Case Study: Wages

Dianne Cook
Iowa State University
Jan 27, 2010

1 Description

The data was collected to track the labor experiences of male high-school dropouts. The men were between 14 and 17 years old at the time of the first survey. This data was compiled [1] from a subset taken from the National Longitudinal Survey of Youth (NLSY) described at `http://www.bls.gov/nls/nlsdata.htm`. The data was collected between 1966 and 1981, on 5225 young men.

The data available are:

Variable	Explanation
id	1–888, for each subject.
lnw	natural log of wages, adjusted for inflation, to 1990 dollars.
exper	length of time in the workforce (in years). This is treated as the time variable, with t_0 for each subject starting on their first day at work. The number of time points and values of time points for each subject can differ.
ged	when/if a graduate equivalency diploma is obtained.
black	categorical indicator of race = black.
hispanic	categorical indicator of race = hispanic.
hgc	highest grade completed.
uerate	unemployment rates in the local geographic region at each measurement time.

The primary question is "How do wages change with workforce experience?"

Secondary questions include: "Does the workforce experience differ by race?" or "Does the unemployment rate have an effect on wages?"

2 Plan

Approach	Reason	Type of questions addressed
Data Restructuring Separate tables created for demographic variables and time-dependent variables Also a new categorical variable for race with levels "white", "black" and "hispanic" created.	To remove redundant information from the time data, and make it clear what information is specific to the subject and what information is collected repeatedly. It will mke it easier to calculate summary statistics and make facetted plots.	
Summary statistics (marginal and conditional)	extract location/scale information	"How many subjects are there?", "How many subjects are black?", "What proprtion of subjects received the graduate equivalency degree?", "What is the range of wages experienced?", "How long is the wage force experience that is studied?"
Time plots	explore overall distribution of wages with time, and conditional on some of the demographic covariates	"Is there a general trend in the wages relative to experience?", "Is there a difference between the trends for different races?"
Scatterplots	explore relationship between time-dependent information	"Is the wage related to unemployment rates in the local region?"
Interactive plots	explore the individual profiles	"What are the differences between individuals in their wages by experience?"
Calculate descriptor variables of time series	organize the subjects profiles according to temporal patterns	"What proportion of the subjects experience increasing wages with experience?", "Are there any subjects who experience a decline in wages, as their experience grows?"
Model fitting	Summarize the association between wages and experience, dependent on the covariates.	"Are wages increasing with experience, generally?"

3 Analysis

3.1 Summary Statistics

There are 888 subjects. The log(wages) range from 0.708-4.304, corresponding to \$2.03-\$74.00, per hour. Experience ranged between 0.001-12.700 years. The number of measurements per subject ranged between 1-13. The correlation between log(wages) and unemployment rate is -0.191.

	white	black	hispanic
count	438	246	204
proportion	0.49	0.28	0.23

Table 1: Counts and proportions for race. Half the subjects are labelled as white, and about a quarter each are described as black or hispanic.

	ged	no ged
count	200	688
proportion	0.23	0.77

Table 2: Counts and proportions for graduate equivalency diploma. About a quarter of the subjects have received a graduate equivalency degree.

	highest grade completed						
	6	7	8	9	10	11	12
count	32	95	239	211	186	110	15
proportion	0.036	0.107	0.269	0.238	0.210	0.124	0.017

Table 3: Counts and proportions for highest grade completed. A reasonable number of subjects have completed between 7-11 years of education, but only a few subjects have 6 or 12 years of education.

	race		
ged	white	black	hispanic
yes	90 (0.21)	68 (0.27)	42 (0.21)
no	348 (0.79)	178 (0.72)	162 (0.79)

Table 4: Cross-tabulation of race and graduate equivalency diploma. Proportionally more blacks have graduate equivalency degrees than white or hispanics.

	race		
hgc	white	black	hispanic
6	15 (0.03)	8 (0.03)	9 (0.04)
7	45 (0.10)	27 (0.11)	23 (0.11)
8	117 (0.27)	66 (0.27)	56 (0.27)
9	122 (0.28)	50 (0.20)	39 (0.19)
10	78 (0.18)	62 (0.25)	46 (0.22)
11	54 (0.12)	28 (0.11)	28 (0.14)
12	7 (0.02)	5 (0.02)	3 (0.01)

Table 5: Cross-tabulation of race and highest grade completed. The distributions of highest grade completed is similar for the three races. Perhaps slightly less blacks reached 9th grade, but proportionally more reached 10th grade.

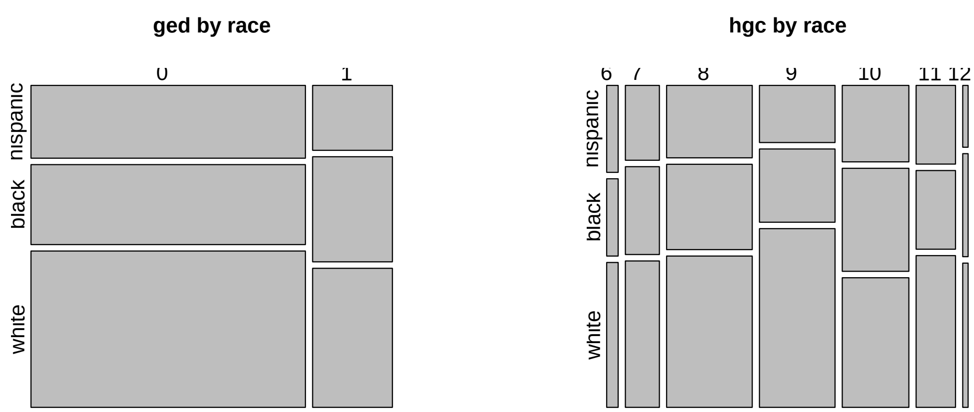


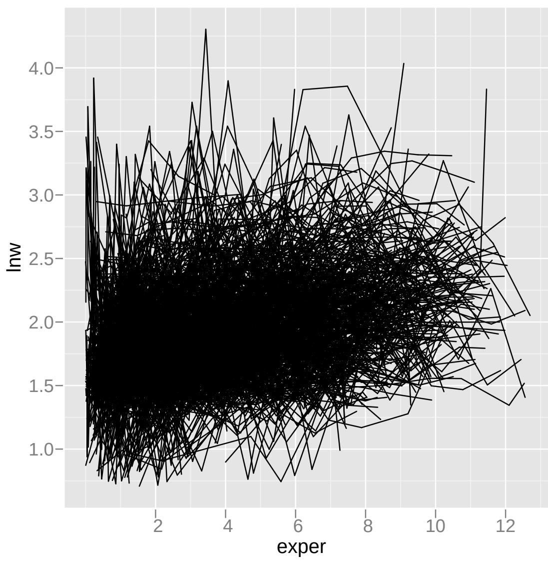
Figure 1: Mosaic plots showing association between race and graduate equivalency diploma (left), and race and highest grade completed (right), reporting the same statistics as Tables 3.1, 3.1. Proportionally more blacks have graduate equivalency degrees than white or hispanics. The distributions of highest grade completed is similar for the three races. Perhaps slightly less blacks reached 9th grade, but proportionally more reached 10th grade.

3.2 Time plots

When the profiles of all 888 subjects are shown together not much can be seen because of the overplotting of the lines (Figure 2). There appears to be a *slight* increasing relationship between log(wage) and experience. A loess smoother (right plot) confirms that there is an increasing trend in wages with more experience.

Exploring the relationship between wage trends and race (Figure 3) indicates that whites and hispanics have similar trends. Black subjects seems to have slightly less association which flattens out around 4 years of experience, but there are also less black subjects with more years of experience.

Figure 4 explores the relationship between wage trends and highest grade completed. The temporal trend appears to increase more as grade completed increases (grades 6 and 12 exempted for lack of data).



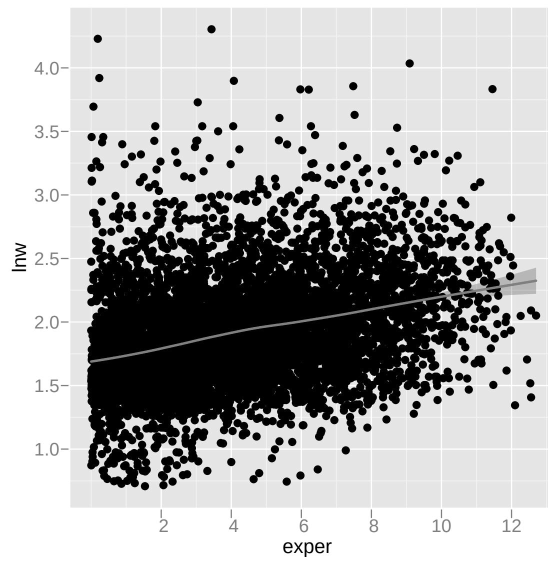


Figure 2: (Left) Profiles of all 888 subjects. Not much can be seen, because there is too much overplotting, but there appears to be a *slight* increasing relationship between log(wage) and experience. (Right) Scatterplot of log(wages) against experience with a loess smoother overlaid. There is an increasing trend in wages with more experience.

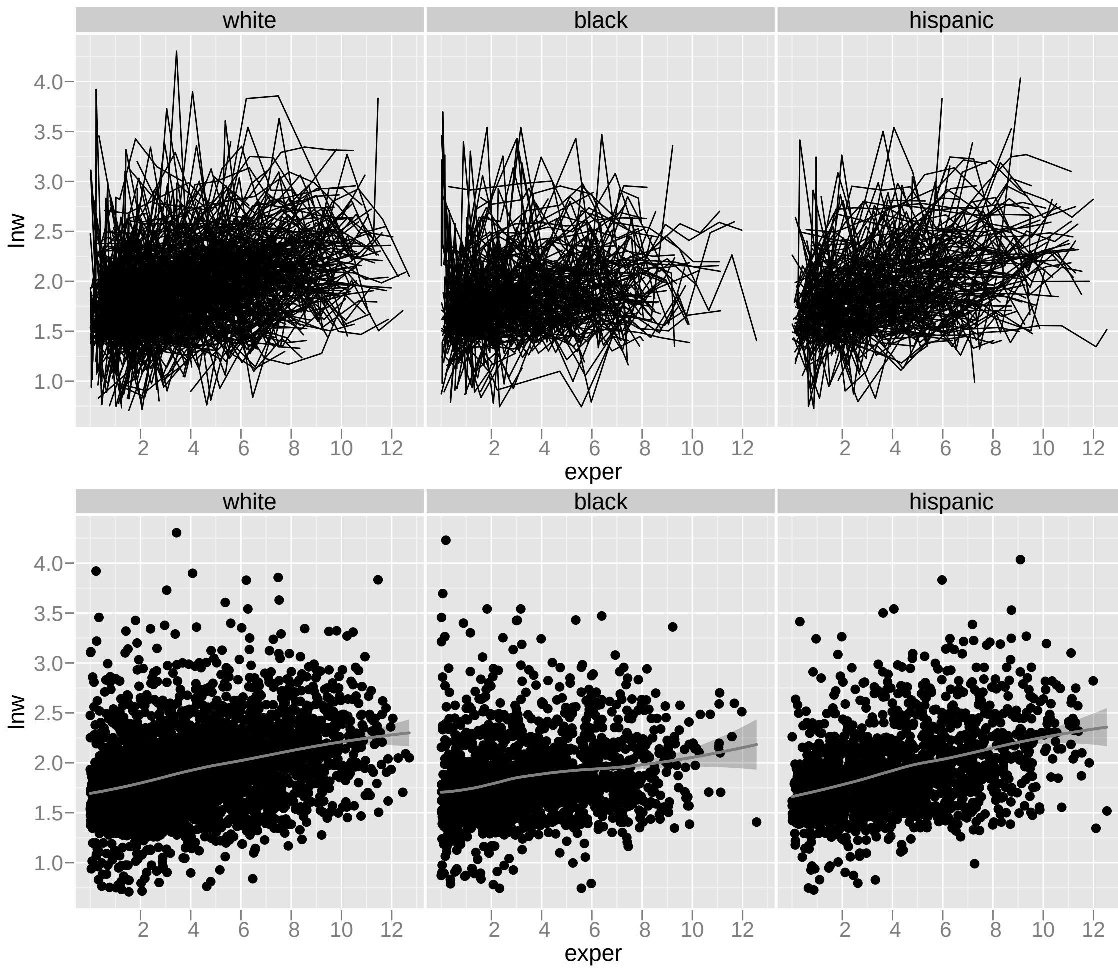


Figure 3: (Top row) Profiles, facetted by race. Again overplotting makes it difficult to see much, but whites and hispanics look similar, while blacks seems to have slightly less association and less subjects with longer experience. (Bottom row) Scatterplot of log(wage) by experience, facetted by race, with loess curves overlaid. Generally wages increase with more experience. The association between log(wages) and experience for whites and hispanics looks similar, while for blacks the temporal trend is flatter for at 4 years and more of experience.

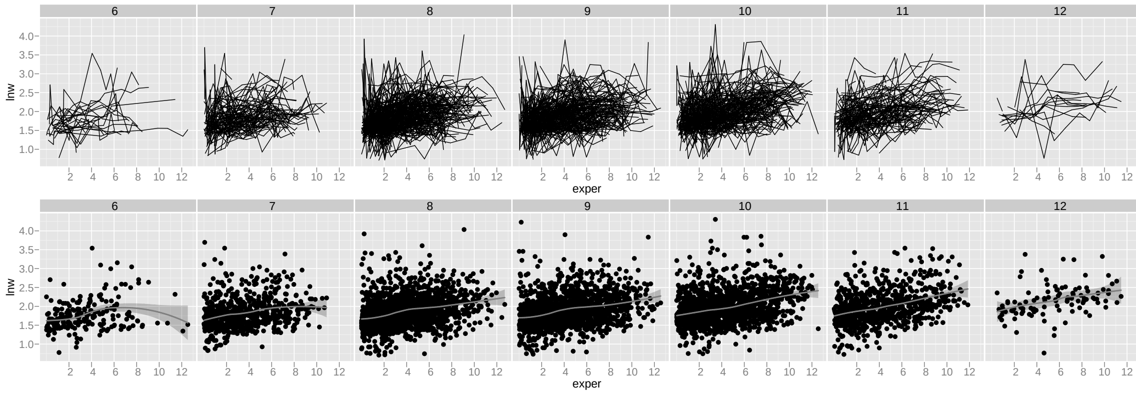


Figure 4: (Top row) Profiles of log(wage) by experience, facetted by highest grade completed. Overplotting makes it difficult to see any trends. (Bottom row) Scatterplot of log(wage) by experience, facetted by highest grade completed, with loess curves overlaid. Temporal trend seems to be increasing more as grade completed increases (grades 6 and 12 exempted for lack of data).

3.3 Scatterplots

Unemployment rate is a time-dependent covariate. For each subject, for each measurement the unemployment rate in the local region was recorded. The relationship between unemployment rate and wages is explored using a scatterplot (Figure 5). There is a very slight negative association which corresponds with the calculated correlation of -0.191. The association is nonlinear, as indicated by the smoother, starting with higher wages, when unemployment is low, which gradually declines as unemployment increases. At some point in unemployment rate the wages are not affected.

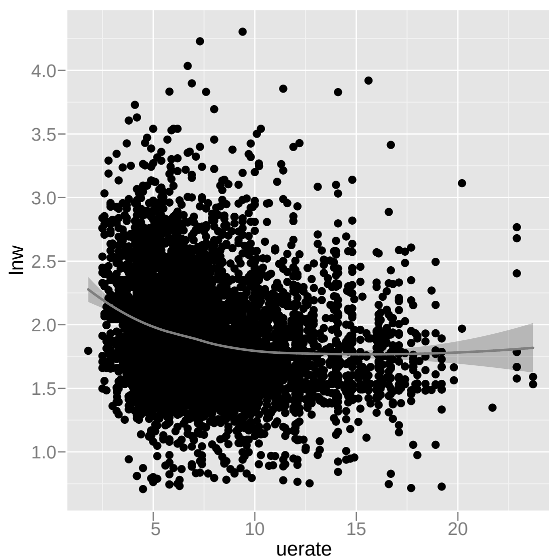


Figure 5: Scatterplot of log(wage) by unemployment rate. A *slight* negative association can be seen: higher unemployment is associated with lower wages.

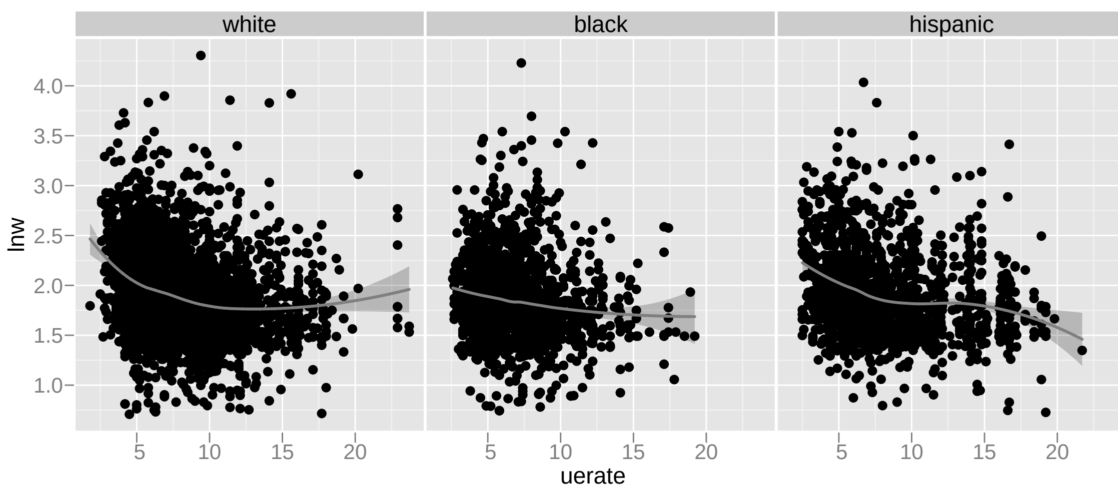


Figure 6: Scatterplot of log(wage) by unemployment rate, facetted by race. Whites and hispanic have higher wages, when unemployment is the lowest. Generally as the unemployment rate increases wages decrease.

3.4 Exploring individuals using interactive plots

To explore the individual variation we use the subject's ID in linked brushing, to explore individual profiles. Moving around the edge of the body of points brushing the outliying points, we notice that there is a lot of difference between individuals: some individuals have a lot of variation in their wages despite increasing experience level, some have large jumps or drops from one time point to another, some experience a drop in wages over time, and a few others have steady wages.

Clearly the average trend is not going to be the most interesting structure in this data. We will want to quantify the variation in the individual profiles. In the next section we calculate a number of simple statistics that might be used to describe the different types of trends.

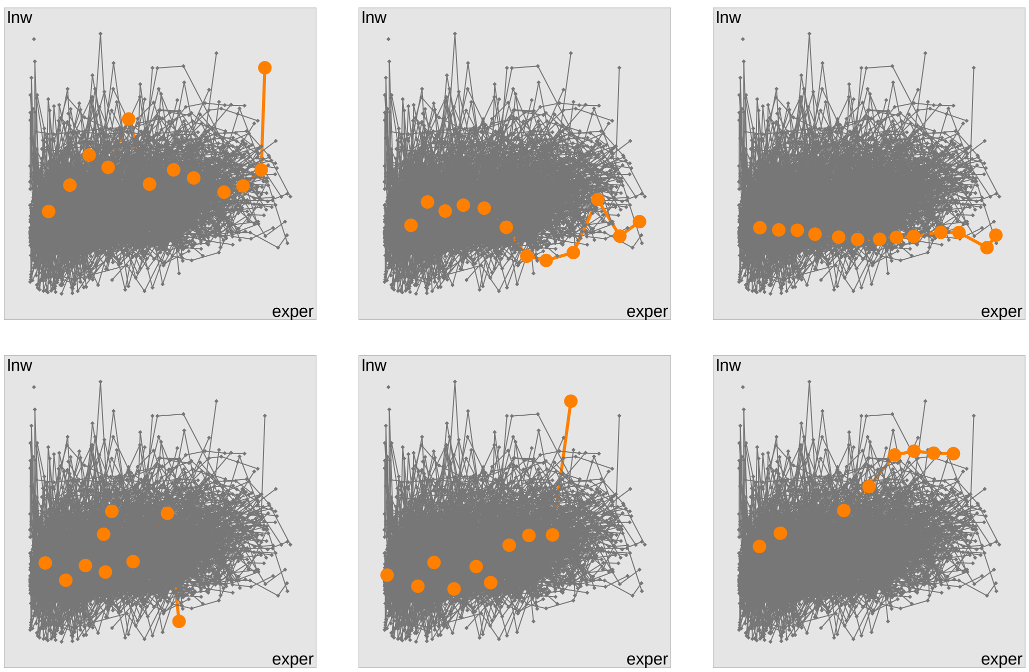


Figure 7: Profiles of six subjects. There's a lot of variability between subjects!

3.5 Descriptor variables

How wages vary with experience is the primary pattern of interest. These descriptor variables are calculated to explore these temporal trends. (The descriptors from the mixed effects linear models are discussed in the next section.)

Number of measurements	count
Average	av
Standard deviation	sd
Minimum	min
Maximum	max
Minimum experience	expmin
Maximum experience	expmax
Start	start
End	end
Average difference between time points	avdif
Standard deviation of difference between time points	sddif
Sum of difference between time points	sumdif
Smallest difference	smalldif
Largest difference	bigdiff
Average difference between time points, adjusted by time	avedif
Standard deviation of difference between time points, adjusted by time	avesd
Sum of difference between time points, adjusted by time	sumedif
Smallest difference, adjusted by time	smalledif
Largest difference, adjusted by time	bigedif
Intercept of linear regression	linint
Slope of linear regression	linear
Intercept of robust linear regression	rlin
Slope of robust linear regression	rlinint
Intercept of mixed effects linear model	lme(2)exper
Slope of mixed effects linear model	lme(2)int
Standard deviation of residuals from mixed effects linear model	lme(2)sd

There are 37 subjects with two or less measurements (Figure 8). These subjects are not used when calculating trend, and their values on these descriptors is set to 0.

The slope for a linear fit for almost all subjects is between -1 and 1 (Figure 8). Many of these subjects, about 400, have slopes between 0 and 0.1. About 300 subjects have negative slopes between -0.1 and 0. Slightly more subjects have positive slopes than negative slopes. The intercepts from the linear fit range between 1-3 for almost all subjects.

The minimum wage on average is about \$4.50 (log(wage)=1.5), and the maximum wages on average is around \$10.00 (log(wage)=2.3) (Figure 9). Starting wages are about \$5.00 (log(wage)=1.6), and ending wages are around \$7.50 (log(wage)=2). Ending wages have a bimodal distribution, with modes around log(wage)=1.6 and log(wage)=2.3.

The distribution of average wages over individuals is fairly symmetric centered at around \$6.00 (log(wage)=1.8) and the distribution of standard deviation of wages for individuals is slightly skewed towards higher values centered around log(wage)=0.3 (Figure 10). The average differences is fairly symmetric centered around log(wage)=0.1, and the distribution of the standard deviation of differences is skewed towards larger values centered around log(wage)=0.4. The distribution of the sum of the differences is mostly positive, which suggests that most of the subjects enjoy some level of increase in their wages. The distribution of the biggest

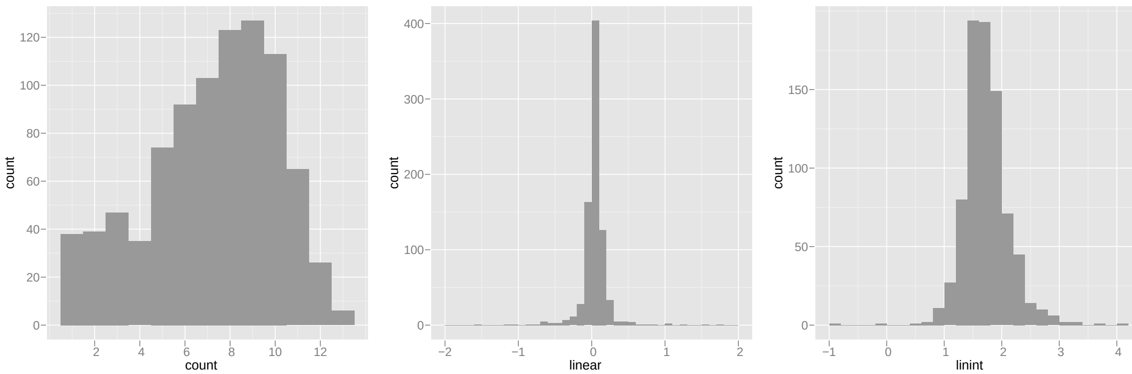


Figure 8: Histograms of several descriptor variables: (Left) Number of measurements for the 888 subjects. Most subjects have around 8 measurements. A little less than 80 subjects have two or less measurements. (Middle) Slopes of linear regression. (Right) Intercept of the linear fit.

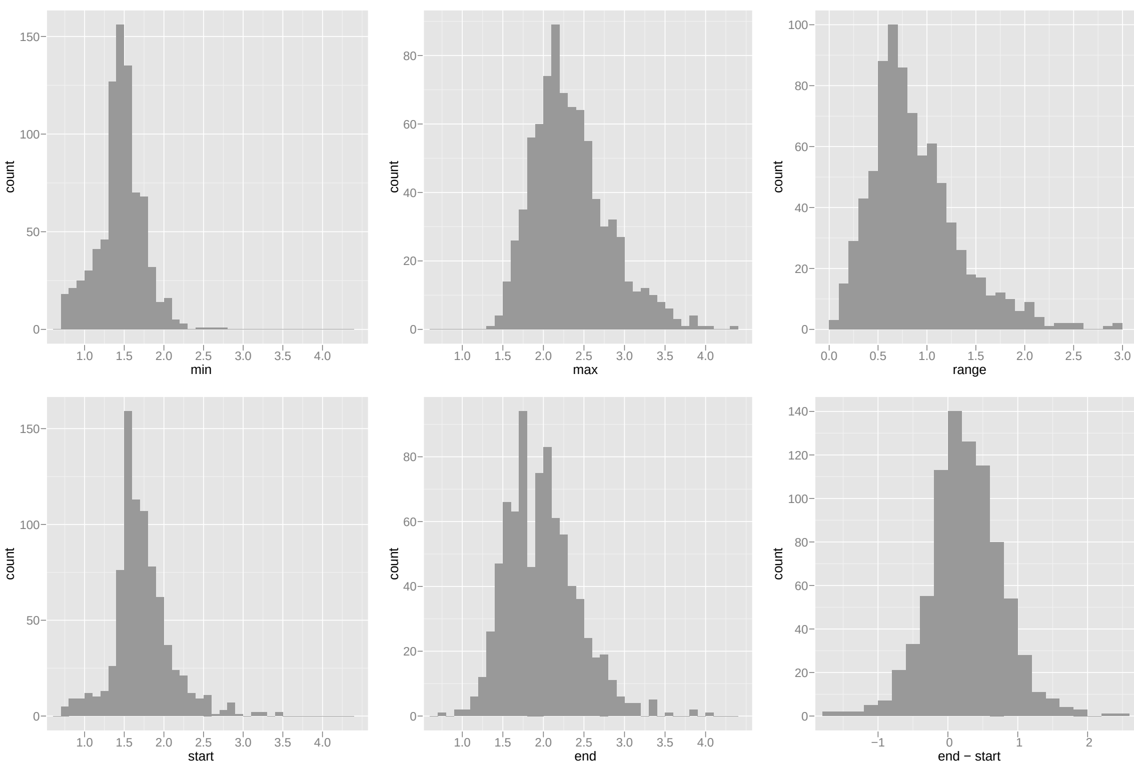


Figure 9: Histograms of descriptor variables: (From top left to bottom right) minimum, maximum, range, starting, ending, ending-start log(wage).

differences is skewed to large values, centered around log(wage)=0.6. One subject experienced a difference of log(wage)=3.0 between consecutive measurements.

Using the descriptor variables we explore the variability among individuals using principal component analysis. The results are summarized in Table 3.5. Interpreting the principal components is assisted by plots of the data (Figure 13). The main sources of variation might be interpreted as volatility (PC 1), trend (PC 2), and starting wage (PC 3). The fourth principal component is mostly the length of the series.

Variable	PC1	PC2	PC3	PC4
av	0.17	-0.19	**-0.34**	-0.04
sd	**0.29**	-0.05	0.15	-0.12
min	-0.02	-0.14	**-0.42**	-0.12
max	**0.30**	-0.12	-0.14	-0.03
range	**0.31**	-0.03	0.13	0.05
start	0.10	0.12	**-0.35**	-0.16
end	0.13	**-0.35**	-0.16	0.02
stend	0.04	**-0.40**	0.12	0.13
expmin	-0.04	-0.03	-0.10	-0.08
expmax	0.09	-0.07	-0.09	**0.53**
avdif	0.02	**-0.30**	0.11	-0.08
sddif	**0.28**	0.11	0.15	-0.05
sumdif	**0.30**	0.07	0.15	-0.05
smalldif	0.07	0.01	0.09	**-0.50**
bigdif	**0.30**	0.07	0.15	-0.05
up	**-0.28**	-0.11	-0.15	0.05
linear	-0.02	**-0.25**	0.11	0.07
linint	0.18	0.15	-0.17	**0.34**
rlin	-0.01	**-0.31**	0.11	0.13
rlinint	0.17	0.11	-0.13	**0.44**
lmeint	0.16	-0.030	**-0.37**	-0.14
lmeexper	0.05	**-0.39**	0.05	-0.04
lmesd	**0.31**	0.07	0.08	0.00
lmeint2	0.16	-0.01	**-0.36**	-0.13
lmeexper2	0.05	**-0.39**	0.06	-0.02
lmesd2	**0.31**	0.07	0.08	0.00
% Total Variance	34.0	52.5	69.6	77.1

Table 6: Summary of principal components analysis.

PC 1 is composed mostly of the descriptor variables sd, max, range, sddif, sumdif, bigdif, up (related to sddif), lmesd, and lmesd2 (summary values of the variance in the random effects from model fitting). The low values on PC ! corresponse to profiles that are very flat, and smooth. The high values on PC1 correspond to profiles that are extremely varied, subjects who have experienced huge changes in their wages as they've gained more experience.

PC 2 is composed mostly of the descriptor variables end, stend, avdif, linear, rlin, lmeexper, lmeexper2. Low values correspond to profiles that are increasing with experience. High values are associated with profiles that are decreasing with experience.

PC 3 is composed mostly of the descriptor variables av, min, start, lmeint and lmeint2. Low values correspond to profiles of subjects that have high values at the start of their workforce experience, and high

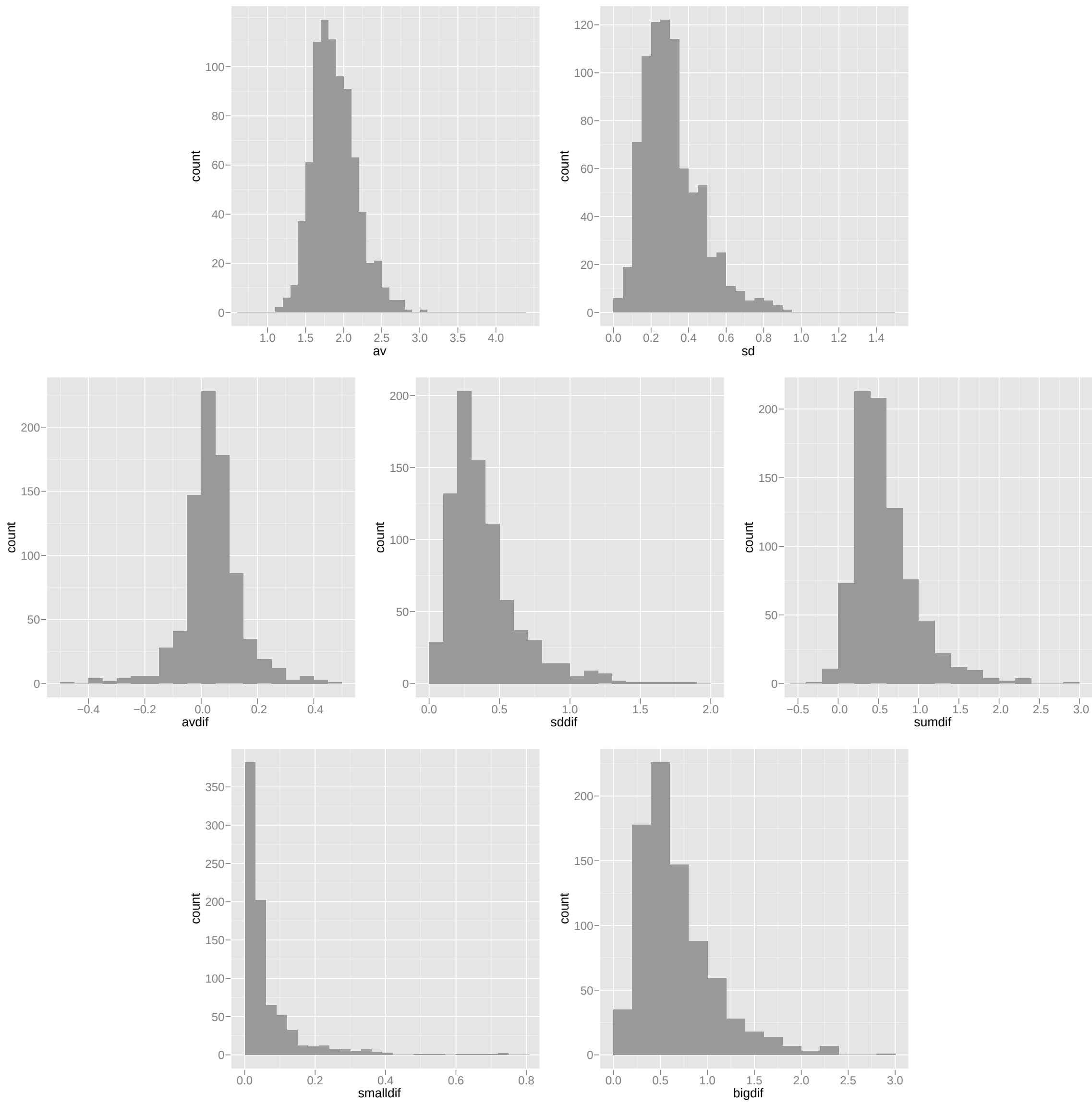


Figure 10: Histograms of descriptor variables: (From top left to bottom right) average, standard deviation, average difference, standard deviation of differences, sum of differences, smallest and largest absolute difference log(wage).

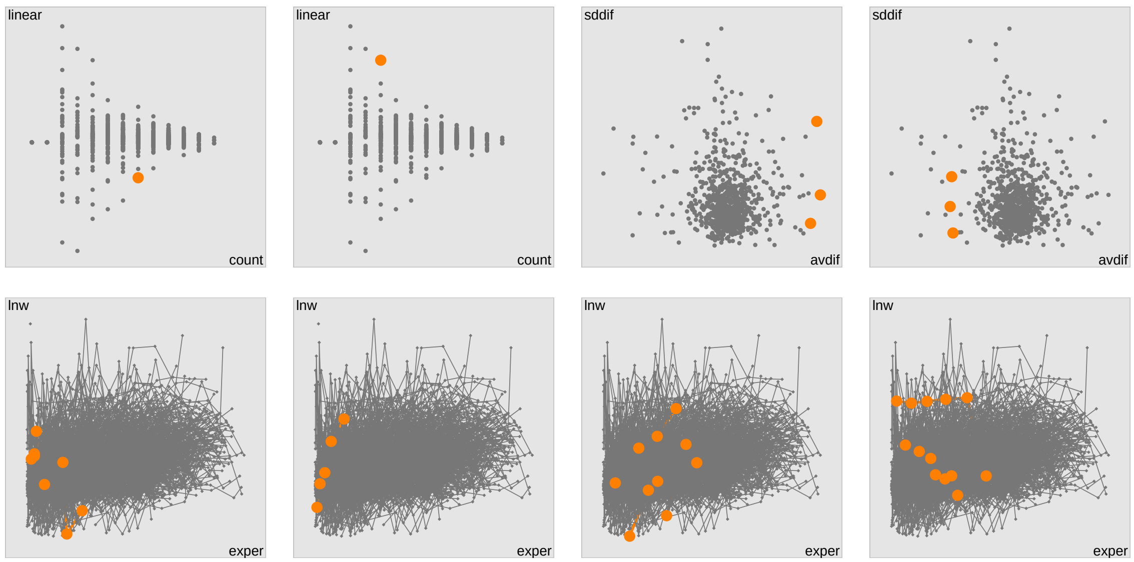


Figure 11: Profiles of subjects with extreme values on selection of descriptor variables.

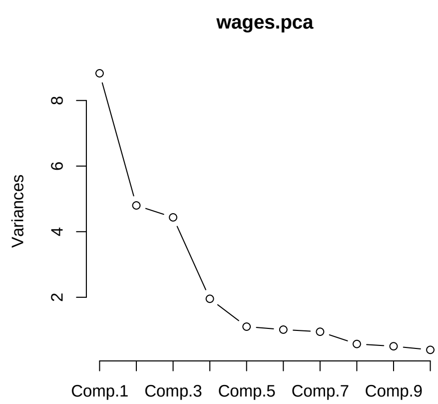


Figure 12: Scree plot for principal components analysis, suggests three or four principal components might be sufficient to describe the variation among descriptors.

values correspond to subjects who have low wages at the start of their workforce experience.

PC 4 explains less (7.5%) of the total variation than the first three principal components (34%, 18.5%, 17.1%). This is a big drop in variance explained. It also has less interesting explanation of the variation: low values being short series and large values correspond to long series. This is a structural aspect of the data - how many measurements on each subject were made.

In summary, the principal components analysis suggests that there are three main types of variability in the profiles: volatility, trend and starting wages. We will pick just one descriptor from these three types to use to summarize subjects experiences.

3.6 Models

To summarize the relationship between wages and experience dependent on the demographic and temporal covariates it would seem that a linear mixed effects model would be appropriate. As a baseline we fit a fixed effects model of log(wages) on experience:

$$\hat{lnw} = 1.691 + 0.052 \times exper$$

which has a residual sum of squares (RSS), $\sum e^2 = 1034.9$. Tables 7 and 8 summarize the range of reasonable models for this data. Figure 14 compares the AIC and BIC values for the selection of models.

	Models							
Variable	1	2	3	4	5	6	7	8
intercept	***1.716	***1.818	***1.700	***1.724	***1.617	***1.800	***1.817	***1.725
exper	***0.046	***0.041	***0.045	***0.046	***0.045	***0.040	***0.040	***0.039
uerate	-	***-0.011	-	-	-	***-0.011	***-0.011	***-0.012
ged	-	-	***0.068	-	-	***0.062	***0.065	***0.062
black	-	-	-	-0.028	-	-	**-0.047	**-0.054
hispanic	-	-	-	-0.004	-	-	-0.004	-0.007
hgc7	-	-	-	-	0.051	-	-	0.054
hgc8	-	-	-	-	0.061	-	-	0.062
hgc9	-	-	-	-	*0.086	-	-	*0.082
hgc10	-	-	-	-	***0.152	-	-	***0.156
hgc11	-	-	-	-	***0.174	-	-	***0.178
hgc12	-	-	-	-	***0.306	-	-	***0.324
RSS	518.1	516.1	517.8	518.1	519.4	515.9	515.9	517.5
loglik	-2460.7	-2442.1	-2453.0	-2459.8	-2441.7	-2435.7	-2433.0	-2411.8
BIC	4974.0	4945.6	4967.3	4989.7	4988.6	4941.5	4953.7	4963.8
AIC	4933.4	4898.2	4920.0	4935.6	4907.5	4887.3	4886.0	4855.6

Table 7: Summary of random effects models, main effects only. There is not a lot of difference in these models, measured by how much variation is explained! What is surprising is that the RSS increases as more explanatory variables are added to exper, uerate and ged. According to AIC the best model includes all of the main effects. (Note: $*\alpha = 0.1$, $**\alpha = 0.05$, $***\alpha = 0.01$.)

The biggest reduction in RSS comes from including random effects in the model: RSS is cut in half, from 1034.9 to 518.1. beyond this it might be argued there is not much difference between the random effects

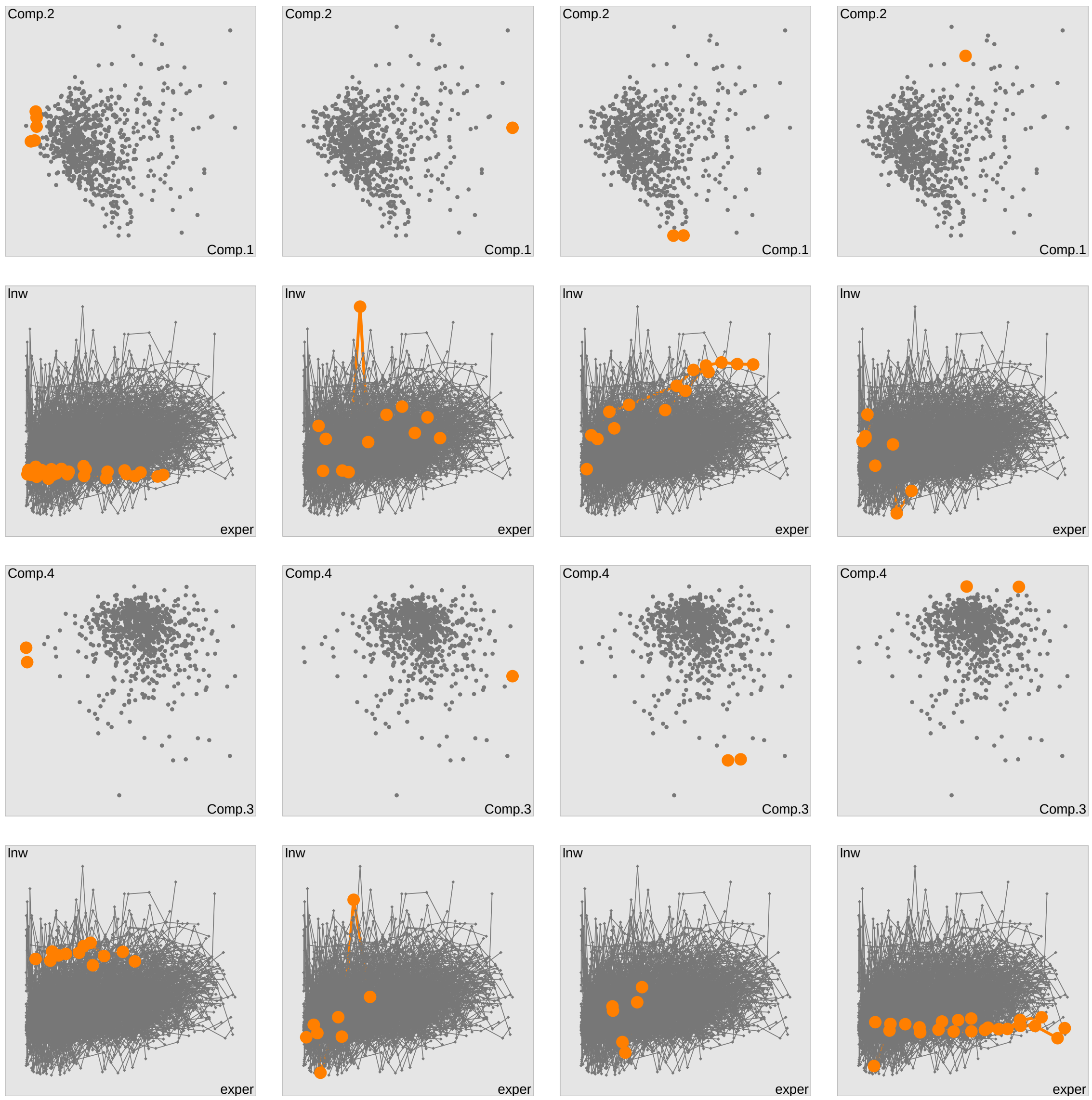


Figure 13: Exploring the principal components using lined brushing between the plots of principal components and profiles: (top two left columns) first PCA explains the volatility of the individual's wage experience, (top two right columns) second PCA describes linear trend, (bottom two left columns) third principal component explains starting wage, (bottom two right columns) and fourth principal component explains the length og the experience.

models. BIC agrees the most with RSS, suggesting that the best model is one that contains only three predictors: exper, uerate and ged. AIC recommends a model with most of the predictors.

Examining the coefficients for the fixed effects suggests that there is a relation between most predictors and log(wages). The predictors uerate, ged, and hgc all have highly significant coefficients in all of the models. Race has a slightly different effect: the coefficients are not significant in the main effects model, although the coefficient for black subjects becomes significant when the education variables are included in the model. When interactions with experience are considered only one is significant: that for black subjects. And interestingly the main effects coefficient for blacks is not significant when the interaction is included.

	Models				
Variable	9	10	11	12	Final
intercept	***1.700	***1.721	***1.607	***1.714	***1.711
exper	***0.045	***0.047	***0.049	***0.044	***0.043
uerate	-	-	-	***-.012	***-0.012
ged	**0.068	-	-	**0.047	***0.061
black	-	0.008	-	-0.016	-0.006
hispanic	-	-0.028	-	-0.035	-
hgc7	-	-	0.066	0.073	0.054
hgc8	-	-	0.067	0.073	0.063
hgc9	-	-	*0.108	*0.111	0.084
hgc10	-	-	***0.160	***0.169	***0.157
hgc11	-	-	**0.163	***0.176	***0.179
hgc12	-	-	***0.319	***0.335	***0.325
*ged	0.004	-	-	0.005	-
*black	-	***-0.015	-	***-0.015	***-0.018
*hispanic	-	0.009	-	0.010	-
*hgc7	-	-	-0.006	-0.008	-
*hgc8	-	-	-0.003	-0.004	-
*hgc9	-	-	-0.009	-0.011	-
*hgc10	-	-	-0.004	-0.005	-
*hgc11	-	-	0.004	0.000	-
*hgc12	-	-	-0.005	0.005	-
RSS	515.9	518.7	519.6	518.5	518.1
loglik	-2435.3	-2453.3	-2440.4	-2402.9	-2406.5
BIC	4975.5	4994.3	5038.5	5024.8	4953.1
AIC	4921.4	4926.6	4916.7	4855.7	4844.9

Table 8: Summary of random effects models, first order effects. The interaction terms do not add much in terms of explanation of the variation in log(wages). The only interaction which appears to be important is that for black subjects. When this term is added the main order effect reduces almost to zero, which suggests that blacks experience a different trend in their wages than other races. This importance is not reflected by the model selection criteria. (Note: $*\alpha = 0.1$, $**\alpha = 0.05$, $***\alpha = 0.01$.)

Although a simple model might be suggested by RSS and BIC we chose a final model that includes race and hgc, because the coefficients have a significant contribution to log(wage), and it looks reasonable for explaining the variation in log(wages) in a little more detail. Thus our final model is:

$$
\begin{aligned}
\hat{lnw} &= 1.711 + 0.043exper - 0.012uerate \\
&\quad +0.061ged + 0.157hgc10 + 0.179hgc11 + 0.325hgc12 \\
&\quad -0.018 * black * exper
\end{aligned}
$$

Figure 14 shows the main effects of this model. As education level increases so does the log(wage), and blacks experience slightly less increase than whites or hispanics. A higher unemployment rate reduces the log(wage) very slightly.

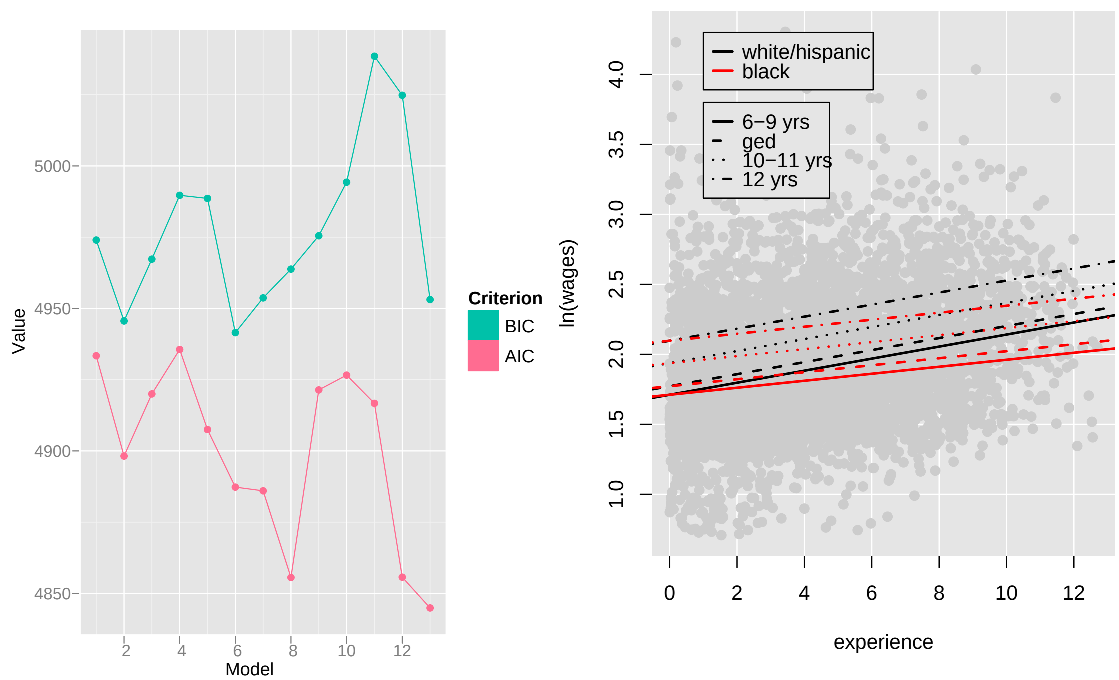


Figure 14: (Left) Comparison of the AIC and BIC values for different selection of random effects models. BIC suggests that the model that contains exper, uerate, ged, black, hgc and the interaction between black and exper is by far the best model. (Right) Fixed effects of the final model.

Getting beyond the fixed effects, to the most interesting part of this data, we examine the random effects. Exhaustively this involves examining all 888 profiles, but the descriptor variables can help in organizing the order in which these are examined. Figure 15 shows profiles of selected subjects. Given the estimates of slope, intercept and residual standard deviations for each individual we can explore the different characteristics of the subjects: ones who have dramatic increases in wages, or steady increases, or even declines in wages, with experience.

There are several subjects where the fit is not at all consistent with the data. And generally from scanning over the profiles, examining both the observed and fitted data, suggests that the model underfits the negative trends. Subjects, who's wages tend to go down with time, have fitted values with a trend that

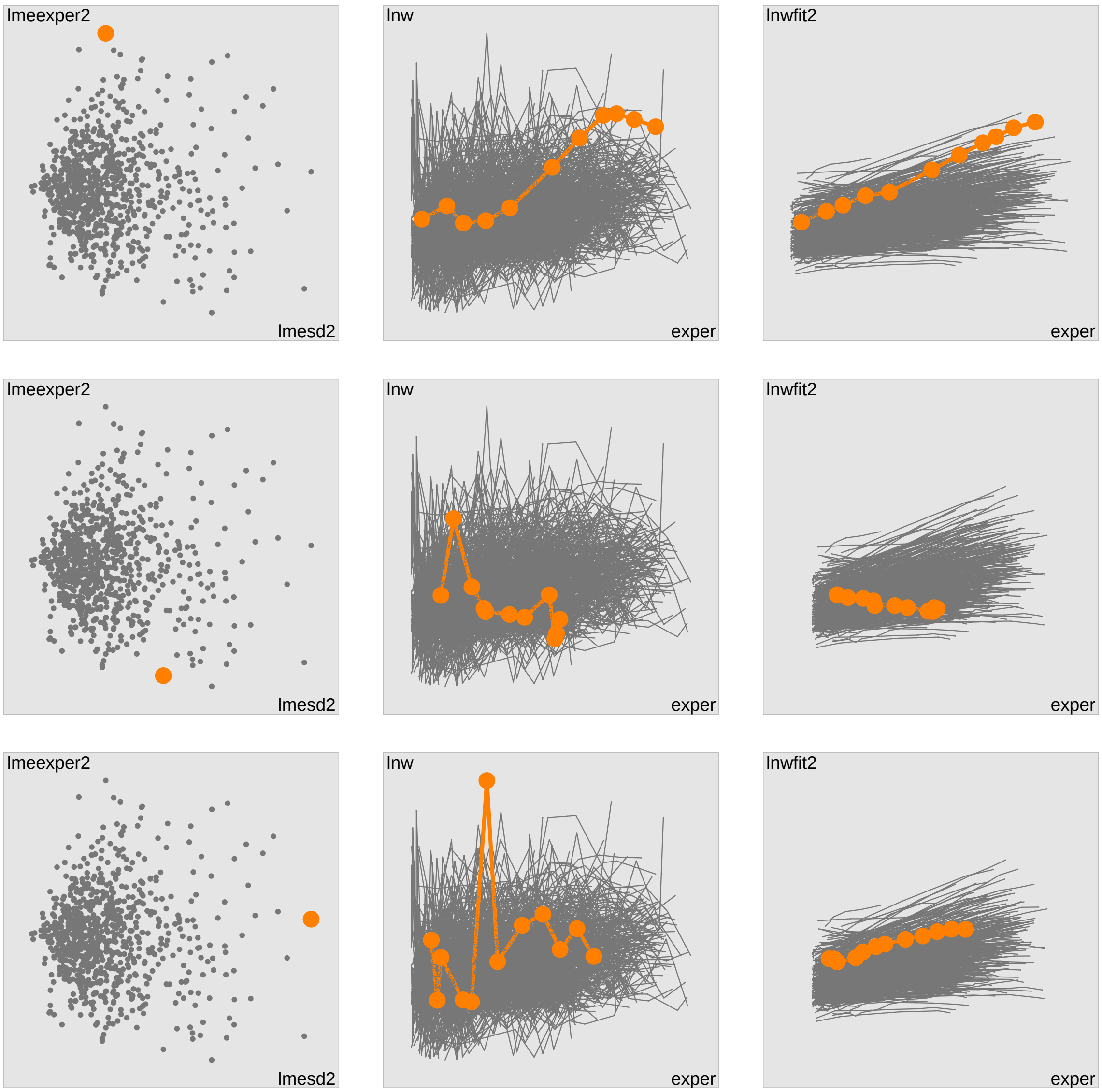


Figure 15: Studying the individual fits: slopes vs standard deviations of the residuals (left column), observed profiles (middle column), and fitted profiles (right column). The fitted profiles are a lot less varied than the observed profiles. An individual with a large positive slope and moderate variability is shown in the top row, one with large negative slope and highish variability is shown in the middle row, and and individual with high variability is shown in the bottom row.

is more positive. We can see this also by comparing the slopes from simple linear regressions with that of the mixed effects model. This is investigated further in Figure 16. The percentiles of the distributions of the slopes from each fit are calculated. These are compared in a side-by-side dot plot, where matching percentiles are connected by lines, and a pattern can be seen: the slopes from simple linear regressions are more varied than those of the linear mixed effects model. Both distributions are centered at the same value. The differences between percentiles of the slopes from the two models are also calculated, and plotted. From this its clear that negative slopes from simple linear regressions are more negative than those of the random effects model, and correspondingly, positive slopes are much bigger.

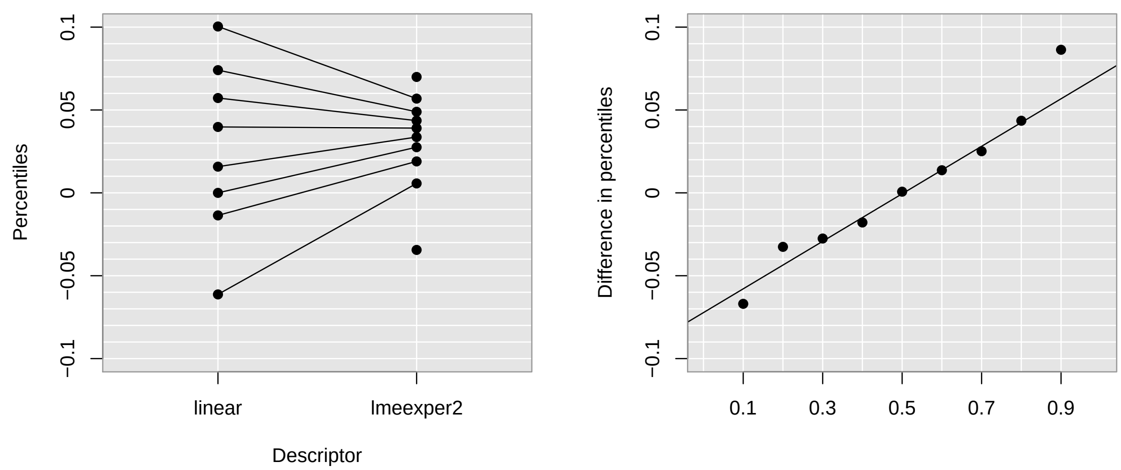


Figure 16: Comparison of simple linear regression with linear random effects model: (left) side-by-side dot plots of, and (right) difference between, the percentiles of the slopes from both models. Lines connect corresponding percentiles in the dot plots. A regression is fit to the differences between percentiles. The simple linear fitted models have more extreme slopes than the random effects model.

We can add several statistics calculated by the model as additional descriptor variables for the data: coefficients for exper, and RSS, for each subject. These were used in the previous section.

3.7 Organizing the individuals

Since the main structure in this data is the individual variability, a further step is taken in the analysis. This is to characterize and organize individuals into similar types, based on their profiles. Perhaps we can group the subjects into a set of archetypes: the "high flyers", who have rapid increaase in wages, or the "steady earners", who have steady increases in wages with experience, or the "inconsistent earners" who's wages bounce around.

The descriptor variables are categorized, to create variables summarizing trend, volatility and starting wages. To measure trend we use the robust linear fit:

decrease	little change	increase	steep increase
< -0.05	$-0.05 - 0.05$	$0.05 - 1.5$	> 1.5

To measure volatility we use the standard deviation of the differences between consecutive time points, weighted by time difference:

NA	smooth	wrinkled	spiky
less than 3 values	$0.02 - 0.25$	$0.25 - 0.46$	> 0.46

and starting wage is measured using the starting value:

low	median	high
< 1.5	$1.5 - 1.9$	> 1.9

Tables 9, 10 and 11 describe the profiles in terms of these categories. Almost half of these men (46%) experience volatility in their wages, and almost half experience very little change in their wages with experience. About a third of the men experience an increase in their wages, with about 10% experiencing steep increases, but the same amount experience a decline in wages.

	trend				
volatility	decrease	little change	increase	steep increase	total
NA	0 (0.00)	80 (0.09)	0 (0.00)	0 (0.00)	80 (0.09)
smooth	9 (0.01)	79 (0.09)	75 (0.08)	9 (0.01)	172 (0.19)
wrinkled	11 (0.01)	104 (0.12)	98 (0.11)	14 (0.02)	227 (0.26)
spiky	69 (0.08)	170 (0.19)	114 (0.13)	56 (0.06)	409 (0.46)
total	89 (0.10)	433 (0.49)	287 (0.32)	79 (0.09)	888 (1.00)

Table 9: Trends by volatility. Most subjects have little change or increasing wages with experience, and about half of the subjects experience a lot of variability in their wages.

	trend				
starting wage	decrease	little change	increase	steep increase	total
low	5 (0.01)	79 (0.09)	64 (0.07)	36 (0.04)	184 (0.21)
median	30 (0.03)	244 (0.27)	180 (0.20)	32 (0.04)	486 (0.55)
high	54 (0.06)	110 (0.12)	43 (0.05)	11 (0.01)	218 (0.25)
total	89 (0.10)	433 (0.49)	287 (0.32)	79 (0.09)	888 (1.00)

Table 10: Trends by starting wage.

	volatility				
starting wage	NA	smooth	wrinkled	spiky	total
low	25 (0.03)	21 (0.02)	40 (0.05)	98 (0.11)	184 (0.21)
median	30 (0.03)	113 (0.13)	146 (0.16)	197 (0.22)	486 (0.55)
high	25 (0.03)	38 (0.04)	41 (0.05)	114 (0.13)	218 (0.25)
total	80 (0.09)	172 (0.19)	227 (0.26)	409 (0.46)	888 (1.00)

Table 11: Volatility by starting wage.

We can also examine these descriptors in association with the demographic information of the men (Tables 12, 13, 14).... more to do here.

ged	trend				total
	decrease	no change	increase	steep increase	
0	68	341	218	61	688
1	21	92	69	18	200
total	89	433	287	79	888

Table 12: Trend by graduate equivalency degree.

race	trend				total
	decrease	no change	increase	steep increase	
white	40	222	136	40	438
black	35	123	68	20	246
hispanic	14	88	83	19	204
total	89	433	287	79	888

Table 13: Trend by race.

hgc	trend				total
	decrease	no change	increase	steep increase	
6	5	13	9	5	32
7	14	42	23	16	95
8	25	120	73	21	239
9	20	109	66	16	211
10	19	84	73	10	186
11	5	56	38	11	110
12	1	9	5	0	15
total	89	433	287	79	888

Table 14: Trend by highest grade completed.

3.8 Notes on the analysis

Sampling and cleaning of data not clear. How has periods of unemployment been handled in the data reporting? Have all of these subjects been continuously employed?

4 Summary

- The predominant structure is volatility in wages.
- Some boys see a decline with longer workforce experience, but most have relatively moderate wage increases, and some have dramatic wage increases.
- Blacks experience lower wages, particularly with more experience. Hispanic and white experience is similar.
- Longer in school additively increases wages, and year 12 is dramatically better than any other.
- GED doesn't add much.
- Unemployment rate has a small effect.

References

[1] J. D. Singer and J. B. Willett. *Applied Longitudinal Data Analysis.* Oxford University Press, Oxford, UK, 2003.