

# ECON 717B: PS 1

Alex von Hafften

April 9, 2022

## 1. Roy Model

- (a) Why can we normalize  $\pi_1 = \pi_2 = 1$ ?

We can normalize  $\pi_1 = \pi_2 = 1$  without any loss of generality because we can change  $\mu_k$  to account for non-unit wages:

$$\begin{aligned} W_k &= \pi_k S_k \\ &= \pi_k \exp(\mu_k + \varepsilon_k) \\ &= \exp(\log(\pi_k)) \exp(\mu_k + \varepsilon_k) \\ &= \exp(\log(\pi_k) + \mu_k + \varepsilon_k) \\ &= \exp(\tilde{\mu}_k + \varepsilon_k) \end{aligned}$$

where  $\tilde{\mu}_k \equiv \log(\pi_k) + \mu_k$

We don't observe skills, so we can just define the skills distribution with a level shift that accounts for wages.

- (b) Program to compute the model

See `run.jl`. The `parameters` structure holds parameters and the data structure holds data. The `simulate_data` function creates simulated data based on the `parameters` structure you pass it.

- (c) Choose parameters such that 60 percent of simulations choose occupation 1

I found  $\theta_0 = (\mu_1 = 1.15, \mu_2 = 1.0, \sigma_1 = 0.4, \sigma_2 = 0.5, \rho = 0.25)$  result in about 60 percent of observations choosing occupation 1. From the formula for the fraction choosing occupation 1 (derived for part g), these parameters analytically result in 60.62 percent choosing occupation 1. I simulate 1,000 observations and 59.55 percent choose occupation 1.

(d) Write estimator for  $\theta$

I use maximum likelihood to derive the estimator. The maximum likelihood estimator is  $\hat{\theta}_{MLE}$ :

$$\hat{\theta}_{MLE} = \arg \max_{\theta \in \Theta} \sum_{i=1}^N \ln L_i(\theta | X_i, Z_i)$$

where  $L_i$  is the observation-level likelihood function. If the model is correctly specified (which we know holds because this is a Monte Carlo) and some regularity conditions, the  $\hat{\theta}_{MLE}$  is consistent for  $\theta_0$ :

$$\hat{\theta}_{MLE} \xrightarrow{p} \theta_0$$

I evaluate the likelihood function per observation two different ways. The first approach uses Julia's built-in approximation for the normal CDF and the second uses simulations. I am interested in both approaches because while the first is "better" (i.e., it does not introduce simulation error), it can only be done in limited circumstances (e.g., when we can evaluate the likelihood as a function of normal pdfs and cdfs).

I use a Nelder-Mead simplex optimizer to find  $\hat{\theta}_{MLE}$ . It works well for the likelihood functions using the built-in normal cdf approximation, but is much slower and performs worse for the simulation-likelihood function. This makes sense because the former is smooth and the latter has simulation error.

### Likelihood Function using Built-In Normal CDF Approximation

For clarity, I use uppercase letters for random variables and lowercase letters for realization of random variables. For each observation  $i$ , we observe their occupation choice  $d$  and their wage  $w$ . Thus, the likelihood function for each  $i$  is:

$$\begin{aligned} L_i(\theta | w, d) &= Pr(W_1 > W_2, W_1 = w | \theta)^d Pr(W_2 > W_1, W_2 = w | \theta)^{1-d} \\ &= Pr(W_1 > W_2 | W_1 = w, \theta)^d Pr(W_1 = w | \theta)^d \\ &\quad \times Pr(W_2 > W_1 | W_2 = w, \theta)^{1-d} Pr(W_2 = w | \theta)^{1-d} \end{aligned}$$

Considering each term separately:

$$\begin{aligned} Pr(W_1 = w | \theta) &= Pr(\pi_1 \exp(\mu_1 + \varepsilon_1) = w | \theta) \\ &= Pr\left(\varepsilon_1 = \log\left(\frac{w}{\pi_1}\right) - \mu_1 \mid \theta\right) \\ &= \phi\left(\frac{\log(\frac{w}{\pi_1}) - \mu_1}{\sigma_1}\right) \\ \implies Pr(W_2 = w | \theta) &= \phi\left(\frac{\log(\frac{w}{\pi_2}) - \mu_2}{\sigma_2}\right) \\ Pr(W_1 > W_2 | W_1 = w, \theta) &= Pr\left(\pi_1 \exp(\mu_1 + \varepsilon_1) > \pi_2 \exp(\mu_2 + \varepsilon_2) \mid \varepsilon_1 = \log\left(\frac{w}{\pi_1}\right) - \mu_1, \theta\right) \\ &= Pr\left(\log \pi_1 + \mu_1 + \varepsilon_1 > \log \pi_2 + \mu_2 + \varepsilon_2 \mid \varepsilon_1 = \log\left(\frac{w}{\pi_1}\right) - \mu_1, \theta\right) \\ &= Pr\left(\log \pi_1 - \log \pi_2 + \mu_1 - \mu_2 + \varepsilon_1 > \varepsilon_2 \mid \varepsilon_1 = \log\left(\frac{w}{\pi_1}\right) - \mu_1, \theta\right) \end{aligned}$$

where  $\phi$  is the standard normal pdf. In this case, we can use the fact that for bivariate normal distributions the conditional distribution is normal:

$$\varepsilon_2 \Big| \varepsilon_1 = a \sim N\left(\frac{\sigma_2}{\sigma_1}\rho a, (1 - \rho^2)\sigma_2^2\right)$$

So,

$$\begin{aligned} Pr(W_1 > W_2 | W_1 = w, \theta) &= \Phi\left(\frac{\log \pi_1 - \log \pi_2 + \mu_1 - \mu_2 + [\log\left(\frac{w}{\pi_1}\right) - \mu_1] - \frac{\sigma_2}{\sigma_1}\rho[\log\left(\frac{w}{\pi_1}\right) - \mu_1]}{\sqrt{1 - \rho^2}\sigma_2}\right) \\ &= \Phi\left(\frac{\log \pi_1 - \log \pi_2 + \mu_1 - \mu_2 + (1 - \frac{\sigma_2}{\sigma_1}\rho)[\log\left(\frac{w}{\pi_1}\right) - \mu_1]}{\sqrt{1 - \rho^2}\sigma_2}\right) \\ \Rightarrow Pr(W_2 > W_1 | W_2 = w, \theta) &= \Phi\left(\frac{\log \pi_2 - \log \pi_1 + \mu_2 - \mu_1 + (1 - \frac{\sigma_1}{\sigma_2}\rho)[\log\left(\frac{w}{\pi_2}\right) - \mu_2]}{\sqrt{1 - \rho^2}\sigma_1}\right) \end{aligned}$$

where  $\Phi$  is the standard normal cdf. So, the likelihood function boils down to:

$$\begin{aligned} L_i(\theta | w, d) &= \left[\phi\left(\frac{\log\left(\frac{w}{\pi_1}\right) - \mu_1}{\sigma_1}\right)\right]^d \left[\Phi\left(\frac{\log \pi_1 - \log \pi_2 + \mu_1 - \mu_2 + (1 - \frac{\sigma_2}{\sigma_1}\rho)[\log\left(\frac{w}{\pi_1}\right) - \mu_1]}{\sqrt{1 - \rho^2}\sigma_2}\right)\right]^d \\ &\quad \times \left[\phi\left(\frac{\log\left(\frac{w}{\pi_2}\right) - \mu_2}{\sigma_2}\right)\right]^{1-d} \left[\Phi\left(\frac{\log \pi_2 - \log \pi_1 + \mu_2 - \mu_1 + (1 - \frac{\sigma_1}{\sigma_2}\rho)[\log\left(\frac{w}{\pi_2}\right) - \mu_2]}{\sqrt{1 - \rho^2}\sigma_1}\right)\right]^{1-d} \end{aligned}$$

### Simulation-Based Likelihood Function

Given  $\theta$ , I simulate  $N$  observations. I split the simulations based occupation choice and I estimate a kernel density for the wages in each occupation. Then, the likelihood of each observation is:

$$L_i(\theta | w, d) = \left[\frac{N_1}{N} kde_1(w)\right]^d \left[\frac{N_2}{N} kde_2(w)\right]^{1-d}$$

where  $N_k$  is the number of the simulations that choose occupation  $k$  and  $kde_k(w)$  is the estimate of the kernel density at  $w$  for occupation  $k$ . Importantly, I fix the random seed as I change parameters, so the  $\varepsilon_k/\sigma_k$  are the same across calls by the optimizer.

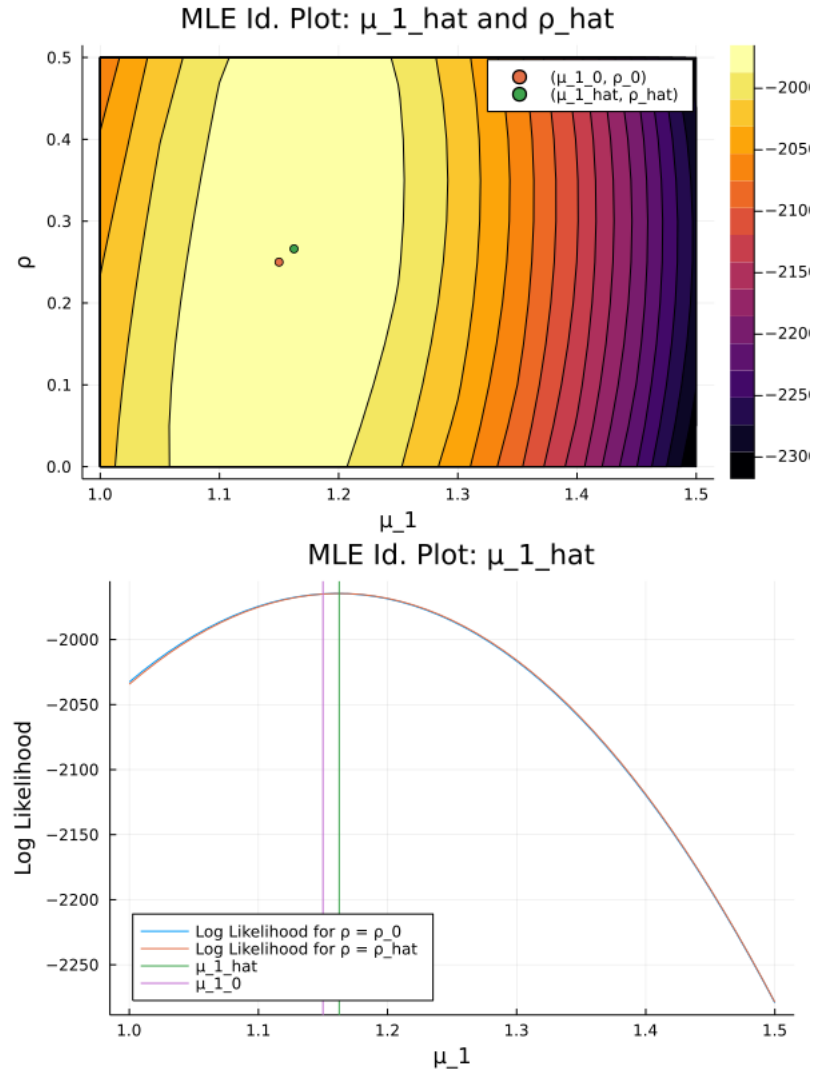
- (e) Write code to estimate  $\theta$  using estimator

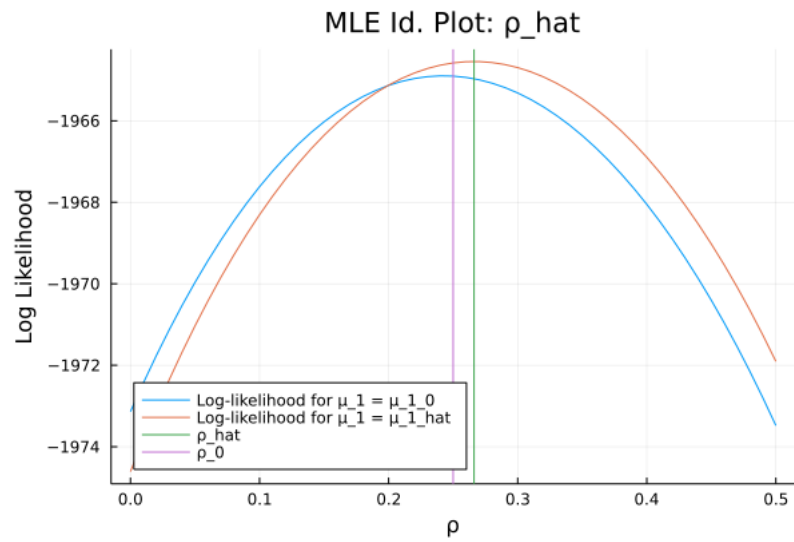
See `model.jl`. The function `likelihood` evaluates the observation-level likelihood based using the built-in normal CDF approximation. The function `log_likelihood` evaluates the full log-likelihood. It has an option to run it using each method for evaluating likelihood. The function `mle` runs the optimization step. I use a scaled and shifted logistic function to make sure  $\rho$  is confined between -1 and 1.

(f) Plot identification figures

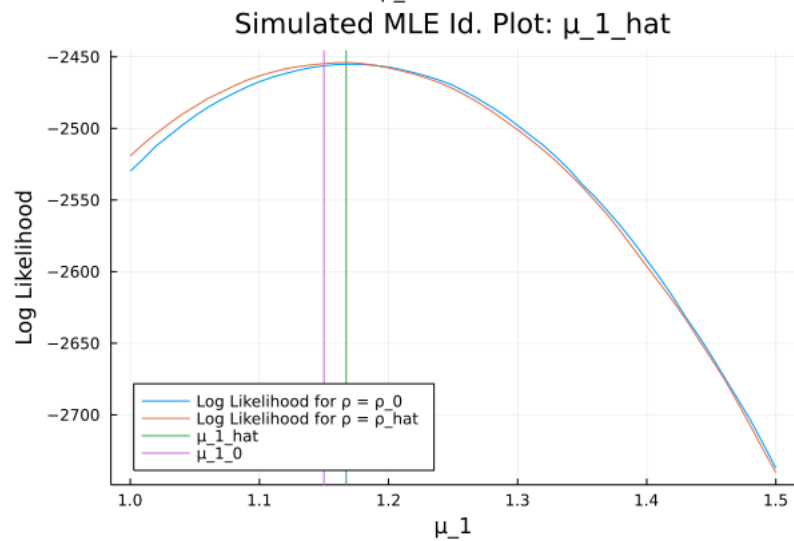
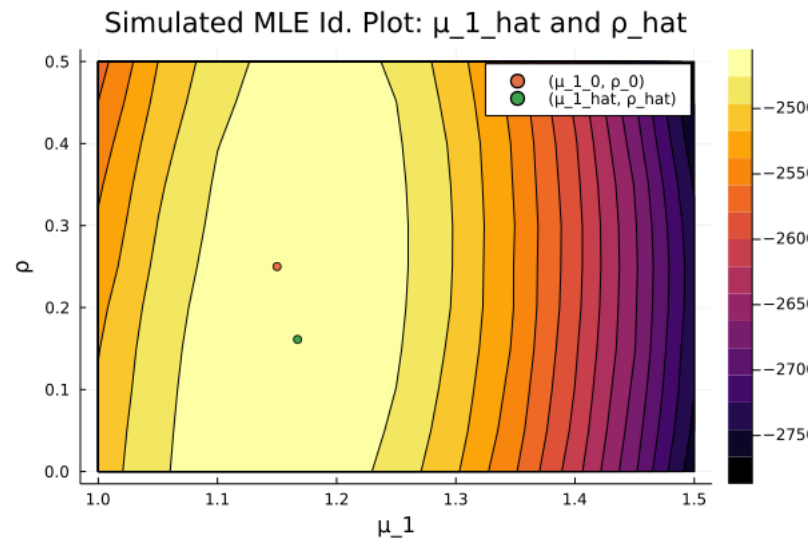
I plot identification figures for each method of computing likelihood. I plot a contour figure where  $\mu_1$  changes on the x-axis and  $\rho$  changes on the y-axis and the color indicates the log-likelihood. The orange dot is the true value of the parameters and the green dot is the ML estimates. In addition, I plot the log-likelihood as we change one parameter and holding the other fixed at both its “true” value (blue line) or the MLE estimated value (orange line). The true value of the parameter is shown by the purple line and the ML estimate is the green line. We can see that  $\mu_1$  is more sharply identified than  $\rho$ . The log-likelihood function is more curved as you change  $\mu_1$ . The  $\hat{\mu}_1$  is very close to the true value, but  $\hat{\rho}$  is further away from its true value. As expected, the simulation-based likelihood function added noise. The effect for  $\mu_1$  is not so bad, but there is much more noise and a much less accurate estimate for  $\rho$ .

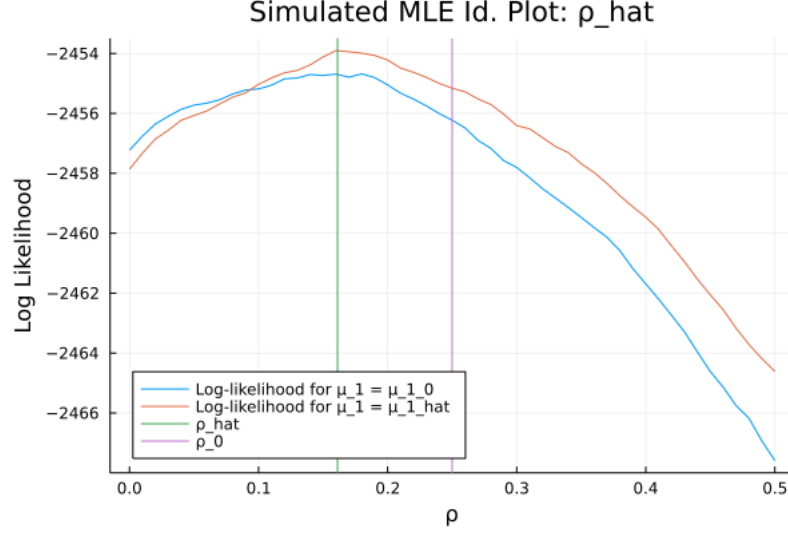
**Likelihood Function using Built-In Normal CDF Approximation**





### Simulation-Based Likelihood Function





(g) Table of parameters and model fit

The fraction of agents that choose occupation 1 implied by the model is:

$$\begin{aligned}
 Pr(W_1 > W_2) &= Pr(\pi_1 \exp(\mu_1 + \varepsilon_1) > \pi_2 \exp(\mu_2 + \varepsilon_2)) \\
 &= Pr(\log(\pi_1) + \mu_1 + \varepsilon_1 > \log(\pi_2) + \mu_2 + \varepsilon_2) \\
 &= Pr(\varepsilon_2 - \varepsilon_1 < \log(\pi_1) + \mu_1 - \log(\pi_2) - \mu_2) \\
 &\text{where } \varepsilon_2 - \varepsilon_1 \sim N(0, \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2) \\
 \Rightarrow Pr(W_1 > W_2) &= \Phi\left(\frac{\log(\pi_1) + \mu_1 - \log(\pi_2) - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}}\right)
 \end{aligned}$$

For the other model fit moments, I just simulate a 100,000,000 observations for each parameter vector and compute the data moments because I was running into troubles deriving an explicit formula.<sup>1</sup> I tried simulating 20 percent more simulations and my estimates did not change, so I think this number of observations were sufficient for at least three significant digits.

The table asked for in the problem set is below. The first column is the simulated data moment and the second column are these moments at the parameters estimated with MLE. Everything is quite close.

<sup>1</sup>My progress on the expected value of observed  $W_1$ :

$$\begin{aligned}
 E[W_1|W_1 > W_2] &= E[\pi \exp(\mu_1 + \varepsilon_1)|W_1 > W_2] \\
 &= \pi \exp\left(\mu_1 + E[\varepsilon_1|W_1 > W_2] + \frac{1}{2}V[\varepsilon_1|W_1 > W_2]\right) \\
 E[\varepsilon_1|W_1 > W_2] &= E[\varepsilon_1|\pi_1 \exp(\mu_1 + \varepsilon_1) > \pi_2 \exp(\mu_2 + \varepsilon_2)] \\
 &= E[\varepsilon_1|\log \pi_1 + \mu_1 + \varepsilon_1 > \log \pi_2 + \mu_2 + \varepsilon_2] \\
 &= E[\varepsilon_1|\varepsilon_1 > \log \pi_2 - \log \pi_1 + \mu_2 - \mu_1 + \varepsilon_2] \\
 &= \dots
 \end{aligned}$$

description	data	mle
Fraction choosing occupation 1	0.609	0.616
Average observed W1	3.881	3.867
Average observed W2	4.228	4.21
Std of observed W1	1.475	1.502
Std of observed W2	1.737	1.816

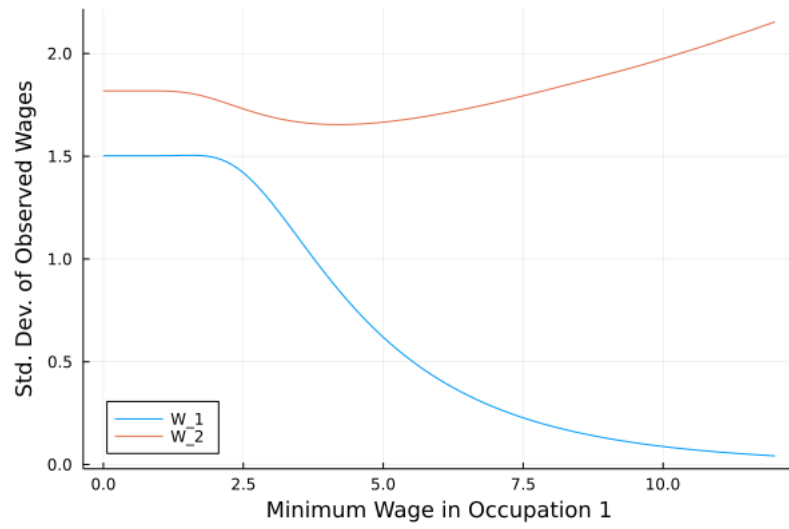
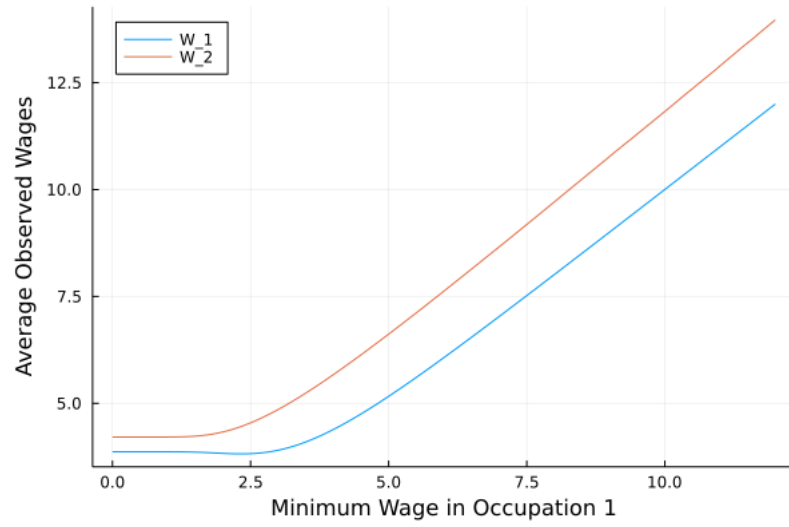
I expanded the previous table to include some addition information. The first column are these moments at the “true” parameter. The second column are the data moments. The difference between these columns stems from simulation error. The third column are the MLE estimates. The difference between these columns is due to the estimation error. The fourth column are these moments using the parameters from the simulation-based MLE estimates. The difference between these columns is due to using simulations in the estimation.

description	truth	data	mle	simulation
Fraction choosing occupation 1	0.606	0.609	0.616	0.612
Average observed W1	3.837	3.881	3.867	3.937
Average observed W2	4.191	4.228	4.21	4.241
Std of observed W1	1.485	1.475	1.502	1.501
Std of observed W2	1.808	1.737	1.816	1.799

(h) Minimum wage counterfactual

I run the counterfactual for a range of minimum wages in occupation 1 from 0.0 to 12.0, at which over 99.9 percent of simulations choose occupation 1. I plot the fraction choosing occupation 1, the average observed wage in each occupation, and the standard deviation of wages in each occupation. I then focus on 2.9, which is the minimum wage in occupation 1, so that 70 percent of simulation choose occupation 1.





W'1'bar	frac'choosing'1	average'W'1	average'W'2	std'W'1	std'W'2
0.0	0.616	3.867	4.212	1.502	1.818
2.9	0.707	3.878	4.786	1.309	1.698