Differences-in Differences (*Diff-in-Diff*): *Two applications*

I: Crime at the Border

Gavrilova, Evelina and Kamada, Takuma and Zoutman, Floris, ***Is Legal Pot Crippling Mexican Drug Trafficking Organizations? The Effect of Medical Marijuana Laws on US Crime*** (December 27, 2014). Available at SSRN: https://ssrn.com/abstract=2350101 or http://dx.doi.org/10.2139/ssrn.2350101 or http://www.cmaxxsports.com/ec228/news.html

Data: MML adoption timeline; violent crime/100,000: 1994-2014, by county by year

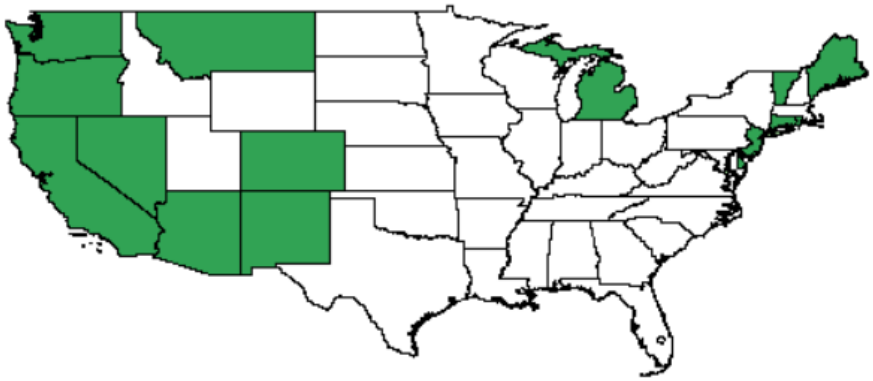


Figure 1: Map of Medical Marijuana Laws

Notes: This graph shows the states in which MML have been introduced. Not shown are Alaska and Hawaii, which have also introduced MML. Dark shade corresponds to states that have introduced MML until the end of 2012, while light shaded are state that have introduced MML after the beginning of 2013.

Table 1: Medical Marijuana Laws

State	Date Active	Home Cultivation	Dispensaries	Dispensaries Open	Number of Dispensaries per 100,000
Alaska	04.03.1999	Yes	No	No	NA
Arizona	14.12.2010	Yes	Yes	2012[a]	0.42
California	06.11.1996	Yes	2004	1997[b]	5.11
Colorado	01.06.2001	Yes	2009	2009[a]	8.79
Connecticut	01.10.2012	No	No	No	NA
DC	27.07.2010	No	Yes	No	NA
Delaware	01.07.2011	No	Yes	No	NA
Hawaii	28.12.2000	No	No	No	NA
Maine	22.12.1999	Yes	2009	2011[a]	0.82
Michigan	04.12.2008	Yes	No	2010[a]	0.85
Montana	02.11.2004	Yes	No	2009[a]	1.27
Nevada	01.10.2001	Yes	No	2011[a]	0.07
New Jersey	18.07.2010	No	Yes	2012[a]	0.07
New Mexico	01.07.2007	Yes	Yes	2009[c]	0.62
Oregon	03.12.1998	Yes	No	2010[a]	1.56
Rhode Island	03.01.2006	Yes	2009	No	0.47
Vermont	01.07.2004	Yes	2011	No	NA
Washington	03.11.1998	Yes	No	2010[a]	1.88

Notes: The Table presents MML and their specific provisions up to the year 2012. The second column presents the date the law became active, the third column shows whether there is a statewide allowance for home cultivation, the fourth column gives the same information about dispensaries, the fifth column shows the date when the first licensed dispensary opened, and the final column gives the number of dispensaries per 100,000 inhabitants in each states. "No" means that the original MML does not allow for the feature in question, while "Yes" means that it does. Whenever some feature is allowed in a later amendment to original law the year is given. For example, in California MML became active in 1996. Home cultivation was immediately allowed, while dispensaries were not allowed statewide until 2004. 1997 is the date in which the first licensed dispensary opened. All information except the final two columns comes from `procon.org`. For the fifth column the sources are listed below. The final column contains self-collected data through the website `findthebest.com` on January 26th 2014.

[a] Source: Anderson and Rees (2014)

[b] Source: Novack (2012)

[c] Source: US Department of Justice (2013)

Differences-in-Differences: ***Diffs-in-Diffs***

@ the Border: Focus on border counties (county centroid < 40 miles from the US-Mex border)

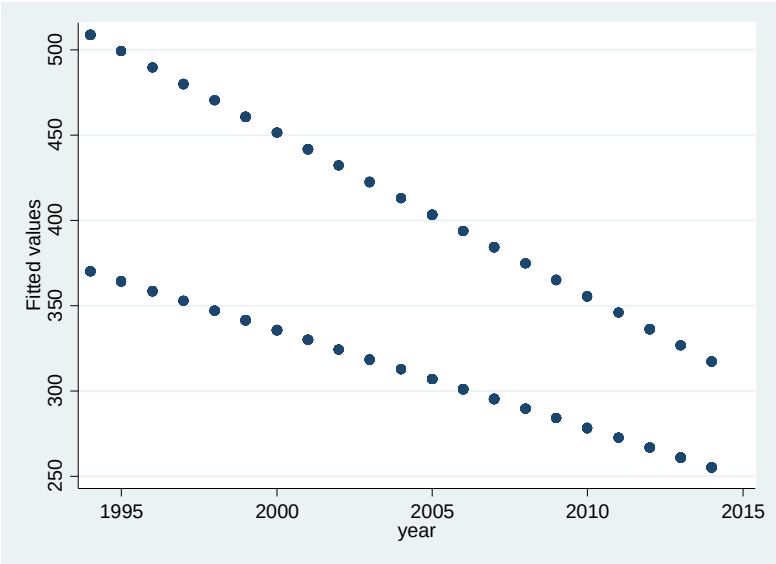
- **Arizona** (4): Cochise, Pima, Santa Cruz and Yuma
- **California** (2): Imperial and San Diego
- **New Mexico** (3): Doña Ana, Luna and Hidalgo
- **Texas** (16): Brewster, Cameron, Dimmit, El Paso, Hidalgo, Hudspeth, Jim Hogg, Kinney, Maverick, Presidio, Starr, Terrell, Val Verde, Webb, Willacy and Zapata

General violent crime trend: 1994-2014

```
. areg viol2k100 trend if bordercty==1, absorb(county)

------------------------------------------------------------------------------
   viol2k100 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       trend |  -7.120794   1.024426    -6.95   0.000    -9.133504   -5.108084
       _cons |   419.9443     11.976    35.07   0.000     396.4148    443.4738
-------------+----------------------------------------------------------------
```

Texas vs. California, Arizona, and New Mexico: 1994-2014

```
. gen trend = year-1994
. gen texas = (state=="Texas")
. gen txtrend = texas*trend
. reg viol2k100 texas trend txtrend if bordercty==1

      Source |       SS           df       MS      Number of obs   =       525
-------------+----------------------------------   F(3, 521)       =     17.64
       Model |  2257720.95         3  752573.648   Prob > F        =    0.0000
    Residual |  22221556.3       521  42651.7395   R-squared       =    0.0922
-------------+----------------------------------   Adj R-squared   =    0.0870
       Total |  24479277.2       524  46716.1779   Root MSE        =    206.52

------------------------------------------------------------------------------
   viol2k100 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       texas |  -138.6135   36.25293    -3.82   0.000    -209.8334    -67.3936
       trend |  -9.573893   2.480857    -3.86   0.000    -14.44761    -4.70018
     txtrend |   3.832967   3.101071     1.24   0.217    -2.259174    9.925108
       _cons |   508.6569   29.00234    17.54   0.000      451.681    565.6328
------------------------------------------------------------------------------
```



Differences-in-Differences: ***Diffs-in-Diffs***

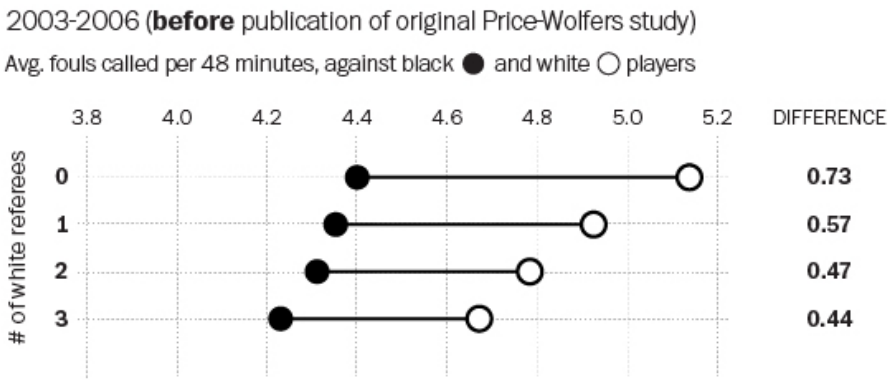
II: Own-Race Referee Bias in the NBA

Joseph Price and Justin Wolfers, ***Racial Discrimination Among NBA Referees,*** The Quarterly Journal of Economics, Volume 125, Issue 4, 1 November 2010, Pages 1859–1887, https://doi.org/10.1162/qjec.2010.125.4.1859, Published: 01 November 2010. Also available: http://users.nber.org/~jwolfers/data.php and http://www.cmaxxsports.com/ec228/ex5v4.html

The Basic Argument: *Difference in Differences*

On average, more fouls are called against white players than against black players, and more fouls are called by referee crews having a higher percentage of black referees. However the spread between the foul rates of white and black players appears to depend systematically on the racial composition of the referee crew. As referee crews skew white (the proportion of white refs increases) the foul rate premium for white players falls… or conversely, the spread widens as the referee crews skew black.

The following chart from the Washington Post is illustrative (note that it focuses on the 2003-06 seasons):[1]



Data: The Price/Wolfers data are at the player/game level, spanning 15,641 games in 14 NBA regular seasons (1991-92 through 2004-05). Their dataset includes the following five general categories of variables:[2]

- **(Regular Season) Game data**: year, date played; game time; attendance; televised; referee lockout dummy (Fall 1995))[3]
- **Player data**: demographics (names, unique player ids, salary, race, foreign born, height, weight, position, age, NBA experience, All Star selections); game stats (starter/substitute, mins played, game stats including called fouls by type of foul); corresponding career stats.

---

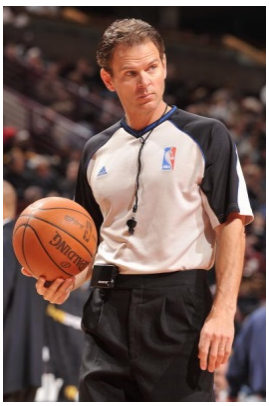
[1] http://www.washingtonpost.com/blogs/wonkblog/wp/2014/02/25/what-the-nba-can-teach-us-about-eliminating-racial-bias/

[2] These are my categorical assignments. I hope I didn't miss anything major.

[3] Note the warning above that the variable **year** in the dataset appears to reflect the year at the start of the season, which is not the general convention.

Differences-in-Differences: ***Diffs-in-Diffs***

- **Referee data**: demographics (names, unique referee ids, race); game stats (which refs were covering which games). (Refs in pics: Leroy Richardson (left) and Mark Ayotte (right))



- **Coach data**: demographics (unique coach ids, race).
- **Team data**: home and visiting teams; scores by quarter; final scores; total stats totals for the game; wins/losses season to date; made playoff dummy; out-of-contention for playoffs; arena capacity.

Price and Wolfers list the following data sources for their data:

- player/game performance stats (minutes played; fouls; points; blocks; steals; etc.): box scores
- game referees: box scores
- referee race: inspection of photographs; assisted by a former NBA referee
- player race: two papers;[4] visual inspection of past issues of the *Official NBA Register*[5] as well as images posted to www.nba.com
- player's height, weight and position: www.basketball-reference.com
- game characteristics (including home team and attendance): box scores
- team characteristics (including coach's race): *Official NBA Register*

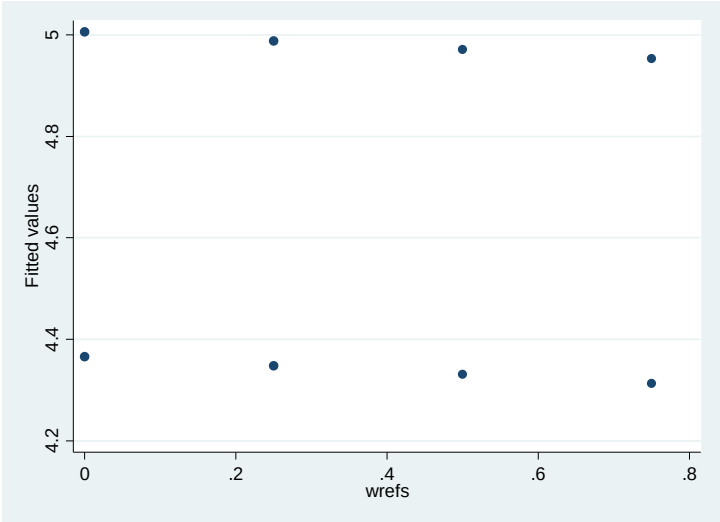
Case I: Same slopes for black and white players

```
. reg foulrate bplayer wrefs [aw=min] if sample == 1
(sum of wgt is   6.4293e+06)

      Source |       SS           df       MS      Number of obs   =   266,984
-------------+----------------------------------   F(2, 266981)    =    691.49
       Model |  15349.9933         2  7674.99665   Prob > F        =    0.0000
    Residual |   2963260.2   266,981  11.0991426   R-squared       =    0.0052
-------------+----------------------------------   Adj R-squared   =    0.0051
       Total |  2978610.19   266,983  11.1565538   Root MSE        =    3.3315

------------------------------------------------------------------------------
    foulrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     bplayer |  -.6397623   .0172337   -37.12   0.000      -.67354   -.6059846
       wrefs |  -.0701963   .0330753    -2.12   0.034     -.135023   -.0053697
       _cons |   5.005556   .0230014   217.62   0.000     4.960474    5.050638
------------------------------------------------------------------------------

. predict fhat0
```

[4] Timmerman (2000) and Kahn and Shah (2005)

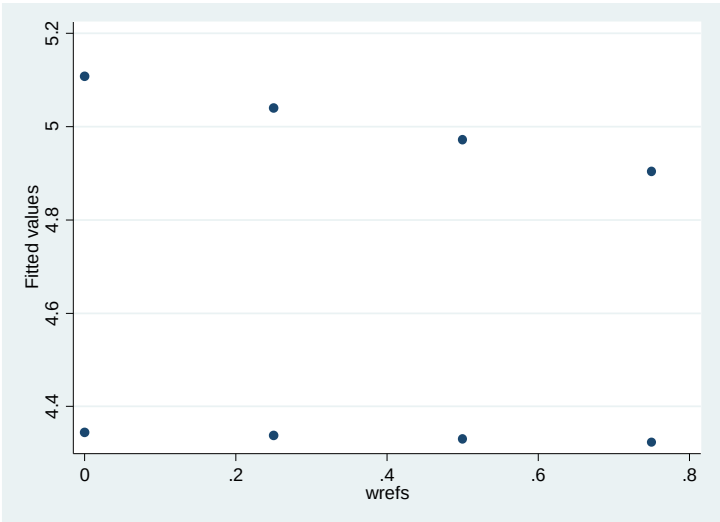
[5] You can access the current editions of the NBA Guide and NBA Register here: http://www.nba.com/news/nba-register-and-nba-guide

Differences-in-Differences: *Diffs-in-Diffs*



Case II: Allow for different slope… for diff-in-diffs

```
. gen b_wrefs=bplayer*wrefs
. reg foulrate bplayer wrefs b_wrefs [aw=min] if sample == 1
(sum of wgt is   6.4293e+06)

      Source |       SS           df       MS      Number of obs   =   266,984
-------------+----------------------------------   F(3, 266980)    =    463.55
       Model |   15434.489         3  5144.82966   Prob > F        =    0.0000
    Residual |  2963175.71   266,980  11.0988677   R-squared       =    0.0052
-------------+----------------------------------   Adj R-squared   =    0.0052
       Total |  2978610.19   266,983  11.1565538   Root MSE        =    3.3315

------------------------------------------------------------------------------
    foulrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     bplayer |  -.7633373   .0479883   -15.91   0.000     -.857393   -.6692815
       wrefs |  -.2722478   .0803521    -3.39   0.001    -.4297357   -.1147599
     b_wrefs |   .2432698   .0881678     2.76   0.006     .0704632    .4160763
       _cons |    5.10815   .0437219   116.83   0.000     5.022456    5.193844
------------------------------------------------------------------------------.
predict fhat1
```



Differences-in-Differences: ***Diffs-in-Diffs***

Testing for Robustness: Add in referee, player and year fixed effects

1. Basic model

```
. qui: reg foulrate bplayer wrefs b_wrefs[aw=min]
```

2. Add referee effects to basic model

```
. qui: reg foulrate rid* bplayer b_wrefs [aw=min]
```

3. Add player effects to basic model

```
. qui: areg foulrate wrefs b_wrefs [aw=min], absorb(player)
```

4. Add referee and player effects to basic model

```
. qui: areg foulrate rid* b_wrefs [aw=min], absorb(player)
```

5. Add in year effects to previous model

```
. qui: areg foulrate rid* i.year b_wrefs  [aw=min], absorb(player)
```

```
. esttab , r2 ar2 scalar(rmse) keep(b_wrefs) compress
```

Note: These results differ slightly from the results above… not sure how that happened.

	(1)	(2)	(3)	(4)	(5)
	foulrate	foulrate	foulrate	foulrate	foulrate
b_wrefs	**0.182****	**0.184****	**0.205*****	**0.200****	**0.198****
	(2.76)	(2.78)	(3.31)	(3.25)	(3.21)
N	266984	266984	266984	266984	266984
R-sq	0.005	0.009	0.164	0.167	0.168
adj. R-sq	0.005	0.008	0.160	0.163	0.164
rmse	3.331	3.326	3.061	3.055	3.053
Fixed Effects					
Referees	**No**	**Yes**	**No**	**Yes**	**Yes**
Players	**No**	**No**	**Yes**	**Yes**	**Yes**
Years	**No**	**No**	**No**	**No**	**Yes**

Weighted by minutes played.

.