Economics 312
Final Project
Dean Young

Factors that Drive Unemployment

Introduction

This project was done as a final project for Professor Jeff Parker's Theory and Practice of Econometrics course at Reed College (Spring 2014). The goal of this project was to determine some of the variables that can help explain state level unemployment rates obtained by the Bureau of Labor Statistics through the Current Population Survey.

Data

The data obtained for this project has been spliced together from many sources into one data set. The majority of the data was obtained from Professor Jeff Parker who obtained some of his data from Professor Jon Rork. State level corporate tax rates were spliced together from data in Tax Foundation as well as data in the tax database at University of Michigan. The data set was then further culled into a panel of the lower 48 states from 1991 to 2010 in order to form a balanced panel. This final data set is named finalproject.dta.

Variable	Units	Description
ed_hs	Percent	Pop. Share with >= HS education
ed_coll	Percent	Pop. Share with >= college education
remw	Dollars	Real (adjusted by CPI) effective minimum wage
per_18to24	Percent	Pop. Share that are between ages 18 and 24
per_25to64	Percent	Pop. Share that are between ages 25 and 64
tot_permem	Percent	Pop. Share of labor force that are members in a union
urbanization	Percent	Urbanization of state
cyclical	Percent	Projected growth in industry based cyclical employment growth
salesrate	Percent	Sales tax rate
cigrate	Percent	Cigarette tax rate
gasrate	Percent	Gas tax rate
topinmtr	Percent	Top marginal personal income tax rate
topcorpmtr	Percent	Top marginal corporate income tax rate
pub_per_emp	Percent	Pop. Share of labor force that are employees of the public sector
urate	Percent	Unemployment rate

To use OLS or not to use OLS?

Given the nature of my panel data, I would not expect a simple OLS model to be sufficient. I do not expect California to have the same dynamics as Rhode Island nor would I expect 1991 to have the same dynamics as 2010. Given that each unit and year probably has its own unique

characteristics not measured by my independent variables, it would seem appropriate to use both unit and year fixed effects which would give me a unique intercept for each year and state. I further assume that the parameters are the same for every year and state in order to save many degrees of freedom. With fixed effects, one would hope that the autocorrelation that inevitably exists in this type of data would be eliminated. A Wooldridge test for autocorrelation in panel data (Wooldridge 2010) shows that there is strong evidence that autocorrelation exists in the data.

```
Wooldridge test for autocorrelation in panel data
H0: no first order autocorrelation
    F(  1,      47) =    742.698
           Prob > F =      0.0000
```

However, I was unable to find a test that tests for autocorrelation in conjunction with fixed effects so I will compare both the OLS pooled data model with the fixed effects model with cluster robust standard errors below. The cluster robust standard errors create 48 clusters (one for every state) and tell Stata that there is autocorrelation through time for each state. The coefficients and t-statistics (in parentheses) have been omitted for the unit and year dummies. This will be the case for all fixed effects Stata outputs shown in this project to save room.

	Urate (OLS)	Urate (Cluster Robust)
ed_hs	-0.004	-0.004
	(0.49)	(0.46)
ed_coll	-0.017	-0.017
	(0.93)	(0.66)
remw	-0.128	-0.128
	(1.77)	(1.12)
per_18to24	0.574	0.574
	(9.21)**	(4.83)**
per_25to64	-0.121	-0.121
	(2.72)**	(1.36)
tot_permem	-0.093	-0.093
	(3.78)**	(2.04)*
urbanization	0.175	0.175
	(5.17)**	(2.27)*
cyclical	-0.840	-0.840
	(4.11)**	(3.27)**
salesrate	0.110	0.110
	(1.36)	(0.79)
cigrate	-0.001	-0.001
	(0.67)	(0.30)
gasrate	0.015	0.015
	(2.06)*	(0.79)
topindmtr	0.137	0.137

	(3.87)**	(1.88)
topcorpmtr	0.105	0.105
	(3.01)**	(0.99)
pub_per_emp	0.078	0.078
	(3.23)**	(1.91)
_cons	-7.030	-9.524
	(2.04)*	(1.24)
R^2	0.85	0.79
N	955	955

* $p<0.05$; ** $p<0.01$

The cluster robust standard errors do change the hypothesis testing for a few coefficients, indicating that there may be autocorrelation even after unit and year dummies are added. OLS is still consistent even with autocorrelation but the standard errors are biased and inconsistent, making hypothesis testing invalid. Using cluster robust standard errors fixes that problem. I will use the cluster robust standard errors in conjunction with the unit and year dummies.

```
Fixed-effects (within) regression               Number of obs      =       955
Group variable: fips                            Number of groups   =        48

R-sq:  within  = 0.7906                         Obs per group: min =        19
       between = 0.0180                                        avg =      19.9
       overall = 0.1913                                        max =        20

                                                F(33,47)           =     96.09
corr(u_i, Xb)  = -0.8076                        Prob > F           =    0.0000

                                 (Std. Err. adjusted for 48 clusters in fips)
------------------------------------------------------------------------------
             |               Robust
       urate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       ed_hs |  -.0042753   .0092578    -0.46   0.646    -.0228997    .0143491
     ed_coll |  -.0173935   .0261573    -0.66   0.509    -.0700152    .0352283
        remw |  -.1277127   .1135877    -1.12   0.267    -.3562217    .1007962
  per_18to24 |     .57385   .1188792     4.83   0.000     .3346958    .8130042
  per_25to64 |  -.1211157   .0893603    -1.36   0.182    -.3008853     .058654
  tot_permem |   -.093459   .0457187    -2.04   0.047    -.1854331   -.0014848
urbanization |   .1752099   .0773002     2.27   0.028      .019702    .3307178
    cyclical |  -.8398345   .2564532    -3.27   0.002    -1.355752   -.3239173
   salesrate |   .1103411   .1390064     0.79   0.431    -.1693037     .389986
     cigrate |  -.0006356   .0021278    -0.30   0.766    -.0049162    .0036449
```

```
     gasrate |   .0150542   .0189814     0.79   0.432    -.0231315    .0532399
   topindmtr |   .1374492    .073071     1.88   0.066    -.0095507    .2844491
  topcorpmtr |   .1048417   .1063636     0.99   0.329    -.1091342    .3188176
 pub_per_emp |   .0775697   .0406688     1.91   0.063    -.0042454    .1593847
       _cons |  -9.523835   7.661121    -1.24   0.220    -24.93602    5.888352
-------------+----------------------------------------------------------------
     sigma_u |  2.4887502
     sigma_e |   .7361973
         rho |   .9195372   (fraction of variance due to u_i)
```

Selecting the important variables

Before continuing, I would like to comment on the resulting hypothesis tests and coefficients. I am not surprised that percentage of HS graduates matter because nowadays, any job you can get as a HS graduate, you can get as a drop out. However, I am more surprised that percentage of college graduates is not statistically significant. Some of the recent literature examining income inequality claims that the wage premium to college graduates due to shortage in supply is the main reason for widening inequality (Goldin et al. 2009). This insignificant coefficient seems to dispute that claim. It is possible that college graduates are simply just more picky about choosing the right job than non-college graduates however.

It is also possible that percentage of HS and college graduates are simply not stationary over time. I would expect that unemployment rate is stationary. It is very difficult to explain a stationary variable with a non-stationary variable. However, a Harris-Tzavalis unit-root test that tests for the value of rho indicates that both variables are indeed stationary (Harris et al. 1999). A value of rho that is less than one indicates if a variable is stationary.

```
Harris-Tzavalis unit-root test for ed_hs
----------------------------------------
Ho: Panels contain unit roots               Number of panels  =     48
Ha: Panels are stationary                   Number of periods =     20

AR parameter: Common                        Asymptotics: N -> Infinity
Panel means:  Included                                   T Fixed
Time trend:   Not included                  Cross-sectional means removed
--------------------------------------------------------------------------------
                     Statistic         z         p-value
--------------------------------------------------------------------------------
 rho                   0.0756      -36.6045       0.0000
--------------------------------------------------------------------------------

. xtunitroot ht ed_coll, demean

Harris-Tzavalis unit-root test for ed_coll
```

```
----------------------------------------
Ho: Panels contain unit roots               Number of panels  =     48
Ha: Panels are stationary                   Number of periods =     20

AR parameter: Common                        Asymptotics: N -> Infinity
Panel means:  Included                                   T Fixed
Time trend:   Not included                  Cross-sectional means removed
--------------------------------------------------------------------------------
                     Statistic         z         p-value
--------------------------------------------------------------------------------
 rho                   0.4645      -18.3908       0.0000
```

Real effective minimum wage did not have a statistically significant effect. This was slightly surprising, as I would think higher real wages without an equal increase in productivity would lead to a deadweight loss where there is more demand for work than there is supply. The population share that was between the ages of 18 and 24 was statistically significant and the coefficient was positive. I suspect that the reason for that is because those between the ages of 18 and 24 are starting to enter the workforce but have not built up enough experience, connections, and motivation to obtain jobs quickly. Percentage of labor force in unions had a statistically significant negative coefficient. This is not surprising since unions tend to protect jobs and decrease unemployment. However, I would argue that this is not an efficient way to lower unemployment because unions may keep underachievers safe in their jobs. Urbanization had a statistically significant positive coefficient. This is likely explained by the fact that people migrate to cities to search for jobs, not the country. If you live in the country, you probably already have a stable (no pun intended) job working in a farm. Projections in cyclical employment had an extremely statistically significant and economically significant negative coefficient. This may or may not be surprising, depending on how much you believe in the abilities of today's economists. The three consumption tax rates had no significant effect. This is likely due to the fact that the goods sold in one state are produced in a different state. It could also be due to the fact that these rates do not vary much over time, which means the fixed effects model will struggle to find significant coefficients. The top personal marginal income tax rate was significant at the 90% confidence level and had a positive coefficient. I suspect that depending on the elasticity of the supply of labor (for second wage earners it is very elastic), employees may ask for higher wages in order to compensate for higher tax rates which puts a burden on corporations which makes them hire less people. The top corporate marginal income tax rate was not significant, which was a little surprising. This rate also does not vary much over time and so the fixed effects model may not be valid here. I will address this issue later. Finally, the percentage of workforce that are public employees had a significant positive coefficient at the 90% confidence level. This seems to suggest that public sector jobs are less safe than private sector jobs, which is counterintuitive. There may be omitted variable bias in this case. In fact, many of these variables may suffer from omitted variable bias where the omitted variable is unemployment insurance. It is plausible that high tax rates, high minimum wages, union and public employee percentage is correlated with a democratic legislature. Democratic legislature tends to make more generous unemployment benefits which would increase unemployment in theory. A better model would include unemployment benefits for each state. However, for the sake of this project, I must assume there is no omitted variable bias.

After eliminating the variables with coefficients not significant at least at the 90% confidence level, my final fixed effects model is displayed below.

```
Fixed-effects (within) regression               Number of obs      =       960
Group variable: fips                            Number of groups   =        48

R-sq:  within  = 0.7829                         Obs per group: min =        20
       between = 0.0185                                        avg =      20.0
       overall = 0.1652                                        max =        20

                                                F(25,47)           =     96.54
corr(u_i, Xb)  = -0.8385                        Prob > F           =    0.0000

                                  (Std. Err. adjusted for 48 clusters in fips)
------------------------------------------------------------------------------
             |               Robust
       urate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  per_18to24 |   .5979566   .1311129     4.56   0.000     .3341914    .8617217
  tot_permem |  -.1022996   .0503713    -2.03   0.048    -.2036335   -.0009656
urbanization |   .1977776   .0797766     2.48   0.017     .0372878    .3582675
    cyclical |    -.83367    .274529    -3.04   0.004    -1.385951   -.2813889
   topindmtr |   .1812596    .081274     2.23   0.031     .0177575    .3447617
 pub_per_emp |   .0794407   .0397425     2.00   0.051     -.000511    .1593923
       _cons |  -17.77273   6.016247    -2.95   0.005    -29.87586   -5.669603
-------------+----------------------------------------------------------------
     sigma_u |  2.7422367
     sigma_e |   .7474165
         rho |  .93084955   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

I would like to highlight a few coefficients. According to this model, for every 1% increase in percentage of population between the ages of 18 and 24, unemployment increases by 0.6%. That means for every 10 young adult in the labor force, on average, 6 of the 10 will be struggling in the job search at any given time. This does not make me, a young adult, feel very good at all. Also, if there is a 1% decrease in top marginal personal income tax, there is a 0.18% decrease in unemployment. This is a jackpot for conservatives. However, it is likely that top marginal personal income tax correlates with the average income tax, which matters more for unemployment.

Addressing some issues

As noted before, the fixed effects model struggles when variables have little variation over time. The random effects model does a much better job with this. The random effects model finds an average intercept across all 48 states and each state has an additional error term to account for being below or above the average. The model's new error term then is a combination of the residual and the error term when estimating each state's intercept

$$^{2}{}_{1i} = \overline{\beta} + u_i$$

$$v_i = u_i + e_i$$

The random effects model is consistent if x_i and u_i are not correlated. However, in many situations this is the case. The unexplained variables that put a state above or below the average likely have correlation with the explanatory variables I have included in my model. The fixed effects estimator is still consistent even when x_i and u_i are correlated, because the u_i will cancel (HGL 2011). This means a Hausman test is appropriate to test if the random effects model can be used here. The Hausman test compares the coefficients from both fixed and random effect models. If they are significantly different, that indicates that there are endogenous explanatory variables that correlate with u_i. The Hausman test has been displayed below.

```
                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |     fixed        random       Difference          S.E.
-------------+------------------------------------------------------------
       ed_hs |   -.0042753    -.0331171        .0288418        .0050292
     ed_coll |   -.0173935    -.0813472        .0639538        .0146599
        remw |   -.1277127     .2487754       -.3764881        .0509394
  per_18to24 |      .57385     .2028317        .3710184        .0422802
  per_25to64 |   -.1211157    -.0313451       -.0897705        .0440755
  tot_permem |    -.093459     .0203563       -.1138152        .0225403
urbanization |    .1752099     .0298107        .1453992         .033546
    cyclical |   -.8398345    -.3579007       -.4819338        .2036959
   salesrate |    .1103411     .2188228       -.1084816        .0742889
     cigrate |   -.0006356      .004305       -.0049406        .0006269
     gasrate |    .0150542     .0357005       -.0206463        .0040001
   topindmtr |    .1374492      .015544        .1219052        .0286732
  topcorpmtr |    .1048417      .029774        .0750677        .0268778
 pub_per_emp |    .0775697     .0852291       -.0076594        .0198242
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                 chi2(14) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      493.41
                Prob>chi2 =      0.0000
```

I must unfortunately conclude that the random effects model is not valid since the Hausman test rejects the null that the two estimators return the same coefficients. This means I will probably never know if the consumption and corporate taxes truly affect unemployment rates.

One could argue with me here and say that the overarching assumption, that x_i and e_i are not correlated, is also violated. In that case, it doesn't matter if x_i and u_i are correlated, both fixed and random effects models would be inconsistent due to endogeneity. That would mean the Hausman test isn't even valid. Indeed, I would agree. I have already admitted my model has some possible omitted variable bias. There is also likely simultaneity bias, which I will address afterwards. That means if both of my models are inconsistent, there's no point in fussing over the validity of the random effects model in comparison.

Sadly, with suspicion of both simultaneity and omitted variable bias in my fixed effects model, I cannot be confident in its validity. I am unable to fix omitted variable bias without collecting more data so I am forced to simply assume that none of my explanatory variables correlate with unemployment insurance. Simultaneity bias is also a problem. It is possible that economists use projections of unemployment rates to decide on minimum wage and taxes. Also, people might make their migration decisions based on projected unemployment rates which mean the age structure variables may be biased as well. The conventional way to address simultaneity bias is the classic instrumental variable two stage least squares regression. However, I cannot think of a good instrumental variable for tax rates, minimum wage or age structure. Instead, I will restrict my data set to the years 2008 through 2010. This period, as most know, is the Great Recession. This recession was largely unforeseen and had an enormous impact to the economy. That means the predictions in unemployment rates probably are uncorrelated with observed unemployment rates. As long as the predictions aren't systematically off (ex. Always 10% over predicted), there can be no effect from unemployment rate onto minimum wage, taxes, and age structures.

Before I actually show the restricted data set regression, I would like to mention that by restricting my data set to the Great Recession period, I have essentially thrown external validity into the dumpster. The coefficients for those variables are likely very different when the economy is in a recession versus when it is in a boom. We are (hopefully) not going to have another recession of such a magnitude any time soon. This means the results from this restricted data set regression cannot be generalized to just any time period. With that disclaimer, I will now show the results.

```
Fixed-effects (within) regression               Number of obs      =       144
Group variable: fips                            Number of groups   =        48

R-sq:  within  = 0.9514                         Obs per group: min =         3
       between = 0.1732                                        avg =       3.0
       overall = 0.0436                                        max =         3

                                                F(16,47)           =    753.58
corr(u_i, Xb)  = -0.9849                        Prob > F           =    0.0000
```

(Std. Err. adjusted for 48 clusters in fips)

urate	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
ed_hs	-.0162115	.0041505	-3.91	0.000	-.0245613	-.0078616
ed_coll	-.0194148	.0802911	-0.24	0.810	-.1809396	.1421101
remw	.2574532	.4358758	0.59	0.558	-.6194157	1.134322
per_18to24	-.5601022	1.017873	-0.55	0.585	-2.607798	1.487593
per_25to64	-4.550095	.9827703	-4.63	0.000	-6.527174	-2.573017
tot_permem	-.0844783	.0933174	-0.91	0.370	-.2722087	.1032521
urbanization	-.6995981	.4035135	-1.73	0.090	-1.511362	.1121663
cyclical	-.0771433	.3175878	-0.24	0.809	-.7160477	.561761
salesrate	.5939233	.208433	2.85	0.006	.1746102	1.013236
cigrate	.0016327	.0035122	0.46	0.644	-.0054329	.0086982
gasrate	-.0467676	.0330001	-1.42	0.163	-.1131552	.01962
topindmtr	.0562893	.0663695	0.85	0.401	-.0772288	.1898075
topcorpmtr	-.0155587	.0379288	-0.41	0.684	-.0918617	.0607442
pub_per_emp	.1434082	.0626854	2.29	0.027	.0173014	.2695151
_cons	298.0641	72.25575	4.13	0.000	152.7043	443.4239
sigma_u	13.491807					
sigma_e	.49280405					
rho	.99866762	(fraction of variance due to u_i)				

There are some noticeable differences in hypothesis testing between the restricted data set regression and unrestricted data set regression. Not surprising. However, whether the difference is attributed to the change in economic setting or due to elimination of simultaneity bias or a combination of both is not known. What I can be more confident in is that there is unlikely to be simultaneity bias in this model. This is confirmed by the abysmally low t-statistic for cyclical. In the unrestricted data set, cyclical had a strong correlation with unemployment because economists did a good job of predicting the future. In the Great Recession period, projections in the business cycle were abysmally off and therefore projected unemployment rate is likely uncorrelated with true unemployment rate.

Another noticeable difference is that HS grad rate is now very statistically significant. The coefficient indicates that a 1% increase in population proportion with HS degrees leads to a 0.02% decrease in unemployment rate. Small but statistically significant. My suspicion is that because the majority of the people counted in ed_hs have a high school degree but no college degree, they can simply drop out of the labor force and go to/return to college. They won't be counted against the state in terms of unemployment.

A huge difference was in the percentage of population between ages 25 and 64. The coefficient

on that variable went from being very insignificant both economically and statistically, to being very significant both economically and statistically. A 1% increase in percentage of population between ages 25 and 64 leads to a whopping 4.55% decrease in unemployment rate. This could mean several things. Perhaps older workers worked more recession proof jobs. Perhaps older workers had the option of retiring early in the recession. I'm not really sure.

Percentage of the labor force in unions became insignificant, possibly indicating that being in a union didn't shelter you from a recession.

Urbanization was still statistically significant, but the coefficient changed from positive to negative. A 1% increase in urbanization leads to a 0.7% decrease in unemployment. This may suggest that the rural sectors were hit hard by the recession.

Most of the tax variables stayed the same in terms of statistical significance. There were two changes. Sales tax rate became statistically significant. A 1% increase in sales tax leads to a 0.59% increase in unemployment. This may be due to the fact that in a recession, consumers become a lot more sensitive to the price of the items they are consuming. This would mean sales tax would have a big effect since consumption falls and jobs are let go. The top personal marginal tax rate became statistically insignificant. In a recession, I suspect that people's supply of labor becomes extremely inelastic. People will work for any wage in order to put food on the table. That would mean the marginal income tax rate wouldn't have an effect since corporations wouldn't have to increase wages to match the income tax rate.

The final regression with only statistically significant variables (at least meeting the 90% confidence level) is shown below.

```
Fixed-effects (within) regression               Number of obs      =       144
Group variable: fips                            Number of groups   =        48

R-sq:  within  = 0.9488                         Obs per group: min =         3
       between = 0.1743                                        avg =       3.0
       overall = 0.0412                                        max =         3

                                                F(7,47)            =   1253.20
corr(u_i, Xb)  = -0.9826                        Prob > F           =    0.0000

                                    (Std. Err. adjusted for 48 clusters in fips)
------------------------------------------------------------------------------
             |               Robust
       urate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       ed_hs |  -.0140519   .0022504    -6.24   0.000    -.0185791   -.0095246
  per_25to64 |  -4.396261   .7019931    -6.26   0.000    -5.808489   -2.984033
```

```
urbanization |  -.6542978   .2885663    -2.27   0.028    -1.234818   -.0737773
   salesrate |   .5535311   .2178889     2.54   0.014     .1151951    .9918671
 pub_per_emp |   .1426926   .0609776     2.34   0.024     .0200215    .2653637
       _cons |   280.9227   45.51694     6.17   0.000     189.3544    372.4909
-------------+----------------------------------------------------------------
     sigma_u |  12.582348
     sigma_e |  .47956033
         rho |  .99854945   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Conclusions/Validity Assessment

I have settled on two different models, both of which use the fixed effects estimator that estimate both unit dummies and year dummies. The first estimate includes a larger time span from 1991 to 2010. The second estimate only includes the period of the Great Recession.

I have made an overarching assumption that omitted variables bias either does not exist or has very little effect. I have acknowledged that it is possible that by omitting unemployment insurance, my estimates may be inconsistent. However, due to lack of data and for the sake of this project, I will assume that it has no correlation with my included variables.

With that assumption set, choosing between the two models puts me at a small dilemma. The full data set model is generalizable to most time periods on average since it includes both booms and busts. However, I have strong suspicious of simultaneity bias. The restricted data set model eliminates simultaneity bias since projections missed the mark significantly, but cannot be generalized to time periods outside of a major recession. I would like to at least have consistent estimates, so selecting the second model seems appropriate to me.

Within a recession as impactful as the Great Recession, it’s really difficult to tell what factors drive unemployment. It would seem that those who have graduated high school were able to return to school and shelter themselves from unemployment. There is a strong effect of having more people between the ages of 24 and 64. The decrease in unemployment is large and likely due to a few factors such as that age group workings more recession proof jobs and also having the option of retiring early. Higher urbanization led to a decrease in unemployment, possibly indicating a strong hit to the rural areas during the recession. An increase in percentage of public employees led to an increase in unemployment. This suggests that the recession was especially rough on tax revenues, therefore meaning more cut jobs in the public sector. Many of these factors can’t be adjusted much by government intervention, especially in a short enough amount of time to respond to a recession.

The one variable that has an effect during a recession and may possibly be adjusted in due time, is sales tax rates. From the data, it seems that states with higher sales tax rates had higher unemployment. This is likely due to the elasticity of demand for consumption changing due to the recession. I remember during the recession, many economists were talking about how the effects of the recession compound on themselves, making the situation worse. Businesses close, people lose their jobs, people spend less, businesses have less business, businesses close and the cycle continues. If states were to lower sales taxes during recessions, it may be helpful since it would spur consumption and bring the economy back towards the right track.

References

Harris, R. D. F., and E. Tzavalis. 1999.
Inference for unit roots in dynamic panels where the time dimension is fixed. *Journal of Econometrics* 91: 201–226.

Wooldridge, J. M. 2010.
Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.

Goldin, Claudia and Lawrence F. Katz. 2009.
The Future of Inequality: The Other Reason Education Matters So Much*The Milken Institute Review* (Third Quarter) 26-33