Regression Based Service Rate Estimation

September 7, 2021

1 Overview

This week I worked on extending the regression model for service rate estimation to account for idleness. The main problem that we are trying to tackle is that the simple model, $\text{Days}_j = \beta_t \text{Trial}_j + \beta_p \text{Plea}_j + \epsilon_j$, doesn't account for idling judges. Hester's qualitative interviews with the judges indicates that harsher judges idle more often than more lenient judges. An overview of the results are below.

- **Iterative Idleness Estimation Using Expected Utilization:** This yields estimates of 3 days per trial, and about 14 pleas per day.
- **Iterative Idleness Estimation Taking Mins:** This yields estimates of 6.2 days per trial and 9.7 pleas per day.
- **Fixed Effects Model:** This yields estimates of 3.7 days per trial, and 10 pleas per day.

2 Iterative Idleness Estimation Using Expected Utilization

Step 0: We estimate the model, $\text{Days}_j = \beta_t \text{Trial}_j + \beta_p \text{Plea}_j + \epsilon_j$.

Steps 1-n: We then use the estimates of $\beta_t^{(1)}$ and $\beta_p^{(1)}$ to estimate the expected number of days it would take each judge to complete their work. Mathematically: Expected $\text{Days}_j^{(1)} = \beta_p^{(1)} \cdot \text{Plea}_j + \beta_t^{(1)} \cdot \text{Trial}_j$. Then, the utilization for each judge would be: $\text{Utilization}_j^{(1)} = \frac{\text{Expected Days}_j^{(1)}}{\text{Days}_j}$. Let $\gamma^1 = \max_j \text{Utilization}_j^{(1)}$, be the maximum utilization amongst all judges. Each judges idleness will be: $\text{Idleness}_j^{(1)} = \frac{\text{Utilization}_j^{(1)}}{\gamma^{(1)}}$. We then set $\text{Days}_j^{(1)} = \text{Days}_j \cdot \text{Idleness}_j^{(1)}$. We then estimate the model $\text{Days}_j^{(1)} = \beta_t \text{Trial}_j + \beta_p \text{Plea}_j + \epsilon_j$ and repeat until convergence.

Table 1: Regression model, utilization method

Dep. Variable:	y	**R-squared (uncentered):**	1.000
Model:	OLS	**Adj. R-squared (uncentered):**	1.000
Method:	Least Squares	**F-statistic:**	4.033e+32
Date:	Tue, 07 Sep 2021	**Prob (F-statistic):**	0.00
Time:	17:44:59	**Log-Likelihood:**	1541.8
No. Observations:	50	**AIC:**	-3080.
Df Residuals:	48	**BIC:**	-3076.
Df Model:	2		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Plea	0.0715	5.92e-18	1.21e+16	0.000	0.071	0.071
Trial	3.1154	3.49e-16	8.94e+15	0.000	3.115	3.115

Omnibus:	29.779	**Durbin-Watson:**	0.611
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	60.328
Skew:	-1.784	**Prob(JB):**	7.94e-14
Kurtosis:	7.027	**Cond. No.**	85.5

Notes:
[1] R is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

3 Iterative Idleness Estimation Taking Mins

Step 0: We estimate the model, $\text{Days}_j = \beta_t \text{Trial}_j + \beta_p \text{Plea}_j + \epsilon_j$.

Steps 1-n: We then use the estimates of $\beta_t^{(1)}$ and $\beta_p^{(1)}$ to estimate the expected number of days it would take each judge to complete their work. Mathematically: $\text{Expected Days}_j^{(1)} = \beta_p^{(1)} \cdot \text{Plea}_j + \beta_t^{(1)} \cdot \text{Trial}_j$. We would then set $\text{Days}_j^{(1)} = \min(\text{Days}_j, \text{Expected Days}_j^{(1)})$ We then estimate the model $\text{Days}_j^{(1)} = \beta_t \text{Trial}_j + \beta_p \text{Plea}_j + \epsilon_j$ and repeat until convergence.

Table 2: Regression model, min method

Dep. Variable:	y	**R-squared (uncentered):**	0.981
Model:	OLS	**Adj. R-squared (uncentered):**	0.980
Method:	Least Squares	**F-statistic:**	1215.
Date:	Tue, 07 Sep 2021	**Prob (F-statistic):**	7.78e-42
Time:	17:45:05	**Log-Likelihood:**	-183.73
No. Observations:	50	**AIC:**	371.5
Df Residuals:	48	**BIC:**	375.3
Df Model:	2		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Plea	0.1035	0.006	17.983	0.000	0.092	0.115
Trial	6.2803	0.339	18.540	0.000	5.599	6.961

Omnibus:	78.231	**Durbin-Watson:**	1.993
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	916.837
Skew:	-4.289	**Prob(JB):**	8.15e-200
Kurtosis:	22.144	**Cond. No.**	85.5

Notes:
[1] R is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

4 Fixed Effects Model

We know from our exploratory analysis that there is large heterogeneity in activity amongst counties. Therefore, it is likely that the county that a judge happens to be in also significantly affects the number of pleas he is able to process. One way we could try to incorporate both judge and county idleness, would be to use a fixed effects model. Here, the unit of observation would be a judge-county combination. For each judge county combination, i with judge j and county c, we could run the regression $\text{Days}_i = \alpha_j + \delta_c + \beta_p \text{Plea}_i + \beta_t \text{Trial}_i + \epsilon_i$. **Pros:** this would flexibly control for both judge and county fixed effects. **Cons:** We only have 248 observations of judge county combinations, and we would be trying to estimate around 96 parameters.

Table 3: Fixed Effects Model

	Dependent variable:
	Days
Plea	0.099***
	(0.009)
Trial	3.714***
	(0.399)
Observations	278
R^2	0.801
Adjusted R^2	0.696
Residual Std. Error	7.380 (df = 181)
Note:	*p<0.1; **p<0.05; ***p<0.01