Summary of Week 4

Contents

1 Things Done Last Week

1.1 Extract the Parameters from Pre-trained Model

I run the latent space model `wages_simple3-9-21.stan` and name the new stan object with 7-13-21, then use the mean of each parameter sample as the estimate and store them in `parm_fit_ifls_7-13-21.rda` for the ease of loading.

1.2 Prepare Parameters for Data Generation

The parameters of worker and sector latent space, β_0^k, β_1^k, β_2^k, wage variance σ_i^2, sector effect ν_k and its hyperparameter τ_k from the pre-trained latent space model are used in the simulation.

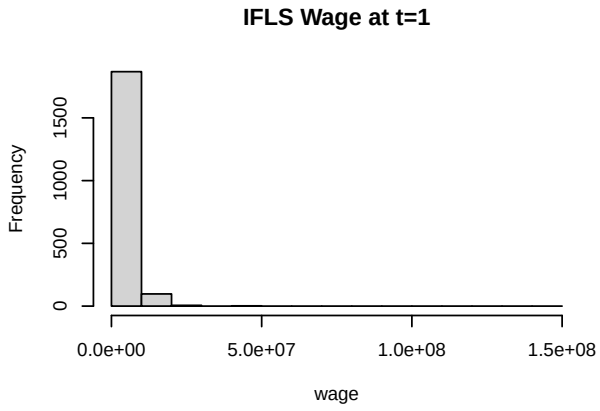
Values of coefficients of edge probability between firm and sector (i.e. γ_0^l, γ_1^l, γ_2^l) might require some further consideration. At this stage, for all $l = 1, \ldots, L_k$, I let γ_0^l to be random sample from β_0^k, $\gamma_1^l = \beta_1^k$ and $\gamma_2^l = \beta_2^k$. Using normal random sample for all γ^l's might sometimes lead to weird edge pattern. (e.g. 99% of workers in each sector work for only one firm over time)

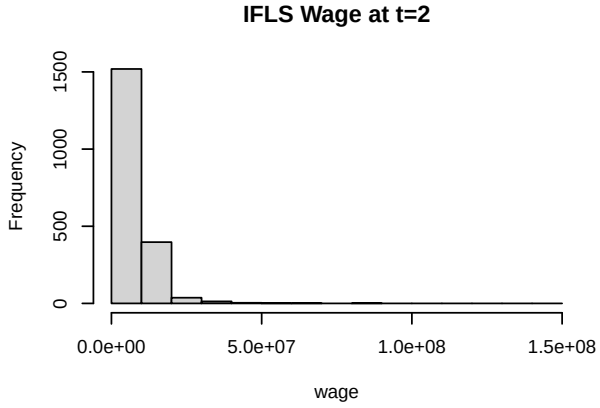
1.3 Data Generation

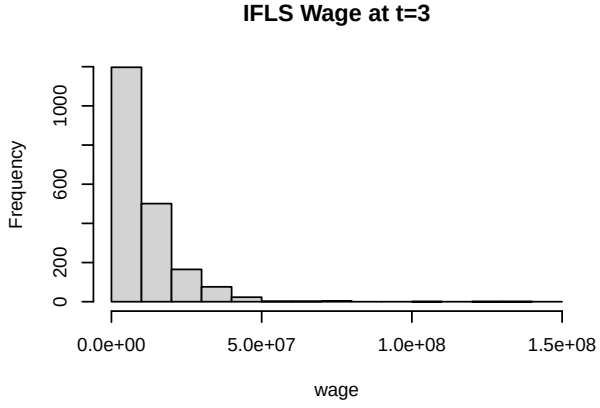
In the new data generation pipeline, to mitigate the issue of simulated wage too large and number of wage increase dissimilar to real data, instead of sampling based on $Y_i^{(t)} \sim N(\hat{\mu}_i + f_i^{(t)}, \hat{\sigma}_i^2)$, I tried $Y_i^{(t)} \sim N(\alpha_1 \hat{\mu}_i + \alpha_2 f_i^{(t)}, \hat{\sigma^2}/c^{(t)})$, where $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ are parameter estimates from the latent space model and α_1, α_2 are fitted from a regression model: $Y_{i_{IFLS}}^{(t)} = \alpha_1 \hat{\mu}_i + \alpha_2 f_i^{(t)} + \epsilon_i$. $c^{(t)}$ serves to reduce the variance. The current values are 2, 2.3, 2.5, which were chosen when the wage increase pattern of simulated data is similar to the real one. These changes are somewhat ad hoc and may need further justification in the future.

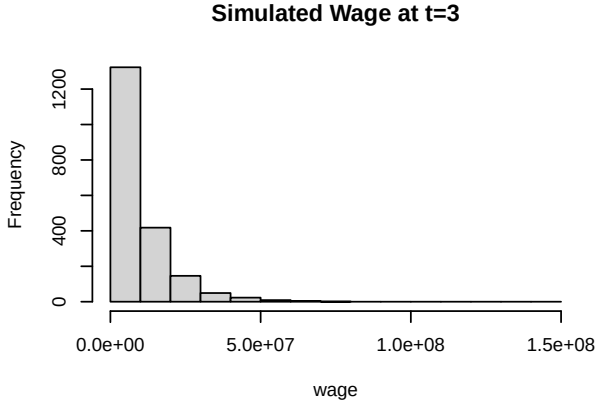
2 Comparison between Simulated and IFLS Data

2.1 Comparison of Wage

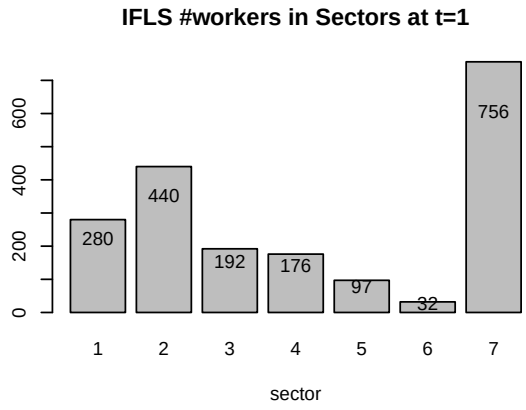


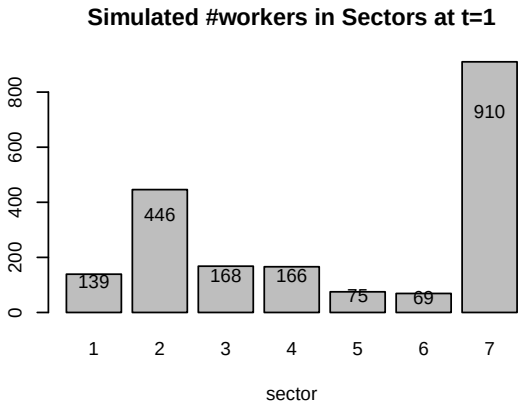


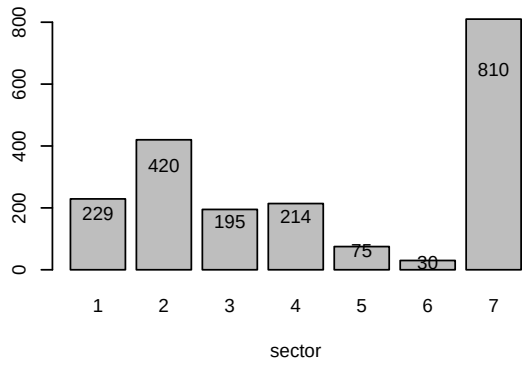


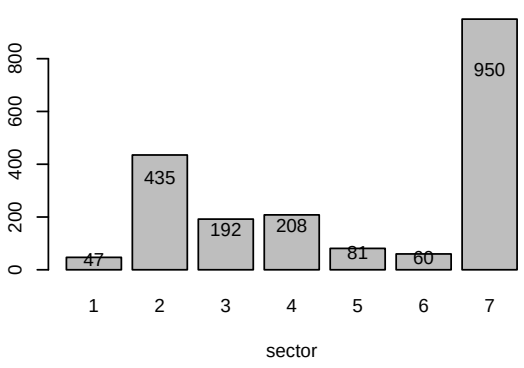


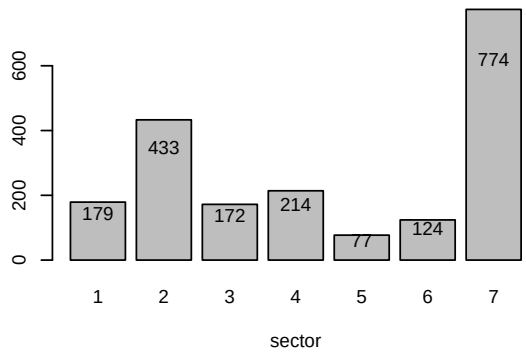
2.2 Comparison of Sector Distribution

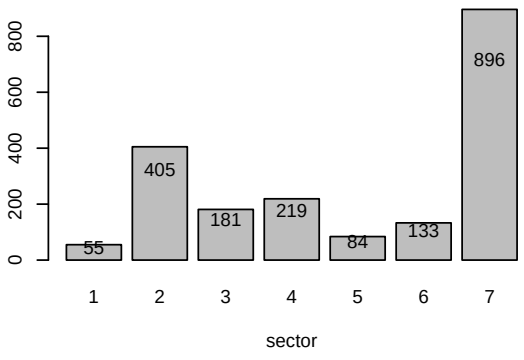




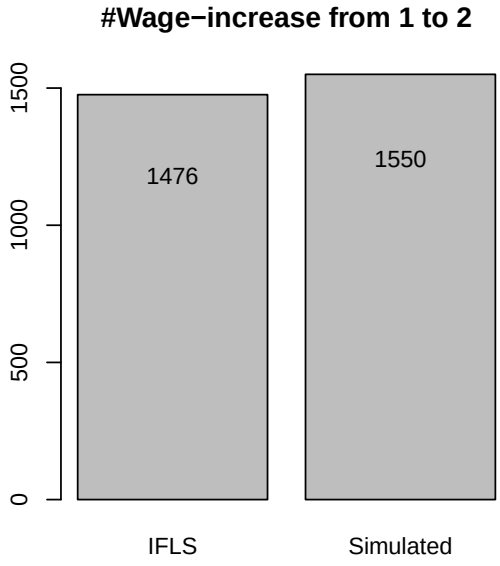


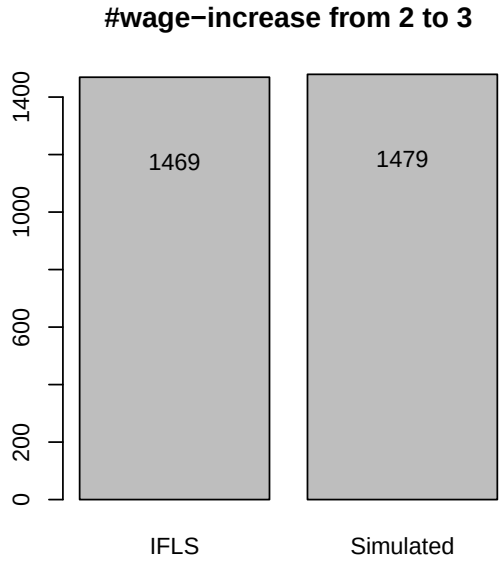


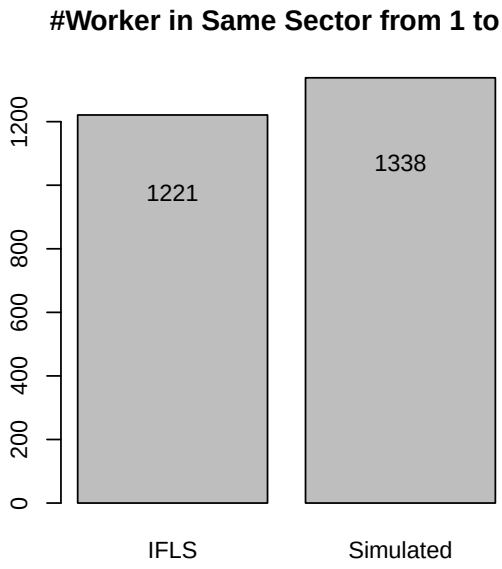


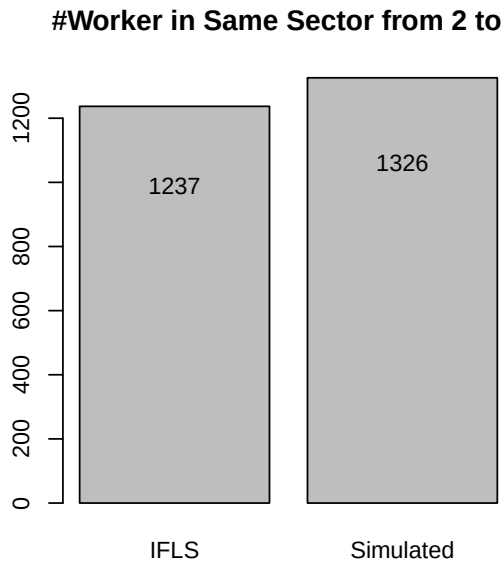


2.3 Comparison of Number of Workers with Wage Increase and Same Sector









3 Question

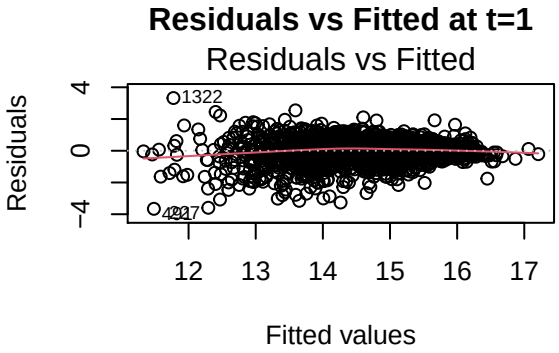
In the regression model between log wage and worker and firm effect, heteroschedasticity exists while the variance parameters from latent space model are rather stable for different workers. Is this reasonable?

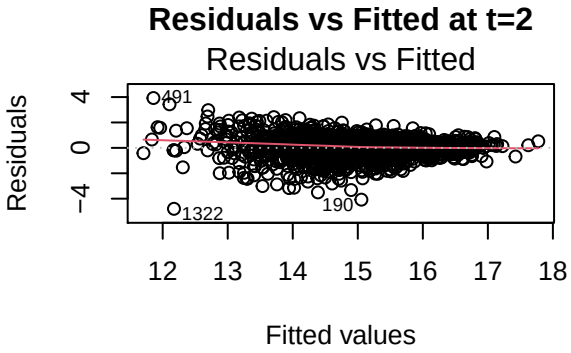
Here are some diagnostic plot and test result showing the non-constantness of variance:

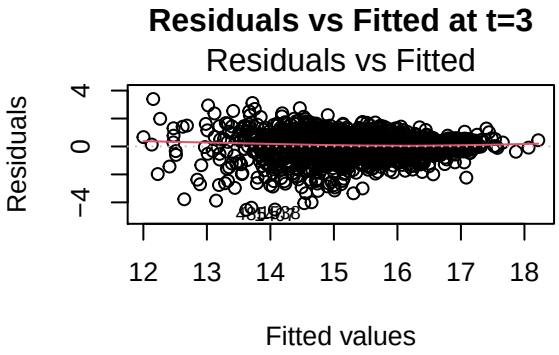
```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 807.32, Df = 1, p = < 2.22e-16
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 807.32, Df = 1, p = < 2.22e-16
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 807.32, Df = 1, p = < 2.22e-16
```







Meanwhile, the variance parameter obtained from the model are relatively stable. Here is a list of variance parameter values for each time:

```
## $`t=1`
## [1] 0.7907794
##
## $`t=2`
## [1] 1.4554529 0.7907794 1.1006559 1.5248514 1.2854718 1.4679018 1.2767479
## [8] 1.7265632 1.6263506
##
## $`t=3`
## [1] 1.4554529 0.7907794 1.1006559 1.5248514 1.4679018 1.7265632 1.2767479
## [8] 1.2854718 1.6263506
```

My thought would be to try to work out a function between log wage variance and work and firm effect for us to obtain better wage samples, is it worth it?