Applied Econometrics Assignment Report

- **Student Name:** Umut Demirhan
- **Student Number:** 46739106
- **Lecturer's Name:** Chris Heaton
- **Unit Code:** ECON8040
- **Unit Name:** Applied Econometrics
- **Assignment Title:** The assignment report
- **Due Date:** 05/Jun/2023
- **Date of Submission:** 05/Jun/2023
- **Word count (excluding the tables, figures, headers, and footers):** 1391.

Contents

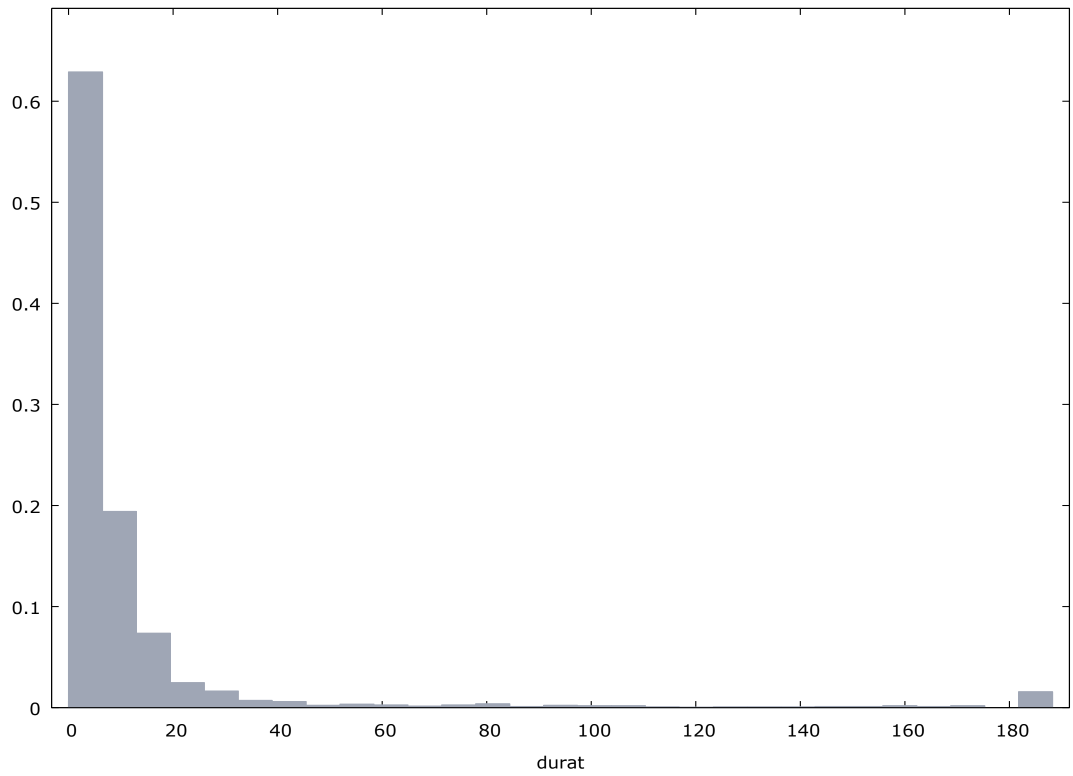
Introduction

Workers' compensation laws offer financial protection to employees with temporary total disability due to workplace injury. In the 1980s, Kentucky and Michigan significantly amended these laws, especially the maximum benefits. In this report, we investigate the impact of these changes on the duration of leave taken by high-income workers. Through statistical analysis of a comprehensive dataset from both states, we aim to deduce whether the policy amendments influenced the leave durations. Our findings could provide valuable insights for future policy formulation.

Data Description and Summary Statistics

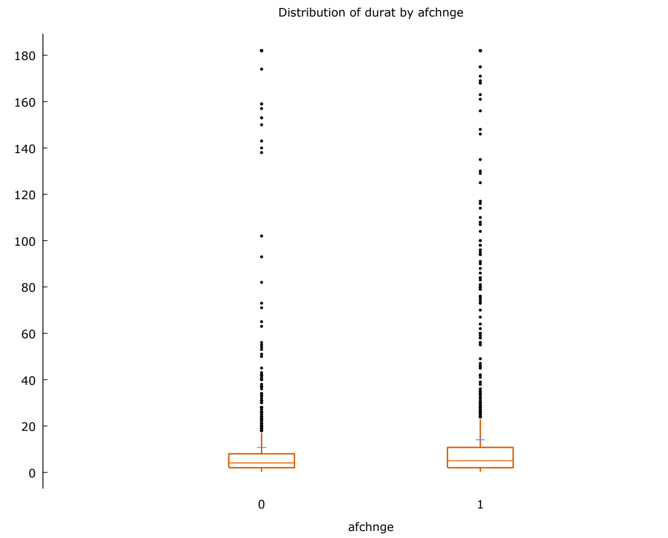
The dataset utilised for this analysis consists of 2679 observations across a range of variables. The primary variable of interest, 'ducat', indicates the duration of unemployment. As shown in Figure 1, the frequency distribution for 'ducat' shows that many unemployment durations fall in less than 6.0583 days. The mean duration of unemployment is approximately 12.35 days, which is skewed by a small proportion of significantly longer unemployment durations, as evidenced by the high standard deviation of 28.82.

Figure 1: Distribution of 'Durant'



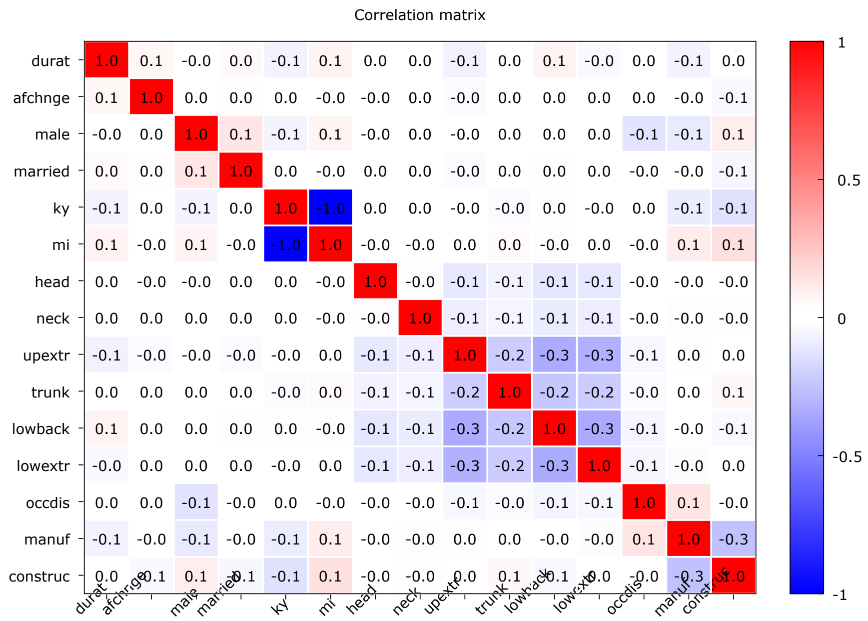
The variable 'archangel' represents whether the observation was made before or after a policy change. In Figure 2, the distribution of 'Durant' is split by 'archangel', illustrating that observations from after the policy change tend to have a broader distribution, higher mean, and more outliers compared to those from before the policy change. I will test if this difference is significant later in this report.

Figure 2: Distribution of ‘Durant’ by ‘afchnge’ (Factorized boxplot)



A correlation matrix for the data, depicted in Figure 3, reveals correlations between 'Durant' and several other variables in the dataset. Notably, 'Durant' exhibits a slight positive correlation with 'archangel', suggesting that the policy change might have a minor impact on unemployment duration.

Figure 3: Correlation matrix



The remaining variables are binary dummy variables denoting specific characteristics or conditions. These variables include gender ('male'), marital status ('married'), injury locations ('head', 'neck', 'exert', 'trunk', 'lowback', 'lowextr'), occupational disability ('cordis'), and industry ('manuf', 'construction').

Table 1 provides summary statistics for all variables, indicating their central tendency, dispersion, and distribution shape. Most binary variables have means around 0 or 1, reflecting their categorical nature, while 'Durant' displays significant variability as reflected in its high standard deviation.

Table 1: Summary Statistics Table

Summary Statistics, using the observations 1 - 2679

Variable	Mean	Median	Minimum	Maximum
durat	12.347	5.0000	0.25000	182.00
afchnge	0.49123	0.00000	0.00000	1.0000
male	0.96043	1.0000	0.00000	1.0000
married	0.85554	1.0000	0.00000	1.0000
ky	0.83315	1.0000	0.00000	1.0000
mi	0.16685	0.00000	0.00000	1.0000
head	0.038447	0.00000	0.00000	1.0000
neck	0.025383	0.00000	0.00000	1.0000
upextr	0.24188	0.00000	0.00000	1.0000
trunk	0.13214	0.00000	0.00000	1.0000
lowback	0.27100	0.00000	0.00000	1.0000
lowextr	0.24748	0.00000	0.00000	1.0000
occdis	0.010078	0.00000	0.00000	1.0000
manuf	0.19746	0.00000	0.00000	1.0000
construc	0.22023	0.00000	0.00000	1.0000

Variable	Std. Dev.	C.V.	Skewness	Ex. kurtosis
durat	28.819	2.3341	4.6868	22.635
afchnge	0.50002	1.0179	0.035093	-1.9988
male	0.19498	0.20301	-4.7239	20.315
married	0.35162	0.41099	-2.0227	2.0913
ky	0.37291	0.44760	-1.7871	1.1936
mi	0.37291	2.2350	1.7871	1.1936
head	0.19231	5.0019	4.8010	21.050
neck	0.15731	6.1977	6.0352	34.423
upextr	0.42830	1.7707	1.2055	-0.54669
trunk	0.33870	2.5632	2.1726	2.7201
lowback	0.44456	1.6405	1.0304	-0.93818
lowextr	0.43163	1.7441	1.1703	-0.63041
occdis	0.099903	9.9126	9.8098	94.232
manuf	0.39816	2.0164	1.5200	0.31032
construc	0.41448	1.8820	1.3502	-0.17689

Variable	5% Perc.	95% Perc.	IQ range	Missing obs.
durat	1.0000	45.000	7.0000	0
afchnge	0.00000	1.0000	1.0000	0
male	1.0000	1.0000	0.00000	0
married	0.00000	1.0000	0.00000	0
ky	0.00000	1.0000	0.00000	0

mi	0.00000	1.0000	0.00000	0
head	0.00000	0.00000	0.00000	0
neck	0.00000	0.00000	0.00000	0
upextr	0.00000	1.0000	0.00000	0
trunk	0.00000	1.0000	0.00000	0
lowback	0.00000	1.0000	1.0000	0
lowextr	0.00000	1.0000	0.00000	0
occdis	0.00000	0.00000	0.00000	0
manuf	0.00000	1.0000	0.00000	0
construc	0.00000	1.0000	0.00000	0

These tables and visualisations are embedded within this section to offer a comprehensive understanding of the dataset we are working with and will serve as a base for our subsequent analysis.

Econometric Methodology

The econometric technique used in this analysis is the Ordinary Least Squares (OLS) method. OLS is chosen because it is one of the best linear unbiased estimators, assuming particular classical linear regression model (CLRM) assumptions hold. In addition, OLS minimises the sum of the squared residuals and provides estimates that are easy to interpret. Further, the application of heteroskedasticity-robust standard errors (variant HC1) allows us to manage potential heteroskedasticity in the models, thus preserving the efficiency of our estimators.

OLS is widely used in econometric analysis due to its simplicity, interpretability, and desirable properties under the CLRM assumptions. When the assumptions hold, OLS is BLUE (Best Linear Unbiased Estimator), providing the most accurate linear prediction of the dependent variable. Moreover, applying robust standard errors is appropriate when dealing with potential heteroskedasticity. It ensures more reliable hypothesis testing by correcting standard errors that might otherwise be underestimated due to heteroskedasticity.

Two models are used: a full model (Model 1) with all variables included and a simplified model (Model 2) containing only a subset of statistically significant variables. Both models employ heteroskedasticity-robust standard errors (variant HC1) to correct for heteroskedasticity, thus strengthening the reliability of the estimated coefficients.

Model 1 includes all variables: afchnge, male, married, ky, head, neck, upextr, trunk, lowback, lowextr, occdis, manuf, and construc. Despite the variety of variables considered, this model only accounts for about 2.92% of the variation in durat, as indicated by the R-squared value. Nevertheless, the F-statistic (5.67) and its corresponding p-value (2.33e-10) denote that the model is statistically significant, suggesting that at least some of the independent variables contribute to the explanation of durat.

Model 1:

Model 1: OLS, using observations 1-2679
Dependent variable: durat
Heteroskedasticity-robust standard errors, variant HC1

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	24.4121	5.55515	4.395	<0.0001	***
afchnge	2.98909	1.09851	2.721	0.0066	***

male	−3.23486	2.90416	−1.114	0.2654	
married	3.33338	1.28787	2.588	0.0097	***
ky	−6.51992	1.79566	−3.631	0.0003	***
head	−5.01472	5.41288	−0.9264	0.3543	
neck	−2.69407	5.48647	−0.4910	0.6234	
upextr	−10.0655	4.13146	−2.436	0.0149	**
trunk	−4.91131	4.38682	−1.120	0.2630	
lowback	−3.41148	4.24068	−0.8045	0.4212	
lowextr	−9.05667	4.14513	−2.185	0.0290	**
occdis	2.32895	7.56197	0.3080	0.7581	
manuf	−6.25597	1.06436	−5.878	<0.0001	***
construc	−0.563023	1.33195	−0.4227	0.6725	

Mean dependent var	12.34686	S.D. dependent var	28.81932
Sum squared resid	2159231	S.E. of regression	28.46433
R-squared	0.029219	Adjusted R-squared	0.024484
F(13, 2665)	5.673530	P-value(F)	2.33e-10
Log-likelihood	−12765.36	Akaike criterion	25558.71
Schwarz criterion	25641.22	Hannan-Quinn	25588.56

In Model 2, the methodology involves a more simplified approach by including only a selection of significant variables from the first model. This reduced model consists of the variables: afchnge, married, ky, upextr, lowextr, and manuf. As a result, the R-squared value is marginally lower than in Model 1, suggesting the model explains approximately 2.72% of the variation in durat. However, despite fewer variables, the F-statistic is larger (11.07) with a smaller p-value (3.14e-12), suggesting the model is still statistically significant.

Model 2:

Model 2 with selected variables:OLS, using observations 1-2679
Dependent variable: durat
Heteroskedasticity-robust standard errors, variant HC1

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	17.4109	2.43522	7.150	<0.0001	***
afchnge	3.05047	1.09835	2.777	0.0055	***
married	3.10801	1.29263	2.404	0.0163	**
ky	−6.17348	1.82698	−3.379	0.0007	***
upextr	−6.53655	1.20306	−5.433	<0.0001	***
lowextr	−5.53049	1.25410	−4.410	<0.0001	***
manuf	−5.71462	1.01481	−5.631	<0.0001	***

Mean dependent var	12.34686	S.D. dependent var	28.81932
Sum squared resid	2163641	S.E. of regression	28.45604
R-squared	0.027237	Adjusted R-squared	0.025052
F(6, 2672)	11.07321	P-value(F)	3.14e-12
Log-likelihood	−12768.09	Akaike criterion	25550.18
Schwarz criterion	25591.43	Hannan-Quinn	25565.10

In both models, the variables afchnge, married, ky, upextr, lowextr, and manuf were found to be significant, as their p-values were less than 0.05. This suggests they have a statistically significant influence on durat. The other variables in Model 1 did not reach statistical significance (p-value > 0.05), providing a rationale for their exclusion from Model 2.

These results illustrate the trade-off between including more variables for a comprehensive approach and reducing to only significant variables to create a more straightforward yet still effective model.

Hypothesis Testing

The Chow test was employed in this study to test the null hypothesis that the change in the maximum benefit, represented by the variable 'afchnge', did not affect the average length of time high-income workers spend away from work due to events causing temporary total disability.

Model 5 is actually Model 2 without 'afchnge' variable for the Chow test.

Model 5 for Chow test:

Model 5: OLS, using observations 1-2679
Dependent variable: durat
Heteroskedasticity-robust standard errors, variant HC1

	Coefficient	*Std. Error*	*t-ratio*	*p-value*	
const	18.8037	2.38742	7.876	<0.0001	***
married	3.24757	1.29458	2.509	0.0122	**
ky	−6.12685	1.82729	−3.353	0.0008	***
upextr	−6.69884	1.20403	−5.564	<0.0001	***
lowextr	−5.56410	1.25626	−4.429	<0.0001	***
manuf	−5.73961	1.01492	−5.655	<0.0001	***

Mean dependent var	12.34686	S.D. dependent var	28.81932
Sum squared resid	2169851	S.E. of regression	28.49151
R-squared	0.024445	Adjusted R-squared	0.022620
F(5, 2673)	12.79913	P-value(F)	2.55e-12
Log-likelihood	−12771.93	Akaike criterion	25555.86
Schwarz criterion	25591.22	Hannan-Quinn	25568.65

Chow test for structural difference with respect to afchnge -
Null hypothesis: no structural difference
Asymptotic test statistic: Chi-square (6) = 9.95645
with p-value = 0.126498

The Chow test for structural difference with respect to 'afchnge' produced F(6, 2667) value of 1.659 with a p-value of 0.126498. The p-value is more significant than the standard 0.05 level, meaning we fail to reject the null hypothesis. Therefore, we conclude there is insufficient evidence to suggest a structural difference in the length of time off work due to disability events for high-income workers before and after the change in benefits.

The Chow test on Model 5:

```
Augmented regression for Chow test
OLS, using observations 1-2679
Dependent variable: durat
Heteroskedasticity-robust standard errors, variant HC1

               coefficient   std. error   t-ratio   p-value
  ----------------------------------------------------------
  const         17.6879       3.50043      5.053    4.64e-07 ***
  married        1.42495      1.82112      0.7825   0.4340
  ky            -6.10273      2.42779     -2.514    0.0120   **
  upextr        -4.98507      1.60430     -3.107    0.0019   ***
  lowextr       -4.34141      1.70221     -2.550    0.0108   **
  manuf         -3.77927      1.55950     -2.423    0.0154   **
  afchnge        1.67483      4.74444      0.3530   0.7241
  af_married     3.69577      2.54660      1.451    0.1468
  af_ky          0.450155     3.66362      0.1229   0.9022
  af_upextr     -3.38144      2.41977     -1.397    0.1624
  af_lowextr    -2.36568      2.50817     -0.9432   0.3457
  af_manuf      -3.91128      2.00261     -1.953    0.0509   *

Mean dependent var   12.34686   S.D. dependent var   28.81932
Sum squared resid     2159289   S.E. of regression   28.45404
R-squared            0.029193   Adjusted R-squared   0.025189
F(11, 2667)          6.866384   P-value(F)           1.61e-11
Log-likelihood      -12765.39   Akaike criterion     25554.78
Schwarz criterion    25625.50   Hannan-Quinn         25580.37

Chow test for structural difference with respect to afchnge
  Chi-square(6) = 9.95645 with p-value 0.1265
  F-form: F(6, 2667) = 1.65941 with p-value 0.1270
```

Subsequently, an augmented regression was conducted for the Chow test. Like the previous, this model included interaction terms for 'afchnge' with all other variables. These interaction terms tested whether the effects of the other variables changed after the benefit adjustment. However, only the interaction term 'af_manuf' could be counted as statistically significant at a little bit higher than the 0.05 level (p-value 0.0509), implying that the effect of manufacturing jobs on duration might have changed with the benefits alteration. The Chow test was then repeated, yielding similar results to before, i.e., we failed to reject the null hypothesis of no structural difference.

Uncertainty and Limitations

Despite our rigorous analysis, the study has its uncertainties and limitations. The models' low R-squared values imply that much of the variability in disability duration isn't explained, suggesting important factors might have been overlooked. Potential omitted variables, such as individual health status or specific job characteristics, might be influencing disability leaves. The assumption that the errors are i.i.d could also be flawed, potentially leading to inefficient estimators. Moreover, if the model's form is wrongly specified, as in if the true relationship is non-linear, our results may be inaccurate. Future research should consider these factors, possibly using a larger dataset and more sophisticated econometric methods.

Conclusion

After thoroughly analysing our dataset, changes to the maximum benefit in workers' compensation laws did not significantly affect how long high-income workers in Kentucky and Michigan took off from work due to temporary total disability in the 1980s. This finding suggests that high-income workers' decision to take time off from work due to injury was not influenced by these legal changes. However, it's important to note that our analysis might not cover all potential factors, such as individual health conditions or specific work environment details. Therefore, while our results provide an essential piece of the puzzle, they should be viewed as part of a larger, complex picture of how policy changes impact workers' behaviours.