

Goal

To estimate the elasticity of demand for tickets to sporting events

Data

- A list of season ticket holders for the 2002 Atlanta Braves and their zip codes
- Which were then added up to find the total number of season tickets sold in each zip code
- The distance each zip code is from the stadium
- Demographic data on each zip code available from census, websites, etc.
- 344 observations (zip codes) within 100 miles of Atlanta
- 4 outliers removed (big post offices in downtown Atlanta)

Could easily do this with survey data at any event?

Key Concepts

- While the Braves offer the same price to customers regardless of where they live (the “ticket price” of attending the game is the same for all customers), the “full price” of attending the game includes travel costs.
- People who live a larger distance from the stadium will pay a higher “full price” to attend the game, and thus will purchase fewer tickets.
- It makes sense that consumers would be as sensitive to a \$1 increase in ticket price as they would be to a \$1 increase in “full price”. Therefore, by finding out how sensitive consumers are to a \$1 increase in “full price”, we can learn how sensitive consumers would be to a \$1 increase in “ticket price.”
- Will accomplish this by finding how the quantity of tickets sold varies with distance from stadium, then translate the distance in difference into “full price” by using simple assumptions about transportation costs and time costs.

Initial Data Check

Do tickets sales fall as we move father away from the stadium?

Percentage of ZIP codes with at least one season ticket sold, by distance.

	Distance to ballpark (mi.)			
	0 - 25 miles	25 - 50 miles	50 - 75 miles	75 – 100 miles
82 games (full season)	96%	81%	44%	30%

Ceteris Paribus Check (Demand Shifters)

Preferences

- Reasonable to assume that preferences are the same for all potential customers (zip codes)? Maybe not. Some zip codes with many children? Some zip codes with many retired people? Different racial / ethnic groups have different preferences?

Include variables for demographic characteristics

Price of Substitutes

- What are substitutes for Atlanta Braves Games? Theatre? Boating? Bars?
- Reasonable to assume the price of complements are the same for all potential customers (zip codes)? Yes.

So don't need to include these in regression.¹

Prices of Complements

- What are complements for Atlanta Braves Games? Beer? Hotdogs?
- Reasonable to assume the price is the same for all potential customers (zip codes?) Yes.

So don't need to include these in regressions.

Population

- Would expect more tickets sold in more densely populated zip codes

Will "handle" by estimating demand curve per 1000 people in the zip code

Income

- Reasonable to assume the income levels are the same for all potential customers (zip codes)? No.

Will need to include this in regression.

Expectations of Future Quality

- Reasonable to assume that expectations are the same for all potential customers? Yes. So don't need to include these in regression.

Others?

¹ Someone suggested other sporting events in the Atlanta area. Technically, if these stadiums were located in the same general area as the Braves stadium, then the relative prices of the complements for each zip code would be the same, and still wouldn't have to be included. If they were located elsewhere, the full price might of attending these events might be different for those closer to these other events, shifting the demand for those consumers. The stadiums in Atlanta are indeed similarly located. Not a problem.

What is the dependent variable?

- Quantity of tickets sold (per 1000 population)

What are the independent variables? What do we include in regression?

- Distance from stadium, as it will proxy full price
- Income
- Demographics for preferences (racial, age, sex groups), working status

Have controlled for population by estimating tickets per 1000 population

Regression Results

Tobit regression results for quantity of tickets sold within a ZIP code per 1000 population²

Distance to ballpark (mi.)	-5.405 *** (0.929)
Median income (\$1000)	6.805 *** (1.814)
Percent ethnic	6.710 *** (2.485)
Percent black	0.213 (1.197)
Percent female	13.618 (12.123)
Percent in labor force	-11.420 ** (4.691)
Percent under-18	-29.574 *** (7.640)
Percent 18 to 24	8.151 (6.963)
Percent 25 to 59	22.738 ** (8.780)
constant	-577.7 (920.9)

n = 344.

standard errors in parentheses

ZIP codes within 100 miles of Turner Field, Atlanta, GA.

Four ZIP codes containing outlier data omitted.

*** - significant at the 99% confidence level,

** - 95% level,

Regression Equation

Quantity of tickets sold =

$$\begin{aligned} & -577.0 - 5.405 * \text{distance} + 6.805 * \text{income} + 6.710 * \text{pct_ethnic} + 0.213 * \text{pct_black} \\ & + 13.618 * \text{pct_female} - 11.420 * \text{pct_labor force} - 29.574 * \text{pct_under18} \\ & + 8.151 * \text{pct_1824} + 22.738 * \text{pct_2559} \end{aligned}$$

² Tobit regression is basically the same as ordinary least squares regression that we did in class, but makes a slight adjustment that is beyond the scope of this class. The interpretation is the same.

Check signs on significant variables

We want to ensure that the signs on the significant variables conform to our economic theories – that they make sense.

- distance: negative and significant – makes sense

As distance from stadium increases, full cost increases. This makes sense.

- income: positive and significant – makes sense

As income increases, more tickets are purchased. Braves tickets a normal good

- pct_ethnic: positive and significant – make sense?

Not sure what to make of this – non white, non black folks like baseball?

- pct_laborforce: negative and significant – make sense?

We have already controlled for income. Holding income constant, zip codes with with more people working buy fewer tickets. Perhaps they are busy working.

- pct_under18: negative and significant – makes sense?

Zip codes with more children purchase fewer tickets. Makes sense.

- pct_2529: positive and significant – make sense?

Zip codes with more young adults purchase more tickets. Perhaps these folks have more time and than other folks who have started a family?

Key Result

- For every one additional mile away from the stadium, there are 5.405 fewer tickets sold (per 1000 population).

- That is, $\frac{\Delta Q}{\Delta Distance} = -5.405$

Elasticity of Demand

The elasticity of demand is given by: $\varepsilon = \frac{\Delta Q}{\Delta P} \frac{P}{Q}$

- But we have $\frac{\Delta Q}{\Delta \text{Distance}}$, so we have a bit of work to do still
- We first need to translate $\frac{\Delta Q}{\Delta \text{Distance}}$, which is measured in units of tickets per mile, into $\frac{\Delta Q}{\Delta P}$, measured in terms of tickets per dollars. That is, we need to figure out how a 1 mile increase in distance from the stadium translates into *dollars*
- But it isn't that hard. If we could figure out what $\frac{\Delta \text{Distance}}{\Delta P}$ was, we could multiply $\frac{\Delta Q}{\Delta \text{Distance}} * \frac{\Delta \text{Distance}}{\Delta P}$ and get $\frac{\Delta Q}{\Delta P}$
- Once we do that, we simply need to grab the average P and Q paid by consumers, and we're home free

Full Cost

For each mile from the stadium, the full costs increase.

Transportation Costs

- As distance from the stadium increase by one mile, this means you have to drive two more miles to go to a game (one mile each way).
- \$0.502 / mile is government estimate of the cost of operating a vehicle
- Therefore, adding a mile from the stadium increases costs by \$1.004 (2 miles)
- But assume people are traveling in pairs and split transportation costs evenly...
- Means adding a mile from the stadium increases costs by \$0.502 per mile (per person).

Time Costs

- As distance from the stadium increases by one mile, this means you still spend more time driving to the game by however long it takes to drive the two miles (round trip)
- Assuming traffic moving at 60 mph, one mile from the stadium adds two minutes in travel time (one minute each way)
- If the average wage rate of the zip codes is \$24.76 per hour, and people value their leisure time at 60% of that rate, the value of time per hour is $\$24.76 * 0.6 = \14.86 per hour.
- Therefore, the extra two minutes cost = $\$14.86 * (2 / 60) = \0.495 , which...
- Means adding a mile from the stadium increases costs by \$0.495 per mile (per person)

Total

- Adding it Up $\frac{\Delta P}{\Delta Distance} = \$0.512 + \$0.495 = \$0.997 / \text{mile}$
- Which mean $\frac{\Delta Distance}{\Delta P} = 1 / 0.997 = 1.003 \text{ miles} / \1

Elasticity Calculation

- $\varepsilon = \frac{\Delta Q}{\Delta P} \frac{P}{Q}$
- And now we know that $\frac{\Delta Q}{\Delta P} = \frac{\Delta Q}{\Delta \text{Distance}} * \frac{\Delta \text{Distance}}{\Delta P}$
 $= 5.405 \text{ tickets/mile} * 1.003 \text{ miles/\$} = -5.421 \text{ tickets /\$}$
- The price of a season ticket is \$30.13 per game
- The average Q in a zip code is 179 tickets³
- Thus, $\varepsilon = \frac{\Delta Q}{\Delta P} \frac{P}{Q} = -5.421 * \frac{\$30.13}{179} = -0.911$

Interpretation / Conclusion

- Slightly inelastic pricing, as $|\varepsilon| < 1$.
- But due to sampling variation, we are not sure that our coefficient on $\frac{\Delta Q}{\Delta \text{Distance}}$ is exactly right, so it could actually be unit elastic.⁴
- Braves may be lowering price to increase concession revenues, etc.
- But bottom line, the Braves seem to be pricing correctly

Applicability to other Events

- Give survey
- Ask zip code / distance traveled, demographic information, wage / income level
- In many regards can do better with survey, as won't have to approximate income level with income level of zip code, or make assumptions about how many people are in the car
- Often done with recreational stuff (demand for national parks)

³ 179 tickets per 1000 people seems like a big number, but remember that a season ticket involves purchasing 81 tickets, so that is about 2.5 season tickets per 1000 people.

⁴ If you are thinking you could calculate a confidence interval, you could. But you'd have to still convert the coefficients +/- two standard errors (measured in tickets / mile) into tickets / \$. If you did so, you'd find the confidence interval ranges from -0.60 to -1.22, perhaps why it has been submitted to the second best sports economics journal, as it isn't so conclusive! Could be inelastic, could be elastic, could be unit elastic.

The paper

“Paper: Is There a Managerial Life Cycle? Evidence from the NFL.”

Authors: Brian Goff and Thomas Wisley.

Journal: *Managerial and Decision Economics* 27: 563-572 (2006)

The big idea

Is there a predictable relationship between coaching productivity and a coach’s age? Is there a “life cycle”?

Why should we care?

1. Applicability to other sports?

If we find this in the NFL, it likely would apply to MLB, NBA, college football, etc...

2. Applicability to other management positions? Consider the description of the head coaching position from the authors.

“Sets basic team policies, allocates decision rights among assistant coaches and players, coordinates, evaluates, and monitors the implementation of these policies. The head coach also assists in selecting, utilizing, and evaluating players. He is also responsible for choices regarding the combining of resources and the choice of ‘technology’ (plays, systems) to be used with these players. The head coach also makes strategic decisions relative to competitors. Further, football involves ongoing strategic decisions by coaches both in preparation for and during games, in determining offensive and defensive schemes, anticipating and reacting to other team schemes, making substitutions, and so on.”

I’d suggest that would be related to regular old managers.

3. Age discrimination legislation? AARP?
4. Other life cycle research.

Levin and Stephan – Scientists’ Publications
 Goodwin and Sauer – Economist’ Publications
 See also the last picture in the text.

Why do this with sports data?

The same reason we do other economics with sports – the data is available. It is “easy” to measure the productivity of a head football coach compared to, say, a university President, a professor, a police officer.

Human Capital Theory

Human capital – knowledge or skills that people have that help them be “productive.” Examples: taxicab driver (knowledge of city streets), brain surgeon (anatomy and steady hand), etc.

First, some human capital investments you all understand – going to school. Say you are trying to decide to spend some money on a college education. If we went to have a theory that predicts who goes to school and who doesn’t, we will be interested are the incremental costs of you decision and the incremental benefits of the decision. We will ignore things that you have to pay whether or not you go to college or not. Take for instance room and board. You have to eat and live somewhere if you go to college, you have

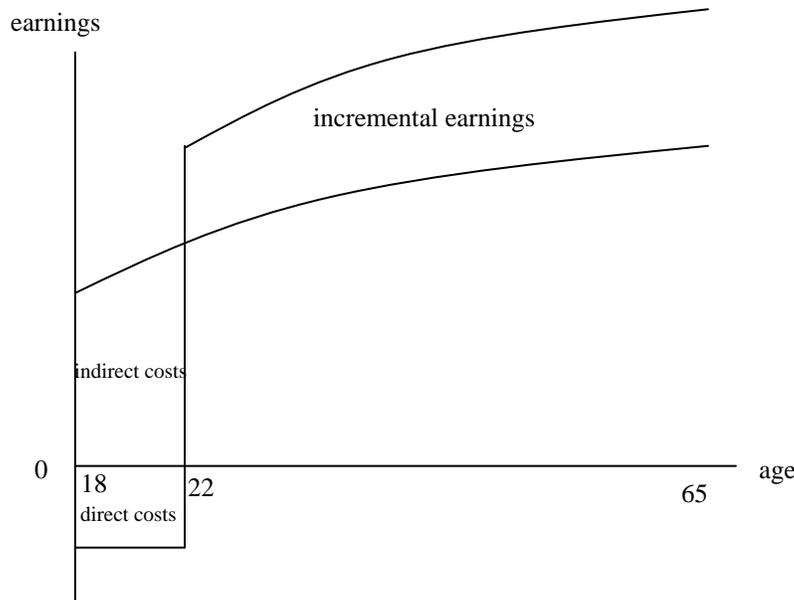
to eat and live somewhere if you go to college, so we will not worry about room and board when we do the calculations.¹

Costs:

- (1) Direct or out of pocket costs – tuition, fees, books, supplies, parking permits
- (2) Indirect or opportunity costs – foregone earnings you give up by not entering the labor force

Benefits:

- (1) Higher future earnings as a result of going to college.



We'll do the math, discount the future earnings (present value) and consider if the investment is "profitable". If the present value of the benefits $>$ present value of the costs, we'll make the investment and for shorthand, we'll say the investment is "profitable". If not, we won't invest in the skills.

Things we know or should be able to figure out pretty quickly...

- The higher the incremental costs of making the investment, the less likely the investment is profitable.
- The higher the incremental benefits (per year) of making the investment, the more likely the investment is profitable.
- The larger the number of years the incremental benefits will accrue, the more likely the investment is profitable.

¹ If your room or board is more expensive in college than you would have otherwise experienced if you hadn't go to college, it would be correct to include "excess" room and board in your costs of going to college. This is only strictly true if the room and board is of the same quality (a dubious assumption given cafeteria food I have experienced).

What is human capital for coaches?

Learn the skills, watching film, developing recruiting relationships, scouting new players, scouting new coaches, learning new motivational techniques, figuring out the Wildcat offense, learning to twitter, etc. The college football game changes over time and coaches have to “keep up”.

Or similarly, human capital depreciates – the stuff you learned three years ago isn’t worth as much. Knowing how to defend the wishbone was important 20 years ago, but not now. Without new “investments” to replace the “old” human capital, the level of human capital (skills) will decline, and hence winning percentage is sure to decline.

How does the likelihood of the investment being profitable change as the coach ages? We have to go back to those three items.

You could argue the first one (incremental costs) increases. Perhaps the coach is getting slower? That is, there is a biological argument here. But another story would be that a coaches time might be more valuable – hit the golf course?

The last one is certainly decreasing. A new unit of human capital that a coach invests in at age 30 will pay off for 35 years or more. A new unit of human capital that a coach invests in at age 60 will pay off for maybe only 5 years.

Whether it is the costs increasing, or the benefits decreasing, investments will be less profitable as a coach ages. Early in their career, a coach will invest a bunch, “learn” a lot and gain experience, increasing human capital. But at some point, the coach will find it no longer is “profitable” to replace the old human capital. Eventually, human capital will fall. Thus, overall, we’d expect a hump shaped human capital profile (an inverted U), and hence a hump shaped winning percentage for coaches!

Now we’ve got a theory of what the age profile of productivity should look like (according to human capital theory). What do we do next?

Grab some Data

Which is exactly what the authors did!

Their dataset: all NFL coaches from 1920 - 2004. 281 coaches, 1600 years of experience.

Oldest NFL coach – George Halas – Chicago Bears 72, 1967

Side note:

While each is a college football coach, Joe Paterno (Penn State) is 83, while Bobby Bowden turns 80 soon. That is old. How are they doing? They’ve been in the news.

Also, a Google search of “Who are the oldest coaches in the NFL” turned up this list. The answer is a bit dated, but that actually is nice for us.

Joe Gibbs (67) - Redskins – Since fired
 Tom Coughlin (61) - Giants – Still with Giants
 Romeo Crennel (60) - Browns – Since fired
 Wade Phillips (60) - Cowboys – About to be fired?

Sean Payton? 46. Remember that number.

What data to gather?

Obviously, need to measure coaches productivity. How?

- Winning percentage.

Of course, our point will be to isolate the effect on age, so we'll keep track of the coaches age.

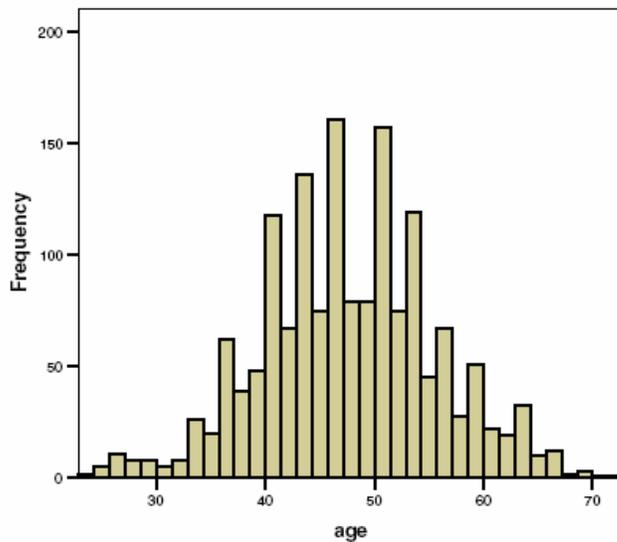
- Age

But what else could determine winning percentage aside from age?

- Differences in ability of coaches? See below.
- Team Endowments?
- Wealth of coaches?
- Biology?²

A peak at data on NFL coaches' age

A histogram of the dataset the authors used. Notice that there aren't that many young coaches, and there aren't that many very old coaches.

**Coaches of different Ability Levels**

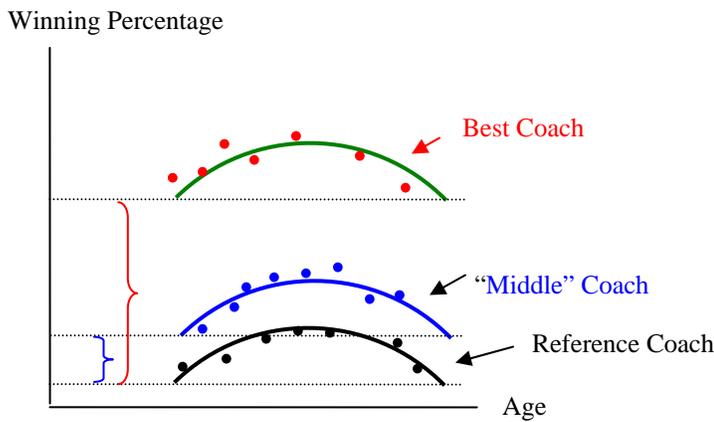
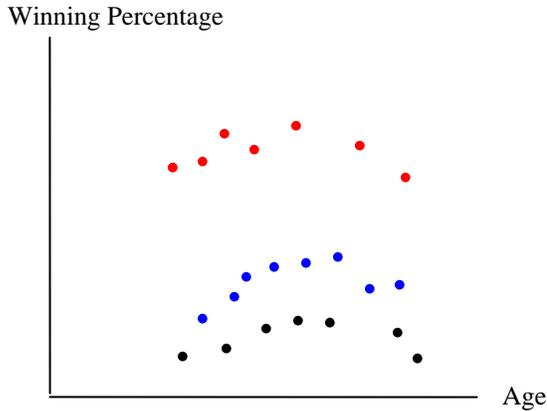
We can't just lump together all coaches and run a simple regression. Consider the first diagram below. Suppose there are three coaches of different ability levels and suppose they all have a hump shaped career profile. If we just threw them all and run a regression, we'd have a mess, even before we take care of the non-linear relationship.

We'll thus want to control for the coaches ability level so we can isolate the effect on age. The way we will do that is have a separate intercept for each coach. We'll have one intercept for some reference coach, say the one with the lowest winning percentage. Then we'll also estimate a differential ability term for each

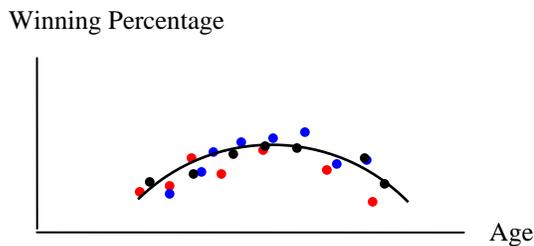
² With coaches, we wouldn't think this is a big deal, but when it comes to professional athletes themselves, there is a considerable amount of research. Folks have looked at changes in oxygen consumption as players age, changes in muscle mass as players age, skeletal changes, neurological changes, etc. The evidence there seems to suggest that most measures of productivity (a combination of learned skills and biology) peak at roughly age 27. Also, a 20-35% falloff between the age of 35 and 56

coach. The ability term for the 2nd best coach would be the small curly bracket in the diagram below (blue). The ability term for the best coach would be the large curly bracket (red) in the diagram below.

Once we are done with that, the age effect will be measured relative to each player’s “baseline” ability level. See the third picture below. And Excel or our statistical program can take care of that easily.



After adjusting for the coach specific intercepts:



When you see the regression results below, keep in mind that we’ll have a whole bunch of dummy variables....one for each **coach**. Again, the regression chooses a “reference” coach, and then estimates the size of the “curly bracket” for all the other coaches. Normally, people don’t report each coefficient separately for each coach as it would take up a bunch of space – but we could. (You won’t see those in the results table.

What else should we include?

Let's control for the winning percentage of the team the year before the coach got there, and call it **Endowment**. If you took over a good team, you'd be expected to do better. Nick Saban inherited an LSU team that went 3-8 in 1999. Miles inherited an LSU team that went 9-3 in 2004. We'd expect Miles to do better. The same would go for NFL coaches that inherited good NFL teams.

Let's control for if the coach was the coach of an **expansion** franchise, the **number of teams** in the league, if the coach got the job **mid-season**, which **league** (AFL/NFL) the coach coached for, and if the coach **switched** teams.

How do we deal with a non-linear relationship?

Thus far, we've only run linear regressions. But here, we think there is a non-linear relationship between age and winning percentage. A linear regression gives a constant slope – that won't do if we think the relationship is a hump (non-linear). (With a linear relationship, a one unit increase in age leads to an X unit increase in winning percentage regardless of age. This can't capture a non-linear relationship.)

Don't worry, all is not lost. How do we capture the non-linear relationship? We simply add an additional variable to the model, a quadratic term. So we'll have both age and age² in our regression. It is easy.

The Regression

$$Win\%_{it} = \beta_0 + \beta_1 age_{it} + \beta_2 Age_{it}^2 + \beta_3 Coach_{it} + \beta_4 Endowment_{it-1} + Other\ stuff$$

β_0 is just fancy notation for the intercept. Then age and the quadratic term on age (age²), the dummy variables for each coach, Endowment, and the other stuff.

If you are interested, you can worry about the subscripts. The i stands for each coach, while the t stands for time (the season). If not, don't worry about it

The results

Note: The authors have a typo in the second column. The coefficient on the quadratic term should be negative.

Note that the authors have p-values in parentheses. "<0.01" means the variable is statistically significant.

Table 2. Regression Results for Winning Percentage and Age, 1920–2004

Variable	Coefficient/(p-value)		
	All coaches	> = 10 Years	All coaches
<i>Age</i>	0.027 (<0.01)	0.029 (<0.01)	0.034 (<0.01)
<i>Age</i> ²	-3.1e-4 (<0.01)	3.3e-4 (<0.01)	-3.2e-4 (<0.01)
<i>Endowment (t - 1)</i>	0.24 (<0.01)	0.27 (<0.01)	0.22 (<0.01)
<i>Expansion</i>	-0.16 (<0.01)	-0.16 (<0.01)	-0.15 (<0.01)
<i>AFL</i>	0.14 (<0.01)	0.17 (<0.01)	0.16 (<0.01)
<i>Nteams</i>	0.008 (<0.01)	0.009 (<0.01)	0.008 (<0.01)
<i>Part Season</i>	-0.16 (<0.01)	-0.05 (0.86)	-0.15 (<0.01)
<i>Switch</i>	-0.06 (<0.01)	-0.06 (0.01)	-0.07 (<0.01)
<i>Tenure</i>			-0.008(0.01)
<i>Coach Effect F</i> ^a /(p-value)	3.37 (<0.01)	3.15 (<0.01)	3.07 (<0.01)
<i>R</i> ²	0.47	0.24	0.47
<i>F-Statistic</i> /(p-value)	3.97/(<0.01)	4.05/(<0.01)	4.01/(<0.01)
<i>N</i>	1600	745	1600

^aThis *F*-Statistic tests the null that coach fixed-effect terms equal zero. An intercept is included.

Interpretation of non-age variables

- Endowment – as we'd expect. Teams that begin with a better team win more games.
- Expansion – as we'd expect. Coaches of expansion teams don't fare well.
- N teams
- Part Season – as we'd expect – coaches beginning in partial seasons are inheriting an unstable situations.
- AFL – some bad teams in the AFL?

How do we estimate the age at which coaches peak?

Consider a simple relationship....

$$y = \beta_1 x + \beta_2 x^2$$

If we calculate the partial derivative with respect to x we get...

$$\frac{\partial y}{\partial x} = \beta_1 + 2\beta_2 x$$

In order to find the value of x where we obtain the maximum value of the function (y), we set the derivative equal to zero. We then solve for x .

$$\frac{\partial y}{\partial x} = 0 \Rightarrow \beta_1 + 2\beta_2 x = 0 \Rightarrow 2\beta_2 x = -\beta_1 \Rightarrow x = -\beta_1 / 2\beta_2$$

That means, that the value of y will be at a maximum at the value of x where $x = \frac{-\beta_1}{2\beta_2}$.

Now, back the situation at hand. Ignore everything but the terms involving age. Concentrate on terms with age get rid of the all the subscripts...

$$Win\% = \beta_1 age + \beta_2 Age^2$$

We just figured out how to take the partial derivative with respect to age...do it again here...

$$\frac{\partial Win\%}{\partial age} = \beta_1 + 2\beta_2 Age$$

To find the value of age where winning percentage reaches its maximum, set that equal to 0...and solve for age...

$$\frac{\partial Win\%}{\partial age} = 0, \Rightarrow \beta_1 + 2\beta_2 Age = 0 \Rightarrow 2\beta_2 Age = -\beta_1 \Rightarrow Age = \frac{-\beta_1}{2\beta_2}$$

A coach will reach his peak coaching ability when $Age = \frac{-\beta_1}{2\beta_2}$

Using the regression in the first column above...

$$\begin{aligned} \beta_1 &= 0.027, \\ \beta_2 &= 3.1e-4 \quad (0.00031) \end{aligned}$$

So peak winning percentage occurs is where:

$$Age = \frac{-\beta_1}{2\beta_2} = \frac{-0.027}{0.00031} = 43.6$$

Perhaps you should calculate the peak ages from the regressions in column 2 and column 3.³

What about the second and third columns

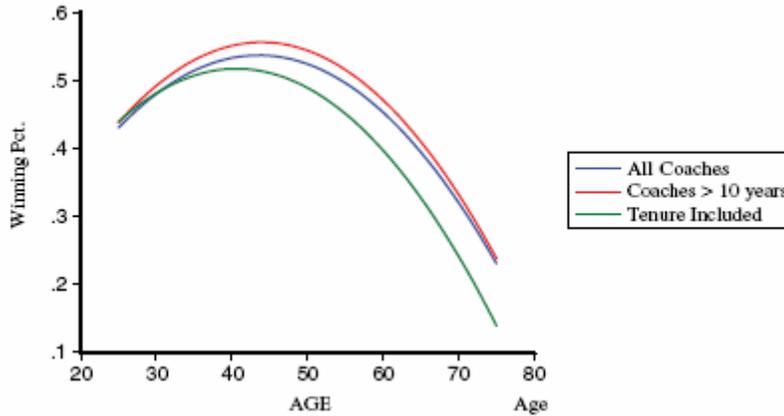
The authors were concerned that if you're a coach that lasts 10 years, there might be something different about the relationship between age and winning percentage that isn't simply captured by the dummy variables. To be "safe", they run the regression on just those coaches that "stuck around" a long time. You'll notice the answer isn't that much different.

The third column adds a variable called tenure, and doesn't do much either.

³ The authors report 43.5 years for the first column – just rounding error.

Could we plot the relationship between performance and age?

We sure could. We’d have to pick out some values for the other independent variables (other than age and age²), but once we did that, we could plug and chug and draw the picture. One for each column of regression table.



Is there more?

What about all those dummy variables, one for each coach, that they estimated? Can they tell us something about how good a coach each person is?

Maybe so.⁴ See below. When the authors did their regression, they chose the “reference coach” to be Sam Wyche. A value of 0.20 means the coach would be expected to win 20% more games than Sam Wyche. There are some familiar names to NFL fans in the list of coaches with “big values”. Not so many in the low values.

Table 5. Summary of Coach Effects for Coaches with 10 Years or More Experience

Range of effect	Coaches
> 0.2	Lombardi (0.47)**, Neale (0.40)**, Shula (0.36)*, Halas (0.36)**, Brown (0.35)*, Landry (0.33)**, Allen (0.32)*, Levy (0.29)**, Parker (0.29)**, Lambeau (0.26)**, Madden (0.26)**, Schottenheimer (0.24)**, Walsh (0.24)**, Ewbank (0.23)**, Grant (0.22)**, Owen (0.22)**, Gibbs (0.21)**, Marchibroda (0.21)**, Noll (0.21)**
0.10–0.19	Knox (0.19)**, Mora (0.19)**, Parcells (0.19)**, Coryell (0.18)**, Holmgren (0.17)**, Phillips (0.17)**, Gillman (0.16)**, Kurahich (0.16)**, Fisher (0.14)*, Vermeil (0.14)**, Ditka (0.13)*, Belichick (0.12)*, Cowher (0.12)*, Green (0.12)*, Saban (0.12)*, Lemm (0.11), Reeves (0.11)*, Van Brocklin (0.11)**,
0.0–0.09	Wilson (0.09), Gregg (0.08), Johnson (0.07), Pardee (0.05), Stram (0.03), Flores (0.01), Nolan (0.00), Wyche (0.03), Wannstedt (–0.03)

Notes: Coefficients are from the regression in the first column of Table 3. The coefficient for Sam Wyche is the reference value. *indicates that the coefficient is significantly different from 0 at or below the 10% level. **indicates that the coefficient is significantly different from 0 at or below the 5% level.

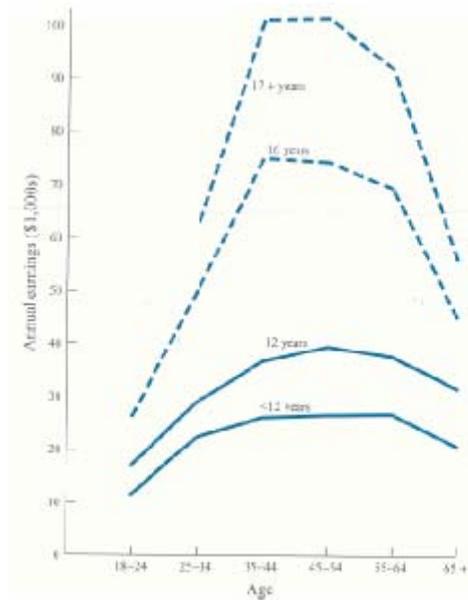
⁴ The coefficients here are actually the results of another regression reported in the paper, not those directly involved in Table 2. Nonetheless, they’ll give us a pretty good idea.

Applicability to the real world?

I hope you are already convinced, but just so you know, you've been tricked into doing some labor economics in this paper. We found above the coaches have a "life cycle" effect. They age. What about real people?

We do find that regular people have a "life cycle" effect, not just managers and professional athletes and coaches. See below for the picture for all people in the economy, classified by education level. Here the variable on the vertical axis is earnings (which ought to correlate with productivity) and the horizontal axis is age. This doesn't print well, but if you look at the file on the computer, it might be easier to see.

The point of course, is that this works for everyone. This whole human capital stuff seems to be important.

**Question**

Can anyone recall a coach that "went out at the top of their game" on their own terms?

In class: Tom Osborne? I agree. Any others?

The paper

Paper: “Automobile Safety Regulation and the Incentive to Drive Recklessly: Evidence from NASCAR”

Authors: Russell Sobel and Todd Nesbit.

Journal: *Southern Economics Journal* 74: 71-84 (2007)

The big idea

Do automobile safety regulations (such as seat belts, passive restraint systems, airbags) cause drivers to drive more recklessly?

Why should we care?

1. 37,000 traffic deaths last year.
2. Seat belt usage varies widely across states from 66% in Massachusetts to 97% in Michigan.

Economic Theory – Peltzman Effect

Peltzman Effect is named for economist named Sam Peltzman, who came up with originally in the 1970s as the nation debated seatbelts.

At its most basic level, we can model automobile injuries as coming from two distinct elements:

$$\text{Injuries} = \text{Probability}(\text{Injury} \mid \text{Accident}) * \text{Accidents}$$

If you're not familiar with the probability notation, when we write “Probability (Injury | Accident)” we are denoting a conditional probability. This means, given an accident occurs, what is the probability an injury will result.

For example, if there are 100 accidents, and the conditional probability of being injured, given an accident has occurred is 0.2, then there will be 20 injuries.

So now we focus on automobile safety. In mind, we have things like seatbelts and airbags. The big question is – do seatbelts (or airbags) reduce the number of injuries?

You'd think that would be an easy question. First, the easy part...

If there is a safety improvement (think seatbelts), it seems fairly obvious the conditional probability term will be reduced. That is, conditional on being in an accident, the probability of injury is reduced. This undoubtedly, by itself, increases driver safety.

However, Peltzman points out something else happens at the same time. Now that there has been a safety improvement, it is now less costly for an individual to be in an accident, and thus drivers will expend fewer resources to prevent being in an accident. That is, they will drive more aggressively. **Drivers will respond to a decrease in the conditional probability of an injury by driving more aggressively, and thus the number of accidents will increase.** The increase in aggressiveness is called “offsetting behavior” or the “Peltzman effect”. If this happens, this increase in the number of accidents, by itself, increases injuries.

Combining these two effects, we see that we are unsure whether injuries actually increase or decrease as a result of the safety innovation! It depends on which of the two effects is bigger.

Let's recap, and return to the formula:

$$\text{Injuries} = \text{Probability (Injury | Accident)} * \text{Accidents}$$

- As a result of a safety innovation Probability (Injury | Accident) decreases
- But according to the Peltzman effect (offsetting behavior), as a result of a safety innovation, accidents increase
- So the number of injuries could go either up or down as a result of a safety innovation. Airbags and seatbelts could, theoretically, make us less safe!

So, what do we have to do? We turn to the data to determine two things

- (1) **Does the Peltzman effect really exist? That is, do drivers get more reckless after the conditional probability of an injury decreases? Basically, are there more crashes after safety improvements?**
- (2) **If the Peltzman effect exists, is it large enough to entirely offset the increased safety from the innovation? Basically, are there more injuries after safety improvements?**

Why do this with sports data?

With real accidents...

- there are different drivers, different roadways, different weather conditions, different traffic volumes, and different cars speed all of which might influence accident rates and are difficult to measure and account for
- there are data collection issues (different counties, states, collection techniques) above and beyond that described above – adding up these injuries is troublesome
- while we know the laws in those states (e.g. speed limit, seatbelts, use of cell phones), it isn't clear which states enforce those laws
- drivers have different insurance policies (e.g. high deductible vs. low deductibles) and hence different incentives

With NASCAR...

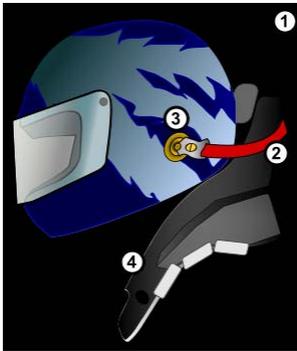
- Weather and racing conditions are known (race in dry conditions). We have a “driving history” for each driver. We can measure “traffic volume” and “car speed”
- Every crash and injury is observable (no data collection issues)
- Don't have to worry about enforcement / non-compliance issues – cars and drivers must pass inspections, no impaired drivers (except Jeremy Mayfield?)
- No differences in insurance policies, same incentives (ignoring start and park)

Background on NASCAR Auto Safety

- Dale Earnhardt Sr. at Daytona (the reason why there are head and neck restraints)
 - <http://www.youtube.com/watch?v=YmwI940wo2g>
- Edwards at Talladega (a reminder why there are restrictor plates)
 - <http://www.youtube.com/watch?v=QDUDd-sTwws>
- Joey Logano at Dover
 - <http://www.youtube.com/watch?v=bGImS2VvfVU>
- Story about Logano, aged 19, after first “big” crash. Is he a chicken afterwards? Michael Waltrip quote. See end of notes.

Modern safety features

- Roll cages (keep driver cockpit from being crushed)
- 5-point harnesses (deluxe seatbelt)
- Window nets (keep people in, stuff out of car)
- Lexan Windshields (don’t shatter)
- Fuel Cells (fuel doesn’t catch on fire up on crash, though engines do)
- Roof Flaps (keep cars from going airborne, slow down if going backwards)
- HANS (keeps head and neck immobilized, see picture below)
- Restrictor Plates (slow down cars)

**Key Assumption**

Drivers risk taking behavior will depend on their perceptions of risk. The authors assume that drivers’ perceptions of risk match the recent injury history. That is, the authors assume that drivers correctly perceive the risk of injuries. This seems reasonable because drivers see the crashes, talk to other drivers, etc.

Grab some Data

Races from 1972 – 1993 seasons, 631 races, 22 seasons

Our focus here is on whether drivers become more reckless after safety innovations. Thus, in our regression, we will use reckless driving as the dependent variable. But how do we *measure* reckless driving?

The basic answer is how many crashes there are. More reckless driving leads to more crashes. Some crashes involve one car, some 10 cars, some are severe, and some are minor. Additionally, because different races have different sized tracks, different number of laps, different number of miles run, it is not so clear cut how to measure reckless driving.

Possibilities:

- **Percentage of cars eliminated from the race because of accident**
- **Percentage of laps run under caution (talk about cautions)**
- **The number of caution laps**
- **The number of race miles run under cautions**

Authors try them all! But at the end, we will get the same answer with each, so we felt pretty good about the story that the regressions are telling.

The main independent variable we want to focus on is Probability (Injury | Accident).

But there is other stuff we should include.

What else should we include?

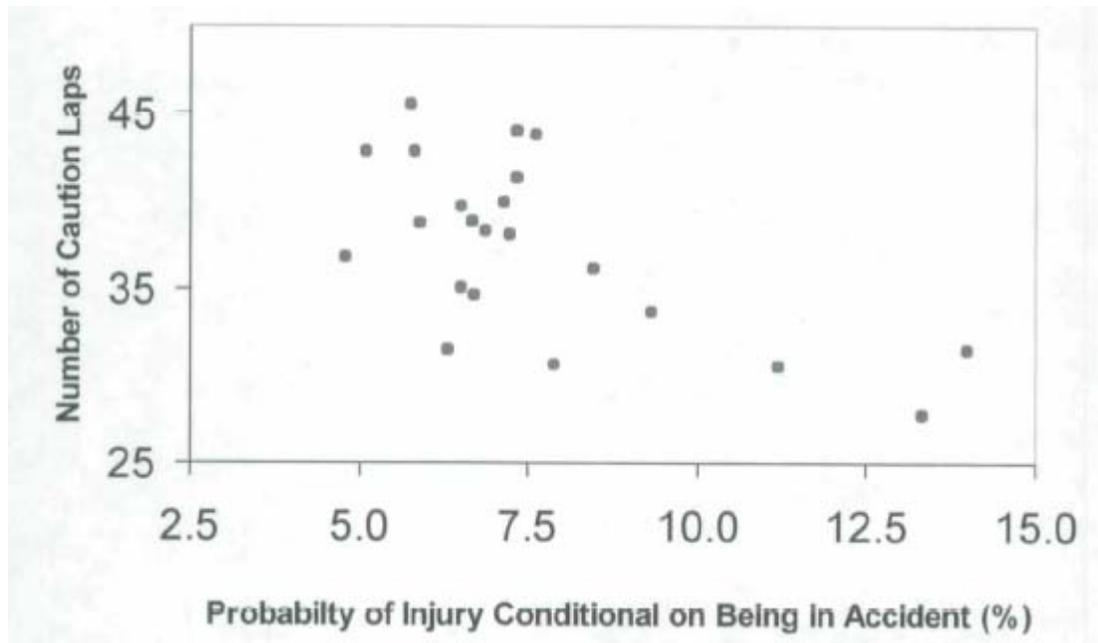
Surely there are things other than recklessness that will affect the number of crashes. To isolate the effect of safety innovations on recklessness, we'll want to control for these other factors. What else should we include?

- Race Distance – depends on dependent variable
- Cars per Mile of Track – heavier “traffic” density leads to more crashes
- Prize Differential between 1st and 2nd – bigger incentive to win and thus incentive to take risks (according to Tournament theory)
- Pole qualifying speeds – higher speeds makes it more likely to avoid accidents
- Percentage of cars that lead race – as more cars lead the race, the cars are more competitively equal, not spread out as much

A quick peak at data

For each season, the authors count up number of cars in accidents. They then calculate the number of injuries. By taking injuries / accident, they have calculated the conditional probability of injury given accident.

The next collect a measure of recklessness – the average number of caution laps per race. And then they graph them, with one observation for each season. See picture below.



This picture is certainly very suggestive of offsetting behavior. The lower the probability of being injured (as we move to the left along the horizontal axis), the larger the number of caution laps (more crashes.)

But as you know, we want to control for other stuff that might matter, so we'll do a regression. Hurray!

The Regression

$$\text{"Crashes"} = \beta_0 + \beta_1 * \text{Prob}(\text{injury} | \text{accident}) + \beta_2 * \text{Race Distance} + \beta_3 * \text{Cars Per Mile} \\ + \beta_4 * \text{Prize Difference} + \beta_5 * \text{Pct Cars Lead} + \beta_6 * \text{Pole Speed}$$

where "Crashes" is either

- Percentage of cars eliminated from the race because of accident
- Percentage of laps run under caution
- The number of caution laps
- The number of race miles run under cautions

The authors also included a separate intercept for each track to control for other differences in tracks not captured by the above variables. (A track might be particularly slick or bumpy leading to more accidents.)

As the purpose is to focus on offsetting behavior, we focus on the coefficient on the first slope term, β_1 . According to Peltzman's theory, the sign should be negative. As the conditional probability on injury decreases, reckless driving should increase.

The Results – Table 1: Race Level Data

[Bigger version of results at end of the notes]

Table 1. Race-Level Track Fixed Effects Model, 1972–1993

	Dependent Variable							
	Percentage of Cars Involved in Crashes		Percentage of Laps Run under Caution		No. of Caution Laps		No. of Race Miles under Caution	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Conditional probability of injury	-0.28*** (3.13)	-0.21** (2.49)	-0.40*** (3.70)	-0.35*** (3.43)	-1.13*** (3.60)	-0.96*** (3.27)	-1.39*** (3.66)	-1.18*** (3.40)
Constant	8.07*** (3.00)	-12.18** (2.47)	20.55*** (7.27)	25.11*** (4.02)	1.05 (0.17)	0.68 (0.05)	10.05 (1.58)	30.82* (1.69)
Race distance (×10 miles)	0.02 (0.34)	-0.05 (1.01)	-0.14** (2.49)	-0.22*** (4.03)	0.47*** (5.61)	0.24*** (2.98)	0.12*** (7.43)	0.88*** (6.34)
Cars per mile of track	0.21 (1.57)	0.21 (1.58)	0.24** (2.23)	0.22* (1.94)	0.84* (1.87)	0.78* (1.70)	0.85** (2.38)	0.77** (2.02)
First-to-second-prize differential (2000 dollars) (×\$10,000)	0.03* (1.85)	0.03* (1.84)	-0.01 (0.29)	0.001 (0.04)	-0.03 (0.28)	-0.01 (0.07)	-0.05 (0.36)	-0.02 (0.15)
Percentage of cars that led race		0.23*** (5.97)		0.34*** (9.97)		1.00*** (9.86)		1.34*** (10.37)
Pole speed for race		0.09*** (3.45)		-0.05 (1.40)		-0.06 (1.21)		-0.20** (2.21)
R ²	0.14	0.22	0.20	0.31	0.38	0.46	0.45	0.54
Observations	631	631	631	631	631	631	631	631

*** indicates statistical significance at the 1% level, ** at the 5% level, and * at the 10% level. Absolute t-ratios appear in parentheses and have been corrected for heteroskedasticity using White's matrix. All regressions include dummy variables for each track, which have been suppressed from the table. Full results are available from the authors on request.

Interpretation of slope coefficient on offsetting behavior variable

In all models, the slope coefficient on conditional probability of injury is negative and statistically significant, confirming offsetting behavior.

NASCAR drives increase recklessness as a result of increases in safety. There is a Peltzman / offsetting behavior effect.

Interpretation of control variables

As we focus on the above variable, not so important, but nonetheless...

- Race Distance
 - Depends on which version of the dependent variable we use
- Cars per Mile of Track
 - Positive (according to theory) and usually statistically significant
- Prize Differential between 1st and 2nd
 - When statistically significant, positive (according to theory), but negative occasionally
- Pole qualifying speeds
 - Results vary – picking up different track types?
- Percentage of cars that lead race
 - Positive (agrees with theory) and nearly always statistically significant

Can we determine if injuries (not accidents) increase?

Recall, the safety improvement is reducing the conditional probability of an injury given an accident (which tends to reduce the number of injuries), but our regression shows there is an increase in accidents as result of the recklessness (which tends to increase the number of injuries).

Which effect is bigger? Can we figure out if, overall, injuries increase?

We can but the math is messy, both because we have to do some total differentials with the chain rule involved, and also because we are dealing with probabilities, which makes the interpretation of a “one unit change more difficult.

Here is the bottom line. We can come up with an elasticity of accident with respect to improved safety.

Roughly speaking, a 10% increase in automobile safety results in about a 2% increase in reckless driving. Or stated a bit differently, 20% of the initial safety increase is offset by the Peltzman effect, but overall, drivers are still, on net, less likely to suffer injuries. They “give back” 20% of the safety increase.

So with that, we can answer our original questions.

- (1) **Does the Peltzman effect really exist? That is, do drivers get more reckless after the conditional probability of an injury decreases? Basically, are there more crashes after safety improvements?**

The answer here is yes, as we see a negative and statistically significant sign on the probability term in the regressions.

- (2) **If the Peltzman effect exists, is it large enough to entirely offset the increased safety from the innovation? Basically, are there more injuries after safety improvements?**

The answer here is no, as while the sign on the probability term in the regression is negative and significant, the size of the sign (coupled with some additional information on the mean number of injuries and probability of being involved in an accident in order to calculate the elasticity) is such that behavior is partially, but not completely offset.

Is there more?

The authors repeated the analysis using whole seasons instead of races as the unit of measure. The results are very similar. See Table 2 if you're interested.

Is there still more?

If you're thinking that these results are bogus because this might be the result of marginal drivers hopping in and out of NASCAR, you might want to do another analysis focusing only on a few drivers with long careers. The authors did such a thing, selecting 5 drivers (Cale Yarborough, Benny Parsons, Bobby Allison, Dave Marcis, and Richard Petty) and performed the analysis just on these 5 racers in 275 races that they completed.

The only difference is rather than using the number of caution laps or crashes as a dependent variable, they let the dependent variable be 1 if the drivers were involved with a crash in the race and 0 if they were not.

It seems funky, but the end result is that the regression can estimate of the probability that these drivers were in a crash (which ranges from 0 to 1). Again, our “main” independent variable is the conditional probability of injury, and throw in the same control variables.

Table 3. Binomial Probit and Logit Models; Marginal Effects Reported.

Variable	Probit		Logit	
Conditional probability of injury	-0.015* (1.746)	-0.015 (1.628)	-0.015* (1.739)	-0.015* (1.664)
Constant	0.002 (0.008)	-0.093 (0.179)	-0.005 (0.025)	-0.091 (0.177)
Race distance (×10 miles)	-0.004 (1.290)	-0.006 (0.629)	-0.004 (1.204)	-0.006 (0.572)
Cars per mile of track	0.0005 (0.225)	0.011 (0.656)	0.001 (0.316)	0.011 (0.655)
First-to-second-prize differential (2000 dollars) (×\$10,000)	0.019 (1.637)	0.014 (0.993)	0.018 (1.639)	0.013 (0.940)
Track fixed effects	No	Yes	No	Yes
Log-likelihood ratio test	10.41**	19.99	10.33**	19.99
Occurrences	275	275	275	275

Dependent variable = 1 if at least one driver in group was involved in an accident. The group of drivers used in the regressions includes C. Yarborough, B. Parsons, B. Allison, D. Marcis, and R. Petty. ** indicates statistical significance at the 5% level and * at the 1% level. Absolute *t*-ratios appear in parentheses. The fixed effects regressions include dummy variables for each track, which are suppressed from the table. Full results are available from the authors on request.

Let's pay only attention to the first variable.¹ What we see again in that in 3 of the 4 models, the sign on the conditional probability of injuries is negative and significant (and in the 4th, it's pretty close.) This means that the probability one of the 5 drivers is involved in an accident is negatively and statistically significantly related to the conditional probability of an injury.

Or in "regular" word, the less likely the chance of an injury given an accident, the higher the likelihood of a driver to be involved in an accident. This confirms that for these drivers, they too display offsetting behavior, which makes us even more confident that offsetting behavior exists.

Applicability to the real world?

The advantage of using NASCAR data, as noted above, allows for the research to control for many factors that are difficult to control for in the real world. It gives us something closer to a "natural experiment" to learn about offsetting behavior / Peltzman effect.

At a basic level, the results here suggest that then when safety features are considered, this offsetting behavior should be taken into account.

Auto manufacturers and government often do cost / benefit analyses of safety features. The results here suggest that these experts should consider that drivers will "give back" a portion of their increase safety by driving more aggressively.

[Talk about Canada and airbags]

- By the way, what happens to pedestrian fatalities as a result of seat belts and airbags?
- In North Carolina, they enforce the seat belt law while in South Carolina, they do not. Where are there more pedestrian fatalities? Where are accidents more severe?
- How safe would you drive if instead of an airbag coming from the center of your steering wheel there was a super sharp dagger pointed at your chest? Would you drink 3 beers and go driving?

[Talk about one for the road]

¹ I'm cheating a bit here. What you actually see in the table are called marginal effects, which are calculate from the actual regression coefficients. If you're interested, come talk to me.

Extensions - Are safety increases good for NASCAR?

Recall that the result is more accidents, but fewer injuries. Will this making racing fans happy? Why are Daytona and Talladega so popular? Because of high speeds? Perhaps. Because of “cool” wrecks? Maybe.

Some folks say that NASCAR fans like accidents (they typically cheer after big accidents). Presumably, fans don't like injuries or when their favorite driver (or even their least favorite driver) dies at the track.

So an improvement in safety is good news on both fronts, no? More crashes and less injuries?

- Is there then a profit-motive for improving safety of drivers for NASCAR?
- Are there an optimal number of accidents for NASCAR? Too few accidents means not enough excitement? Too many and we're watching bumper cars?
- Do team owners feel differently? Who suffers the costs of accidents?
- What does this have to do with the prize structure of NASCAR races?

Quote**From the authors of the paper:**

“How do drivers react to having cars so safe that they can generally walk away with no injuries when they crash it into a concrete wall or another car at very high speeds? The answer is that they race them at 200 miles per hour around tiny oval racetracks only inches away from other automobiles – and have lots of wrecks.”

From Michael Waltrip regarding how Joey Logano would react to his big crash:

"Generally, crashes that look terrible and you don't get hurt in actually makes a driver more confident. It makes you more aggressive because you know you are going to fine."

Both seem to me to be rather clear statements of offsetting behavior.

Means of variables, if we need them**Table A1.** Descriptive Statistics, 1972–1993

Variable	Race-Level Data				Season-Level Data			
	Mean	Minimum	Maximum	Standard Deviation	Mean	Minimum	Maximum	Standard Deviation
Conditional probability of injury	7.63	3.65	14.69	2.50	7.62	4.79	13.98	2.39
Percentage of cars involved in crashes	7.62	0.00	36.67	7.24	7.50	4.06	10.71	2.08
Percentage of laps run under caution	12.77	0.00	46.00	7.22	12.81	9.91	16.27	2.10
No. of caution laps	38.26	0.00	133.00	23.90	37.43	27.84	45.56	5.00
No. of race miles under caution	49.12	0.00	169.50	31.81	57.12	41.78	69.60	8.11
Race distance ($\times 10$ miles)	38.42	12.50	60.00	13.31	38.06	36.93	39.23	0.66
Cars per mile of track	32.33	10.40	68.57	16.29	32.08	30.50	35.39	0.96
First-to-second-prize differential (2000 dollars) ($\times \$10,000$)	3.26	0.00	170.16	7.28	3.27	1.79	8.41	1.68
Percentage of cars that led race	20.44	2.94	65.00	7.82	20.11	12.11	25.23	3.36
Pole speed for race	145.63	84.12	212.23	34.97	145.14	136.44	151.66	5.22

Sources: Fielden (1989, 1990, 1994) and Golenbock and Fielden (1997).

Slightly More Readable Version of Table 1**Table 1.** Race-Level Track Fixed Effects Model, 1972–1993

	Dependent Variable							
	Percentage of Cars Involved in Crashes		Percentage of Laps Run under Caution		No. of Caution Laps		No. of Race Miles under Caution	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Conditional probability of injury	-0.28*** (3.13)	-0.21** (2.49)	-0.40*** (3.70)	-0.35*** (3.43)	-1.13*** (3.60)	-0.96*** (3.27)	-1.39*** (3.66)	-1.18*** (3.40)
Constant	8.07*** (3.00)	-12.18** (2.47)	20.55*** (7.27)	25.11*** (4.02)	1.05 (0.17)	0.68 (0.05)	10.05 (1.58)	30.82* (1.69)
Race distance (×10 miles)	0.02 (0.34)	-0.05 (1.01)	-0.14** (2.49)	-0.22*** (4.03)	0.47*** (5.61)	0.24*** (2.98)	0.12*** (7.43)	0.88*** (6.34)
Cars per mile of track	0.21 (1.57)	0.21 (1.58)	0.24** (2.23)	0.22* (1.94)	0.84* (1.87)	0.78* (1.70)	0.85** (2.38)	0.77** (2.02)
First-to-second-prize differential (2000 dollars) (×\$10,000)	0.03* (1.85)	0.03* (1.84)	-0.01 (0.29)	0.001 (0.04)	-0.03 (0.28)	-0.01 (0.07)	-0.05 (0.36)	-0.02 (0.15)
Percentage of cars that led race		0.23*** (5.97)		0.34*** (9.97)		1.00*** (9.86)		1.34*** (10.37)
Pole speed for race		0.09*** (3.45)		-0.05 (1.40)		-0.06 (1.21)		-0.20** (2.21)
R^2	0.14	0.22	0.20	0.31	0.38	0.46	0.45	0.54
Observations	631	631	631	631	631	631	631	631

*** indicates statistical significance at the 1% level, ** at the 5% level, and * at the 10% level. Absolute t -ratios appear in parentheses and have been corrected for heteroskedasticity using White's matrix. All regressions include dummy variables for each track, which have been suppressed from the table. Full results are available from the authors on request.

Background terminology

- Ex-ante: before the fact (speculation, estimates, guesses, predictions)
- Ex-post: after the fact (verifiable, statistics, facts, results)

Why do we see ex-ante studies instead of ex-post studies?

- To plan and make decisions we need to estimate, predict, forecast
- It would be nice to “go back”. But rarely happens, and politicians are long gone.

Rough estimates of economic impacts

- NFL Superbowl: \$300 - \$400 million
- MLB All Star Game: \$75 million
- MLB World Series: \$250 million
- NCAA Basketball Final Four: \$30 - \$100 million

Why do we see inflated ex-ante estimates?

To justify public subsidies!

- Grab any article where Superbowl was dangled as prize for Saints stadium improvement / lease.
- 13 / 15 new MLB stadiums built between 1970 and 1997 got All-Star Game

Who does it benefit to use of tax payer dollars to inflate estimates to fund mega-events?

- Owners of franchises?
- Politicians? Why? Obama / Chicago? Jindahl / Saints?

How are ex-ante estimates done?

- Take some surveys, make some estimates
- Try to figure out how many people will travel to the event, what they will spend, do the math, and then hit it with a multiplier (see below)
- May use past attendance, past surveys, hotel occupancy information, etc.

More terminology: direct vs. indirect effects and the “multiplier”

All economic impact reports will report a total economic impact, and will break it down into a “direct” effect and an “indirect” effect

- The direct effect is the spending (economic activity) brought about by people spending on the event. Examples: Hotel rooms, tickets, souvenirs, restaurants, bars
- The indirect effect is the spending (economic activity) brought about by people who received the direct effect then going out and spending money. Examples: a bartender earns tips from mega-event fans, spends their tips on a Doors album; a t-shirt seller spends their profits on a Nintendo Wii
- Total effect is a “multiple” of direct effect. The indirect effects and hence the multiplier come from complicated economic models. But in the end, just a number, like 1.8, or 2.4
- Example: Superbowl XXVIII in Atlanta. Alleged economic impact was 2,736 jobs and \$166 million. Direct effect \$76 mm. Indirect \$90 mm. Total \$166 million. Multiplier is roughly 2.2. 306,680 visitor days and expenditures per visitor was \$252 per visitor day

Easy to understand problems with ex-ante measures of direct and indirect effects: lies / overstatements

Problem 1. Overestimation of number of guests

- Example: 2005 Denver NBA All-Star game estimated as 100,000 visitors. Stadium has only 20,000 seats and city has only 6,000 hotel rooms

Problem 2. Overestimation of economic impact per guest

- US Open Tennis tournament \$420 million (3% of direct impact of all tourism for NY)
- 2006 World Cup South Africa \$6 billion (4% of GDP)
- 2002 World Cup \$24.8 billion for Japan (0.9% of GDP) and \$8.9 billion for South Korea (2.2% of GDP)

Problem 3. Wildly inconsistent estimates suggests something is amiss

- Different studies of impact of 1992 NBA All-Star game range between \$3 and \$35 million
- 1997 Women's Final Four NCAA Basketball (Cincinnati) estimated at \$7 million, while 1999 Women's Final 4 NCAA Basketball (San Jose) estimated at \$32 million
- Ask them where the numbers come from. "Um, ah, er,"

Harder to understand problems with ex-ante measures of direct and indirect effects:**Problem 1. Substitution Effect**

There are three main types of substitution effects

- Often a “**local**” will spend money at the event, rather than on other goods and services within the economy local economy. If this is the case, this resident’s spending on the event is not new spending – simply a reshuffling.
 - Solution for “good studies” – don’t include spending by locals.
 - Which will involve a bigger substitution effect? A Superbowl in the Superdome or a regular Saints game in the Superdome? Why?
 - Which will involve a bigger substitution effect? A Mardi Gras parade in New Orleans or a Mardi Gras parade in Thibodaux? Why?
- Often people show up in town for some other reason, but then attend the event. These persons would likely have spent money on something else in town. We call these people **casual visitors**.
 - At the very least, it would not be appropriate to include and spending on hotels and restaurants for these individuals in the economic impact, as they are not directly attributable to event.
- Occasionally, a person traveling for some other purposes may plan their visit to coincide with an event. I’m going to Hawaii this winter, I may as well go the weekend of the Pro-Bowl. We say these people are engaging in “**time-shifting**”. Person was going to go to Hawaii regardless – event didn’t affect *if* they were coming, only affected *when* they were coming.

All of the above are referred to as “substitution effects”, and failing to them into account will result in an overstated economic impact.

- In the case of locals spending on mega-event, they miss the fact the local is not spending elsewhere, overstating the economic impact.
- In the case of the casual visitor, they miss the fact that the guest would be spending elsewhere in the local economy, overstating the economic impact.
- In the case of the time-shifting individual, they miss the fact that the guest would have been spending on some other date in the local economy, overstating the economic impact.

Problem 2. Crowding Out

- Congestion from mega events dissuades recreational and business visitors from coming at the time of the mega-event.
- If hotels are near capacity, simply replaces other tourists.
- 2002 World Cup in South Korea – soccer tourists high, business tourists low. No net change.
- Superbowl in New Orleans pre Katrina?
- Superbowl in New Orleans post Katrina?

Problem 3. Leakages

- Money may be spent in local economies, but may not end up in pockets of local residents. But what is certain is that all that most (if not all) of the taxes will be collected from local residents.
- During mega-events, the relationships from which the economic models are based might be atypical. If you estimate a multiplier based on a run-of-the-mill economic situation, it might not be the same a mega-event.
- Consider spending on \$100 worth of beer at the Cat's Meow. It provides tips to the wait staff and bartenders, income to the owners, etc. Think of the extra economic activity (the indirect effects) of that spending. Then, imagine a mega-event where there is now \$300 worth of beer purchased. The indirect effects might be 3 times bigger. The bartender might actually spend three times more money.
- Same drill with a \$100 hotel room. It provides some income for the hotel owners, etc. Now, imagine a mega-event and a \$300 hotel room. Do the hotel maids earn 3 times more? The clerks? Does the \$200 stay in the local economy? Are the indirect effects as big as the model says? Or does it go back to Hilton's headquarters and shareholders?
- Reggie Bush makes a bunch of money, something touted when the economic impact of the Saints are discussed. Kim Kardashian was dating Reggie Bush. How often did you see Reggie and Kim in New Orleans when you watched her show? Is Reggie spending all of his salary in New Orleans?

Problem 4. Ignored Costs

- Costs of New Stadiums / Improved Stadiums
- Stadiums are costs, not benefits. They are stimulatory – do create jobs, but this money could have been spent elsewhere. An increase in spending on a stadium means less spending elsewhere (or higher taxes or more borrowing). While there are employment gains in the stadium construction, there are employment losses elsewhere (whatever else the money would have been spent on), that go unseen.
- In the state of Louisiana, there is a real possibility that the Saints deal will come in part, at expense at jobs at higher educational institutions such as Nicholls State University (hopefully not mine!).
- Let's also not forget, particularly with the Olympics, that such buildings can be very useless. 5,000 seat swimming buildings. Bicycle tracks?
- Transportation and Public Safety. For the Superbowl, said to be \$1.5 million. For the Greece Olympics, \$1.5 billion. Look up Chicago's Olympic bid?
- Looting / Violence (Chicago, Detroit, West Virginia)

How are ex-post estimates done? (How do you know the ex-ante estimates were inflated)?

- Compare economic performance of region that had event vs. similar regions that didn't have events
- Compare economic performance of region that had event vs. same region in different time period (with no event)
- Variables include: per capita income, employment, sales tax collections, hotel occupancy rates, airport arrivals

Problems

- Compared to size of economy, impacts are small, and thus hard to isolate.
- Easier to isolate the smaller the area (Bourbon Street vs. New Orleans vs. New Orleans Metro area vs. Louisiana)
- Monthly data better than quarterly data, quarterly data better than yearly data

Results on employment and income

- NBA All-Star game – reduces employment growth
- Olympics in Atlanta created between 4% to 55% of jobs promised
- Incomes in NCAA Final Four host cities lower in years when hosting tournament
- No statistical significant increase in incomes in cities holding major sports playoffs (NFL, MLB, NHL, NBA)
- Increase in incomes in Superbowl cities is 25% of what is promised by NFL (\$91 million vs. \$300 million)

Results on sales tax collections

- Better data to use, as sales tax collections data is gathered more frequently (monthly) and smaller geographic areas can be examined
- NFL claim of \$670 million in South Florida's tax able sales because of 1999 Superbowl, yet ex-post estimates find \$37 million
- MLB All-Star games in California – drop in taxable sales
- Generally, no positive and statistically significant results
- One study finds many jobs in Atlanta as a result of Olympics, but is a bit sketchy

What is the Bottom Line?

- Hold on to your hat. Ex-ante estimates run way higher than the ex-post estimates. Some times exaggerations are as large as a factor of 10.
- Most studies take care of crowding out fairly well (by excluding local spending), but use lousy multipliers (based on normal situations), and don't do well on leakages and crowding out.
- We (sports and other reputable economists and practitioners) could do better estimates. But they don't ask us! Why?
- The source of the information is important.
- Examine the incentives of the messenger!

Side Issue: Are there intangible (unmeasurable) benefits to mega-events?

Maybe. In fact, no economist claims there are not. For example...

- 1995 Rugby World Cup in South Africa (we're not so racist)
- 2006 NFL game in New Orleans post Katrina (we're not so wet)

Interesting study – cities that have professional franchises have higher rents.

- A potential implication is that people might be willing to pay more to live in cities with sports franchises?
- Ceteris paribus?
- Are we sure they are sports franchises and not other opportunities also in that city?
- Other economists shot it down

Other stories on intangible benefits

- Later visits
- Businesses relocate headquarters after visiting
- Television viewers might go after watching on TV (NFL Saints post Katrina?)

Could happen, but no evidence for any of the above, not even stories.

Side Issue: Are there intangible (unmeasurable) costs?

- Fires in Detroit after Piston's Championship
- Bribery in Salt Lake City after 2002 Winter Olympics
- Bombings in Atlanta during 2006 Summer Olympics
- NFL Saints game post-Katrina? (we're still wet)

Quote

Allen Sanderson (University of Chicago – one of the good ones).

“Anytime anybody uses the world invaluable, they are usually too lazy to measure it or don't want to know the answer.”

Is this only a sports economics problem?

Impact on Higher Education in Louisiana. \$8 return for each \$1 invested. Bullshit!

Daily Comet article on impact of Thibodaux Mardi Gras. Bullshit!

What should I focus on?

What you should worry about is...

- Ex-ante vs. ex-post
- Direct vs. indirect
- What are the incentives?
- In general, three easy problems
- In general, four tougher problems
- What is meant by each type of substitution (local substitution, time shifting, casual visitors)

Do not worry about any particular event or its impact.

Come test time – I'll likely give you a popular press article. You'll analyze / evaluate the claims.

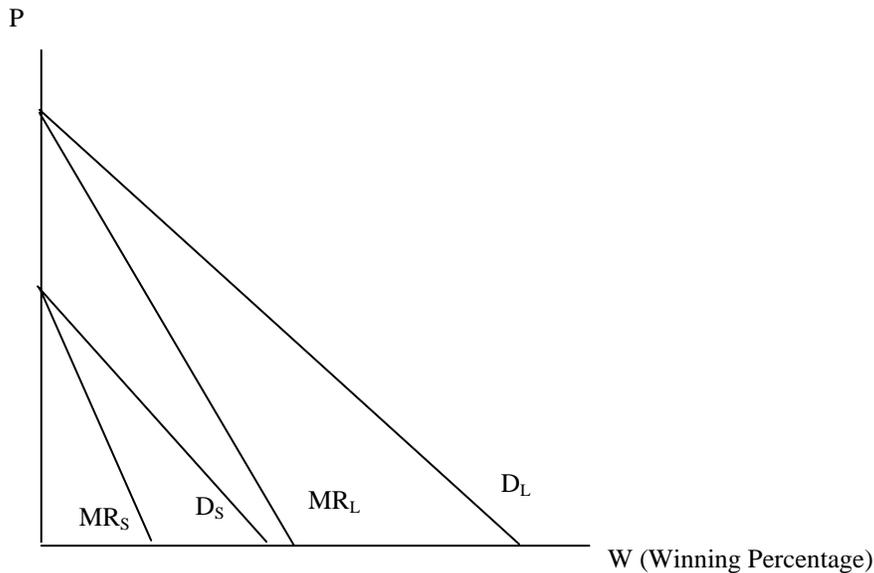
Demand and Marginal Revenue

Let's go way back. We know that once we have a demand curve, we can find a MR curve. We also know that shifters of the demand curve are items like population and income level. Let's put the quality of the team aside for a moment (we'll get to that I promise), and focus on the consequences of differences in "market size".

We know that not all cities are alike. Some are "large markets" and some are "small markets". When it comes to baseball, New York, despite having two teams, is a larger market than, say, Kansas City.

Because large markets have higher demand (let's say because of higher population), it stands to reason their marginal revenue curve is "higher". See the picture below. Here I have drawn the demand curve for a large market (D_L) and its associated MR curve (MR_L). I've also drawn a demand curve for a small market (D_S) and its associated marginal revenue curve (MR_S).

What we'll be after today is to examine the impact on this difference of market size on **competitive balance** – how closely matched are teams. If a league is competitively balanced, the winning percentages of teams in the league will be near 0.500, and the variability in winning percentage will be relatively low. If the league is competitively unbalanced, there will be a larger disparity in winning percentage across teams.



Let's make some simplifying assumptions. Let's assume that the thing teams produce (the measure of output) is winning games. More specifically, let's measure it as winning percentage (the fraction of the games they win). That will really help us to get at competitive balance.

Measuring Talent

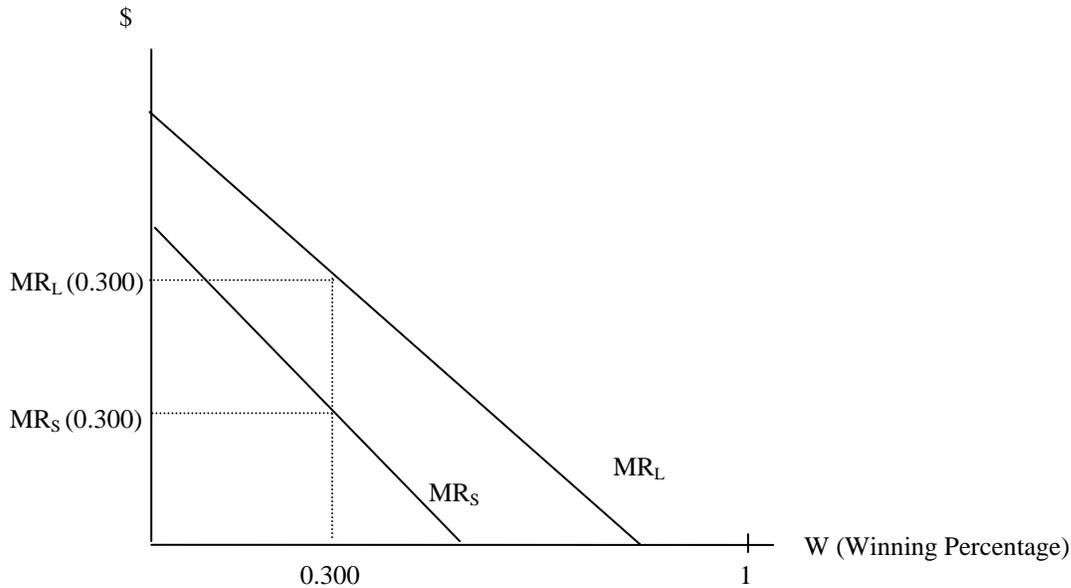
We'll also do one more trick. Let's measure talent in a peculiar way. Let's say that a unit of talent is the amount of talent it takes to increase a team's winning percentage by one point. What does this mean? Suppose you know if you hired Alex Rodriguez (one of the very best players in baseball), your team's winning percentage would increase by 5%. In this case, Alex Rodriguez would be 5 units of talent. A player that would increase your team's winning percentage by only 0.5% would be 0.5 units of talent.¹

Why do we do this? Because then we can measure both talent and winning percentage on the horizontal axis. A one unit increase in talent leads to a 1% increase in winning percentage.

¹ Later we will talk about the price per unit of talent. If a unit of talent costs \$4,000,000, then Alex Rodriguez will get paid \$20,000,000, while a player who is only half a unit of talent will get paid \$2,000,000.

Now, you may recall that we found there wasn't always a strong statistical relationship between payroll and winning percentage in baseball. For this story, we are suggesting, at least on average, there is a statistical relationship between *talent* and winning percentage.²

So let's start with the picture above, and get rid of the demand curves, just to de-clutter our graphs, and "zoom in" to make thing a little easier to see. Let's also re-label the vertical axis simply \$, as MR is just the additional revenue from winning a few more games (and so is measured in \$). See below.



In drawing the MR_L curve above the MR_S curve, we're suggesting here that every level of winning is more valuable (in terms of revenue) in the large market than in the small market. For example, at a winning percentage of 0.300 (30% of the team's games), the additional revenue associated with hiring an additional unit of talent is higher for the large market team (MR_L) than for the small market team (MR_S).

Leagues and the Setup

To get us to where we want to go – which is to examine competitive balance in leagues – we need to think about leagues. To make our life as simple as possible, we'll imagine a league with only two teams. And as the point it to get at competitive balance, we'll assume there is one large market team (L) and one small market team (S).

As we're ready to proceed, you might be tempted just to draw in a MC curve for each team, choose the level of talent / winning percentage for each team, and be done. This wouldn't be quite right. There is a constraint you'd be missing. Can it be that both (all) teams in a league can simultaneously increase their winning percentage? No. Every game has a winner and every game has a loser. The average winning percentage in a league, therefore, has to be 50%.

² How can I square those two statements? I've got two "outs". First, it might be the case that the actual payroll figures aren't accurate reflections of the talent level. For instance, teams with lousy GMs might overpay for players. In that case, payroll might not be an accurate predictor of talent, and thus payroll might not be an accurate predictor of winning percentage. A second story is that even while there is great deal of random variation in one season (injuries, bad luck, bad calls from umpires), on average, or in the long run, teams with higher payrolls will tend to perform better. Even if you don't buy either of those, we'll still get the point here from assuming that talent is related to winning percentage.

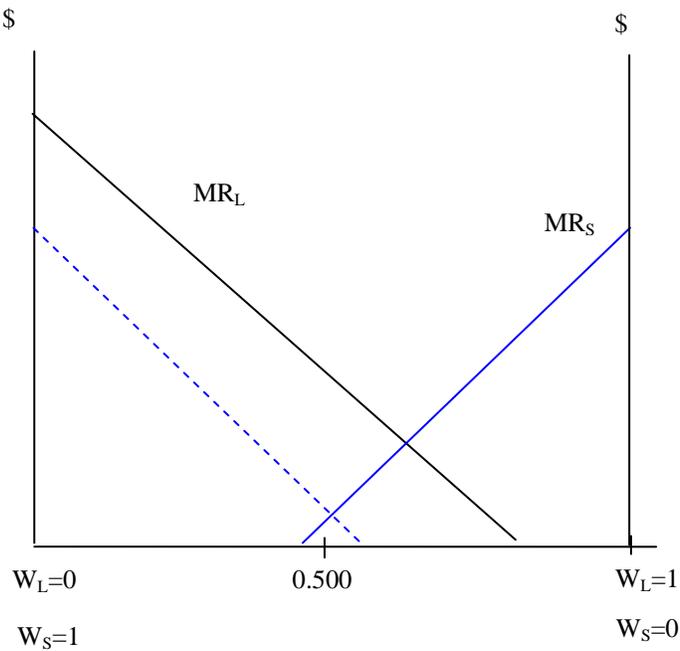
If team A goes 73-27, it has to be the case that team B goes 27-73. Team A’s winning percentage would be 0.730, while team B’s would be 0.270. In fact, $W_L = 1 - W_S$, or if we add up the winning percentage in our two-team league, it has to add up to 1.

So to exploit this fact, we make a box that allows winning percentage to vary from 0 to 1 and then we make one more trick and “flip around” the MR curve of the small market team. In the picture below, we remove the original MR_S curve (now shown as a blue dotted line) and replace it with the new MR_S curve (the solid blue line). Both convey the same information – though we have to be careful.

Just as before, as we move to the right on the graph, the large market team’s winning percentage increases. But at the same time, as we move to the right on the graph, the small market’s team winning percentage decreases.

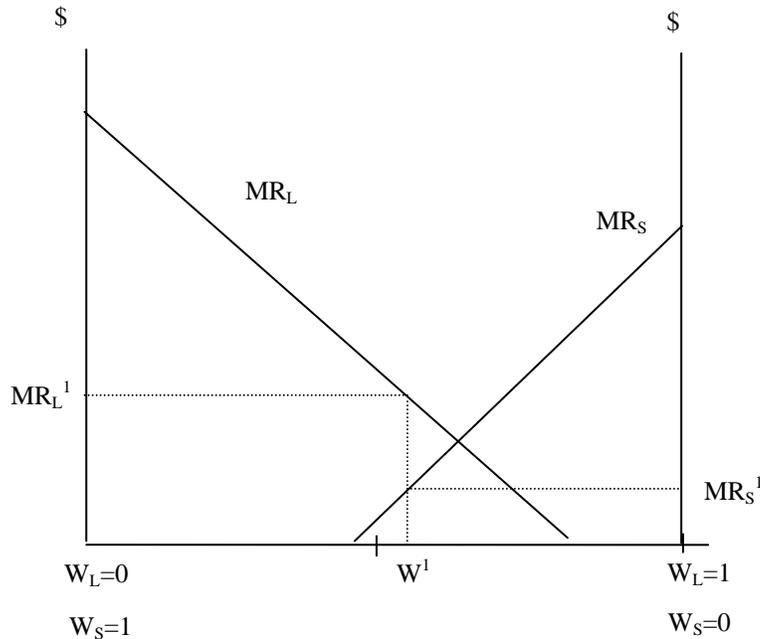
Another way to state this is as follows: as we move to the right, the large market team’s winning percentage increases. As we move to the left, the small market team’s winning percentage increases.

The nice thing about this trick is that now, if we choose any spot along the horizontal axis, it will satisfy the idea the winning percentages must sum to 1. One spot is where $W_L = 0.730$, $W_S = 0.270$. If we move a bit further to the right so that $W_L = 0.750$, our “box” ensures that $W_S = 0.250$.



Competition and Equilibrium

Now, we assume that each of the teams must compete for talent. We'll tryout a particular winning percentage and see if it is an equilibrium. See the picture below.



Is W_1 the equilibrium winning percentage? From the large market team's MR function (MR_L), we know the additional revenue for this team of hiring the last unit of talent is MR_L^1 . From the small market team's MR function (MR_S), we know the additional revenue for this team of hiring the last unit of talent is MR_S^1 . And in eyeballing these two amounts, it is clearly the case that $MR_L^1 > MR_S^1$.

What does this mean? It means the large market team could make more additional revenue from the last unit of talent hired than the small market team. Therefore, the large market team could make an offer to the last unit of talent that is larger than the offer the small market team can. We'd then expect that unit of talent will move to the large market team, increasing the large market team's winning percentage, reducing the small market team's winning percentage, and thus moving us to the right on the graph.

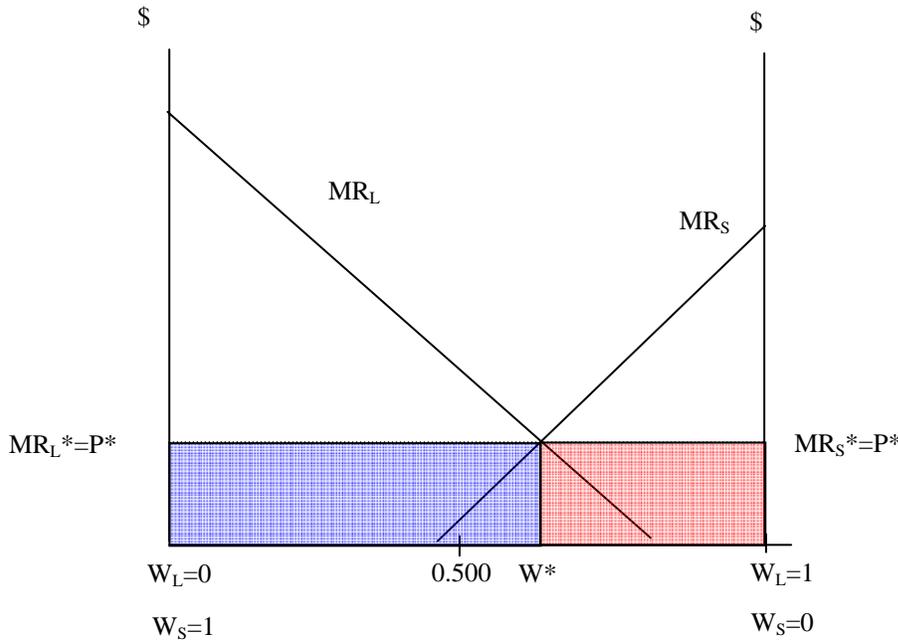
This means that W_1 was not an equilibrium. An equilibrium is a situation that is stable, or where teams don't want to change their behavior. As we saw above, the large market team will act to acquire more talent.

As we move to the right (as the large market teams acquires more talent), the discrepancy between MR_L and MR_S will decrease, but the large market team will still desire to bid away talent from the small market team. This will continue until we reach the point where teams neither team wishes to try to bid away talent from each other, which occurs where $MR_L = MR_S$.³

³ If for some reason, the winning percentage was to the right of the intersection of MR_L and MR_S , in that case $MR_S > MR_L$, and the small market team would be able to offer more money to the last unit of talent. The talent would move to the small market team, and the winning percentage of the small market team would increase, the winning percentage of the large market team would decrease, and we'd move to the left on the graph.

Thus, the equilibrium in this market is the winning percentage where $MR_L = MR_S$. See the picture below, where the equilibrium winning percentage is W^* .

Believe it or not, we've learned a lot of stuff, which will jot down in a moment. But note that in the competitive bidding process for talent, the teams will have to pay a price of P for each unit of talent, which will be equal to the MR each team.⁴



So we know a bunch already:

1. First, in equilibrium, $MR_L^* = MR_S^*$. That is, in equilibrium, the MR for an extra unit of talent will be the same for each team.
2. **The original revenue imbalance (the big market team having a higher demand curve and thus being a large revenue team) will result in a payroll imbalance. Large market teams will have higher payrolls.⁵**

How do we know? We can shade in the amount each team spends on talent. The price per unit of talent is P^* , and the amount of talent they hired is given by W^* (both winning percentage and the number of units of talent), so the spending on talent is given by P^* multiplied by W^* . It turns out

⁴ Students are always cranky about this point. Generally speaking, labor is paid how much it is “worth” – how productive it is. In the current context, the worth of the player is the additional revenue the team earns from hiring that player, simply MR. Competition amongst teams to attract the talent will result in a situation where the price of talent is equal to its marginal revenue ($P = MR$). Students always ask: How can the firm make any money if they pay the last unit of talent a salary equal to the additional revenue from hiring that worker? The firm doesn't earn any surplus on the *last* unit of talent hired, but does on all *previous* units of labor hired. Compare the MR on the 1st unit of talent to P .

⁵ Remember, the number of units of talent is not the same as the number of players. The large market team can still have the same number of players as the small market team and have a larger number of units of talent (the average player will be more talented).

graphically this is easy to shade in, as it forms a rectangle.⁶ I've shaded in the large market team's spending in blue, and the small market's team spending in red. Clearly, the blue rectangle is larger than the red rectangle.

3. **The original revenue imbalance (the big market team having a higher demand curve and thus being a large revenue team) has result in a competitive imbalance.** Note the equilibrium winning percentage (W^*) is to the right of 0.500. This means the large market teams win more than half the games and the smaller market team will win less than half their games. Because every unit of talent is more valuable to the large market team, they will hire more talent and win more games.

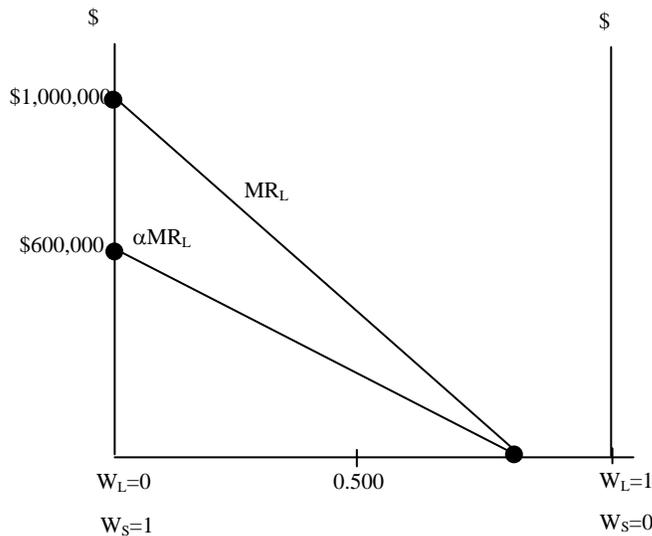
So what do I say to the people who are mad at the Yankees? They should be mad at New York. The reality of the matter in the Yankees being able to "buy" more players is that is the population and income level of New Yorkers to blame. It is exclusive territories lead to geographic monopolies, which are different in size. It is this difference in size and income levels that lead to differences in demand, and thus competitive imbalance.

Potential Remedy – Gate Sharing

What happens when teams only get to keep a fraction (α) of their gate revenue? Their new MR curve is simply the old MR curve multiplied by the fraction, α . How do we incorporate this in our picture? We draw a new MR curve that is 60% as tall as the old one. It is not that hard.

First imagine the vertical intercept of the MR curve. Perhaps it used to be the case that the additional revenue associated with a 1% increase in winning percentage was \$1,000,000. Now the team only keeps 60% of gate revenue, so the new additional revenue associated with a 1% increase in winning percentage is now \$600,000. The vertical intercept of the MR curve will be shifted down by 40%.

It seems silly, but if it used to be the case that the additional revenue associated with a 1% increase in winning percentage was \$0, but now the team only keeps 60% of gate revenue ($\alpha = 0.60$), then the new additional revenue associated with a 1% increase in winning percentage is still zero. So the horizontal intercept of the MR curve hasn't changed.

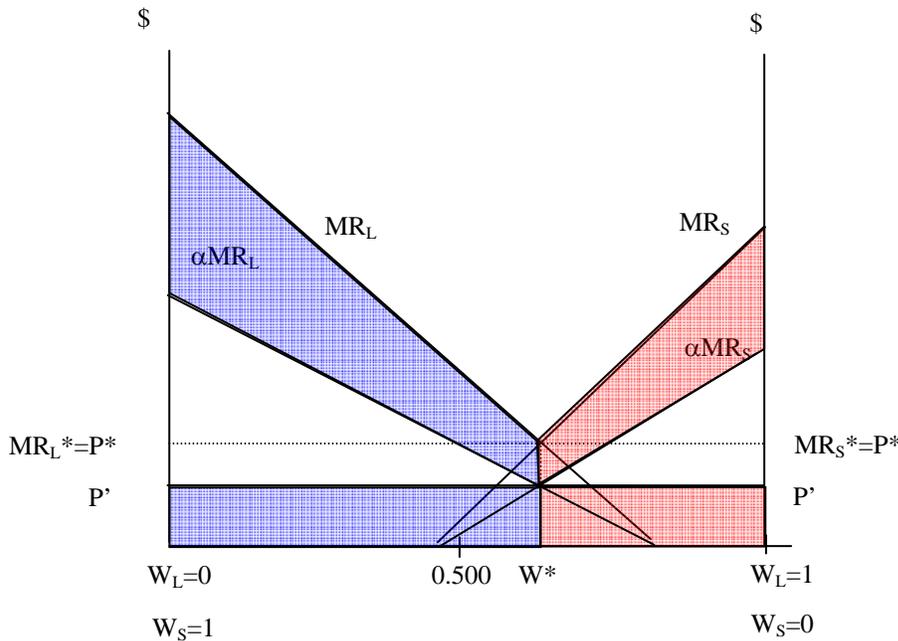


Of course, this happens for the small market team as well. So we shift down the MR curve each curve, and label the results αMR_L and αMR_S , respectively. See the picture below.

⁶ For example, suppose $W^* = 0.600$. This means the large market team hired 60 units of talent and the small market team hired 40 units of talent. The large market has spent $60 * P^*$ on talent, while the small market team has spent $40 * P^*$ on talent.

But does the level of talent change? Actually, no! (I lied before!) Notice that the equilibrium winning percentage is the same both before and after gate sharing is imposed, still labeled W^* . **Gate revenue sharing does not change competitive balance.**

Don't believe me? Let say, before revenue sharing, the equilibrium winning percentage where $MR_L = MR_S$ was previously 0.600, and at that winning percentage, $MR_S = MR_L = \$500,000$. Now, suppose $\alpha = 0.6$. At a winning percentage of 0.600, MR_L is now $= 0.6 * \$500,000 = \$300,000$, and MR_S is now $= 0.6 * \$500,000 = \$300,000$, and once again they are again equal.



So why did I lie to you before? It wasn't that I lied to you, I just didn't tell the whole truth. While revenue sharing reduces the incentive for big market teams to invest in talent, it also reduces the incentive for small market teams to invest in talent. Since each team is having the "return" to their investment reduced by the same percentage, these changes "cancel out".

What happens to the price paid for talent? Recall that the price of talent was equal to the MR of the last unit of talent hired. In the numerical example above, equilibrium MR falls from \$500,000 to \$300,000, meaning the price per unit of talent falls from \$500,000 to \$300,000. So while the allocation of talent is unchanged, the price of talent falls! On the graph, compare P' (the price of talent after revenue sharing) to P^* (the price before revenue sharing.) You can also notice that the amount spent by each team is smaller. Just as before, the amount spent by the large market team is a blue box, while the corresponding measure for the small market team is in red. Competitive balance here is a smokescreen from snatching money from players!

What about the revenue that is shared? The payments each team must pay to the other team are the gap between the old MR and the new MR.⁷ Thus, the revenue shared by the large market team is the trapezoid shared in blue, while the revenue shared by the small market team is the trapezoid in red. Because the blue trapezoid is larger than the red trapezoid, we know that the large market team pays more to the small market team than it receives from the small market team. Overall, the small market team receives revenues on net.

⁷ This is simply $(1 - \alpha)MR$. The only curve was MR, the new curve αMR , so the gap between the two is $MR - \alpha MR = (1 - \alpha)MR$.

Results

1. Gate revenue sharing does not change the allocation of talent across teams
2. Gate revenue sharing does not affect payroll imbalance
3. Gate sharing does not affect competitive balance
4. Gate sharing does affect the price per unit of talent, reducing it
5. Gate sharing does affect the wealth of the teams (the small market teams get some cash)

Potential Remedy: Draft?

Surely a backwards draft must alter competitive balance, right? Not so, says our theory.

Does having a draft change the marginal revenue associated with a unit of talent?⁸

So what happens then as a result? Both small market teams and large market teams draft players. Absent restrictions to prevent it, at the end of the day, the talent will get bid away to the large market teams until $MR_L = MR_S$.

Results

1. Players who get drafted get paid less than they would absent a draft because they must work for the team that drafted them (and there is no competition between teams that pull this player's salary up to MR).
2. Small market teams will get paid by large market teams for talent they bring in through the draft.

We'll come back to this idea later. It is called the Rottenberg Invariance principle.

Actual Remedy: Luxury Tax

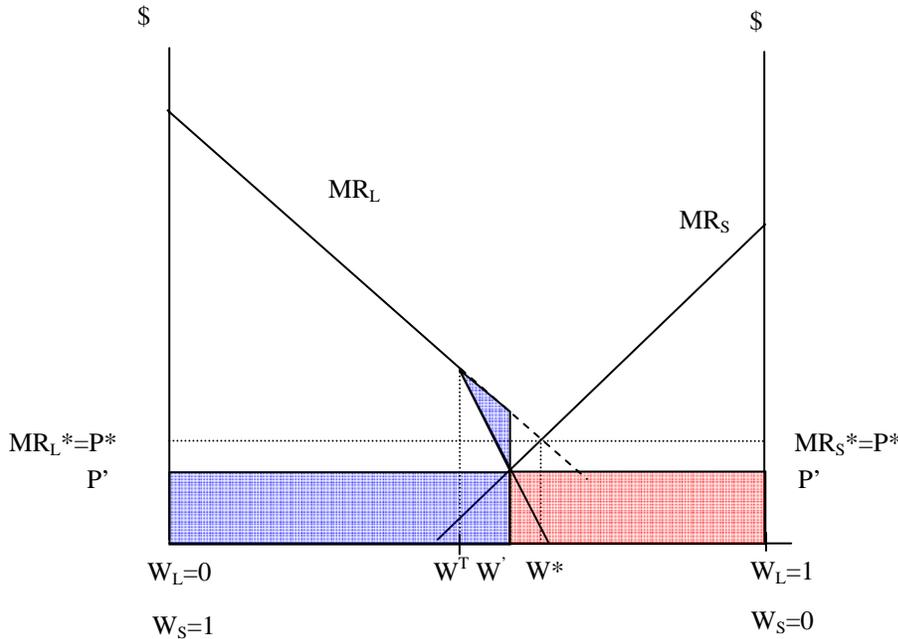
The idea of the luxury tax is that after some threshold level of payroll, a team exceeding that payroll level (a large market team) must pay a luxury tax. The revenue that is collected by the luxury tax will be distributed to low revenue teams.

For example, consider a luxury tax threshold was a payroll of \$150 million and a luxury tax rate of 0.50, and a team whose payroll was exactly a \$150 million payroll. If the team was to sign an additional player at a salary of \$10,000,000 the team must pay a luxury tax of \$5,000,000.

While it may seem natural to model this change as a difference in the price of acquiring talent, alternatively, we can model this by considering what happens to the MR of hiring an additional unit of talent. Consider what happened with revenue sharing – the team kept only 60% of the marginal revenue associated with the unit of talent, and thus the MR curve shifted down and became steeper.

This situation is similar. By having to pay a lump of money to the other team to hire the player, it is like a portion of the MR of player must be paid to the small market team. We can think of the luxury tax as reducing the MR of hiring an additional player after the threshold is met. In fact, it is almost the same story we had with gate revenue sharing, except that it doesn't "kick in" until the threshold. See the picture below. W^T is the level at which the luxury tax kicks in, W' is the new equilibrium winning percentage.

⁸ Have you ever said, "Oh, I'll pay more to watch my favorite baseball team because that particular player was drafted as opposed to acquired with a free-agent contract"? Me neither.



Notice that winning percentage falls from W^* to W' . Also, the price per unit of talent again falls. Just as before, the amount transferred to the small market team is the gap between the old MR curve and the new MR curve that reflects the luxury tax.

Note that here, the incentive to invest in talent is reduced, but only for the big market team.⁹

Results

1. Gate sharing does change the allocation of talent across teams
2. Gate sharing reduces payroll imbalance
3. Gate sharing improves competitive balance
4. Gate sharing does affect the price per unit of talent, reducing it
5. Gate sharing does affect the wealth of the teams (the small market teams get some cash)

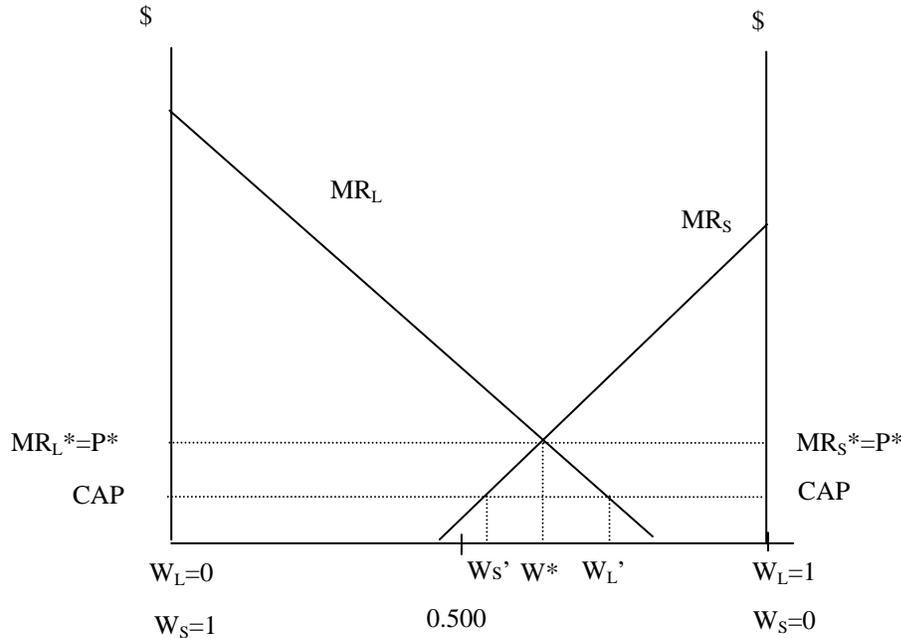
Potential Remedy: Salary Caps

In a perfect world, salary caps would be set up so each team would have a roughly 0.500 winning percentage. What you don't always see in the newspaper is that often salary caps (the maximum allowed to be spent) are accompanied with a minimum amount that must be spent.

Of course, for the salary cap to be effective, it has to reduce the price of talent. So we set a cap level that reduces the price of talent. See the picture on the next page.

But at the low CAP price, the small market team prefers to have less talent and the large market team prefers to have more talent, so **both** teams have an incentive to circumvent the cap at winning percentages between 0.500 and W_S' . The large market teams wish to spend more (until $W = W_L'$) and the small market teams wish to spend less (until $W = W_S'$)! As a result, there are all sorts of shenanigans that ensure. Teams find clever ways to go over the cap. We'll talk about the Larry Bird rule later on. If the cap can be avoided (it is not a "hard cap"), competitive balance is very unlikely to be altered. The casual evidence on the NBA suggests the cap can easily be circumvented.

⁹ Technically, we should draw in the steeper section for the small team, for as if they cross the threshold, they too will pay the luxury tax. But in our two team setup, we know that is unlikely that the threshold will be set up so that small market team has to pay a luxury tax, so we skip this step for simplicity.



Results

1. The price per unit of talent falls.
2. The ultimate affect on competitive and payroll balance depends on the how tightly the salary cap is enforced. With loopholes, competitive and payroll balance are unaffected.

Salary CAP (NBA)

- In NBA, cap was established in 1984/85
- Look at standard deviation of winging percentage before and after
- Naïve theory: Salary cap improves competitive balance, standard deviation of winning percentage will decline
- Our theory: Salary cap will not change competitive balance
- Average standard deviation from 1975/76-1983/84: 0.128
- Average standard deviation from 1984/85-1992/93: 0.156
- Is Naïve theory correct? No – moving the wrong way! Suggests that our theory is correct.

Free Agency in Baseball

- In MLB, free agency was established in 1976
- Look at standard deviation of winging percentage before and after
- Naïve theory: free agency will reduce competitive balance, standard deviation of winning percentage will increase
- Our theory: Free agency will not change competitive balance
- Average 1966-75: 0.068-NL, 0.068-AL
- Average 1976-85: 0.068-NL, 0.075-AL
- Is the Naïve theory correct? NL: unchanged, suggests our theory is right. AL: Moving in the right direction for the naïve theory, but not a statistically significant difference. Suggests that our theory is right.

Free Agency in Baseball – Championships

Some people claim that it isn't regular old competitive balance we should worry about with free agency, it's the championships we should worry about. Let's check those out.

- Number of League Championships for Each Team (10 Years Prior to Free Agency, 1966-75)

<u>NL</u>		<u>AL</u>	
Cincinnati	3	Baltimore	4
LA	2	Oakland	3
St Louis	2	Boston	2
New York	2	Detroit	1
Pittsburgh	1		

- Number of League Championships for Each Team (10 Years Prior to Free Agency, 1976-85)

<u>NL</u>		<u>AL</u>	
LA	3	New York	4
St Louis	2	Baltimore	2
Philadelphia	2	Kansas City	2
Cincinnati	1	Milwaukee	1
Pittsburgh	1	Detroit	1
San Diego	1		

- Don't see much of a difference in comparing big markets to small markets, suggesting our theory is right.

MLB Rookie Draft:

- In MLB, draft was established in 1964
- Look at standard deviation of winning percentage before and after
- Naïve theory: amateur draft will improve competitive balance, standard deviation of winning percentage will decrease
- Our theory: Rookie drafts will not change competitive balance
- Average 1952-63: 0.078-NL, 0.084-AL
- Average 1964-75: 0.069-NL, 0.071-AL
- Is the Naïve theory correct? It is -- we lose one! An interesting challenge for our theory! It appears that balance may have been enhanced by rookie draft in MLB. Why? Can't trade MLB draft picks?

NFL Rookie Draft:

- In NFL, draft was established in 1936
- Look at standard deviation of winning percentage before and after
- Naïve theory: amateur draft will improve competitive balance, standard deviation of winning percentage will decrease
- Our theory: Rookie drafts will not change competitive balance
- Average 1930-35: 0.245
- Average 1935-41: 0.266
- Is the Naïve theory correct? No going wrong way. Suggests our theory is right.

The Paper

<http://hubcap.clemson.edu/~sauerr/working/moneyball-v2.pdf>

The People

- Billy Beane, General Manager of Oakland A's (small market team)
- Michael Lewis, author of Moneyball. Formerly has written other books on financial markets and sports issues. (Liar's Poker – bond trading, Blind Side – NFL linemen, Panic – market bubbles).
- Bill James, long time statistical guru in baseball
- SABR (Society of American Baseball Researchers)

Lewis' Thesis

- There are many skills that go into producing output in baseball (winning games)
- Players have different combinations of skills
- Billy Beane believed that one such skill “getting on base” (walking, measured with on-base percentage, OBP) was underpriced in the baseball labor market
- By acquiring players who were “underpriced”, Beane believed the Oakland A's would be able to win games (produce output) at a lower cost than other teams. They A's / Beane though they could exploit an inefficiency
- Would the information spread? Would it do so quickly?

Baseball Productivity Measures

- The goal is to win games, and to win games you have to score runs (points). Batters try to score runs; pitchers and defense try to prevent that from happening.
- Focus on productivity measures of hitters

Productivity Measure 1: Batting Average

Batting Average = Hits / At Bats

- Basically, a ratio of successes to attempts (as hitter)
- A crude measure of how often the player is successful

Very good player's AVG = 0.300

Mediocre player's AVG = 0.260

Bad player's AVG = 0.220

- Flaw is that it treats all hits the same. Some hits are singles (one base), while other hits are homeruns (four bases).
- Also leaves out some productive activities (walks, sacrifices)

Productivity Measure 2: Slugging Percentage

Slugging Percentage = (Singles + 2 * Doubles + 3 * Triples + 4 * Home Runs) / At Bats

- Starts with idea of batting average, but awards more “points” for different hits
- Counts a double (2-base hit) as 2, a triple as 3, and a home run as 4.

Very good player’s SLUG = 0.550

Mediocre player’s SLUG = 0.400

Bad player’s SLUG = 0.320

- Improvement over batting average, but still doesn’t include everything. There are things that help teams score runs that aren’t included. Sacrifices, but importantly ignores walks!

Productivity Measure 3: On Base Percentage

On Base Percentage \approx (Hits + Walks) / (At Bats + Walks)

- Starts with same idea as batting average, but includes walks in the analysis. Basically, of the number of opportunities that a hitter had, how many of them involve the hitter getting on base.¹

Very good player’s OBP = 0.420

Mediocre player’s OBP = 0.340

Bad player’s OBP = 0.260

Productivity Measure 4: OPS (On Base Plus Slugging)

OPS = On Base Percentage + Slugging Percentage

- A composite of ability to get on base, and ability to hit with power
- Generally accepted as being the single most effective metric for evaluation players

Very good player’s OPS = 0.950

Mediocre player’s OPS = 0.800

Bad player’s OPS = 0.650

We glossed over this in terms of the presentation, to focus on the Moneyball story, but if you wanted a good overall metric, this is the one.²

Back to the big picture?

- Lewis’ story. Even though GMs (and SABR guys and Bill James) knew that OBP and SLUG were both important, Beane’s claim was that OBP is undervalued.

¹ The approximation comes with the treatment of rare events such as being hit by pitch, etc.

² For those of you who really get into this stuff (SABR types), there are some slightly better metrics, but OPS is pretty darn good. VORP, RC27, etc.

Do these statistics actually predict team's winning percentage?

Dependent variable: team winning percentage

- Model 1: Regression with only OBP predicts well $R^2 = 0.825$
- Model 2: Regression with only SLUG slightly less well $R^2 = 0.786$
- Model 3: Regression with OPS (both) do better $R^2 = 0.871$
- Model 4: Regression with OBP and SLUG separately $R^2 = 0.882$

Table 1 – Productivity Estimates:
The Impact of On Base and Slugging Percentage on Winning

	Model			
	1	2	3	4
Constant	0.506 (0.087)	0.593 (0.066)	0.558 (0.073)	.500 (0.002)
OPS			1.261 (0.061)	
OPS Against			-1.337 (0.055)	
On Base	3.297 (0.189)			2.032 (0.180)
On Base Against	-3.315 (0.171)			-2.032 ^R
Slugging		1.753 (0.116)		0.900 (0.105)
Slugging Against		-1.971 (0.107)		-0.900 ^R
Number of Observations	150	150	150	150
R^2	.825	.786	.871	.882

Hypothesis Test of Model 4, H^0 : On Base = Slugging
 $F(1, 147) = 17.21$, p-value = 0.0001

- Model 4 also nice because it allows us to see the relative value of a “point” of each statistic³

$$\text{Winpct} = .500 + 2.032 \cdot \text{OBP} - 2.032 \cdot \text{OBPAgainst} + 0.900 \cdot \text{SLUG} - 0.900 \cdot \text{SLUGagainst}$$

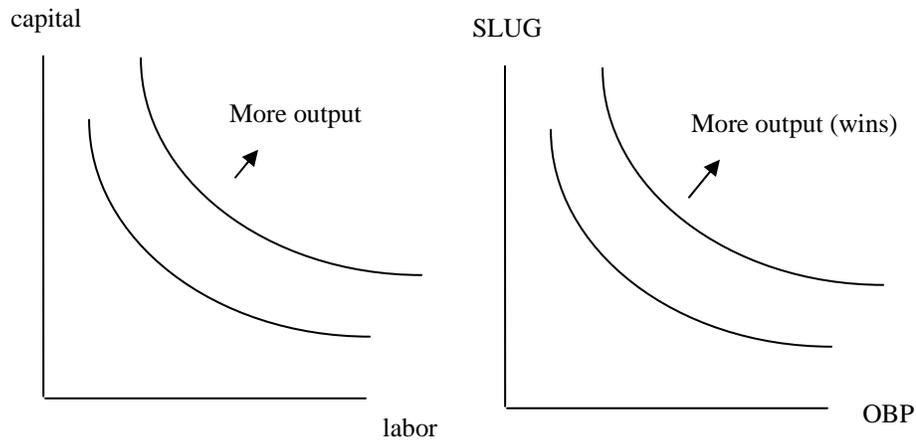
- Ceteris paribus, a one point (0.001) increase in OBP leads to a 2.032 point (0.00232) change in winpct
- Ceteris paribus, a one point (0.001) increase in SLUG leads to a 0.900 point (0.0009) change in winpct
- Roughly speaking, a point of OBP is worth twice as much as a point of SLUG.⁴

³ You might have noticed that the coefficients on OBP and OBP Against are the same. This is a constraint the authors imposed when running the regression. Talk to me if you are interested.

⁴ You may have noticed a hypothesis test at the bottom of the table. That tests the hypothesis that the coefficient on OBP is the same as the coefficient on SLUG. It is rejected, suggesting there is strong evidence that the true slope coefficients are not the same.

Aside – Isoquants / Equi-marginal principle

- Let's talk about digging a hole. Suppose there are three ways to dig a hole.
 - 1 worker, 1 steam shovel (small amount of labor, large amount of capital)
 - 10 workers, 10 shovels (medium amount of labor, medium amount of capital)
 - 100 workers, 100 spoons (large amount of labor, small amount of capital)
- Suppose all of above combinations can dig the hole we want. Which do you choose? What does it depend on?
- The answer is that it depends on the price of labor and capital! The decision that will be made depends on how productive each input is, but also their price.
- The point is there are various ways to combine inputs to get the same amount of output. We could plot these combinations and call it an isoquant – see below and to the left. Each spot on an isoquant involves the same amount of output. The curve connects the various ways this level could be produced.
- Of course, with more K and more L, more output can be produced.
- See also, the isoquants for baseball below and to the right.



-
- How will we determine whether we use OBP or SLUG to win baseball games?
 - It will depend on the prices of each.

Equi-marginal principle

- Suppose the **marginal product of labor** – the additional output associated with hiring an extra unit of labor was 3 units of output. Suppose the price of hiring an extra unit of labor is \$10. We could calculate:

$$\bullet \frac{MP_L}{P_L} = \frac{\frac{3 \text{ units of output}}{\text{unit of labor}}}{\$10} = \frac{0.3 \text{ units of output}}{\$}$$

- That tells us how much extra output we get *per dollar spent on labor*.
- Suppose the **marginal product of capital** – the additional output associated with hiring an extra unit of capital was 5 units of output. Suppose the price of hiring an extra unit of capital is \$4. We could calculate:

$$\bullet \frac{MP_K}{P_K} = \frac{\frac{5 \text{ units of output}}{\text{unit of capital}}}{\$4} = \frac{1.25 \text{ units of output}}{\$}$$

- That tells us how much extra output we get *per dollar spent on capital*.

What should the firm do?

- They should use more capital. More “bang” for their buck on money spent on capital than labor.

As firm hires more capital, MP_K falls, while MP_L rises.

- In fact, the firm should continue to hire capital until the firm finds that:

$$\frac{MP_L}{P_L} = \frac{MP_K}{P_K}$$

which is called equi-marginal principle.

What does this have to do with Billy Beane and the As?

- Believe it or not, we already know MP_{OBP} and MP_{SLUG} .
- But we don't know the price P_{OBP} and P_{SLUG} .
- No problem. Go run some salary regressions!

Side Issue / Log Salaries

- Smart folks don't like to run regressions of salaries, primarily due to outliers and some statistical stuff that isn't so interesting.⁵
- So instead run regression on log salaries.
- Example:

Salary = \$1,000,000	Natural Log Salary = 13.82
Salary = \$1,100,000	Natural Log Salary = 13.91

- 10% increase in salary increases natural log salary by just less than 0.10.
- | | |
|-----------------------|----------------------------|
| Salary = \$10,000,000 | Natural Log Salary = 16.12 |
| Salary = \$11,000,000 | Natural Log Salary = 16.21 |
- Again, 10% increase in salary increases natural log salary by just less than 0.10!
 - So roughly speaking, we can think of the change in natural log salaries as percentage increases in salaries.

⁵ The dependent variable is supposed to be normally distributed (bell shaped). Salary distributions are often right-skewed (a few large outliers on the right side of the distribution – high salaries). Transforming salaries into log salaries reduces these problems.

Table 2 – The Baseball Labor Market's Valuation of On Base and Slugging Percentage

	All Years	2000	2001	2002	2003	2004
On Base	1.360 (0.625)	1.334 (1.237)	-0.132 (1.230)	0.965 (1.489)	1.351 (1.596)	3.681 (1.598)
Slugging	2.392 (0.311)	2.754 (0.628)	3.102 (0.613)	2.080 (0.686)	2.047 (0.850)	2.175 (0.788)
Plate Appearances	0.003 (0.000)	0.003 (0.000)	0.003 (0.000)	0.003 (0.000)	0.003 (0.000)	0.003 (0.000)
Arbitration Eligible	1.255 (0.047)	1.293 (0.102)	1.106 (0.100)	1.323 (0.100)	1.249 (0.111)	1.323 (0.115)
Free Agency	1.683 (0.044)	1.764 (0.096)	1.684 (0.092)	1.729 (0.097)	1.663 (0.107)	1.575 (0.105)
Catcher Dummy	0.152 (0.056)	0.137 (0.124)	0.065 (0.116)	0.208 (0.122)	0.343 (0.134)	0.059 (0.133)
Infielder Dummy	-0.029 (0.040)	0.060 (0.087)	0.069 (0.083)	-0.087 (0.086)	-0.054 (0.095)	-0.100 (0.098)
Intercept	10.083 (0.170)	10.078 (0.360)	10.347 (0.321)	10.490 (0.358)	10.289 (0.387)	9.782 (0.414)
Number of observations	1736	353	357	344	342	340
R ²	0.675	0.676	0.728	0.695	0.655	0.635

2000 Salary regression

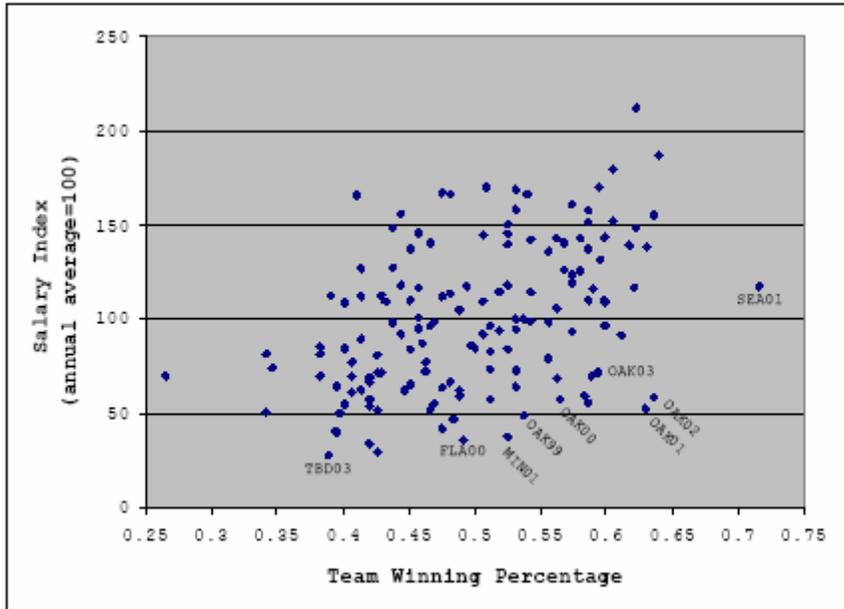
- Coefficient on OBP = 1.334
- A one point (0.001) increase in OBP leads to a 1.33% higher salary
- Coefficient on SLUG = 2.754
- A one point (0.001) increase in SLUG leads to a 2.75% higher salary.
- Roughly speaking, a unit of slugging percentage is twice as expensive as a unit of OBP. Flipped around, a point of OBP is half as expensive as a point of SLUG.

Now be Billy Beane and apply the equi-marginal principle

- We know from before that a point of OBP is worth about twice as much as a point of SLUG
 - $MP_{OBP} = 2.032$ (wins per point)
 - $MP_{SLUG} = 0.900$ (wins per point)
- We just found out that a point of OBP is a bit more than half as expensive as a point of SLUG.
 - $P_{OBP} = 1.33$ (log dollars per point)
 - $P_{SLUG} = 2.75$ (log dollars per point)
- OBP is twice as productive, but half as expensive. What do you do?
- Buy OBP! Which is exactly what Billy Beane did!
- We could do the math....

$$\frac{MP_{OBP}}{P_{OBP}} = \frac{2.032}{1.33} \approx 1.5 \quad \frac{MP_{SLUG}}{P_{SLUG}} = \frac{0.900}{2.75} \approx 0.33$$

- More bang for buck in OBP. Hire high OBP players! Beane / As should be able to produce more (win more games) at lower cost than other teams.
- Did it work? See picture below.⁶ Notice the high winning percentage low salary Oakland A's observations.

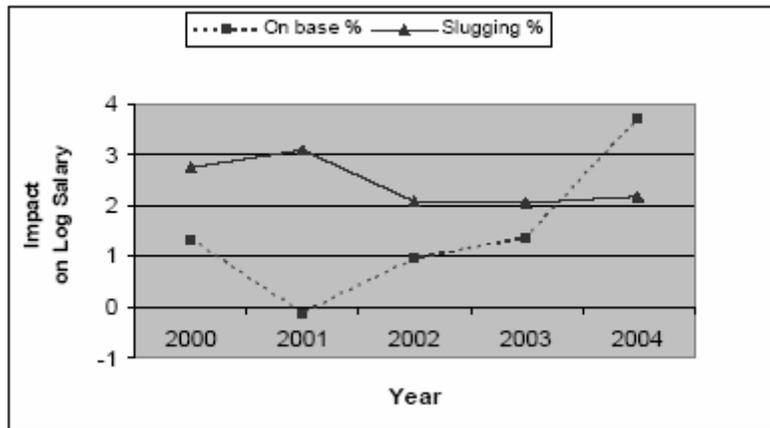


Next Step – what happens as people start to catch on?

- Suppose the productivity doesn't change. That is suppose we expect that OBP remains roughly 2 times as productive as SLUG.
- Then what must be true about the relative prices of OBP and SLUG after people adjust? They must end up so that OBP is twice as expensive as SLUG.
- Back to 2004 regression. Coefficient on OBP is 3.681 while coefficient on SLUG is 2.175. Ratio is 1.7. Fairly close, moving in the right direction, and have standard errors to worry about.⁷ Check out the picture on the next page as well, which graphs the coefficients on OBP and SLUG.

⁶ As pointed out in class, the main thrust of the Moneyball story, as it pertains to offense, was to focus on high OBP players. The A's also engaged in other strategies, such as avoiding drafting high school players, avoiding power pitchers, trading "overvalued" closers, and others. In that regard, what you see here is only part of the story (it is a whole book!), but an important part of the story.

⁷ The long-term nature of many contracts is also an issue. Some of the players in the 2004 salary regressions were playing under the terms of contracts that were signed in 2003, some in 2002, and so on. To the extent that only new contracts can be adjusted, the coefficients you see in the regressions underestimate the relative increase in P_{OBP} . That would make a nice project for someone to look at. If you are interested, come talk to me. I'd bet someone a six-pack if only players with new contracts are included, the ratio of P_{OBP} / P_{SLUG} is larger than MP_{OBP} / MP_{SLUG} (teams over-adjusted) in 2004.



Is Moneyball Dead?

- No. Some people like to suggest that because you can no longer win by gathering high OBP players, that somehow the Moneyball approach (which I view as simply using quantitative analysis to make business decisions) is dead. I disagree.

I would claim (I'd be willing to be corrected) that neither Beane nor Lewis claimed this strategy would last forever. As we saw above, others caught on, mimicked the strategy of the As, and in doing so drove up the price of OBP to the point where this was no longer an "inefficiency". The inefficiency was corrected. It was a good ride!

Again, I'd claim that Beane and Lewis say there was inefficiency and it was exploited. Sauer and Hakes would suggest that it is no longer there. In other work they do, they confirm this.

- Also mentioned in class is that the A's have changed their strategy. This too, makes sense, as the inefficiency has been corrected. Inefficiencies don't last for long in markets. As prices change, so too does the mix of inputs. Will Beane always find a big inefficiency? Likely not. He found one.
- Two GMs hired that used to be Beane assistants have been fired, but Theo Epstein is still working for the Red Sox. Also, the Red Sox hired Bill James to help them analyze quantitative stuff.

Paul DePodesta (Dodgers) - fired
 JP Riccardi (Blue Jays) – fired this weekend!
 Theo Epstein (Red Sox)

Why does Beane agree to discuss his strategy with Lewis?

- Either he was an ego-maniac (possible), he was foolish (doubtful), or he thought that everyone else had already picked up on what he was doing, so the efficiency was gone (most likely). I'll see if I can't dig up an interview with Beane where I believe he answered this question. Is it in the book?

Other curiosities

- Billy Beane does screw a few things up. I think he makes a pretty substantial mistake in his drafting strategy by selecting his "favorite" players in early rounds. Would a better strategy be to select the players that other teams liked in those early rounds, then trade them away, and draft the undervalued players in later rounds? I think so.

The Sherman Act (1890)

Section 1:

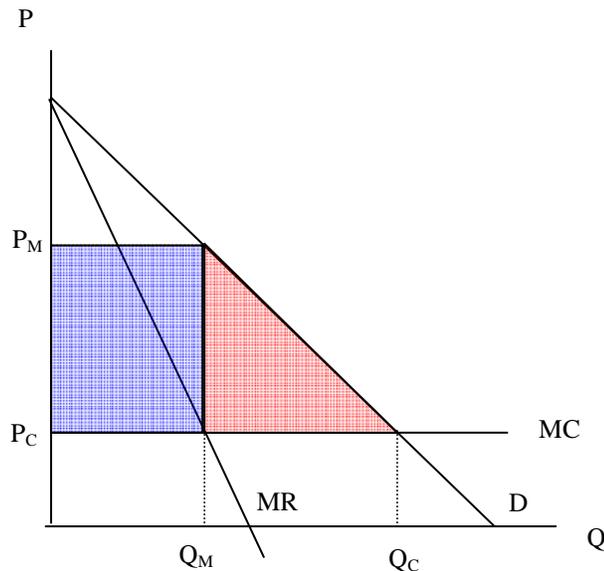
“Every contract, combination in the form of a trust or otherwise, or conspiracy, in restraint of trade or commerce among the several states, or with foreign nations is hereby declared to be illegal.”

- Target is firms: “combination or conspiracy in restraint of trade”. The term “trust” is synonymous with “cartel”. Target is firms that *ought* to be competing with one other that agree *not* to compete.
- Examples: price fixing, bid rigging at auctions, agreements not to compete, to act like a monopolist would

Section 2:

“Every person who shall monopolize or attempt to monopolize any part of the trade or conspire with any other person or persons to monopolize any part of the trade or commerce among the several states or with foreign nations, shall be deemed guilty of a misdemeanor...”

- Target is business practices designed to kill competition for sake of establishing or maintaining monopoly.
- Examples: predatory pricing, lock up of essential facilities, boycotts of potential competitors, merger for monopoly

Why is monopoly bad?

Let's compare and contrast the results of a competitive market with that of a monopolist market.

Competitive Market $P = MC$:	P_C and Q_C
Monopolistic Market $MR = MC$:	P_M and Q_M

As a result of switching from competitive to monopolistic market

- Price Rises
- Quantity falls

- Some wealth transferred from consumers to producers (CS to PS). Blue rectangle on the graph. Economists don't care about the transfer. So long as somebody gets the surplus.
- Some wealth destroyed (DWL). The units between Q_M and Q_C are worth more than the cost of production, but are not made available. Size of the dead weight loss is red triangle.

Thus, monopoly is inefficient and society as a whole is better in a competitive situation. Thus, the Sherman Antitrust Law

Monopsony

[Not now!]

Are leagues more like competition or monopolies?

Jointness of production (leagues) adds value through coordination (schedules and rules), but is also conducive to suppression of competition (entry restrictions).

Competition is generally between leagues. Occasional league "wars:"

- AL-NL (1901-2),
- AFL vs. NFL 1966 (full merger)
- NBA vs. ABA 1976 (Nuggets, Pacers, Nets, Spurs)
- NHL vs. WHA (Oilers, Whalers, Nordiques, Jets)
- NFL vs. USFL

Both of these ended in merger, enhancing monopoly. Legal?

National League (1876) set mold in US. Two basic tenets of National League were

- Monopoly Power- limited number of teams w/ restricted entry (exclusive territories)
- Monopsony Power – system bound players to teams

Contrast with soccer in Europe.

- Play your way in...relegation and promotion
- No Monopsony Power
- Many teams in large cities. For example, in London: Arsenal, Chelsea, Tottenham, West Ham., Charlton, Fulham, Milwall, Queens Park Rangers

How do leagues protect their monopoly power?

Block entry by strategic location

- League may expand to reduce entry threat.
- The NFL purportedly did this in the 1960s with Dallas and Minnesota to block AFL teams.
- In MLB, the New York Mets and Houston Astros were a response to a threat from the continental league.
- Possibility of entry increases league size above level that maximizes profits per team, which may erode the negative effect on society.

Lock up TV?

- With no TV revenue, no ability to compete for talent.

- ABC/NBC saved AFL in 1960s
- USFL (1980s) failed w/o network commitment
- Exclusive contracts an antitrust concern

Price Discrimination

Successful Price Discrimination requires information. Generally speaking must know differences in MV across consumers or differences in MV across different quantities and must be able to separate consumers. If high MV consumers can buy at lower P, this will thwart the price discrimination efforts.

First degree

Know WTP of all consumers for all Q

Second degree

Know demand curve slopes down (MV falls as Q increases)

Third degree

Know different groups behave differently

First Degree Price Discrimination

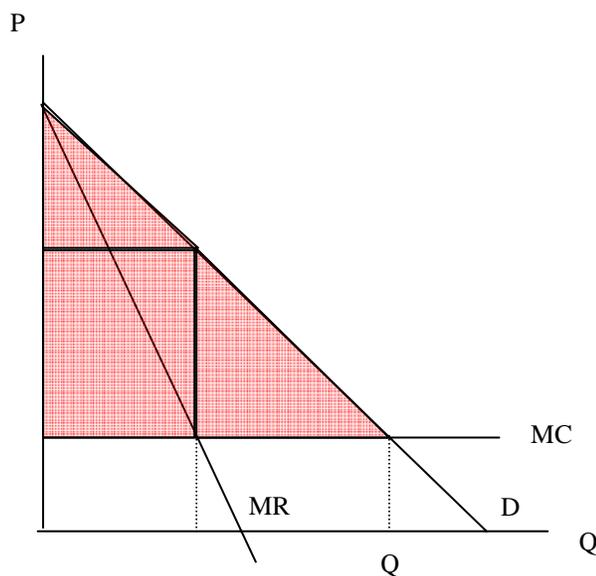
Suppose you knew what everyone is willing to pay.

In that case, you can, if you can prevent resale, charge everyone a different price. If so, the seller captures all the consumer surplus. Revenue is the entire red area. The price is the same as MV for each unit.

Perhaps surprisingly, this is efficient, as $P = MC$ for last unit, so no DWL (same Q as competitive situation).

Very hard to do in practice as requires tons of information.

Real world examples? Car Salesman? Financial Aid Packages? Tax Return Preparation Fees?

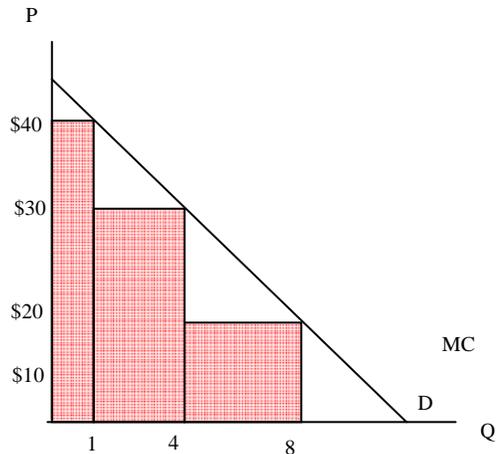


Second Degree Price Discrimination

Don't know MV for everyone, but do know that the demand curve slopes down. That is, consumers will be willing to pay less for each additional unit. Thus, the firm will want to charge less for additional tickets.

In doing so, second degree price discrimination captures more consumer surplus than simply charging one price for all tickets.

Real World Examples: Group sales, season tickets



In the example above, we could set up a ticket-price for an eight game season-ticket package. How much should the firm charge?

The demand curve tells us we should charge \$40 for the first game, \$30 for each of the next three games, and \$20 for each of the next 4 games. Thus, the season ticket package should cost:

$$\$40 * 1 + \$30 * 3 + \$20 * 4 = \$40 + \$90 + \$80 = \$210.$$

Third Degree Price Discrimination

Can separate groups with different demands (elasticities)

Real world examples: student discounts for tickets, airline ticket prices.

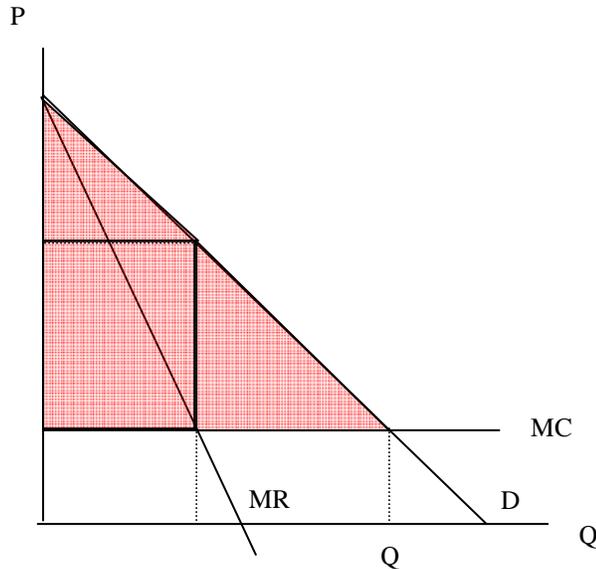
We've done this already.

Personal Seat Licenses

A bit of a puzzle to economists. Relatively new innovation in pro sports. First used by Carolina Panthers in NFL to finance stadium construction costs, but has been used for a long time in colleges.

Personal Seat License is a payment for the right to buy season tickets. So you'd think this would increase revenue. Can't just increase PSL – must either provide additional benefit or offset the increase in higher costs for PSL with lower ticket prices. Thus, PSL don't increase revenue flowing to team, but simply changes the way they collect money.

There is a tax angle – a loophole in the tax law that provides favorable tax treatment to revenue from PSLs.



Charge the competitive price ($P = MC$), but then extract all consumer surplus in the way of charging the PSL fee.

Also, would still be better than charging a simple monopoly price. Pretty similar story to 1st degree price discrimination, no?

Real World Examples: College football, Golf membership

What's good about Monopolies?

How are baseball teams different than TV stations? Does CNN like it if Fox News goes out of business? Does LA Dodgers like it the Milwaukee Brewers go out of business? Stability.

Big question?

Is a league a group of firms that compete? Or is it a multi-plant factory?

Antitrust in Sports

Most sports leagues have monopoly aspects, and thus antitrust litigation is common in US sports.

MLB: exemption from quirky Supreme Court decision, with minor modifications along the way

NFL: has limited legislative exemption for TV contracts

NCAA: beneficiary of monopoly tendencies in US pro sports, loser of numerous cases from 1984 on

Baseball's Antitrust Exemption

Only industry to enjoy *blanket* exemption. Even patents expire, author's rights disappear, and public utilities (while allowed to be a monopoly, are regulated).

Curt Flood Act of 1997 modified exemption. After strike of '94-95 strike

- MLB subject to antitrust when labor negotiations fail.
- MLB gained explicit exemptions as quid pro quo
- League control of team locations

- Joint marketing of product & broadcast rights

The Federal League Case

- AL had just made it after starting in 1901, so seemed possible
- Combination of AL and NL had reduced salaries
- Federal League was rival league to the NL and AL
- Played 1914 & 1915
- FL established in 8 cities to compete w/ AL & NL
- Sought players from existing teams
- AL & NL responded w/ boycotts to enforce reserve clause in player contracts
- 1914 filed Federal lawsuit (district)
- 1922 Federal League Baseball Club of Baltimore vs. National League
- FL sued under Sherman Act
- Claim: Boycott & reserve clause were joint actions by competitors designed to deny entry of rivals

District Court

- Kennesaw Mountain Landis presiding
- Federal League knew he was a “trustbuster” - once ruled against Standard Oil
- Didn’t know he was a *huge* baseball fan
- Landis sat on case for a year (wanted case to settle)
- FL folded before ruling issued
- 7 teams settle: 4 teams are bought out, two were allowed to buy existing teams (St. Louis Browns – now Baltimore Orioles and Chicago Cubs), one team folded bankrupt. Baltimore rejected offer.
- FL club from Baltimore continued & won at trial
- Awarded damages of \$80,000, trebled to \$240,000.

Supreme Court

- NL appealed & won; FL Club of Baltimore appealed that to S. Court
- Unanimous Supreme Court Decision (Justice Holmes)
- Baseball “a public exhibition, *not commerce*”
- Hence Sherman did not apply
- Bizarre opinion.... much speculation
- Subsequent decisions in MLB tortured by logic in Federal League case
- Court did not use as precedent for other leagues

Impact of FL Case on MLB

- Illogic of the case led to uncertainty
- Periodic Congressional Review
- Politicians get publicity from sports hearings
- Cellar Hearings, 1951
- Kefauver Hearings, 1961
- Sisk Hearings, 1976, etc., etc, most recently Steroids...

Toolson vs. NY Yankees (1953)

- Toolson played for Yankees
- Sent down to minor leagues, refused assignment, blacklisted
- Claimed reserve clause (can only play for Yankees) violated Sherman
- In interim congressional hearing (Cellar Commission) stalled by likelihood Supreme Court would decide
- Supreme Court split decision upheld Federal League exemption
- Majority noted congress' review & inaction & stuck with Federal League precedent:

- "We think that if there are evils in this field which now warrant application to it of the antitrust laws it should be by legislation."
- Dissenters agreed that Federal League should be overturned

Flood v. Kuhn (1971)

- Flood traded from Cards to Phillies
- Bowie Kuhn is commissioner of baseball
- Reserve clause applied (had to play for Phillies and Phillies only)
- Refused to play for Phillies, sued claiming reserve clause violated Sherman
- Why opposed to Philadelphia – racism? Flood quote: "I do not feel I am piece of property to be bough and sold irrespective of my wishes."
- District court and appeals court found for Kuhn
- Supreme Court upheld Federal League precedent, but stated...
- "Baseball is a business and it is engaged interstate commerce...."
- and stated its "exemption from antitrust is an exception and an anomaly"
- But don't overturn exemption. "If there is any inconsistency or illogic in this, it... is to be remedied by the Congress"

More on Flood v. Kuhn / Curt Flood Act of 1998

- Judge Marshall collective bargaining could change nature & legal status of reserve system (it did)
- Curt Flood Act of 1998 – didn't change much. Players can file suit so solve labor disputes, but only if players vote to decertify players union.
- League vs. League union a fair fight.

Messersmith / McNally

- Messersmith and McNally were players unhappy with contract offers for 1975.
- They refused to sign and were renewed by their teams (paid the same amount as in 1974).
- The head of the MLB player's union, Marvin Miller argued they were no longer under contract to their teams after the 1975 season and should be declared free agents.
- The union filed grievances on behalf of Messersmith and McNally. An independent arbitrator ruled for the union Dec. 23, 1975.
- No more reserve clause.
- Curt Flood has an impact here.
- Future negotiations in collective bargaining agreements

Spencer Haywood v. NBA (1971)

- NBA by-laws Section 2.05: A person who has not completed high school or who has completed high school but has not entered college, shall not be eligible ... until four years after he has been graduated Similarly, a person who has entered college... shall not be eligible [for four years]
- Spencer Haywood led US team to 1968 Olympic gold medal
- Left U. Detroit after sophomore year to ABA
- Signed w/ Seattle following year (senior)
- NBA office nullified contract à anti-trust suit
- Allegation: Ban on underclassmen was an illegal "group boycott" in restraint of trade
- Decision: "The harm resulting from a "primary" boycott such as this is threefold. First, the victim of the boycott is injured by being excluded from the market he seeks to enter. Second, competition in the market in which the victim attempts to sell his services is injured. Third, by pooling their economic power, the individual members of the NBA have, in effect, established their own private government."
- Result: High school graduates became eligible for the NBA draft
- New Development: can't enter until you're 19

Antitrust & the NFL

- Radovich vs. NFL (1957)
- Radovich was an all-pro lineman for Lions
- NFL blacklisted him after signed w/ AAFC (rival league)
- Lower courts applied FL decision
- Supreme Court ruled against NFL in 6-3 decision
- Antitrust exemption for baseball, but not for football
- Dissenting justices applied *stare decisis*

The "Rozelle Rule"

- Post-Radovich, players could sign *in principle* with others after contracts expired, but very few did so. Why?
- Teams had gentlemen's' agreement not to sign "free agents;" violated in 1963
- NFL response was the Rozelle Rule: compensation must accompany signing of free agent; if none agreed, terms imposed by commissioner Rozelle. Made it effectively impossible for players to switch teams.
- NFL Players association sues in 1972.
- Court ruled Rozelle Rule violated Sherman: "the Rozelle rule significantly deters clubs from negotiating with players [who] are denied the right to sell their services in a free and open market"
- Led to negotiations between NFL & NFLPA
- Cash settlement for past harm from restrictions
- Terms of free agency negotiated by NFLPA
- Result: Compensation scheme similar to Rozelle rule
- Maurice Clarett sues NFL in 2003 & loses

Antitrust & Labor Rules: Overview

- Inconsistent application across sports
- Legacy of Fed League Case: Blacklists, reserve clause etc ok in MLB
- Ultimately resolved by Miller & MLBPA (Messersmith & McNally)
- MLB rules tried & failed to pass muster elsewhere
- Ultimately all restrictions passed into CBAs

Antitrust & Media Rights

- Old way: teams negotiate own deals
- New way: leagues negotiate deal & split revenues
- Why switch?
- NY Giants games compete w/ NY Jets
- League control reduces effects of competition among teams for TV deals (FSU / Miami Game theory)

Are jointly negotiated media contracts legal?

- No, if independents jointly act jointly cut Q & increase price
- Yes, if league is viewed as a single entity
- DOJ issued 1953 injunction barring NFL from joint negotiating of TV contracts
- Rozelle lobbied Congress for relief
- Sports Broadcasting Act of 1962
- exempted professional league media contracts
- quid-pro-quo: NFL would not play on Saturdays
- Huge increase in NFL TV revenue

Joint Contracts & NCAA

- NCAA was contracting agent w/ networks
- Scheduled football telecasts, & distributed revenue to schools
- 2 games each Saturday
- Max of 3 appearances every 2 years per school
- Each of 12 conferences to get TV appearance
- Problem: some schools not satisfied with exposure & their share of \$
- Formed CFA (football schools not in Big 10)
- Set out to negotiate own TV contract
- NCAA threatened expulsion from NCAA
- CFA drew back, UGA & OU took fight to court

NCAA vs. Board of Regents of the Univ. of Oklahoma (1984)

- OU's case: NCAA suppressed competition in the CFB TV market
- TV exposure desired by schools, cut off by NCAA
- Threat of expulsion denied essential facility to compete
- NCAA
- Purpose was to protect live gate & promote competitive balance*
- Pro-competitive joint venture w/ no monopolistic intent
- Court & Appeals Court: OU is right on all counts
- Contract was an illegal restraint of trade that limited viewing of games on TV;
- No evidence of pro-competitive effects
- Consequences
- TV now under conference control [why is this ok?]
- Q increased & price fell; revenues off by 1/3!
- Big 8 conference revenues increased significantly; OU's share fell

NCAA & Media Contracts

- Contrast current NCAA w/ NFL policy:
- NCAA: conferences & independents negotiate own contracts w/ TV
- CFB on 12am-12pm; multiple games & channels compete
- NFL: games in specific slots - little or no choice of game in each time slot. Max exposure w/ minimizing competition between games
- How can NCAA "get away" with being the contracting entity for its basketball tournament?

What these notes hope to do is to do a quick review of supply, demand, and equilibrium, with an emphasis on a more quantifiable approach. These, in some cases, are more mathy than we need.

Demand Curve (Big Picture)

The whole point of a demand curve is to find the relationship between the price of a good and the quantity that consumers wish to purchase (the quantity demanded).

Why does it matter? First off, a solid understanding of supply and demand is generally necessary to have an economic understanding of the world around you. Second, and as important, if firms have knowledge of the demand curve facing their firm, they will have the ability to make more informed pricing and output decisions. These decisions that will directly affect the firm's profitability. This will be particularly true of firms with pricing power (monopoly power), and sports franchises will have pricing power.

I am going to assume that you have some understanding of the 1st Law of Demand. The 1st Law of Demand states, "ceteris paribus (holding other relevant factors constant), as the price of a good falls, the quantity demand of that good increases." Basically, it states that demand curves are downward sloping.

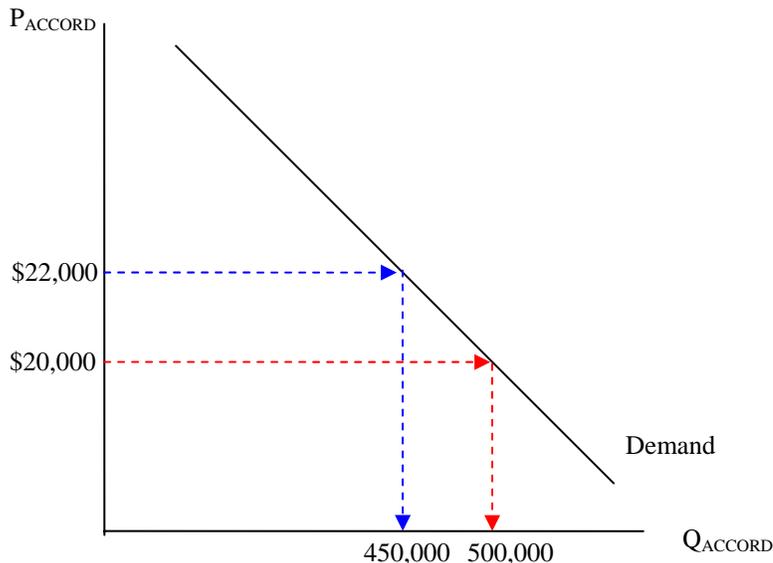
Interpretations of the Demand Curve

There are two different ways you can interpret the information given in a demand curve.

Horizontal Interpretation of Demand Curve - this is the interpretation of the demand curve you are most likely familiar with. The idea here is to pick a price, then move (horizontally) over to the demand curve.

For example, if the price of a Honda Accord is \$22,000, the quantity demanded of Honda Accords might be 450,000. That is, if the price is \$22,000, consumers will wish to purchase 450,000 cars. Likewise, if the price of a Honda Accord is \$20,000, the quantity demanded might be 500,000.

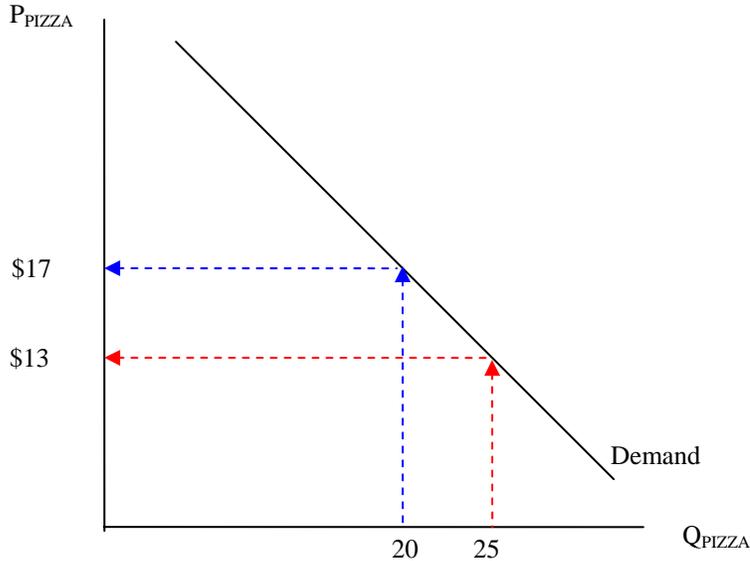
See the graph below. The reason we call this the horizontal interpretation should be apparent.



Vertical Interpretation of Demand Curve – this will be less familiar, and will come in handy when discussing consumer surplus, and later when we discuss price discrimination and other pricing strategies. The idea here is to pick a quantity, then move (vertically) up to the demand curve. We sometimes call the result the "height of the demand curve". See the picture below.

For instance, if the height of the demand curve facing a specialized pizza store at $Q = 20$ is \$17, this means that the most a consumer will be willing to pay for the 20th pizza is \$17.¹

Likewise, if the height of the demand curve at $Q = 25$ is \$13, this means the most a consumer will be willing to pay for the 25th pizza is \$13.²



As eluded to above and in the previous footnote, price discrimination is a strategy where firms will try to charge different customers different prices. If the firm can come up with a way to charge the first consumer \$17 and the other consumer \$13, it will earn more revenue than if it charges a price of \$17 (only the 1st consumer will purchase the pizza) or a price of \$13 (both consumers purchase the good).

Of course, we could use either interpretation on a demand curve, depending in which type of problem we are interested. We could have asked what is the most a consumer will be willing to pay for the 450,000th or the 500,000th Honda Accord.³ Likewise, we could ask how many pizzas that consumers will wish to purchase at a price of \$13 or \$17.⁴ Which interpretation we choose does not change the underlying demand curve. As you progress, you will find it makes more sense to use one interpretation or the other depending on the problem we are dealing with.

¹ This consumer would be willing to purchase the 20th pizza for any price less than \$17, but will not pay any price higher than \$17. We sometimes call the height of the demand curve the marginal benefit, marginal value, or marginal willingness to pay.

² If you recall the difference between an individual consumer's demand curve and the market demand curve, we are discussing the market demand curve in this case. We know that individual's demand curves are downward sloping. Individually, you would be willing to pay more for the 20th pizza than you would for the 25th pizza. Are you still hungry? However, when we look at the market (overall) demand curve for a product, we have combined every person's individual demand curve. The consumer who was willing to pay \$17 for the 20th and the consumer who was willing to pay \$13 for the 25th pizza are likely to be different people, a point we shall revisit when we discuss price discrimination.

³ Some consumer would be willing to pay \$22,000 for the 450,000th Honda Accord and some other consumer would be willing to pay \$20,000 for the 500,000th Honda Accord. In this case, I am very sure these are different consumers. Do you know anyone who owns 50,000 Honda Accords?

⁴ Hopefully not surprisingly, the answers here are 25 and 20, respectively.

Ceteris Paribus Conditions for Demand Curves

Thus far, we have glossed over the 1st Law of Demand and the ceteris paribus conditions for demand. Some more very deep background that you should have learned before...

If, we were interested in the important factors determining the number of cars sold, surely the price of cars would be important. This is exactly what we hope to capture with a demand curve. A demand curve illustrates how the quantity of cars consumers wish to purchase changes as the price of cars changes, holding other relevant factors constant.⁵

On the other hand, many other things (aside from the price of cars) affect the number of cars people wish to purchase. We call these other things “demand shifters” or “ceteris paribus conditions for demand”.

In order to draw a demand curve (to isolate the relationship between the own price of a car and the quantity demanded of cars), we must hold the ceteris paribus conditions constant.

On the other hand, when one of these ceteris paribus conditions changes, we must shift the demand curve in the appropriate direction.

What are these “demand shifters” or ceteris paribus conditions for the demand for cars?

Just to name a few: Price of gasoline / insurance / tires, price of trucks / bicycles / public transportation, incomes of consumers, quality and characteristics of cars, expectations about future car prices, ...

More in the next section...

Mathematical Expressions of Demand Curves

We can learn something from mathematicians. Here is what they would write:

$$Q_x^d = f(P_x, P_y, M, H)$$

What does it mean? It means the quantity demanded of good X (Q_x^d) is a function of, or depends on, the price of good X (P_x) the price of related goods (P_y), the income level of consumers (M) and other stuff (H).

In fact, we can get even more specific about the form of the relationship. While it is not necessarily the case, we often model demand curves as being linear.⁶ If so, we can write out a very simple numerical expression of a demand curve. For example:

$$Q_x^d = 100 - \frac{1}{2}P_x - 2P_y + 6M$$

We will call this a **demand function**. The demand function contains a whole slew of information.

⁵ If you recall from your previous economics classes, in this case, we would call the price of the car the “own price”. The own price is the price of the good for which we are drawing the demand curve.

⁶ A linear demand curve is one with a constant slope of the demand curve. Mathematically, this means the power (exponent) on the P_x term is (implicitly) 1. We will focus on linear demand curves in this class. As an example of a non-linear demand curve, consider $Q_x^d = 100 - \frac{1}{2}P_x^2 - 2P_y + 6M$.

First off, suppose you were told to draw the demand curve for good X. You would pull out a piece of graph paper, label the vertical axis P_x , the horizontal axis Q_x , and then you would be stuck.⁷ In fact, you cannot draw the demand curve without knowing the values of P_y and M , which are the ceteris paribus conditions for demand.

Suppose you were told that $M = 10$ and $P_y = 30$. In that case, you could simplify the demand function to the point where you could draw it.⁸

$$\begin{aligned} Q_x^d &= 100 - \frac{1}{2}P_x - 2P_y + 6M \\ Q_x^d &= 100 - \frac{1}{2}P_x - 2(30) + 6(10) \\ Q_x^d &= 100 - \frac{1}{2}P_x - 60 + 60 \\ Q_x^d &= 100 - \frac{1}{2}P_x \end{aligned}$$

Now, your demand function is expressed with only Q_x^d and P_x . You can graph it, which we will do in a bit. Before doing this, though, let us take an aside on the inverse demand function.

Aside Inverse Demand Function

It turns out it will save us some trouble to find what is called the **inverse demand function**. To find the inverse demand function, simply start with the demand function, and solve it for P_x . In the case where $M = 10$ and $P_y = 30$, we start with:

$$Q_x^d = 100 - \frac{1}{2}P_x$$

Now we just do the algebra. You can take whatever steps get you to the end. I will begin by adding $\frac{1}{2}P_x$ to both sides:

$$Q_x^d + \frac{1}{2}P_x = 100$$

Then subtract Q_x^d from both sides:

$$\frac{1}{2}P_x = 100 - Q_x^d$$

And multiply both sides by 2:

$$P_x = 200 - 2Q_x^d$$

⁷ Why not label it Q_x^d ? You could. Eventually, we will have the quantity of good X supplied, as well, so often we will just be lazy and label it Q_x .

⁸ It is just a coincidence in this case that the last two terms in the demand function cancel out. This will not always be the case. See also below the examples for when M and P_y change.

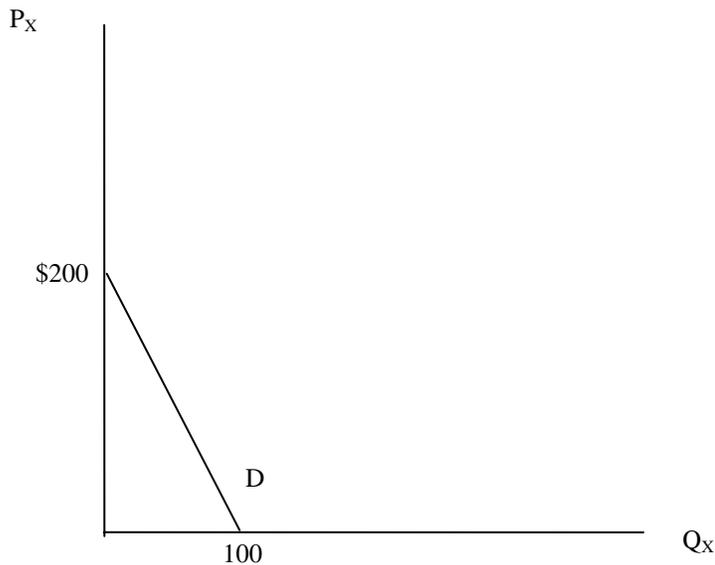
Why have I done this? First off, you will see in a second that it is now very easy to find a second point on our demand curve. In addition, this expression will come in handy when we find the marginal revenue curve for the firm.

Graphing the Demand Function

Recall from our demand function that $Q_x^d = 100 - \frac{1}{2} P_x$. By plugging in $P_x = 0$ in the expression above, we find that $Q_x^d = 100$. This is one point on our demand curve. In fact, this is the horizontal intercept of the demand curve.

We also solved for the inverse demand function, finding that $P_x = 200 - 2Q_x^d$. By plugging in $Q_x^d = 0$, we find that $P_x = 200$. It turns out this is the vertical intercept of the demand curve.

Because we have a linear demand function, to draw the line, we need only find two points on the line, which we have just done. Put these two points on a graph, connect the dots, and we are finished. See below.



Shifting the Demand Curve

What if M increases from \$10 to \$20? What happens to the demand curve? From a mathematical perspective, we can just plug and chug. But we would also like to go back to the intuitions as well. First the math...

$$Q_x^d = 100 - \frac{1}{2}P_x - 2P_y + 6M$$

$$Q_x^d = 100 - \frac{1}{2}P_x - 2(30) + 6(20)$$

$$Q_x^d = 100 - \frac{1}{2}P_x + 60$$

$$Q_x^d = 160 - \frac{1}{2}P_x \quad (\text{compare this to } Q_x^d = 100 - \frac{1}{2}P_x)$$

When we change the demand function, we will also get a new inverse demand function. Taking the same steps as before (but leaving out the explanation of these steps), we get:

$$Q_x^d = 160 - \frac{1}{2}P_x$$

$$Q_x^d + \frac{1}{2}P_x = 160$$

$$\frac{1}{2}P_x = 160 - Q_x^d$$

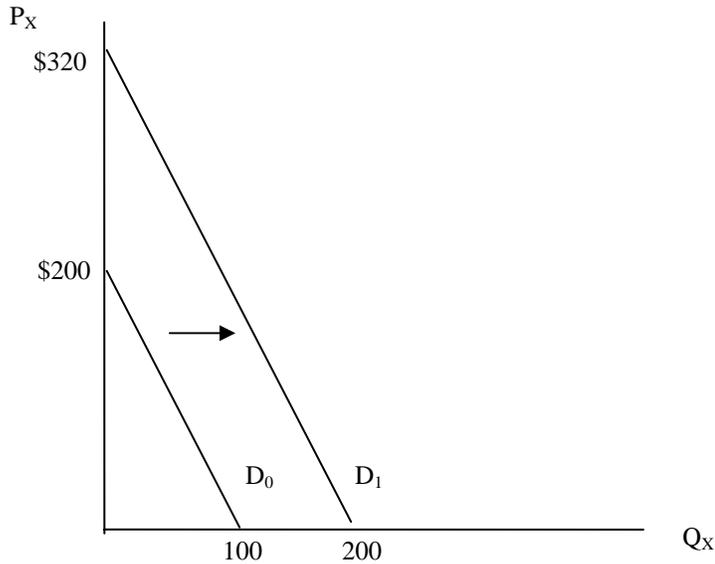
$$P_x = 320 - 2Q_x^d$$

Now, we draw the new demand curve. From the demand function, I see that setting $P_x = 0$ in the expression above results in $Q_x^d = 160$. From the inverse demand function, I see that setting $Q_x^d = 0$ results in $P_x = 320$. Now we stick these new points on our graph paper.

Below is a picture with the original demand curve ($M = \$10$) labeled D_0 and the new demand curve ($M = 20$) labeled D_1 . We see immediately what has happened is that the demand curve has shifted to the right because of the increase in income. We call a rightward shift of the demand curve an “increase in the demand curve”.⁹ Likewise, we call a leftward shift of the demand curve a “decrease in the demand curve”.

Again, any change in a ceteris paribus conditions shifts the demand curve. That is, a change in anything but the own price, causes a shift in the demand curve.

⁹ Why is this called an increase in the demand curve? From the horizontal interpretation of the demand curve, notice that at any (and every) price, there is a larger quantity demanded on the new demand curve than there is on the old demand curve.



Some exercises, you ask?

Start with original demand curve and $M = \$10$ and $P_y = \$30$. What would happen to the demand curve if M fell to $\$5$?¹⁰

Start over at the original values. What would happen if the price of good Y fell to $\$20$?¹¹

Start over at the original values. What would happen if the price of good Y rose to $\$40$?¹²

More on Ceteris Paribus Conditions

In our original demand shift, an increase in M from $\$10$ to $\$20$ results in an increase in the demand for the good. That is, an increase in income has led to an increase in income. In fact, this tells us that good X is a normal good.

But we also want to hone your intuition. We can make predictions about what happens to the demand curve without knowing anything about the actual mathematics of the demand curve.

Now is the time for a review of our ceteris paribus conditions. Our focus will be on how changes in our ceteris paribus conditions shift the demand curve.

¹⁰ The demand curve decreases (shifts left). The new vertical intercept would be $\$140$ and the new horizontal intercept would be 70.

¹¹ The demand curve increases (shifts right). The new vertical intercept would be $\$240$ and the new horizontal intercept would be 120.

¹² The demand curve decreases (shifts left). The new vertical intercept would be $\$160$ and the new horizontal intercept would be 80.

Generally speaking, the ceteris paribus conditions can be classified into a few major groups”

1. Price of Related Goods – Substitutes and Complements.
2. Income of Consumers – Normal or Inferior Goods
3. Expectations of Future Prices
4. Other Stuff

Prices of Related Goods

Complements are things that are used together. The classic example is peanut butter and jelly. In the car example, cars and gas and cars and insurance would be complements. Beers and baseball tickets are complements. Parking and football games are complements.

Substitutes are alternatives. The classic example is butter and margarine. In the car example, car and trucks and are substitutes and cars and public transportation would also be substitutes. Baseball tickets and opera tickets are substitutes. Soccer games and fishing trips are substitutes.

Substitutes

You tend to know them when you see them, but the definitions are as follows:

Goods A and B are called **substitutes**, if, when the price of good A changes, the quantity demanded (demand curve) for good B changes in the **same** direction.

Say we examine Coke and Pepsi, which are substitutes. If the price of Coke increases, the demand for Pepsi will increase. Why? Because people will substitute from drinking Coke (whose price has increased) to drinking Pepsi.

When the price of Coke rises, at each and every price of Pepsi (the horizontal interpretation of a demand curve), there will be a larger quantity demanded of Pepsi on the new demand curve than the original demand curve. The change in the price of coke has caused the demand curve for Pepsi to shift to the right (an increase in demand).

In the car example, if the price of trucks rise, the demand for cars will increase (shift right).

If the price of public transportation falls, the demand for cars will decrease (shift left).

Complements

Goods A and B are called **complements**, if, when the price of good A changes, the quantity demanded (demand curve) for good B changes in the **opposite** direction.

Consider ink cartridges and printers. If the price of ink cartridges increases, the demand for printers will shift to the left, or decrease. Why? People will realize the overall cost of printing will have increased, and thus will cut back on their printer purchases

In the car example, an increase in the price of gas will decrease the demand for cars.

A decrease in the price of car insurance would increase the demand for cars.

The Numerical Example

Recall our original expression of the demand function:

$$Q_x^d = 100 - \frac{1}{2}P_x - 2P_y + 6M$$

Believe it or not, that expression tells us if the goods X and Y are complements. How can you tell?

For every \$1 increase in the price of good Y, the quantity demanded of good X falls by 2 units.

This means the price of good Y and the quantity demanded of good X are changing in the opposite direction. From this, we can conclude the goods are complements. In fact, it is the sign (not the magnitude) on the P_y term that tells us this. In this case, we have $-2P_y$ in the expression (the sign is negative), indicating complements.

If for example, the demand function were instead

$$Q_x^d = 100 - \frac{1}{2}P_x + 3P_y + 6M$$

then X and Y would be substitutes.¹³

Not convinced? Go back to the example where the P_y decreased to \$20. What happened to the demand curve for cars? What happened to the demand curve when P_y increased to \$40?

Income

Normal Goods

Good A is called a **normal** good, if, when the incomes of consumers changes, the quantity demanded (demand curve) for good A changes in the **same** direction.

Examples of normal goods include Saints tickets, steak dinners, SUVs, almost everything else.

An increase in the incomes of Saints fans will increase the demand for Saints Tickets (shift right). A decrease in the incomes of SUV consumers will decrease the demand for SUVs (shift left).

Inferior Goods

Good A is called an **inferior** good, if, when the incomes of consumers changes, the quantity demanded (demand curve) for good A changes in the **opposite** direction.

Examples of inferior Goods include Ramen Noodles, SPAM, Mad Dog 20/20, used underwear.

¹³ If you are wondering about the meaning of the magnitude of the coefficient on the P_y term, that is a good thing to wonder. The magnitude of the coefficient indicates how closely related the goods are. A large coefficient (in absolute value or further from zero) indicates the goods are closely related. If we considered the demand for Pepsi, substitutes for Pepsi that might be included are the price of Coke and the price of orange juice. But the demand for Pepsi will be more sensitive to the price of Coke than the price of orange juice. Thus, we would expect a larger coefficient on P_{COKE} than P_{OJ} , even though both would be positive. More when we get to elasticities.

An increase in the incomes of consumers will decrease the demand for Ramen Noodles (shift left). A decrease in the incomes of consumers will increase the demand for SPAM (shift right).

The Numerical Example

The logic is the same as above. A positive coefficient on the M (in this case we have $6M$) in the demand function indicates that a one unit increase in income leads to a 6 unit increase in the quantity demanded of good X. This is consistent with a normal good.

A negative coefficient on the M term in the demand functions indicates inferior goods, as M and the demand curve for good X change in the opposite direction.¹⁴

Not convinced? See the example above about M increasing from \$10 to \$20 or the exercise decreasing from \$10 to \$5.

Expectations of Future Prices

Quite simple. When people expect prices to fall in the future, the (current) demand falls. People delay their purchases.

When people expect prices to rise in the future, the (current) demand rises. People try to act ahead of the price increases.

This becomes interesting for firms that regularly schedule sales. On the one hand, the sale itself would tend to increase purchases (while the sale occurs) according to the 1st Law of Demand. On the other hand, if people expect the sale, they may curtail their purchases in the period leading up to the sale. Automakers with their model year-end closeout? “Last-year’s” fashions? The hardback version of a book before the paperback comes out? Microsoft Vista? Dollar movie theaters? Personal Seat Licenses of up and coming college football teams?

Everything Else

I do not mind writing notes, but I do not want to write forever! Many other things can shift a demand curve. When it comes to sports franchises, both the absolute quality of the team and the relative quality of the team (parity, uncertainty of outcome) will come in to play. More later on these issues.

Consumer Surplus

The height of the demand curve tells us the maximum amount a consumer is willing to pay for a unit of the good. But very seldom does this consumer actually pay this amount. If there is a gap between the two, we call this gap consumer surplus.

$$\text{Consumer Surplus} = \text{Amount consumer is willing to pay} - \text{the amount they had to pay}$$

¹⁴ Again, the magnitude determines how sensitive consumption of that good is to changes in income. Consider Kraft Mac & Cheese and restaurant meals. I believe both Mac & Cheese and restaurant meals are normal goods, and thus both will have positive coefficients on M in their respective demand functions. I would also expect the coefficient on M in the demand function for Mac & Cheese to be close to zero, while the coefficient on M in the demand function for restaurant meals to be larger. When people win the lottery, they do not a bunch more Mac & Cheese than they used to, but people likely do buy a bunch more restaurant meals than they used to. More when we get to elasticities.

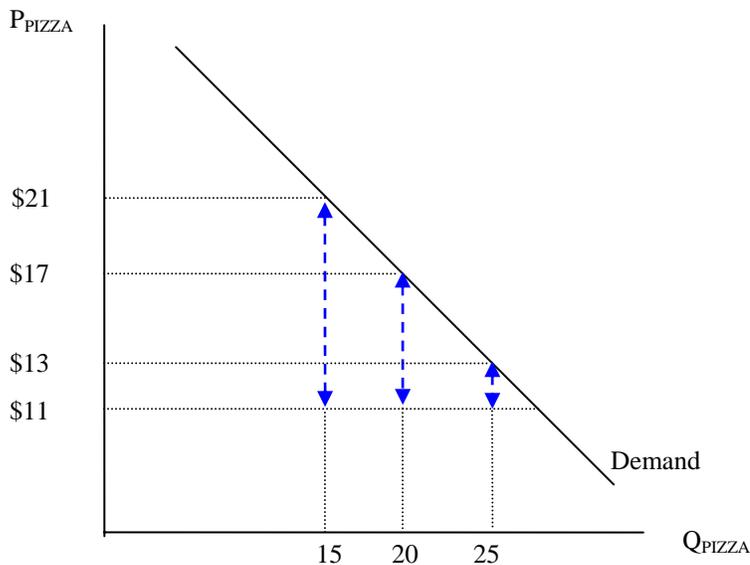
Can we find this graphically? We can. From the vertical interpretation of a demand curve, we know that the height of a demand curve at some quantity tells us the maximum amount a consumer was willing to pay. If we compare this to the price on the graph, we have consumer surplus.

Let's go back to the pizza example, only we will add in some additional information and assume the price of pizza is \$11. We can calculate the consumer surplus enjoyed for each individual unit of pizza consumed.

For example, at $Q = 15$, the height of the demand curve is \$21, while the price is \$11, so consumer surplus is \$10 ($\$21 - \$11 = \10). The idea is this consumer is getting a "deal". They would have paid \$10 more than they had to. The larger the consumer surplus, the larger is the welfare (happiness) of consumers.

At $Q = 20$, the demand curve tells us the consumer is willing to pay \$17, while the price is only \$11, leaving consumer surplus of \$6.

It should be easy for you to calculate consumer surplus on the 25th pizza.¹⁵

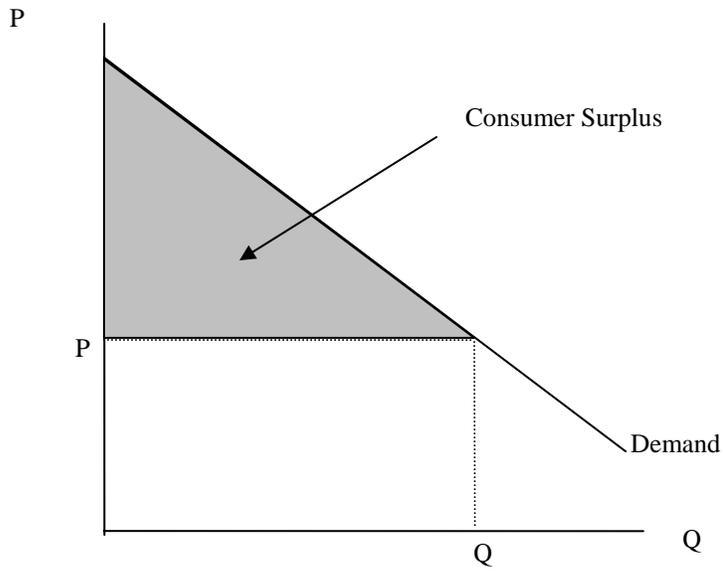


Adding consumer surplus up for each individual unit of the goods gets very tedious. Fortunately there is a calculus trick.

In general, if you want to add up the total consumer surplus for consuming Q units (where Q can be any quantity), simply add up the entire area that is (1) under the demand curve, (2) above the price that consumers pay, and (3) out to the quantity Q . See the picture below.

This will come in very handy when we talk about price discrimination. As a firm, wouldn't you want to try to charge everyone the maximum they are willing to pay?

¹⁵ Consumer surplus is \$2, as the consumer is willing to pay (the height of the demand curve) \$13, while the price is \$11. Consumer surplus is $\$13 - \$11 = \$2$.



Supply Curve (Big Picture)

The whole point of a supply curve is to find the relationship between the price of a good and the quantity that firms wish to produce (the quantity supplied).

I am going to assume that you have some understanding of the 1st Law of Supply. The 1st Law of Supply states, “ceteris paribus (holding other relevant factors constant), as the price of a good rises, the quantity supplied of that good increases.” Basically, it states that supply curves are upward sloping.

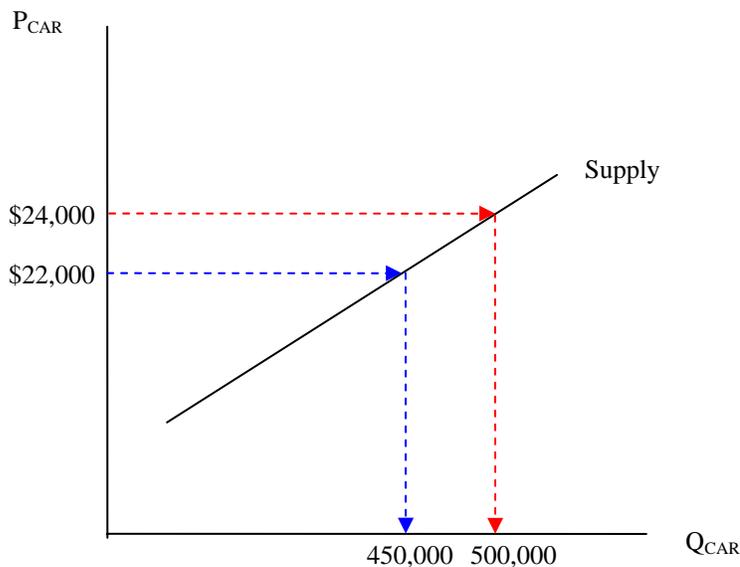
Interpretations of the Supply Curve

There are two different ways you can interpret the information given in a supply curve.

Horizontal Interpretation of Supply Curve - this is the interpretation of the supply curve you are most likely familiar with. The idea here is to pick a price, then move (horizontally) over to the supply curve.

For example, if the price of a car is \$22,000, the quantity supplied of cars might be 450,000. That is, if the price is \$22,000, firms will wish to produce 450,000 cars. Likewise, if the price of a car is \$24,000, the quantity supplied might be 500,000.

See the graph below. The reason we call this the horizontal interpretation should be apparent.

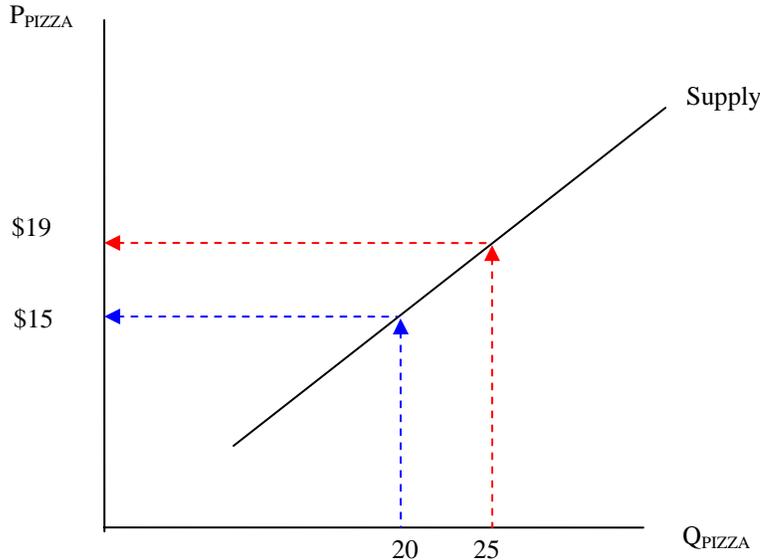


Vertical Interpretation of Supply Curve – again, less familiar, and will be discussed in greater detail when it comes to cost curves. Just as the demand curve illustrated the highest price a consumer would be willing to pay, a supply curve illustrates the *lowest* price a firm would be willing to accept to produce the good. The idea here is still to pick a quantity then move (vertically) up to the supply curve. We sometimes call the result the “height of the supply curve”. See the picture below.

For instance, if the height of the supply curve facing a specialized pizza store at $Q = 20$ is \$15, this means that the least a firm will be willing to accept to produce the 20th pizza is \$15.¹

¹ This producer would be willing to produce the 20th pizza for any price above \$15, but will not produce for any price less \$15. The height of the supply curve is called the marginal cost of producing, a term we will return to in our discussion of cost curves.

Likewise, if the height of the supply curve at $Q = 25$ is \$19, this means the least a firm would be willing to accept for the 25th pizza is \$19.²



And as was case for demand, we can use either interpretation of a supply curve, depending in which type of problem we are interested. We could have asked what is the least a firm would have been willing to accept to produce the 450,000th or the 500,000th car.³ Likewise, we could ask how many pizzas that firms will wish to purchase at a price of \$15 or \$19.⁴

Which interpretation we choose does not change the underlying supply curve. As you progress, you will find it makes more sense to use one interpretation or the other depending on the problem we are dealing with.

Ceteris Paribus Conditions for Supply Curves

Thus far, we have glossed over the 1st Law of Supply and the ceteris paribus conditions for supply. Some more very deep background that you should have learned before...

If, we were interested in the important factors determining the number of cars produced, surely the price of cars would be important. This is exactly what we hope to capture with a supply curve. A supply curve illustrates how the quantity of cars that firms wish to produce changes as the price of cars changes, holding other relevant factors constant.⁵

² If you recall the difference between an individual firm's supply curve and the market supply curve, I want to be discussing the market supply curve in this case. We know that individual's supply curves are upward sloping. When we look at the market (overall) supply curve for a product, we have combined every firm's individual supply curve. It may or may not be the case that the firm that was willing to produce the 20th unit for \$15 is the same firm that is willing to produce the 25th pizza. More later...

³ Some firm would be willing to produce the 450,000th car for \$22,000 and some firm (could be the same firm or a different firm) would be willing to produce the 500,000th car for \$24,000.

⁴ Hopefully not surprisingly, the answers here are 20 and 25, respectively.

⁵ If you recall from your previous economics classes, in this case, we would call the price of the car the "own price". The own price is the price of the good for which we are drawing the supply curve.

On the other hand, many other things (aside from the price of cars) affect the number of cars firms wish to produce. We call these other things “supply shifters” or “ceteris paribus conditions for supply”.

In order to draw a supply curve (to isolate the relationship between the own price of a car and the quantity supplied of cars), we must hold the ceteris paribus conditions constant.

On the other hand, when one of these ceteris paribus conditions changes, we must shift the supply curve in the appropriate direction.

What are these “supply shifters” or ceteris paribus conditions for the supply of cars?

Just to name a few: Price of steel / wages of workers, price of trucks, technology of producing, expectations about future car prices.

More in the next section...

Mathematical Expressions of Supply Curves

We revisit our friends the mathematicians. Here is what they would write:

$$Q_x^s = f(P_x, P_r, W, H)$$

What does it mean? It means the quantity supplied of good X (Q_x^s) is a function of, or depends on, the price of good X (P_x) the price of technologically related goods (P_r), the price of inputs (W) and other stuff (H).⁶

While it is not necessarily the case, we often model supply curves as being linear.⁷ If so, we can write out a very simple numerical expression of a demand curve. For example:

$$Q_x^s = 70 + \frac{1}{3}P_x - 2P_r - 5W$$

We will call this a **supply function**. The supply function contains a whole slew of information.

If you were told to draw the supply curve for good X, again you would pull out a piece of graph paper, label the vertical axis P_x , the horizontal axis Q_x , and would be stuck.⁸ In fact, you cannot draw the supply curve without knowing the values of P_r and W , which are the ceteris paribus conditions for supply.

⁶ Why W for the price of inputs? One of the main inputs that firms use is labor, and the price of this input is called a wage. In short, W is to remind us of wages. Baye’s textbook slips once and uses P_w to refer to wages. Also, the items in H , the other stuff, will be different things for supply curves than they were for demand curves.

⁷ A linear supply curve is one with a constant slope of the supply curve and therefore the power (exponent) on the P_x term is (implicitly) 1. We will focus on linear supply curves in this class. As an example of a non-linear supply curve, consider $Q_x^s = 70 + \frac{1}{3}P_x^3 - 2P_r - 5W$.

Suppose you were told that $W = 8$ and $P_r = 5$. You could simplify the supply function to the point where you could draw it.

$$Q_x^s = 70 + \frac{1}{3}P_x - 2P_r - 5W$$

$$Q_x^s = 70 + \frac{1}{3}P_x - 2(5) - 5(8)$$

$$Q_x^s = 70 + \frac{1}{3}P_x - 10 - 40$$

$$Q_x^s = 20 + \frac{1}{3}P_x$$

Now, your supply function is expressed with only Q_x^s and P_x as unknowns.

Aside Inverse Supply Function

You guessed it, you can solve for an **inverse supply function**. Not as useful as the inverse demand function, but it will help a bit on the graphing. To find the inverse supply function, simply start with the supply function, and solve it for P_x . In the case where $W = 8$ and $P_r = 5$, we start with:

$$Q_x^s = 20 + \frac{1}{3}P_x$$

Now we just do the algebra. You can take whatever steps get you to the end. I will begin by subtracting 20 from both sides:

$$Q_x^s - 20 = \frac{1}{3}P_x$$

Then multiply each side by 3.

$$3Q_x^s - 60 = P_x$$

And then just rearrange the terms:

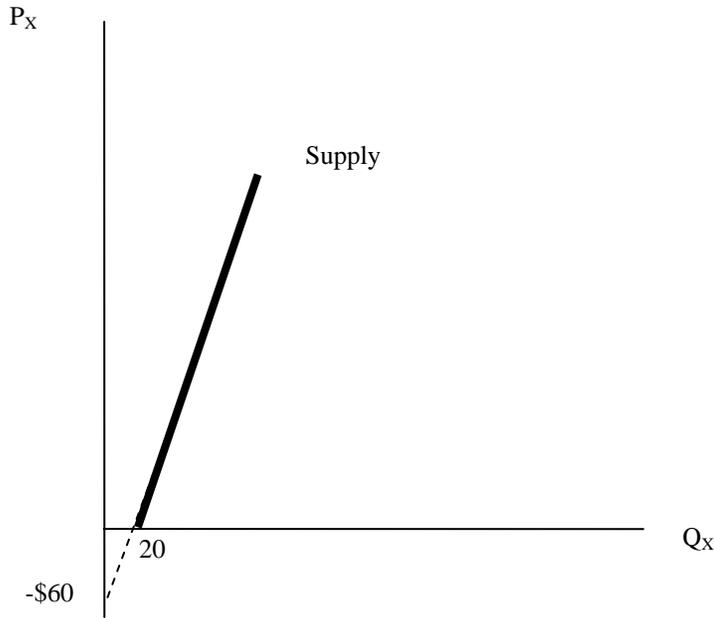
$$P_x = -60 + 3Q_x^s$$

Graphing the Supply Function

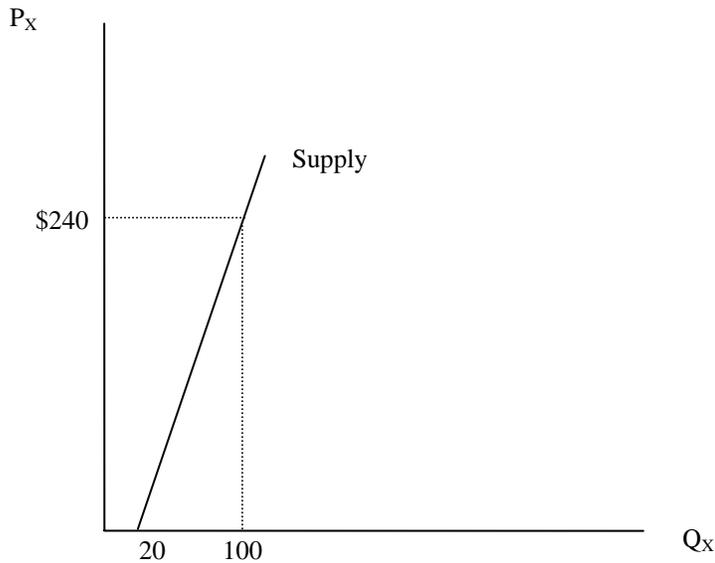
A bit trickier than demand functions in most cases. Recall from our supply function that $Q_x^s = 20 + \frac{1}{3}P_x$. By plugging in $P_x = 0$ in the expression above, we find that $Q_x^s = 20$. This is one point on our supply curve. In fact, this is the horizontal intercept of the supply curve.

We also solved for the inverse supply function, finding that $P_x = -60 + 3Q_x^s$. By plugging in $Q_x^s = 0$, we find that $P_x = -60$. Technically, this is the vertical intercept of the supply curve. But is any firm going to **pay** their customers \$60 for their product? There is no such thing as a negative price. Let's look at what we'd have on a graph thus far.

⁸ Why not label it Q_x^s ? You could. Eventually, we will combine the demand curve and supply curve, and thus we will just lazy and label it Q_x .



However, as you can see, this information still helps us out in drawing the picture. But as we hinted at above, only the portion of the supply curve in the positive quadrant (with positive price and positive quantity) is going to be relevant. So we'll end up discarding that lower (dotted) portion of the supply curve.



Sometimes it helps to throw in one more point. You'll see above, I've chosen $Q = 100$, at random, and stuck that on the graph as well. If we plug in $Q = 100$ to the inverse supply function, we see:

$$P_x = -60 + 3Q_x^s$$

$$P_x = -60 + 3(100)$$

$$P_x = 240$$

Shifting the Supply Curve

What if W decreases from \$8 to \$4? What happens to the supply curve? First the math, then the intuition.

$$Q_x^s = 70 + \frac{1}{3}P_x - 2P_r - 5W$$

$$Q_x^s = 70 + \frac{1}{3}P_x - 2(5) - 5(4)$$

$$Q_x^s = 70 + \frac{1}{3}P_x - 10 - 20$$

$$Q_x^s = 40 + \frac{1}{3}P_x \quad (\text{compare this to } Q_x^s = 20 + \frac{1}{3}P_x \text{)}$$

When we change the supply function, we will also get a new inverse supply function. Taking the same steps as before (but leaving out the explanation of these steps), we get:

$$Q_x^s = 40 + \frac{1}{3}P_x$$

$$Q_x^s - 40 = \frac{1}{3}P_x$$

$$3Q_x^s - 120 = P_x$$

$$P_x = -120 + 3Q_x^s$$

Now, we draw the new supply curve. From the supply function, I see that setting $P_x = 0$ in the expression above results in $Q_x^s = 40$. From the inverse supply function, I see that setting $Q_x^s = 0$ results in $P_x = -120$.

Not much help there. So let me stick in $Q_x^s = 100$ in the inverse supply function and I'll get $P_x = 180$.

Now we stick these new points on our graph paper.

Below is a picture with the original supply curve ($W = \$8$) labeled S_0 and the new supply curve ($W = 4$) labeled S_1 . We see immediately what has happened is that the supply curve has shifted to the right because of the decrease in wages. Just as we did with demand curves, we call a rightward shift of the supply an "increase in the supply curve".⁹ We call a leftward shift of the supply curve a "decrease in the supply curve".

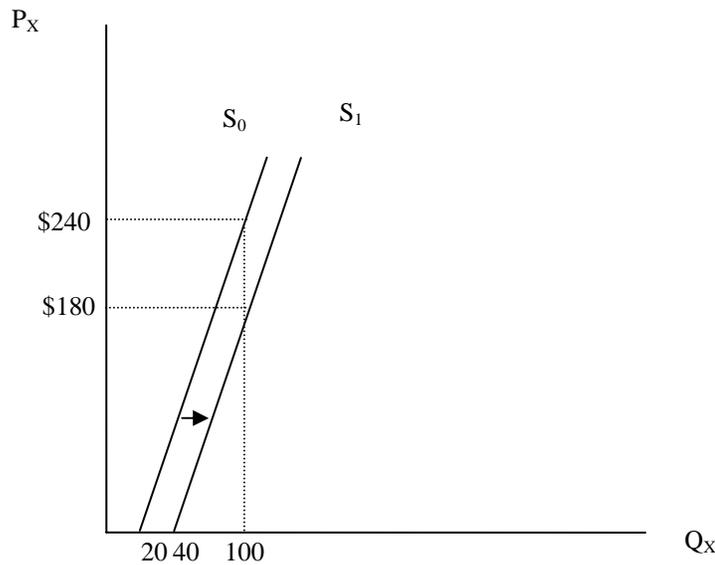
Again, any change in a ceteris paribus conditions shifts the supply curve. That is, a change in anything but the own price, causes a shift in the supply curve.

One more note: having discussed both demand curves and supply curves at this point, it is worth noting that most changes affect only the demand curve or the supply curve. There are separate lists of ceteris paribus conditions for demand and ceteris paribus conditions for supply. A change in wages will shift the supply curve, but not the demand curve. A change in the income of consumers will shift the demand curve, but not the supply curve.

However, every so often something comes about that shifts both the supply curve and the demand curve, but these are fairly rare.¹⁰

⁹ Why is this called an increase in the supply curve? From the horizontal interpretation of the supply curve, notice that at any (and every) price, there is a larger quantity supplied on the new supply curve than there is on the old supply curve.

¹⁰ An example might be the discovery of AIDS in the market for prostitution. This would affect both suppliers of prostitution services and the customers of prostitution services. Expectations about future prices will also affect both supply and demand.



Some exercises, you ask?

Start with original supply curve and $W = \$8$ and $P_r = \$5$. What would happen to the supply curve if W rose to \$12?¹¹

Start over at the original values. What would happen if the price of the related good (P_r) fell to \$2?¹²

Start over at the original values. What would happen if the price of the related good (P_r) rose to \$8?¹³

More on Ceteris Paribus Conditions

In our original supply shift, a decrease in W from \$8 to \$4 resulted in an increase in the supply for the good. That is, a decrease in an input price has led to an increase in supply.

But we also want to hone your intuition. We can make predictions about what happens to the supply curve without knowing anything about the actual mathematics of the supply curve.

Now is the time for a review of our ceteris paribus conditions. Our focus will be on how changes in our ceteris paribus conditions shift the supply curve.

¹¹ The supply curve decreases (shifts left). The new vertical intercept would be $P_x = \$0$ and the new horizontal intercept would also be $Q_x^s = 0$ (the supply curve starts at the origin). $P_x = \$300$ and $Q_x^s = 100$ would be another point.

¹² The supply curve increases (shifts right). The new vertical intercept would be $-\$84$ and the new horizontal intercept would be $Q_x^s = 26$. $P_x = \$216$ and $Q_x^s = 100$ would be another point.

¹³ The supply curve decreases (shifts left). The new vertical intercept would be $-\$42$ and the new horizontal intercept would be $Q_x^s = 14$. $P_x = \$258$ and $Q_x^s = 100$ would be another point.

Generally speaking, the ceteris paribus conditions for Supply can be classified into a few major groups”

1. Input Prices
2. Prices of Technologically Related Goods (substitutes and complements in production)
3. Expectations of Future Prices
4. Other Stuff

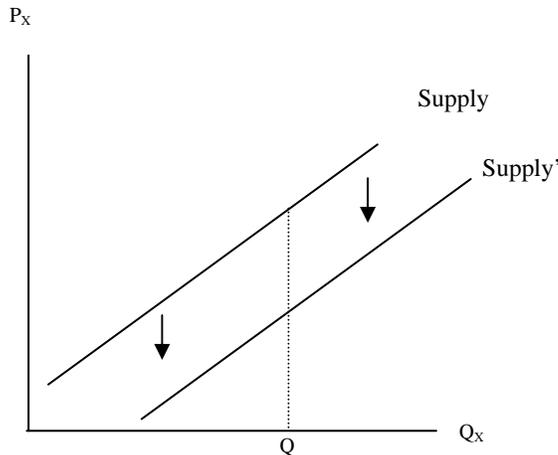
Input Prices

With input prices, a mix of the vertical interpretation of the supply curve and the horizontal interpretation of the supply curve is necessary.

Consider, as we did above, a decrease in wages, one of the firm’s input prices. That is, consider a decrease in an input price. It stands to reason that a firm would now be willing to accept a lower price to produce the good. Why? We know a firm, broadly speaking, won’t produce a good unless it covers its costs. Because the cost of production is decreasing, the firm can accept a lower price (and still cover its costs).

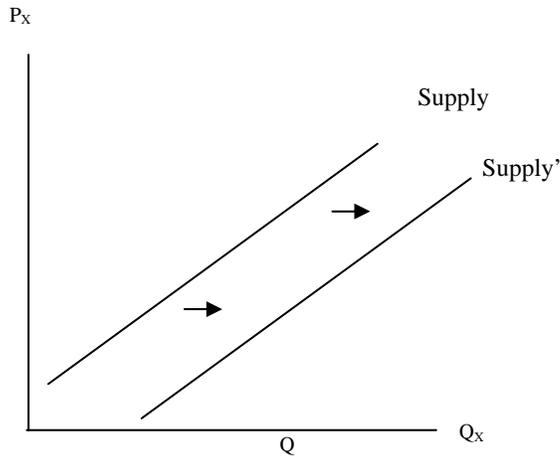
For example, if the height of the old supply curve at some quantity Q was \$12, and now wages fall, the firm might accept \$11 to produce the Q^{th} unit of the good. More generally, because the height of the supply curve tells us the lowest price a firm will accept to produce a good, a decrease in an input price will cause the supply curve to shift down, vertically.¹⁴

See the picture below.



But as soon as we see this picture, we want to go back to the horizontal interpretation of supply curve. When we discuss increases in supply or decreases in supply, we want you think of these as changes as shifts to the right or left (not up or down). So if you were presented with the picture shown below (the same two supply curves), and you were forced to describe it using the horizontal interpretation of a supply curve, hopefully you would conclude there has been a shift to the right of the supply curve, which we call an increase in supply.

¹⁴ If you are comfortable interpreting supply curves as marginal cost curves, then all we are saying is the marginal cost of production is reduced at each level of output. The supply curve (marginal cost curve) shifts down vertically.



At the end of the day, a decrease in an input price leads to a rightward shift of the supply curve (an increase in supply). An increase in an input price leads to a shift to the left of the supply curve (a decrease in supply).

Don't believe me? Go back to the example above where we change the wage from \$8 to \$4. You'll see we had a rightward shift of the supply curve. Finally, for an increase in an input price, do the exercise where I suggested an increase in the wage from \$8 to \$12.

A "shortcut" string of logic to get to this answer is as follows. If an input price decreases, firms will find it more profitable to produce each unit. This will further cause the firm to wish to produce more units. Therefore, a decrease in an input price leads to an increase in the supply of the good.

More examples?

What happens to the supply curve for pizza if there is a decrease in the price of tomato sauce?¹⁵

What happens to the supply curve for cars if there is an increase in the price of steel?¹⁶

Prices of Technologically Related Goods – I'd skip this section – no important for Econ 485

Technological Complements

Technological complements are things that are produced together.

Let's imagine for expositional purposes that every time a donut is produced, a donut hole is also produced. Another example would be that at a slaughterhouse, every time a cow is slaughtered, there is both beef and a hide produced. Any situation in which a "byproduct" is involved would be an example of technological complements.

The main idea is that production of the one good might be influenced by the price of the other, because the production of both goods is linked. A more formal definition:

Goods A and B are called technological **complements**, if when the price of good A increases, the quantity supplied of Good B changes in the **same** direction.

¹⁵ The supply curve increases (shifts right).

¹⁶ The supply curve decreases (shifts left).

Examples:

If the price of donut holes increases, the firm will react by increasing the quantity supplied of donuts (increase in supply).

If the price of beef decreases, a firm would decrease the quantity of hides it will produce (decrease in supply).

Technological Substitutes

Technological substitutes are alternative products. General motors can use its equipment to produce compact cars or SUVs. A donut shop could produce crème filled donuts or danishes.

Goods A and B are called technological **substitutes**, if when the price of good A increases, the quantity supplied of Good B changes in the **opposite** direction.

Examples:

If the price of SUVs decreases, General Motors will increase the quantity supplied of compact cars (increase in supply). The logic here is that because the equipment is capable of producing both, producing SUVs will be less relatively less lucrative, and thus the firm will switch to producing compact cars. Sound familiar?

If the price of danishes increases, a donut shop will decrease its quantity supplied of donuts (decrease in supply). Now producing danishes will be more lucrative, so the firm will substitute from producing donuts to danishes.

The Numerical Example

Recall our original expression of the supply function:

$$Q_x^s = 70 + \frac{1}{3}P_x - 2P_r - 5W$$

Recall that P_r indicates the price of the related good. Are goods X and the related good technological substitutes or technological complements?

Again, it will come down to the sign on the P_r term. In this case, the coefficient on the P_r term is -2.

This means if P_r increases by \$1, Q_x^s will fall by 2 units.¹⁷

This is consistent with technological substitutes, as the price of the one good and the quantity supplied of the other good are moving in the opposite direction.

Not convinced? Go back to the exercise of changing P_r from \$5 to \$2 and then changing P_r from \$5 to \$8. Confirm that the changes in the supply curve are what you'd expect.

¹⁷ If you are thinking that the magnitude of the coefficient tells you how closely technologically related the two goods are you are correct. The larger the size of the coefficient (in absolute value, further from zero) the more technologically related the two goods. And of course, a positive coefficient on the P_r term would indicate technological complements.

Expectations of Future Prices

When suppliers expect prices to fall in the future, current supply will increase. Firms will attempt to sell their products at the high current price.

When firms expect prices to rise in the future, the current supply decreases. Firms will hold back production.

One detail here...take for example the second situation in which firms expect prices to rise in the future. While we say that current supply will decrease, we could be a bit more precise. What is likely to happen is that firms will continue to produce goods, but will not sell as many of these goods today, instead choosing to accumulate inventories, which they then expect to liquidate when prices rise in the future. In this regard, when we say supply, we are talking about those products that are brought to market, not the quantity of goods that are manufactured.¹⁸

You may also have noticed that if firms hold back production today in anticipation of future prices, this act itself may cause prices to increase now. More later...

Everything Else

I liked writing notes more last week... Many other things can shift a supply curve.

Peculiarities with Sports

One of the peculiar things with sports is the structure of the supply curve. Take for example a New Orleans Saints Game. The players and coaches are under contract, and whether 35,000 people show up or 35,001 people up, the ushers, parking staff, etc are already hired. The additional cost of service one customer, is more or less, \$0. Thus, the Saints stand ready to sell as many units of output (tickets) as the stadium will hold, at any price. We could imagine a vertical supply curve at stadium capacity.

Producer Surplus

My hope is that you can anticipate everything that is forthcoming here.

The height of the supply curve tells us the minimum amount a producer is willing to accept to produce a unit of the good. But seldom does this producer actually get paid this amount. If there is a gap between the two, we call this gap producer surplus.

Producer Surplus = Amount producer actually is paid - the minimum amount they would have accepted

From the vertical interpretation of a supply curve, we know that the height of a supply curve at some quantity tells us the minimum amount a producer was willing to accept. If we compare this to the price on the graph, we have producer surplus.

Let's go back to the pizza example, and assume the price of pizza is \$22.

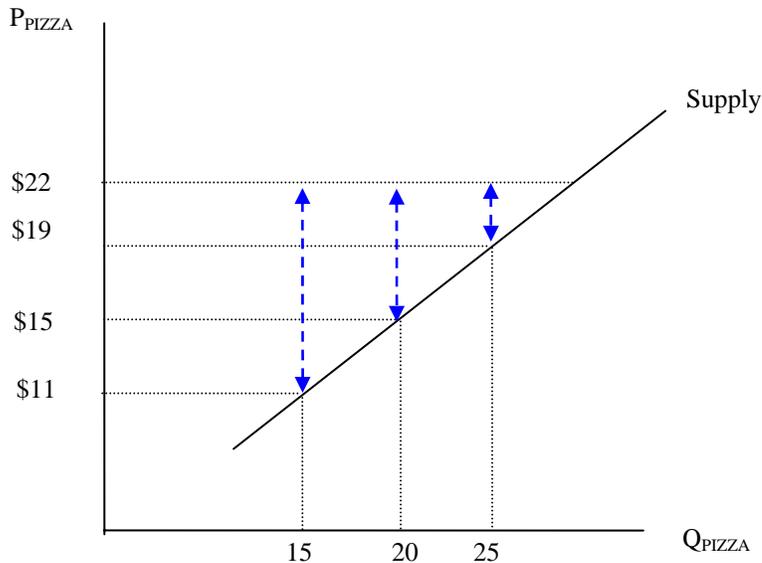
At $Q = 15$, the height of the supply curve is \$11, while the price is \$22, so consumer surplus is \$11 ($\$22 - \$11 = \11). The idea is that the producer is getting a deal. They would have accepted \$11 for the good, but they received \$22. If this \$11 sounds something like profit, you're on the right track.¹⁹

¹⁸ Not all firms have this option. For example, Domino's pizza can not produce pizzas in August, and sell them in September.

¹⁹ If you are comfortable with interpreting supply curves as a marginal cost curve, producer surplus is the difference between P and marginal costs, sometimes called operating profit.

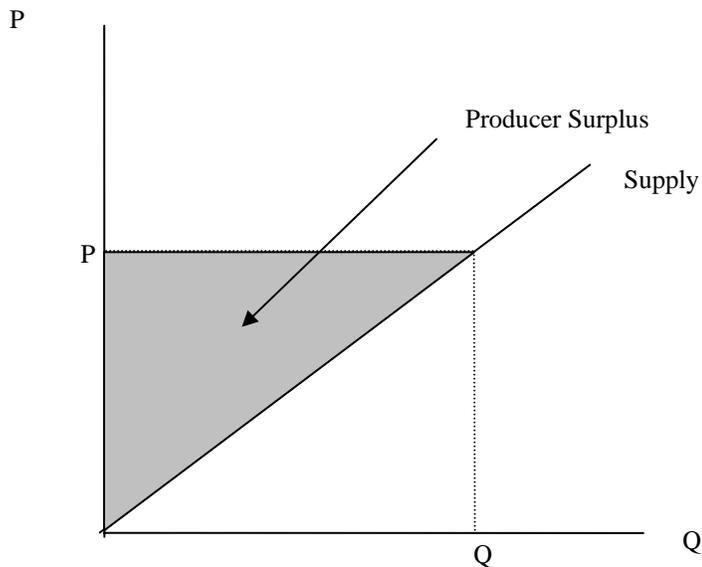
At $Q = 20$, producer surplus is $\$22 - \$15 = \$7$.

It should be easy for you to calculate producer surplus on the 25th pizza.²⁰



Just as before, adding producer surplus up for each individual unit of the goods gets very tedious. We'll use the same calculus trick.

In general, if you want to add up the total producer surplus for producing Q units (where Q can be any quantity), simply add up the entire area that is (1) above the supply curve, (2) below the price that suppliers receive, and (3) out to the quantity Q . See the picture below.



²⁰ Producer surplus is \$3, as the producer is willing to accept (the height of the supply curve) \$19, while the price is \$22. Producer surplus is $\$22 - \$19 = \$3$.

Markets and Equilibrium

Market – a process or location in which equilibrium is established and the otherwise inconsistent aspirations of demanders and suppliers are reconciled.

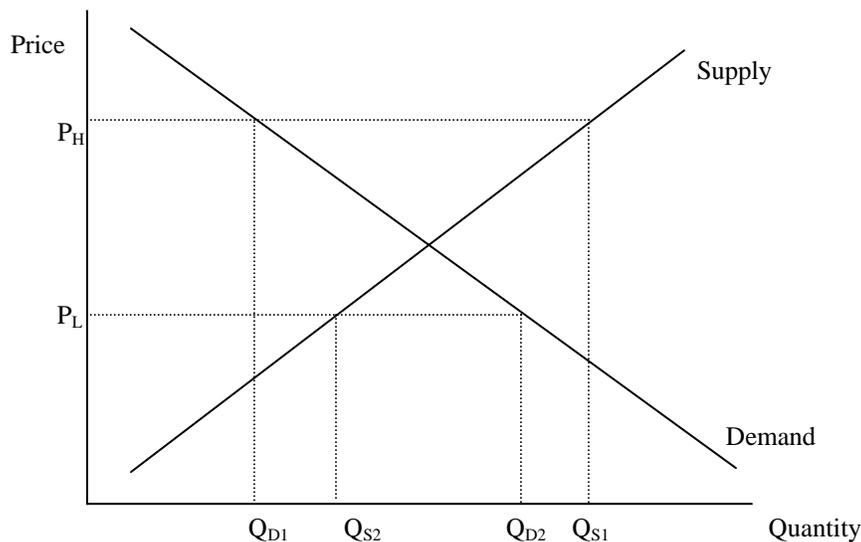
Equilibrium - an outcome or state that will tend to persist unless disturbed by a change in one or more of the ceteris paribus conditions.

Is the equilibrium price P_H ?

I bet you guys know the drill, but as long as we've wandered down this path, we may as well wander a bit more. Let's combine a supply curve (firm behavior) with a demand curve (consumer behavior) and see if we can't see how things play out. Basically, what is the equilibrium price and quantity we will observe?

Let's try out P_H . At this price, we find out how much suppliers want to produce by looking at their supply curve (Q_{S1}), and how much demanders want to consume by looking at the demand curve (Q_{D1}). At this price, $Q_{S1} > Q_{D1}$. This is called an excess quantity supplied or **surplus**. Suppliers would like to supply more than demanders want to purchase (suppliers and demanders aspirations are not consistent). Suppliers will not be able to sell all that they wish to at this high price.

The surplus will cause downward pressure on price. As price falls, two things happen simultaneously. Suppliers desire to produce less as price lowers, and demanders desire to purchase more as price lowers. This causes the size of the surplus to decrease. The downward pressure on price will continue until the equilibrium price is reached and there is no longer a surplus.



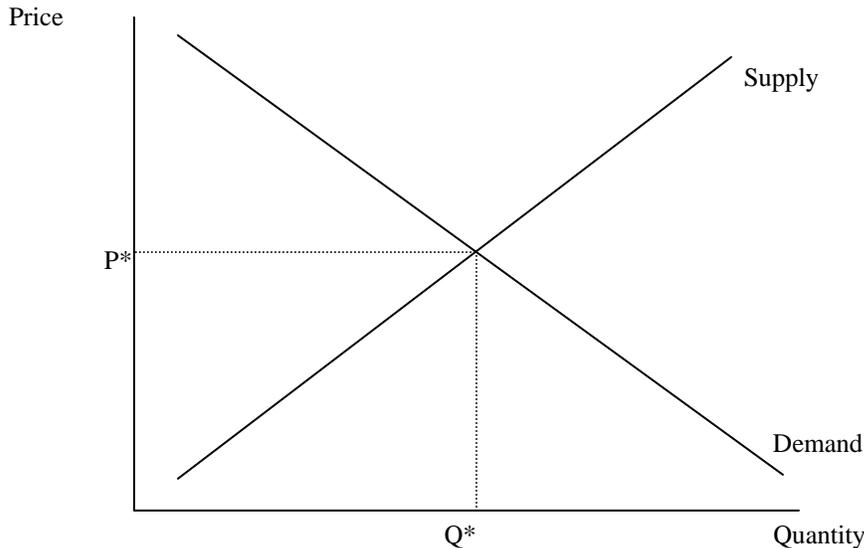
Is the equilibrium price P_L ?

Let's try P_L . At this price, we again find out how much suppliers want to produce by looking at their supply curve (Q_{S2}), and how much demanders want to consume by looking at the demand curve (Q_{D2}). At this price, $Q_{D2} > Q_{S2}$. This is called an excess quantity demanded or **shortage**. At this low price, many consumers want to purchase the product, but suppliers will not be willing to produce as much as consumers desire (the aspirations of suppliers and demanders are inconsistent).

This will cause upward pressure on price. As the price rises, two things simultaneously happen. Suppliers desire to produce more as price rises, and demanders desire to purchase less as price rises. These two factors cause the size of the shortage to decrease. This upward pressure on price will continue until the equilibrium price is reached and the shortage disappears entirely.

The equilibrium price is P*

Let's try P*. Refer to the picture below. At P*, demanders want to purchase Q* units. At P*, suppliers want to supply Q* units. Everyone who is willing to pay P* can buy all the goods they want. Everyone who is willing to supply goods at P* is supplying what they want. Everyone's aspirations are consistent and everyone is happy. Let's all hold hands and sing. No one wants to change their behavior, hence an equilibrium.



P* and Q* are the equilibrium price and equilibrium quantity. Equilibrium occurs where the supply curve intersects the demand curve. **In other words, at the equilibrium price, quantity supplied equals quantity demanded.** There is neither a shortage, nor a surplus - we say that the market clears.

Even if we find the market price temporarily away from the equilibrium price, these pressures will cause the price to tend to return to equilibrium price. This is why we say equilibrium will “tend to persist” unless there is a change in a ceteris paribus condition.

Comparative Statics

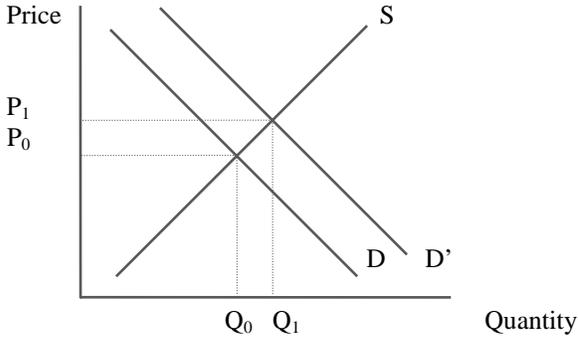
Finding equilibrium prices and quantities are nice, but the most useful thing we'll get out of supply and demand analysis will be what we call **comparative statics**. Basically, we change a ceteris paribus condition for some good then we see what impact this will have on the equilibrium price and quantity of that good.

You can think of comparative statics exercises as a four part process. As your get more comfortable, you'll breeze through these without doing them step by step.

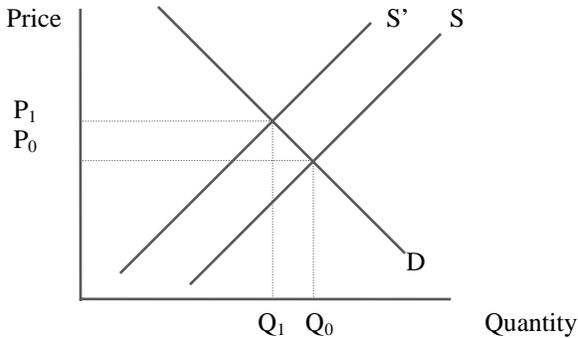
1. Start in initial equilibrium (draw an initial S & D curve to start)
2. Change one or more of the ceteris paribus conditions
3. Examine the impact on demand or supply (shift the appropriate curve)
4. Examine new equilibrium (compare)

Examples of Comparative Statics

Suppose we are looking at the market for oranges. The initial equilibrium is (P_0, Q_0) . There is an increase in income. Oranges are a normal good. This causes an increase in demand (demand curve to shift to the right). At the original price, P_0 , there is now an excess quantity demanded, putting upward pressure on the price. The new equilibrium will be (P_1, Q_1) . Both equilibrium price and quantity will increase.

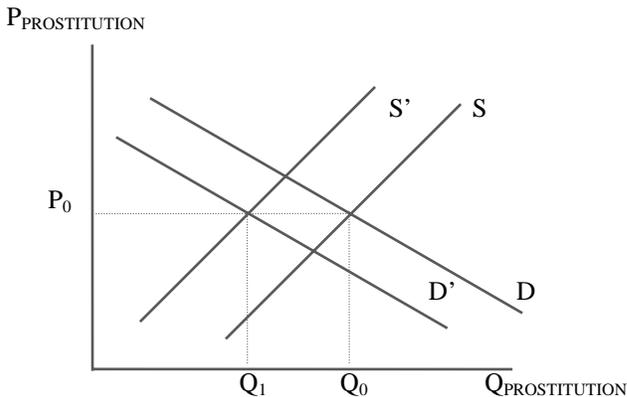


Now, suppose unusually cold weather occurs in Florida, destroying some of the orange crop. This causes a reduction in the supply of oranges (supply curve shifts left). The initial equilibrium is (P_0, Q_0) . The new equilibrium is (P_1, Q_1) . The equilibrium price will increase and the equilibrium quantity will decrease.



What if more than one curve shifts?

Suppose we are looking at the market for prostitution. Suddenly, AIDS is developed. What happens to the prostitution market? Again, the initial equilibrium is (P_0, Q_0) . AIDS decreases the demand for prostitution, as well as decreases the supply of prostitution. The new equilibrium is (P_1, Q_1) . You can think of AIDS as a sort of increase in the cost of producing “prostitution services”. Prostitutes will need to be paid a higher price to incur the risk of contracting AIDS.



As it is drawn, the price does not appear to have changed. However, the change in price is ambiguous. This can be seen two ways. First, take the two changes one at a time.

<u>Change</u>	<u>Effect on P</u>	<u>Effect on Q</u>
Decrease in supply	increases	decreases
Decrease in demand	decreases	decreases
Total	ambiguous	decreases

The other way this can be shown is to use supply and demand shifts of different magnitudes. Draw this with a large demand curve shift and a small supply curve shift. Check your answer. Now draw it again with a large supply curve shift and a small demand curve shift. Compare. You should see that in one case the price rises, while in the other case, the price falls.

In general, if you simultaneously change both a ceteris paribus condition for both supply and for demand, either the direction in the change of price or the direction of the change in quantity will be ambiguous.

What about the math?

If we had a demand function and a supply function, can we solve for the equilibrium price?

The answer is yes. Hurray!

Here is how. We noted above that the equilibrium price was the price where quantity demanded is equal to quantity supplied. The nice thing is the demand function tells us what quantity demanded will be at various prices, while the supply function tells us what quantity supplied will be at various prices. By setting quantity demanded equal to quantity supplied and solving for the price, we can determine the equilibrium price. Once we do this, we can plug the equilibrium price back into the demand function (or the supply function) to determine the equilibrium quantity. Let's do an example with our hopefully now familiar demand function and supply function.

Recall we started with:

$$Q_x^d = 100 - \frac{1}{2}P_x - 2P_y + 6M$$

then assumed that $M = 10$ and $P_y = 30$, which resulted in:

$$Q_x^d = 100 - \frac{1}{2}P_x$$

On the supply side, we started with:

$$Q_x^s = 70 + \frac{1}{3}P_x - 2P_r - 5W$$

then assumed that $W = 8$ and $P_r = 5$, which simplified to:

$$Q_x^s = 20 + \frac{1}{3}P_x$$

Now, set quantity demanded equal to quantity supplied, and solve for P_x .

$$Q_x^d = Q_x^s \Rightarrow 100 - \frac{1}{2}P_x = 20 + \frac{1}{3}P_x \Rightarrow 80 - \frac{1}{2}P_x = \frac{1}{3}P_x$$

$$\Rightarrow 80 = \frac{5}{6}P_x \Rightarrow 80\left(\frac{6}{5}\right) = P_x \Rightarrow 96 = P_x$$

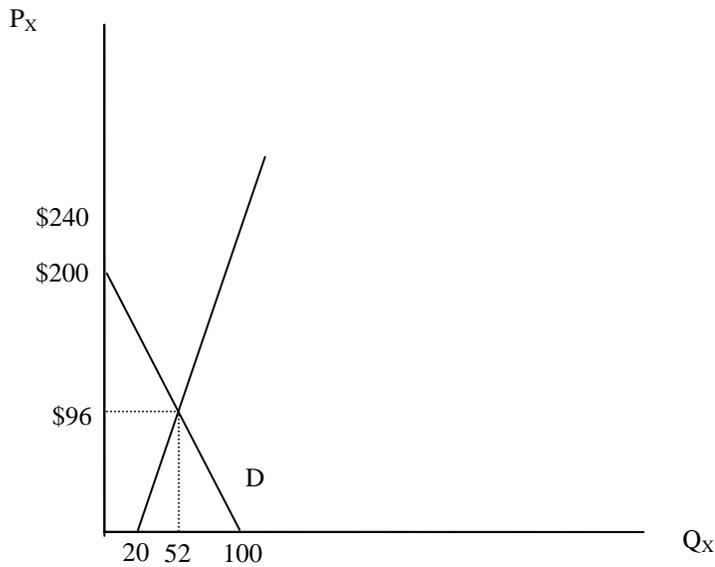
Can we be sure we are correct? Let's indeed confirm that at a price of \$96, the quantity demanded is equal to the quantity supplied.

$$Q_x^d = 100 - \frac{1}{2}P_x = 100 - \frac{1}{2}(96) = 100 - 48 = 52$$

$$Q_x^s = 20 + \frac{1}{3}P_x = 20 + \frac{1}{3}(96) = 20 + 32 = 52$$

Indeed, the quantity demand equals quantity supplied.

One last thing we can do is to check the graph. Visually, our answer looks reasonable.



Intro to Game Theory

Game theory is the theory of strategic behavior. Its development is largely related to military applications, but game theory is now extensively utilized by those studying economics, management, political science, and even biology and genetics.

Broadly speaking, players (**participants**) are placed in strategic situations called **games**. Players are given choices they can take, rules (when they can take them), and develop certain courses of actions (**strategies**). The game is then played, and players either win or lose, or receive some payment (**payoffs**). For example, in the prisoner's dilemma, the payoffs will be the number of years in prison Bonnie and Clyde serve. In most games, the payoffs will be profits of some hypothetical firm. While games can have more than two players, we will focus on games with only two participants.

Some games are **simultaneous** – people make their choices at the same time (e.g. rock, scissors, paper). Other games are **sequential** – people take turns (e.g. tic, tac, toe). Most games we'll look at are **one-shot games** – games that are played only once, while other games are **repeated games** – games that are played more than once with the same opponents.

Solution Methods

Depending on how far you go in game theory, there are number of different ways to solve games.

Nash equilibrium -

One of the most simple is called a **Nash equilibrium**. A Nash equilibrium is called an "equilibrium" because no player wishes to change his or her strategy. Each person is making the best response to their opponent's strategy and these strategies are consistent.

Some games have what are called **pure strategy Nash equilibria**. An example of a game of this type is the prisoner's dilemma.

Other games have no pure strategy Nash equilibrium. These games are often solved with a method called **mixed strategy equilibrium**. In these games, some sort of randomness is involved. An example of this type of game is the employee monitoring game.

Other games have multiple Nash equilibrium. Determining which of the multiple Nash equilibrium will occur in a simultaneous game is a difficult problem. An example of this type of game is the product choice game.

Backwards Induction –

Backwards induction is a way to solve sequential games (players take turns). The idea is – surprise – to work backwards. We won't too much about these, but the solutions you'd get are special cases of Nash equilibrium called **sub-game perfect equilibria**. By going backwards, we are effectively able to eliminate strategy combinations that would not occur.

Minimax / Maximin

We'll see if we want to talk about this. More later?

Prisoner's Dilemma

First off, note that this particular formation of the prisoner's dilemma wasn't discussed in class.¹ This classic game theory exercise is called the **prisoner's dilemma**. The story is just what it sounds like – two criminals are put in separate cells and have been accused of committing a crime, which they have committed. They cannot communicate and will make their decisions simultaneously. It is a one-shot game.

If both confess, they both will go to prison for 5 years. If neither confesses, they will be charged with a lesser crime and each go to prison for 2 years. The prosecutor offers a plea bargain to each – if one confesses and the other does not, the one who confesses gets a break – only one year in prison – while the one who does not gets a long sentence, going to prison for 10 years.

When this might at first glance seem like textbook BS, this will be very much related to when we discuss pricing and output decisions for firms. We'll see a bunch of examples below.

The game can be expressed in the following chart:

		Clyde	
		Confess	Don't Confess
Bonnie	Confess	Bonnie: 5 years Clyde: 5 years	Bonnie: 1 years Clyde: 10 years
	Don't Confess	Bonnie: 10 years Clyde: 1 years	Bonnie: 2 years Clyde: 2 years

Bonnie's strategy choices are reflected in the rows of the game. Clyde's strategy choices are reflected in the columns of the game. Each of the four shaded boxes represents a combination of Bonnie's strategy and Clyde's strategy. For instance, if Bonnie chooses not to confess and Clyde chooses to confess, we end up with Bonnie going to prison for 10 years and Clyde going to prison for 1 year.

Solving the Prisoner's Dilemma / Best Responses / Dominant Strategies

Where you should always start when looking for the solution to a game is to look for dominant strategies. A **dominant strategy** is the case where one player's best response does not depend on their opponent's strategy.²

First, what is Bonnie's best response if Clyde chooses to confess?

If Clyde were to confess and Bonnie chooses to confess, she goes to prison for 5 years.
If Clyde were to confess and Bonnie chooses not to confess, she goes to prison for 10 years.

Going to prison for 5 years is better than going for 10 years. So, if Clyde chooses to confess, Bonnie's best response is to confess.

¹ The game we used in class also in these notes in a few pages. Treat this one as another example you can practice on.

² Sometimes students think that games must have either no players or both players with dominant strategies. A game can be such that only one player has a dominant strategy.

Next, what is Bonnie's best response if Clyde chooses not to confess?

If Clyde were not to confess and Bonnie chooses to confess, she goes to prison for 1 year.
If Clyde were not to confess and Bonnie chooses not to confess, she goes to prison for 2 years.

So when Clyde chooses to confess, Bonnie's best response is to confess.

In this case, we say Bonnie has a **dominant strategy**. Her best response does not change depending on what strategy her opponent chooses. Whether Clyde chooses to confess or Clyde chooses not to confess, we find that Bonnie's best response is to confess. We would expect Bonnie to Confess.

You should now work through what Clyde's best responses are.

What is Clyde's best response if Bonnie chooses to confess?

What is Clyde's best response if Bonnie chooses not to confess?

Does Clyde have a dominant strategy? If so, what is it?³

Solving the Prisoner's Dilemma / Nash Equilibrium

Back to the idea of a Nash equilibrium. At this point, you will hopefully be inclined to think that the solution to this game is that both Bonnie and Clyde confess. In fact, if both players have dominant strategies, the Nash equilibrium (outcome of the game) will be both play their dominant strategy.

It is instructive, however, to confirm that the combination of Bonnie confesses and Clyde confesses is in fact a Nash equilibrium. If it is in fact a Nash equilibrium, neither player will have an incentive to change their strategy. Let's check.

Given that Bonnie is choosing to confess, is Clyde making his best response by confessing? Yes. Does he want to change his strategy? No.

Given that Clyde is choosing to confess, is Bonnie making her best response by confessing? Yes. Does she want to change her strategy? No.

Neither wants to change – and their strategies are consistent. Thus, the Nash equilibrium of this game is for both to confess. Both will confess, and both will spend 5 years in prison.

Now, at this point hopefully you're yelling "wait a minute". If they both had chosen not to confess, they would both go to prison for only 2 years. True. Why can't they pull it off?

Suppose for a second that Bonnie was reasonably convinced that Clyde will "cooperate" and keep his mouth shut.⁴ What should she do at this point? She will again want to take the plea and rat him out. Her best response is to confess. The same goes for Clyde. Regardless of what their opponent does, they are better off confessing.

In a prisoner's dilemma type game, there will be a powerful individual incentive to confess ("cheat"), even though what is best for the group ("cooperate") is for everyone to keep their mouth shut.

³ You should have found that Clyde also has a dominant strategy to confess. We'd expect Clyde to confess.

⁴ I mean cooperate from the point of view of Bonnie and Clyde (not cooperate as in cooperating witness).

This type of situations shows up all the time in the real world. Provision of public goods, studying for exams, and importantly cartels (pricing decisions) all can be thought of as prisoner’s dilemma situations.

I tried to put my money where my mouth was in class one time with undergraduates – 80% of them actually played the Nash equilibrium strategy when playing for extra credit points.

Sports Example of Prisoner’s Dilemma (TV cartel)

Here is the game we discussed in class. The story here was that there are two “powerhouse” football teams both located in Florida. They are trying to determine how many of their football games to televise.

If both schools limit the number of times they appear on TV, they will “split the market” and both will get large payments for their broadcast rights thus and high “payoffs”.

If both schools televise many games, they will both get low prices (too much output) and thus low “payoffs”.

However, if one team chooses to limit their broadcasts while the other does not, the team that does not limit its broadcasts will earn more than the team that limits their broadcasts. The team that broadcasts many games will capture a large fraction of the market.

This can be distilled down to the following game, which has the same fundamental structure of the prisoner’s dilemma.

		Miami	
		Many Game	Few Games
FSU	Many Games	FSU: \$5 million Miami: \$5 million	FSU: \$20 million Miami: \$3 million
	Few Games	FSU: \$3 million Miami: \$20 million	FSU: \$10 years Miami: \$10 years

What is FSU’s best response if Miami chooses many games? To choose many games.
 What is FSU’s best response if Miami chooses few games? To choose many games.

FSU has a dominant strategy to choose many games.

What is Miami’s best response if FSU chooses many games? To choose many games.
 What is Miami’s best response if FSU chooses few games? To choose many games.

Miami also has a dominant strategy to choose many games.

Thus, the Nash Equilibrium is for both choose many games.

Basically, this is a cartel game. Cartels seek to restrict output and drive prices up. If both cooperate (televise few games), both will do well. If both cheat (televise many games), they will not do as well. But each has the incentive to cheat the cartel in order to capture a higher market share. And as we see here, we expect the cartel to have a hard time holding up, as each has a dominant strategy to cheat.

The Prisoner’s Dilemma as a repeated game

The analysis above assumes that this game is what we call a **one-shot game** – a game played only once. The problem with the prisoner’s dilemma, from the viewpoint of the participants, is that it is difficult to get cooperation in the face of the self-interest of the participants. What is good for the individual is not good for the group. If this game is played repeatedly with the same participants, what we call a **repeated game**, however, we get a much more interesting story.

If Bonnie and Clyde know each other, have been in this situation many times in the past, and have always kept their mouth shut, it is quite likely that trust may factor into the game and we might see both keep their mouth shut (both don't confess).

In the sports context, leagues / NCAA might be able to control these problems.

More below.

But before you get too excited about cooperating, one interesting aspect of this part of the story is that when we know that the game is in its "last period", there is no tomorrow and cooperation is very unlikely. That is cooperation induced by repeated games can break down when the end is known.

More below. But first back to some other games.

Product Choice Game

Here is one we didn't talk about in class. You should try it out.

The story here is two breakfast cereal companies are thinking of introducing a new cereal. One is a "crispy" cereal and the other is a "sweet" cereal. Check out the payoffs (profits) in the table below. The game has been set up so the firms don't care which product they produce, so long as both firms don't introduce the same product. The idea is there "isn't enough room" in the market for two new crispy cereals or two new sweet cereals. The firms make their decisions simultaneously without communicating.

		Firm 2	
		Crispy	Sweet
Firm 1	Crispy	Firm 1: -5 Firm 2: -5	Firm 1: 10 Firm 2: 10
	Sweet	Firm 1: 10 Firm 2: 10	Firm 1: -5 Firm 2: -5

First check for dominant strategies.

- What is Firm 1's best response if Firm 2 chooses crispy? Choose sweet
- What is Firm 1's best response if Firm 2 chooses sweet? Choose crispy

Does Firm 1 have a dominant strategy? No. Their best response changes depending on their opponent's strategy.

- What is Firm 2's best response if Firm 1 chooses crispy? Choose sweet
- What is Firm 2's best response if Firm 1 chooses sweet? Choose crispy

Does Firm 2 have a dominant strategy? No. Their best response changes depending on their opponent's strategy.

Since we don't have any dominant strategies, we'll have to work harder to come up with the solution to this game. While a brut force method, you can always check "manually" to see if a certain combination is a Nash equilibrium. Let us try out a couple. Then we'll show you an easier way.

Possibility #1 - Firm 1 chooses crispy, Firm 2 chooses crispy:

Is it a Nash equilibrium? To find out, check to see if each player is making the best response to their opponent's strategy.

Given that firm 1 chooses crispy, is Firm 2's best response to choose crispy? The answer is no. Firm 2 would want to change their strategy. That is enough to tell you this is not a Nash equilibrium, but let's pile

on. Given that firm 2 chooses crispy, is Firm 1’s best response to choose crispy? The answer is no. Firm 1 would want to change their strategy.

Again, this is not a Nash equilibrium.

Possibility #2 - Firm 1 chooses sweet, Firm 2 chooses crispy:

Is it a Nash equilibrium? Again, to find out, check to see if each player is making the best response to their opponent’s strategy.

Given that firm 1 chooses sweet, is Firm 2’s best response to choose crispy? Yes. They don’t want to change their strategy. Given that firm 2 chooses crispy, is Firm 1’s best response to choose sweet? Yes. They don’t want to change their strategy.

Neither wants to change. Firm 1 choosing sweet and Firm 2 choosing crispy is a Nash equilibrium.

You try out the other two possibilities.

Possibility #3 - Firm 1 chooses sweet, Firm 2 choosing sweet

Possibility #4 - Firm 1 chooses crispy, Firm 2 chooses sweet⁵

In this game, there are two Nash equilibria – both involve one firm choosing crispy and the other firm choosing sweet. How do we know which one will happen? In a strict game theory sense, we don’t.

But let’s go outside the game and think about the “real world”. This is a game about coordination. What might we expect the firms to do in this case? Perhaps one firm will send out a press release that says they are introducing a sweet cereal. This would be an indication to the other firm that they should choose sweet. Might this have something to do with trade group meetings?

Back to sports. Suppose there were to be two new expansion teams in a league. Surely the coordination would happen through leagues office, right? The league would grant each franchise an exclusive territory. The exclusive territory essentially prevents this type of problem and coordinates franchise location.

Aside: Is there an easier way to find Nash Equilibria?

There is. Put yourself in the shoes of Firm 1 for a moment. Now, suppose you know Firm 2 was going to choose crispy. Put a circle around your best response. (I’ve labeled this 1.) Now, suppose you know Firm 2 was going to choose sweet. Put a circle around your best response. (I’ve labeled this 2.)

Now, put yourself in the shoes of firm 2 for a moment. Imagine you know Firm 1 is going to choose crispy. Put a circle around your best response. (I’ve labeled this 3.) Now imagine you know Firm 2 is going to choose sweet. (I’ve labeled this 4.) Put a circle around your best response.

		Firm 2	
		Crispy	Sweet
Firm 1	Crispy	Firm 1: -5 Firm 2: -5	Firm 1: 10 ² Firm 2: 10 ³
	Sweet	Firm 1: 10 ¹ Firm 2: 10 ⁴	Firm 1: -5 Firm 2: -5

⁵ Possibility 3 is not a Nash Equilibrium, while Possibility 4 is a Nash Equilibrium

If you have a combination of strategies that has two circles (one indicating player 1’s best response and the other indicating player 2’s best response), that combination of strategies is a Nash Equilibrium. See the chart above.

You should try out this trick on the prisoner’s dilemma game above. Confirm the Nash equilibrium is for both to confess. Likewise, the Nash equilibrium for the TV cartel game is both to choose many games. This is a bit quicker than trying them all out.

Employee Monitoring Game

The story here is the workers can either work hard (work) or be lazy (shirk). The manager can either check on the employee to see if they are working (monitor) or not (don’t monitor).

- If the manager monitors a worker that was already working, the worker wins and the manager loses (wastes time monitoring an employee that was working).
- If the manager monitors a worker that was shirking, the workers loses and the manager wins.
- If the manager doesn’t monitor a worker that was working, the manager wins and the employee loses (the worker could have gotten away with shirking).
- Lastly, if the manager doesn’t monitor a worker that wasn’t working, the manager loses and the employee wins.

All this boils down to the following game:⁶

		Worker	
		Work	Be Lazy
Manager	Monitor	Manager: L Worker: W	Manager: W Worker: L
	Don’t Monitor	Manager: W Worker: L	Manager: L Worker: W

First, check for dominant strategies. You’ll find there is none. Do the circle trick or write them out.

- | | |
|--|---------------|
| What is the manager’s best response if the worker works? | Don’t monitor |
| What is the manager’s best response if the worker shirks? | Monitor |
| What is the worker’s best response if the manager monitors? | Shirk |
| What is the worker’s best response if the manager doesn’t monitor? | Work |

Now, check for Nash equilibrium. There are none. Don’t believe me? Pick any combination, for example, the worker monitors and the worker works. Is it a Nash equilibrium? Given the worker is working, manager’s best would be don’t monitor – the manager would want to change their behavior. You’ll find if you pick any other combination, it too, will not be a Nash equilibrium.

How then do we “solve the game”? What would be the outcome? The best answer here is that the boss should randomly monitor the worker. If the worker knows when the boss will be monitoring, the monitoring won’t be effective. However, if the worker doesn’t know the boss will be monitoring, the effort level may increase all the time.

⁶ I’m pretty sure this is the same way we wrote it down in class, but it is still the same game. Even though the players and strategies may appear in a different order in your notes, it won’t change the results.

When randomness is involved, the solution is called a **mixed strategy equilibrium** (as opposed to a pure strategy equilibrium).

What does this have to do with the real world? I'd argue that monitoring employees is pretty real world. DUI checkpoints? Auditing theory? In the sports world? Poker? Penalty kicks in soccer? Where to serve in tennis? Passing or throwing on first down?

Infinitely Repeated Games

Thus far, we have seen a number of prisoner's dilemma style games where the "cheating" outcome prevailed even though there was a "cooperative" outcome that would have been better for the participants. However, this has occurred in one-shot games. We'd like to change the story now by having the same players repeatedly interact with one another.

When we play repeated games, we assume that the game is played once a time period (once a year). Time period 0 refers to the current period, time period 1 refers to the period 1 year from today, and so on. As I'm sure you know, dollars in the future are not the same as dollars today, so we'll have to divide by a discount factor.

Firms will want to maximize the discounted present value of payoffs to the firm. Letting π stand for the payoff, i stand for the interest rate, and subscripts numbering the year, we have:

$$PV = \pi_0 + \frac{\pi_1}{1+i} + \frac{\pi_2}{(1+i)^2} + \frac{\pi_3}{(1+i)^3} + \frac{\pi_4}{(1+i)^4} + \dots + \frac{\pi_T}{(1+i)^T}$$

Before we get too far, it turns out there will be a handy math trick for us trick – the formula for perpetuity. As long as the payments are the same amount, the first payment comes one year from today, and the payments are repeatedly infinitely (forever), the formula for the present value simplifies:

$$\frac{\pi}{1+i} + \frac{\pi}{(1+i)^2} + \frac{\pi}{(1+i)^3} + \frac{\pi}{(1+i)^4} + \dots + \frac{\pi_T}{(1+i)^T} = \frac{\pi}{i}$$

Again, note that this formula assumes the first payment comes a year from today. If there is a payment now ($t = 0$), we'll have to add that in manually, as it is not included in (π / i) .

Trigger Strategy

We saw above that when prisoner's dilemma games are played as one-shot game, the Nash equilibrium is for each player to cheat. But with repeated games, the player might do better.

In order to discuss trigger strategies, you might find it is helpful to translate your game's options into "cooperating" and "cheating".

In the context of the prisoner's dilemma proper (two people locked up in separate cells), cooperating was to not confess while cheating was to confess. In the context of the TV cartel game, cooperating was choosing televising few games (high price) while cheating was choosing many games (low price).

With repeated games, there is a possibility the players can do better than repeatedly choosing to cheat. One such possibility is a **trigger strategy**. The basic idea of trigger strategy is to choose the cooperative strategy so long as your opponent also plays the cooperative strategy. If, however, at any time your opponent plays the non-cooperative strategy, this "triggers" you to switch to the "cheating" strategy, forever. You never return to the cooperative strategy.

Figuring Out If Trigger Strategies Are Profitable:

Begin by assuming that each player has agreed to play the trigger strategy. Now think about if cooperating is an equilibrium strategy? That is, does anyone have the incentive to change his or her strategy?

Either you can continue to play the cooperative strategy, or you can cheat.

If you continue to play the cooperative strategy, you will repeatedly get the payoff associated with the cooperative outcome.

In the TV game, the payoffs (in millions) from always cooperating would be:

$$PV_{COOPERATE} = \pi_0 + \frac{\pi_1}{1+i} + \frac{\pi_2}{(1+i)^2} + \frac{\pi_3}{(1+i)^3} + \frac{\pi_4}{(1+i)^4} + \dots + \frac{\pi_T}{(1+i)^T}$$

$$PV_{COOPERATE} = \$10 + \frac{\$10}{1+i} + \frac{\$10}{(1+i)^2} + \frac{\$10}{(1+i)^3} + \frac{\$10}{(1+i)^4} + \dots + \frac{\$10_T}{(1+i)^T}$$

The payoffs from cheating immediately would be:

$$PV_{CHEATNOW} = \pi_0 + \frac{\pi_1}{1+i} + \frac{\pi_2}{(1+i)^2} + \frac{\pi_3}{(1+i)^3} + \frac{\pi_4}{(1+i)^4} + \dots + \frac{\pi_T}{(1+i)^T}$$

$$PV_{CHEATNOW} = \$20 + \frac{\$5}{1+i} + \frac{\$5}{(1+i)^2} + \frac{\$5}{(1+i)^3} + \frac{\$5}{(1+i)^4} + \dots + \frac{\$5}{(1+i)^T}$$

It will be optimal to cheat if $PV_{CHEATNOW} > PV_{COOPERATE}$. If you compare the two, you can see that there is a one time game from cheating (\$10 in the first period), but the penalty is \$5 less in profits beginning in one year from today and lasting forever (as cooperating no longer occurs.)

So it is optimal to cheat so long as:

$$\$10 > \frac{\$5}{1+i} + \frac{\$5}{(1+i)^2} + \frac{\$5}{(1+i)^3} + \frac{\$5}{(1+i)^4} + \dots + \frac{\$5}{(1+i)^T}$$

Using the math trick for the formula of perpetuity, this boils down to cheat as long as:

$$\$10 > \frac{\$5}{i}$$

For example, with an interest rate of 10%, $i = 0.10$, $\$5 / i = \$5 / 0.10 = \$50$. So it would not be wise to cheat.

Back to the big picture.

We started by assuming that each player would play the trigger strategy. If they did so, they'll want to continue to cooperate so long as the cost of cheating is higher than the benefit of cheating. And with the trigger strategy (if you cheat one time I'll never cooperate again), the cost of cheating is steep. As long as this is the case, each player will find it optimal to cooperate, and they'll do better than the one-shot Nash equilibrium which was to cheat.

At this point, let's generalize. Players repeatedly playing a prisoner's dilemma game are more likely to cooperate:

- The lower the one time gain from cheating
- The higher the future payoff losses from switching to cheating
- The lower the interest rate

Essentially, each thing on that list makes it less likely that the payoff to cheating is higher than the cost of cheating.

And again, having a league or the NCAA enforce the cartel will make it much more likely.

Repeated games with a known end

If the end period of the game is known with certainty, the game can unravel to the extent that neither player will ever cooperate.

The reason cooperation works in the infinitely repeated game is because there is a cost to cheating. When someone deviates from the trigger strategy, they receive the one-time game from doing so. But the deviation results in the loss of all the incremental payments that would have occurred if they cooperated instead of cheated. That is, they lose the difference between the present value of the cooperative outcome forever and the present value of the cheating outcome forever. With low enough interest rates, the cooperative outcome can occur.

Consider a game that will be played exactly 5 times. You might think that a trigger strategy might work. But consider the incentives in the 5th (last period).⁷ In the 5th period, there is no future. There is no cost to cheating, as there is no future cooperation to consider. At this point, the game is a one-shot game, and each will cheat in the 5th period.

But now, consider what happens in the 4th period. Given that each player knows the other will cheat in the 5th period, is there any cost to cheating in the 4th period? Again there is "no future" and there is no cost to cheating in the 4th period. So each player will then cheat in the 4th period. And then the 3rd period. And it continues on up to the 1st period. So the story is, in a game with a known end, cooperation is very unlikely to occur.

Some of you won't buy it – suggesting that cooperation might occur. You might be right.

However, I will say this. The trick on sequential games is to go backwards. What I can say is that the sub-game perfect equilibrium in a repeated game with known end is to cheat every period (in the context of a trigger strategy).

Does this have anything to do with firms firing employees who give their two weeks notice? If you are not offended by somewhat crude hand gestures, click on the link below to see my t-shirt. Way too much build up for you to be impressed now. Don't forget about T.

[My t-shirt idea](#)

⁷ In game theory notation, the last period is usually indicated by a T (a capital T). Don't forget this point if you look at my t-shirt!

Repeated games with unknown end

One more extension.

The story on the repeated games was that if the last period is known, each player will have the incentive to cheat in the last period, and thus have the incentive to cheat in the period before that, and so on, until the game unravels and each person cheats throughout. Remember, to solve sequential games, you need to work backwards. The sub-game perfect equilibrium is to cheat the whole time.

A caveat. A game theoretician would tell you that this is the equilibrium in the game. As I said, my undergraduates tend to cooperate for a while, but they usually pull the trigger before the last period. I'm not convinced that people would unravel to the extent to which the game theory guys say they would. But nonetheless, the idea is meaningful. And maybe some time, we can talk about tit-for-tat. But I digress...

On to the new part...

There is another way to get cooperation without having the game infinitely repeated you might find appealing.

Instead of playing the game once a year for ever and ever, we can imagine a situation where we play today, and there is a probability (less than one) that we play again tomorrow. So instead of Coke and Pepsi being around forever, we can say there is a 0.995 probability that the two companies will be around tomorrow and play again. That means there is a 0.995^2 probability they will be around the year after that.

So long as there is a sufficiently high probability that the game will be played again, we can get cooperation. Repeating, even if we know there is say a 1% chance the game will not be played next period, we can still get cooperation.

The reason this make sense is because with a probabilistic end to the game, the players don't *know* exactly when the game will end (until after it actually ends). When they are in the last period, they don't *know* they are in the last period. And thus they can't cheat in the last period.

Perhaps that is more realistic and more intuitive than thinking of infinitely repeated games. I don't like thinking about infinite amounts of times. Coke and Pepsi won't be around until the end of time, but there is a very high probability they will both be around text time.

The discounting we had in our problem $1 / (1 + i)$ isn't that much different from the probability the game will continue. In fact, instead of having to calculate the discounted present value of the future games, we could instead just calculate the expected value of the future payments - the payments multiplied by the probability they occur. They are both numbers less than one, and as we go out further into the future, they are going to be raised to higher and higher powers.

Just as when the interest rate rises it makes players less likely to cooperate (because it "shortens" the future or makes those future payments less valuable), when the probability the game will *not* be repeated increases (it "shortens" the expected future or makes those future games less likely), players will be less likely to cooperate.

As the probability the game will end becomes 1 (the probability the game continues is zero), we are back to a one-shot game and cheating becomes certain in (the only) first period.

Last thing

Go watch the princess bride for yet another lesson on game theory. Recall that scene about which glass of wine has the poison?