Panel regression in R

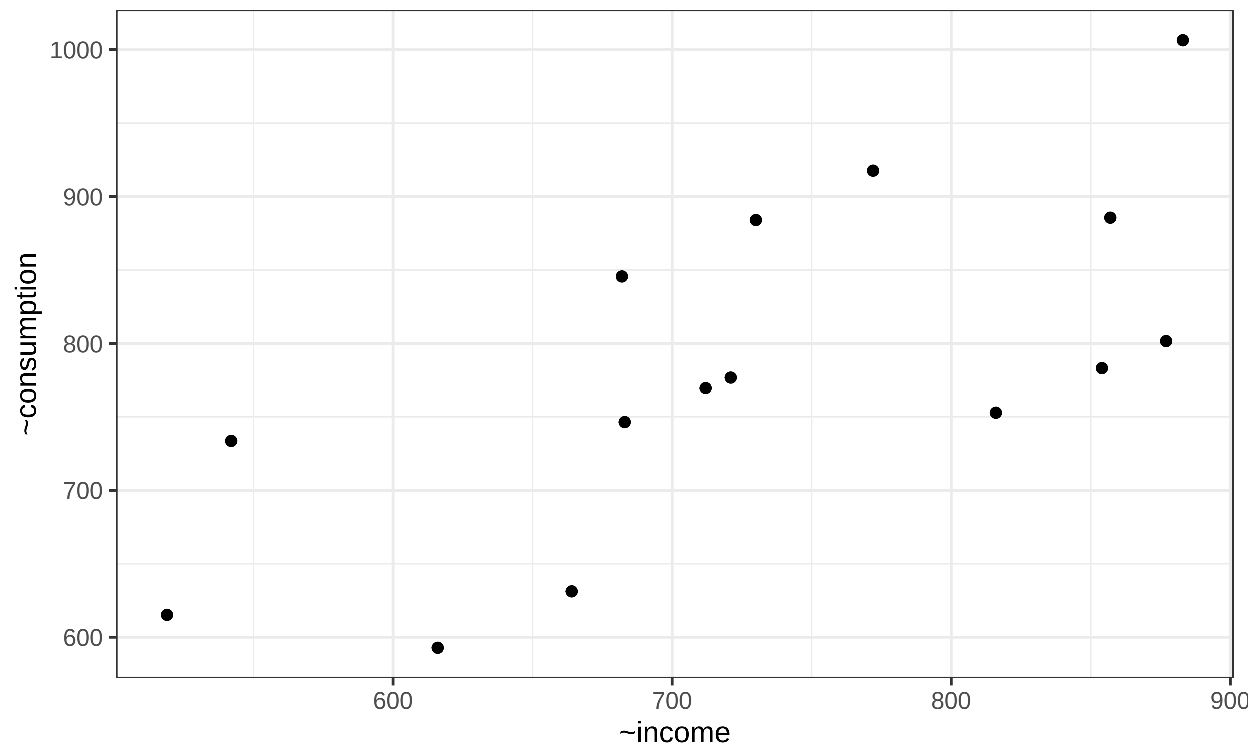
Panel regression refers to a data with a cross section as well as time dimension. The example we'll work with is the monthly expenses (as function of income)of Alice,Bob and Eve.

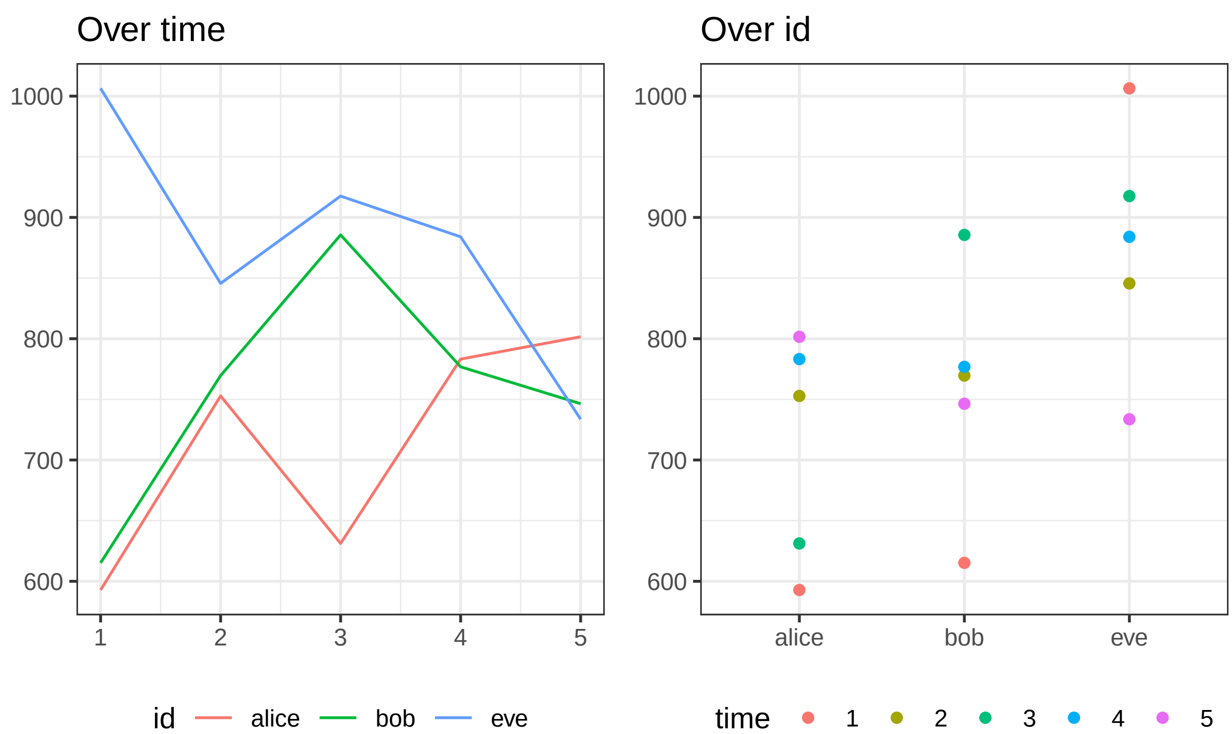
Cross section fixed effects

The model that we are looking at is:

$$Y_{it} = \alpha_i + \beta x_{it}$$

Each state has a unique level of consumption and if we don't account for it we get the wrong estimates. That is probably because






```
all.equal(coefficients(cross_section_case$pooling_reg)[2],
          coefficients(cross_section_case$individual_reg))
```

```
## [1] "Mean relative difference: 0.1697441"
```

Demeaning estimation

One method of dealing with this issue is look at the deviation from the average consumption by state. That will take care of the level difference. The “demeaning” should be performed on all the variables.

```
all.equal(coefficients(cross_section_case$demean_reg)[2],
          coefficients(cross_section_case$dummy_reg)[2],
          coefficients(cross_section_case$individual_reg))
```

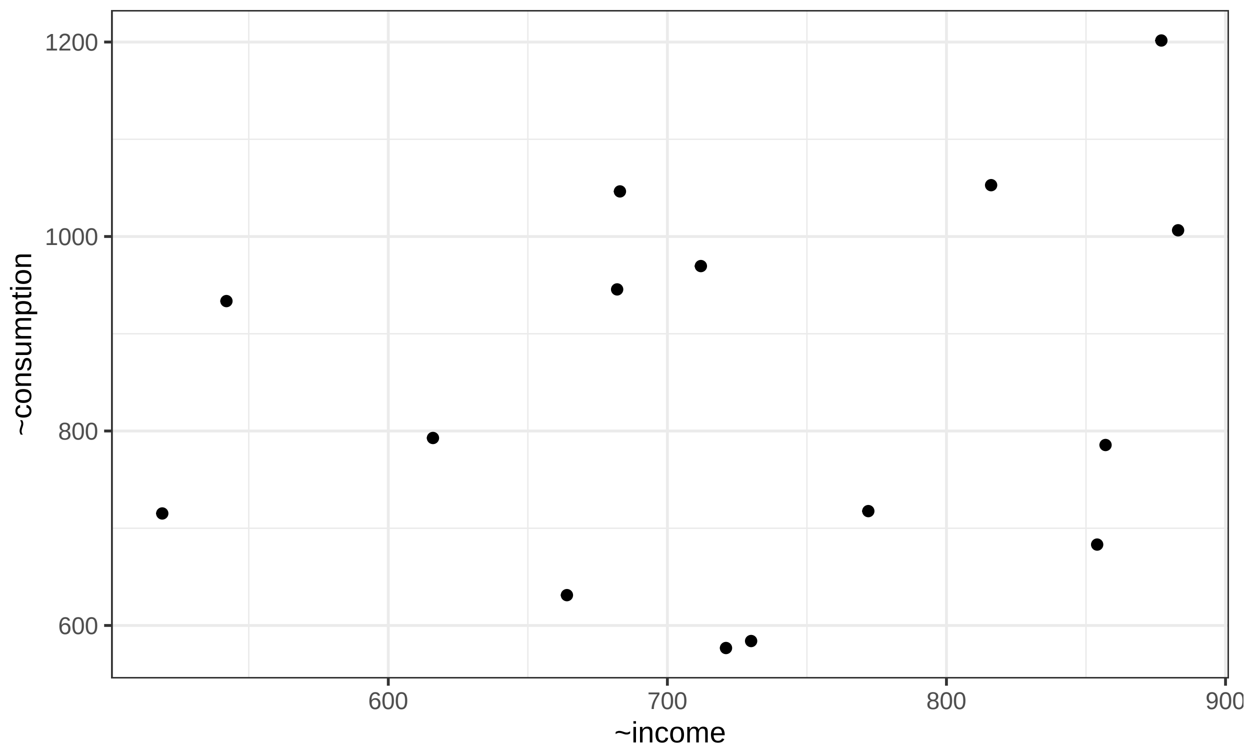
```
## [1] TRUE
```

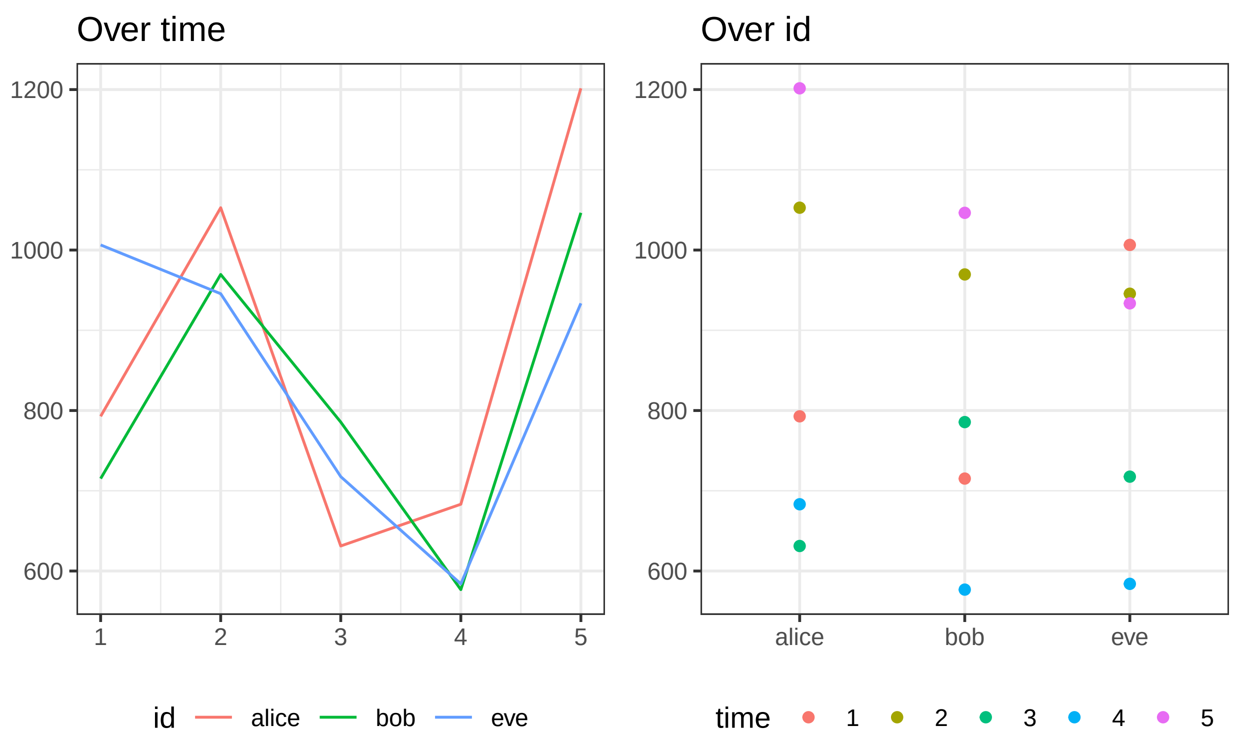
Time fixed effects

The model that we are looking at is:

$$Y_{it} = \alpha + \gamma_t + \beta x_{it}$$

Now each time period has a unique level of consumption and if we don’t account for it we get the wrong estimates.






```r
all.equal(coefficients(time_case$pooling_reg)[2],
          coefficients(time_case$time_reg))
```

```
## [1] "Mean relative difference: 0.9101244"
```

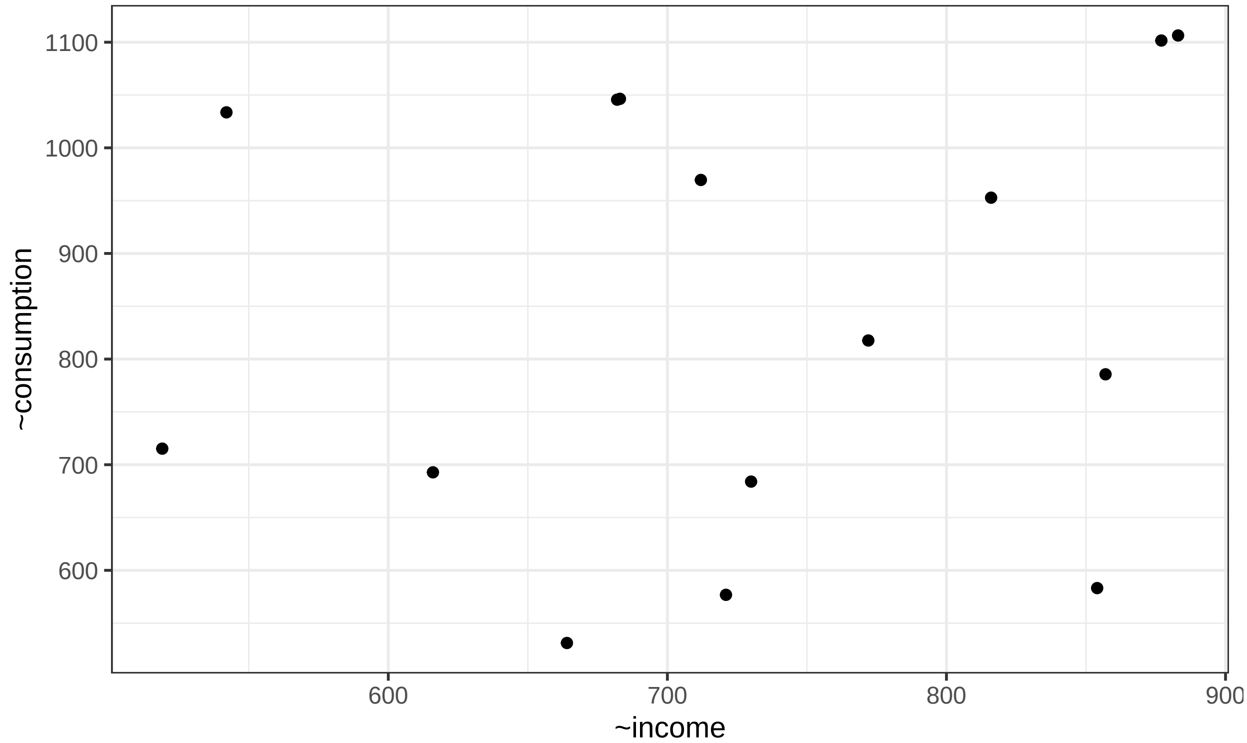
Demeaning estimation

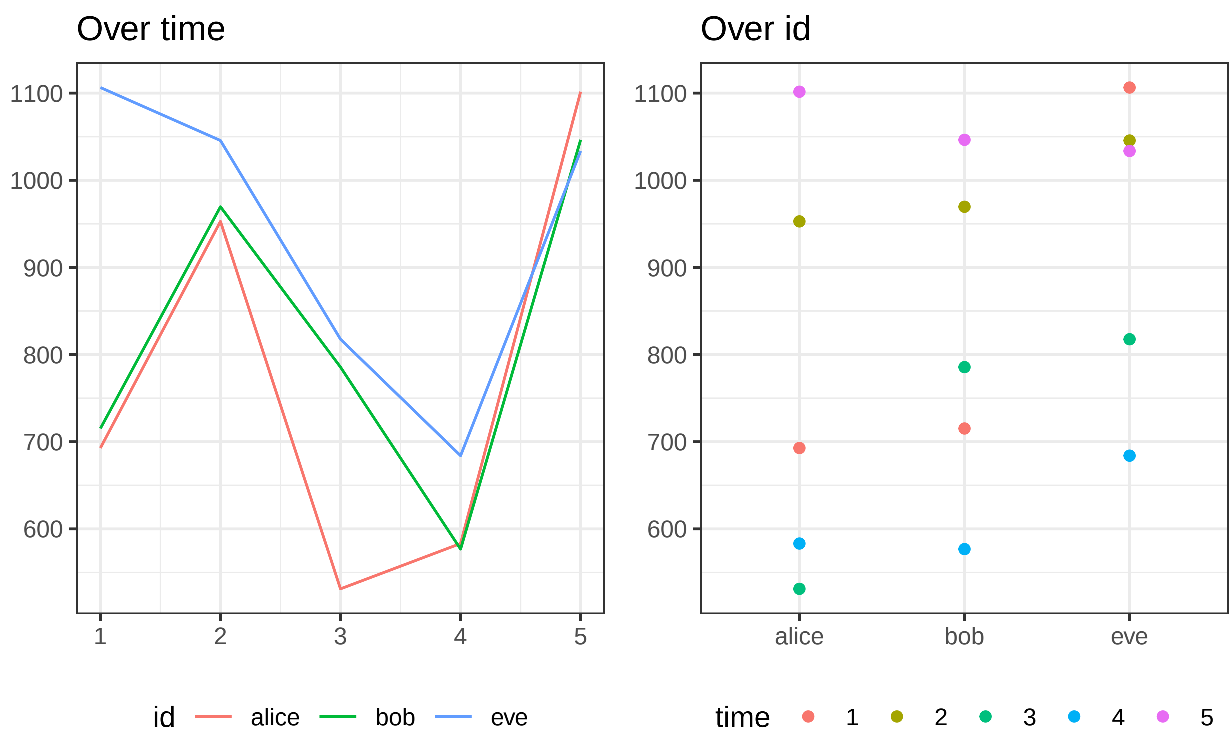
One method of dealing with this issue is look at the deviation from the average consumption by state. That will take care of the level difference. The “demeaning” should be performed on all the variables.

```
all.equal(coefficients(time_case$demean_reg)[2],
          coefficients(time_case$dummy_reg)[2],
          coefficients(time_case$time_reg))
```

```
## [1] TRUE
```

Cross section and time effect






```
all.equal(coefficients(cross_section_time_case$pooling_reg)[2],
          coefficients(cross_section_time_case$fe_reg))
```

```
## [1] "Mean relative difference: 1.642609"
```

```
all.equal(coefficients(cross_section_time_case$demean_reg)[2],
          coefficients(cross_section_time_case$dummy_reg)[2],
          coefficients(cross_section_time_case$fe_reg))
```

```
## [1] TRUE
```

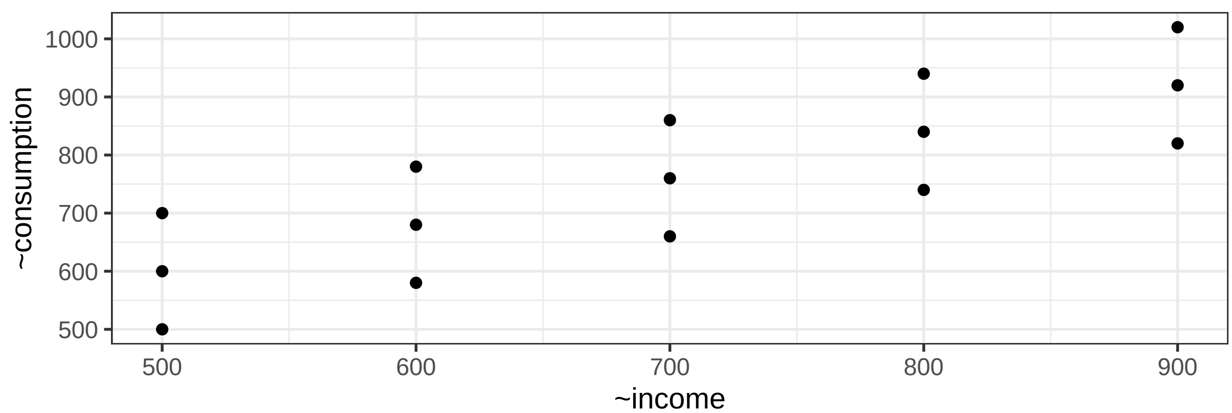
Specific plm issues

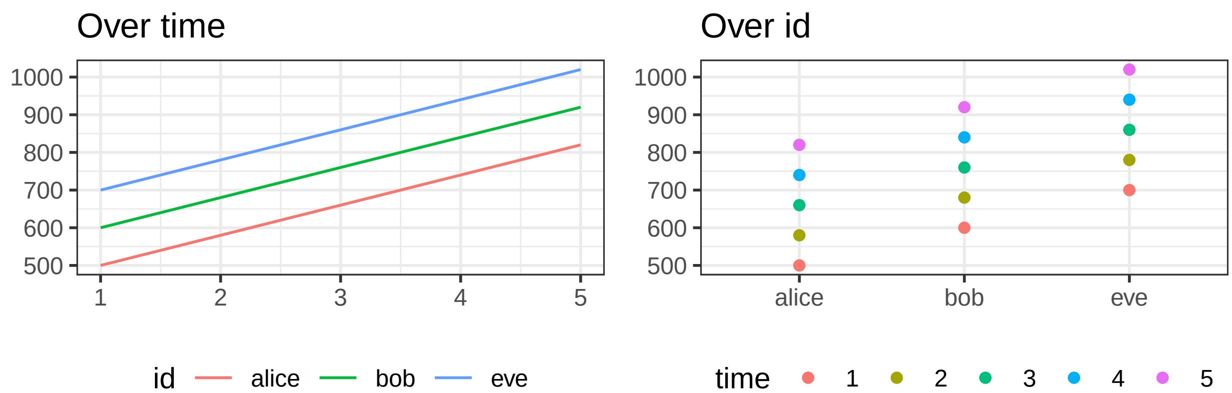
```
all.equal(coefficients(lm_reg)[c("income","income:id_fe","income:time_fe")],
          coefficients(plm_reg)[c("income","income:id_fe","income:time_fe")])
```

```
## [1] TRUE
```

Degenerate case

In case where the income (independent var) is the same for all the individuals we'll get an identity between income levels and time periods (in each period) all the individuals have the same income. In that case pooling regression and individual effect regression will give the same coeffs.





```
all.equal(coefficients(cross_section_degenerate_case$pooling_reg)[2],
          coefficients(cross_section_degenerate_case$individual_reg))
```

```
## [1] TRUE
```

As for the time effects aspect the degenerate case will produce perfect linear combination since we'll essentially get three (by number of id's) "identical" observations for each time period.