attract consumers who know that the low search cost between store makes competition fiercer.

Ayres and Siegelman Evidence on Dealer Price Discrimination

- Price Discrimination in Automobiles: Large Differences in the Prices paid for the same cars.

- To what extent are these differences based on racial or gender discrimination? This is the point of "Race and Gender Discrimination in Bargaining for a New Car" by Ayres and Siegelman.

- Ayres is an **audit** study: sends out pairs of interviewers with the same script for buying a car. Tries to control for as many differences between people as possible. This was done for a number of auto dealerships in the Chicago Area.

- Record the first price and the final price for a car reached.

TABLE 1—SUMMARY STATISTICS ON PROFITS AND COSTS, BY TESTER TYPE

Tester type	Initial profit	Final profit	Concession[a]
White males (18 testers; 153 observations)			
Mean	1,018.7	564.1	454.6
Standard deviation	911.3	708.0	(44.6 percent)
Average markup (percentage)	9.20	5.18	
White females (7 testers; 53 observations)			
Mean	1,127.3	656.5	470.8
Difference from white male average	108.6	92.4	(41.8 percent)
Standard deviation	785.3	472.4	
Average markup (percentage)	10.32	6.04	
Black females (8 testers; 60 observations)			
Mean	1,336.7*	974.9*	361.8
Difference from white male average	318.0	246.1	(27.1 percent)
Standard deviation	887.8	827.8	
Average markup (percentage)	12.23	7.20	
Black males (5 testers; 40 observations)			
Mean	1,953.7*	1,664.8*	288.9
Difference from white male average	935.0	1,100.7	(14.8 percent)
Standard deviation	1,122.7	1,099.5	
Average markup (percentage)	17.32	14.61	
All nonwhite males (20 testers, 153 observations)			
Mean	1,425.5*	1,045.0*	380.5
Difference from white male average	406.8	481.0	(26.6 percent)
Standard deviation	973.6	989.9	
Average markup (percentage)	12.99	9.40	

[a]Average initial profit minus average final profit; average percentage concession is given in parentheses.

*Significantly different from the corresponding figure for white males at the 5-percent level.

TABLE 2—OLS AND FIXED-EFFECTS (ONE DUMMY PER AUDIT) REGRESSIONS OF INITIAL AND FINAL PROFITS AND MARKUPS ON RACE AND GENDER DUMMIES AND CONTROL VARIABLES

	Initial dollar profit		Final dollar profit		Initial percentage markup		Final percentage markup	
Variable	OLS	Fixed effects	OLS	Fixed effects	OLS	Fixed effects	OLS	Fixed effects
Race/gender dummies:								
Constant	1,014.95*		607.51*		0.114*		0.072*	
	(4.12)		(2.98)		(5.44)		(4.19)	
White female	192.38	55.10	174.68*	129.09	0.017	0.007	0.014	0.013
	(1.23)	(0.39)	(1.35)	(1.05)	(1.26)	(0.63)	(1.29)	(1.25)
Black female	404.28*	281.05*	504.64*	404.65*	0.039*	0.027*	0.045*	0.037*
	(2.75)	(2.13)	(4.15)	(3.49)	(3.09)	(2.42)	(4.45)	(3.87)
Black male	1,068.24*	1,061.17*	1,242.85*	1,061.27*	0.094*	0.091*	0.107*	0.090*
	(6.10)	(6.56)	(8.57)	(7.47)	(6.31)	(6.60)	(8.78)	(7.70)
Controls								
SPLIT[a]	20.30		−57.36		−0.02		−0.02	
	(0.15)		(−0.51)		(−1.52)		(−1.95)	
Time[b]	−1.73		−2.47		−0.0004*		−0.0004*	
	(−0.88)		(−1.52)		(−2.29)		(2.70)	
Experience[c]	−3.58		−0.50		0.00		0.00	
	(−0.43)		(0.07)		(0.10)		(0.56)	
First[d]	203.32		192.18		0.01		0.01	
	(1.69)		(1.93)		(1.30)		(1.48)	
$F_{[3, 298]}$:	12.91*		26.52*		14.04*		27.98*	
Adjusted R^2:	0.10	0.44	0.19	0.43	0.11	0.45	0.21	0.47
Standard error of the estimate:	914.35	723.2	757.1	635.6	0.078	0.06	0.064	0.05
Degrees of freedom:	298	150	298	150	298	150	298	150
N:	306	306	306	306	306	306	306	306

Note: The numbers in parentheses are t statistics.
[a]Dummy variable: 1 if tester used a split-the-difference bargaining strategy; 0 otherwise.
[b]Number of days between this test and the first day of testing.
[c]Number of prior tests by this tester.
[d]Dummy variable: 1 if tester was first in the pair; 0 otherwise.
*Statistically significant at the 5-percent level.

See that the markup difference is similar for initial and final price. This should worry you.

TABLE 4—PROPORTION OF TESTS IN WHICH WHITE MALE OBTAINED THE BETTER RESULT

Test	Percentages Initial profits	Final profits
White males vs. all others (153 pairs)	68.0	66.7
White males vs. white females (53 pairs)	58.4	56.6
White males vs. black males (40 pairs)	87.5	85.0
White males vs. black females (60 pairs)	63.3	61.7

Notes: All values are significantly different from 50 percent at the 1-percent level using a likelihood-ratio test ($\chi^2_{[1]}$); in 43.5 percent of the tests, white males received an *initial* offer that was lower than the *final* offer made to the nonwhite male tester.

How to Interpret the Evidence

- Note that all this tells us is what dealers think about consumers.

- Do women and blacks have higher reservation values (i.e. the highest price that I will accept for a car)? Is this due to less information about transaction prices for cars, or simply being intimidated by auto dealers?

- In other words: does the bargaining or search theory explain these facts.

- Most auto dealers are white males: could we have a simple discrimination story instead.

- Goldberg uses an **observational** study: gets transactions prices for a survey called the Consumer Expenditure Survey (CES). (Note: the CES has huge amounts of data on what people spend on, not just cars).

- The idea is that the Ayres and Siegelman evidence may leave out something about "real world" interactions.

- This is always something that you need to be careful about with lab evidence: if biases are large, there is a strong economic incentive to overcome them (like a dealership that promises a "no-haggle" policy attracting black males.

Study Design

- Goldberg runs car fixed effect (i.e. controlling for the car purchased j) regressions of prices on characterics of the purchasers:

$$p_{ij} = \alpha_j + X_{ij}\beta$$

- Really we are looking at explanatory variables for the **differences** in prices.

TABLE 2

OLS Estimation Results (N = 1,279)

Dependent Variable: Absolute Dealer Discount

	Regression			
	1	2	3	4
R^2	.180	.177	.187	.189
Adjusted R^2	.14	.14	.14	.15
Durbin-Watson statistic	1.81	1.81	1.81	1.81
F-value	4.80	5.41	4.37	4.44
Intercept	−1,168.33 (−3.21)	−1,225.59 (−4.71)	−1,146.99 (−3.14)	−1,135.57 (−3.06)
AGE	4.02 (1.06)	...	3.80 (.98)	3.64 (.94)
MINOR	−274.62 (−1.04)	...	−537.70 (−.66)	−158.42 (−.38)
FEMALE	−129.62 (−1.10)	...	−126.03 (−1.17)	−133.76 (−1.25)
MINFEM	−21.96 (−.05)	...	−127.13 (−.24)	−1.45 (.03)
ASSET	−.15E−02 (−.91)	...	−.15E−02 (−.94)	−.14E−02 (−.89)
FINCATAX	−.82E−03 (−.33)	...	−.66E−03 (−.27)	−.87E−03 (−.35)
EDUCA	−25.23 (−.25)	...	−15.36 (−.15)	−33.21 (−.33)
WHITEC	−117.12 (−1.12)	...	−118.69 (−1.13)	−128.51 (−1.24)
RURAL	−216.89 (−1.90)	−192.85 (−1.69)	−204.45 (−1.78)	−236.63 (−2.05)
NE	57.00 (.45)	106.53 (.85)	30.48 (.24)	48.13 (.38)
NE * MINOR	...	...	784.59 (1.14)	...
MW	444.23 (3.92)	468.33 (4.15)	391.98 (3.42)	421.70 (3.71)
MW * MINOR	...	...	1,050.84 (1.74)	...
WE	−70.87 (−.56)	−43.67 (−.35)	−73.55 (−.58)	−82.64 (−.65)
WE * MINOR	...	...	−528.79 (−.57)	...
FINAN	63.24 (.49)	25.07 (.21)	62.10 (.48)	60.13 (.47)
FINAN * MINOR	...	...	−131.70 (−.18)	...
DEALERF	294.97 (2.58)	272.54 (2.39)	303.99 (2.64)	284.09 (2.50)
DEALERF * MINOR	...	...	−200.04 (−.31)	...
FIRSTB	444.29 (2.51)	340.59 (2.11)	414.13 (2.31)	440.80 (2.49)
FIRSTB * MINOR	...	...	791.08 (1.04)	...
TRADIN	−597.77 (−6.84)	−589.95 (−6.73)	−617.38 (−6.96)	−585.00 (−6.72)
TRADIN * MINOR	...	...	355.58 (.78)	...
BRANDF	−20.98 (−.20)	−12.31 (−.12)	−29.14 (−.28)	−43.73 (−.43)
Q1P	44.41 (.11)	9.70 (.02)	107.05 (.26)	9.81 (.02)

What's going on?

- There is no evidence on dealer price discrimination at the statistically significant level.
- Maybe Chicago is different from the rest of the country: particularly discriminatory.
- However, there aren't large regional differences in price discrimination.
- My own pet theory is that different gender and racial groups use different protocols for bargaining: When I bargain (say over when I come over for Christmas) with my father, I use much tougher initial offers that when I bargain with my mother. The same could be true for white and black differences.

Quantiles:

- Let's look not at the mean price received, but at the distribution of prices.
- Controlling for other characteristics, this is called quantile regression.

TABLE 3

RESULTS FROM QUANTILE REGRESSIONS (N = 1,279)

Dependent Variable: Absolute Discount

	OLS (1)	MEDIAN (2)	10% Quantile (3)	90% Quantile (4)
Intercept	−1,168.33	−1,598.95	−2,117.75	−66.82
	(−3.21)	(−6.02)	(−4.75)	(−.19)
AGE	4.02	6.95	4.04	7.72
	(1.06)	(2.63)	(.90)	(1.74)
MINOR	−274.62	−48.73	−784.35	453.14
	(−1.04)	(−.27)	(−2.87)	(1.81)
FEMALE	−129.63	−115.02	190.00	1.11
	(−1.10)	(−1.39)	(1.52)	(.08)
MINFEM	−21.96	−98.02	445.97	−379.54
	(−.05)	(−.34)	(1.06)	(−.86)
ASSET	−.15E−02	−.22E−02	.13E−02	−.11E−02
	(−.91)	(−2.01)	(.80)	(−.66)
FINCATAX	−.82E−03	.12E−02	−.44E−02	−.17E−02
	(−.33)	(.79)	(−1.87)	(−.71)
EDUCA	−25.23	−150.03	62.04	−120.69
	(−.25)	(−2.12)	(.56)	(−1.01)
WHITEC	−117.11	−81.42	−232.37	48.75
	(−1.12)	(−1.10)	(−2.07)	(.39)
RURAL	−216.89	−199.87	−222.44	−166.03
	(−1.90)	(−2.49)	(−1.91)	(−1.30)
NE	57.00	−51.18	311.31	164.35
	(.46)	(−.59)	(2.43)	(1.19)
MW	444.23	322.82	1,054.67	361.21
	(3.92)	(3.78)	(8.37)	(2.57)
WE	−70.87	−216.66	126.59	−145.66
	(−.56)	(−2.40)	(.89)	(−1.00)
FINAN	63.24	96.98	−45.23	251.83
	(.49)	(1.05)	(−.30)	(1.73)
DEALERF	294.98	299.92	251.41	128.65
	(2.58)	(3.66)	(2.06)	(.98)
FIRSTB	444.29	547.81	652.19	34.57
	(2.52)	(3.72)	(2.92)	(.14)

- Notice that women and blacks have **More Dispersed** prices.

- Perhaps the story is that there is more variation in the reservation prices of blacks and women.

- Dealers would respond to this higher variation by starting off with high prices to pick out the people with unusually high reservation values.

Chad Syverson on Productivity Dispersion:

Concrete plants compete in prices and competitors are spatially differentiated. A Salop model can capture this structure, with N identical firms located equidistantly along a unit circle. A mass D of consumers is distributed uniformly on the circle. They have transportation costs t and have a high enough reservation price r that they will purchase from at least one firm. Firms draw marginal costs c from a distribution $F_c(\cdot)$. Firms can charge a different price to each consumer.

Suppose that all firms have marginal costs c to illustrate the equilibrium. In equilibrium, variable profits π^V are:

$$\pi^V(N,D) = \begin{cases} tD\left(\frac{1}{N}\right)^2 & \text{If } N > 1 \\ D\left(r - c + \frac{t}{4}\right) & \text{If } N = 1 \end{cases} \tag{2}$$

I can rewrite this equation for variable profits as:

$$\pi^V(N,D) = \underbrace{\eta(N)}_{\text{markup}} \frac{D}{N} \tag{3}$$

Where $\eta(N)$ is the markup over marginal cost, and $\frac{D}{N}$ is the number of consumers purchasing concrete from each firm, which can be rewritten as $\pi^V =$

$ppc(N)D$. Then the number of firms in the market will be determined by N such that:

$$\begin{aligned} \pi^V(N, D) &\geq f \\ \pi^V(N+1, D) &< f \end{aligned} \tag{4}$$

i.e. firms enter until it is not longer profitable to do so.

Note that as market size increases, the set of firms who choose to enter will also increase. However, profits become more sensitive to marginal costs in larger markets.

Figure 1. Total Factor Productivity Kernel Density Estimates, Plants in Markets above and below Median Demand Density

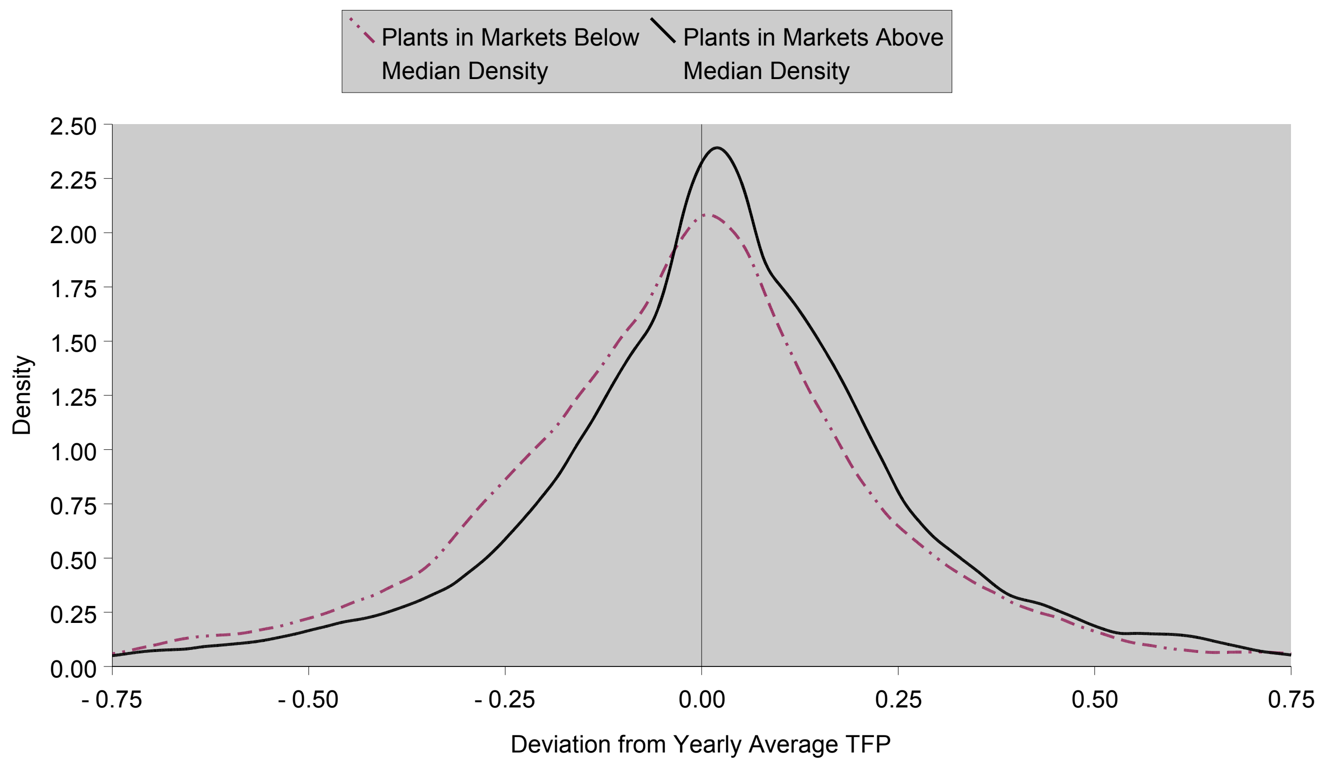

Figure 2. Output (Logged Shipments in Cubic Yards) Kernel Density Estimates, Plants in Markets above and below Median Density

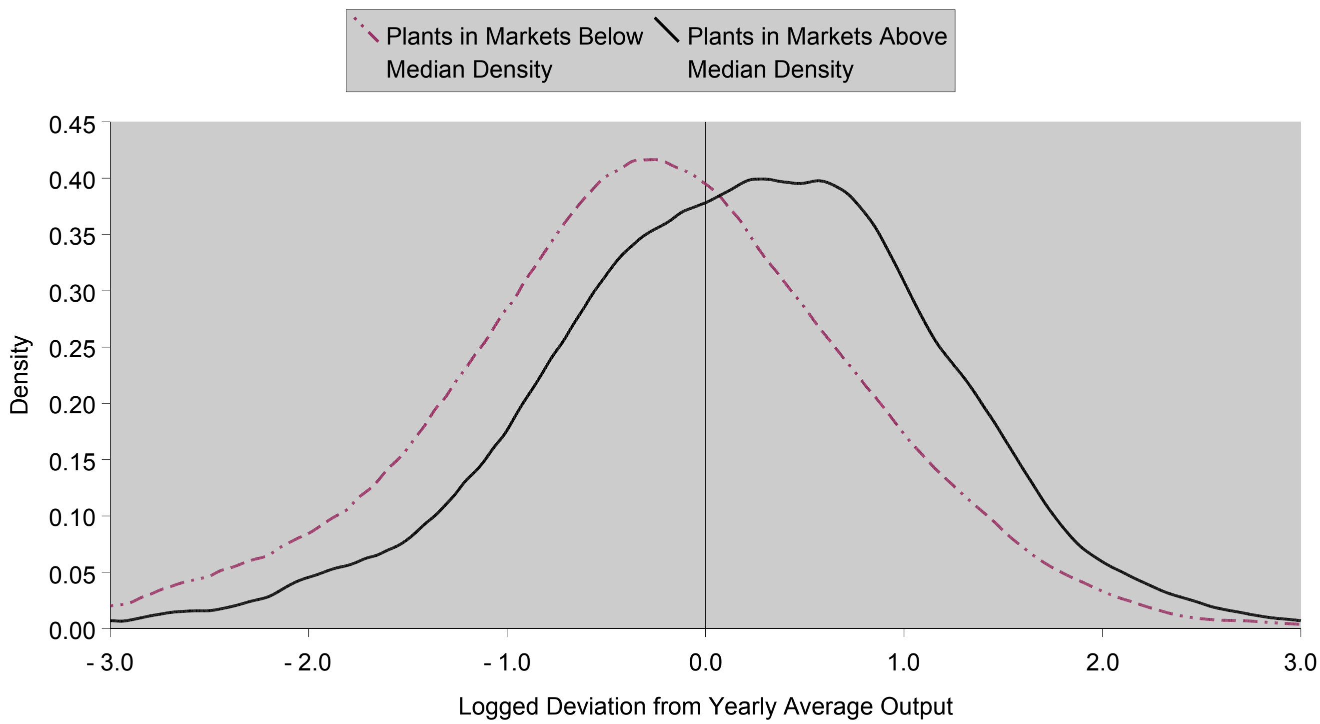

Table 2. Main Regression Results—Local Productivity and Size Moments

A. Demand Density Coefficients

Dependent Variable	Regression Statistic	Model [1]	Model [2]	Model [3]	Model [4]
TFP Dispersion (Interquartile Range)	R^2	0.018	0.036	0.092	0.092
	Demand Density Coef.	-0.014*	-0.015*	-0.029*	-0.031*
	(Standard Error)	(0.004)	(0.004)	(0.008)	(0.010)
Median TFP	R^2	0.059	0.289	0.321	0.322
	Demand Density Coef.	0.021*	0.018*	0.012*	0.008
	(Standard Error)	(0.003)	(0.003)	(0.005)	(0.006)
Output-Weighted Average TFP	R^2	0.045	0.125	0.162	0.162
	Demand Density Coef.	0.026*	0.024*	0.016*	0.012
	(Standard Error)	(0.004)	(0.004)	(0.008)	(0.010)
10^{th} Percentile TFP	R^2	0.033	0.033	0.058	0.059
	Demand Density Coef.	0.057*	0.056*	0.065*	0.056*
	(Standard Error)	(0.010)	(0.010)	(0.019)	(0.022)
Producer-to-Demand Ratio	R^2	0.570	0.584	0.708	0.711
	Demand Density Coef.	-0.369*	-0.363*	-0.313*	-0.278*
	(Standard Error)	(0.015)	(0.015)	(0.022)	(0.030)
Average Plant Output	R^2	0.334	0.376	0.557	0.563
	Demand Density Coef.	0.218*	0.211*	0.184*	0.142*
	(Standard Error)	(0.012)	(0.012)	(0.017)	(0.023)
Year Dummies		No	Yes	Yes	Yes
Demand Controls		No	No	Yes	Yes (+CH)

This panel shows the estimated coefficients on demand density when various moments of the local productivity and size distributions are regressed on demand density and, when applicable, a set of demand controls. Specifications are by column and dependent variables by row. The sample consists of 665 region-year observations with at least five plants for which I have non-imputed production data. "+CH" indicates that Ciccone-Hall measure of overall density was included in controls (see text for details). Reported standard errors are robust to heteroskedasticity, and an asterisk denotes significance at the 5 percent level.