Homework assignment 4

Question 1:

1.The best linear regression model of the explored models has the following independent variables: Age, edu, hr, married, salaried, selfempl, kid1

Check for multicollinearity was tested using VIF and COLLIN function. The variance inflation was < 10 for all the variables and the condition index was <100 for all the variables. Hence, we conclude that there is no multicollinearity in the data. R-squared value of 0.25 implying only 25% of variation in the dependent variable is explained by the independent variables and hence, this suggests that, there are many factors which explain the wage variable which are unobserved in the model as well as the data, this could possibly suggest unobserved heterogeneity.

Model: MODEL1
Dependent Variable: lwage

Number of Observations Read	1002
Number of Observations Used	1002

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	126.23726	18.03389	47.45	<.0001
Error	994	377.81577	0.38010		
Corrected Total	1001	504.05303			

Root MSE	0.61652	R-Square	0.2504
Dependent Mean	2.59348	Adj R-Sq	0.2452
Coeff Var	23.77192		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	1.32637	0.14999	8.84	<.0001	.	0
age	1	0.01088	0.00199	5.47	<.0001	0.95296	1.04936
edu	1	0.06704	0.00718	9.34	<.0001	0.85259	1.17289
hr	1	-0.00014204	0.00002510	-5.66	<.0001	0.86876	1.15107
married	1	0.18367	0.07327	2.51	0.0123	0.95109	1.05143
salaried	1	0.29607	0.04405	6.72	<.0001	0.79809	1.25300
selfempl	1	-0.34544	0.05182	-6.67	<.0001	0.89089	1.12248
kid1	1	-0.09945	0.04725	-2.10	0.0356	0.97266	1.02811

Collinearity Diagnostics

Number	Eigenvalue	Condition Index	Proportion of Variation							
			Intercept	age	edu	hr	married	salaried	selfempl	kid1
1	5.73711	1.00000	0.00049397	0.00143	0.00111	0.00310	0.00216	0.00715	0.00568	0.00663
2	0.84945	2.59883	0.00003378	0.00008669	0.00014815	0.00001022	0.00001764	0.12145	0.63757	0.00818
3	0.75725	2.75250	0.00005512	0.00037008	0.00038119	0.00109	0.00003313	0.01365	0.03788	0.90922
4	0.42474	3.67522	0.00185	0.00713	0.00104	0.00142	0.01101	0.68785	0.25321	0.04332
5	0.10995	7.22365	0.00193	0.08130	0.00735	0.79114	0.00472	0.03109	0.03433	0.00007662
6	0.06491	9.40151	0.00668	0.05624	0.07686	0.02407	0.88824	0.04658	0.00377	0.01800
7	0.04404	11.41349	0.00337	0.50763	0.43917	0.08845	0.01383	0.04155	0.00225	0.00153
8	0.01256	21.37406	0.98558	0.34581	0.47394	0.09071	0.07999	0.05069	0.02531	0.01304

Heteroscedasticity Test					
Equation	Test	Statistic	DF	Pr > ChiSq	Variables
lwage	White's Test	127.5	31	<.0001	Cross of all vars

2. Age variable tends to have a non-linear effect of wage. It suggests that wage will start to decrease after reaching the saturation point for age and hence, we can see this effect on the coefficient of agesq variable in the model.

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Tolerance	Variance Inflation
Intercept	1	0.52572	0.41231	1.28	0.2026	.	0
age	1	0.05133	0.01951	2.63	0.0086	0.00987	101.32239
edu	1	0.06551	0.00720	9.10	<.0001	0.84371	1.18524
hr	1	-0.00014464	0.00002509	-5.77	<.0001	0.86662	1.15391
married	1	0.18256	0.07314	2.50	0.0127	0.95104	1.05149
salaried	1	0.29346	0.04399	6.67	<.0001	0.79744	1.25402
selfempl	1	-0.34821	0.05175	-6.73	<.0001	0.89030	1.12322
kid1	1	-0.10529	0.04726	-2.23	0.0261	0.96924	1.03174
agesq	1	-0.00046560	0.00022340	-2.08	0.0374	0.00984	101.58725

Collinearity Diagnostics											
			Proportion of Variation								
Number	Eigenvalue	Condition Index	Intercept	age	edu	hr	married	salaried	selfempl	kid1	agesq
1	6.57307	1.00000	0.00004979	0.00001162	0.00083296	0.00232	0.00164	0.00529	0.00418	0.00476	0.00003479
2	0.85064	2.77979	0.00000365	9.813314E-7	0.00012504	6.458231E-7	0.00000741	0.11630	0.63654	0.01516	0.00000344
3	0.76480	2.93164	0.00000372	0.00000381	0.00022168	0.00054047	0.00000259	0.00499	0.04033	0.90511	0.00002578
4	0.46172	3.77307	0.00009783	0.00006251	0.00007935	0.00028451	0.00459	0.63588	0.22347	0.01636	0.00036334
5	0.19115	5.86406	0.00030435	0.00020749	0.00594	0.19574	0.02426	0.09639	0.05099	0.02583	0.00380
6	0.07990	9.07024	0.00086737	0.00003518	0.04382	0.70597	0.27199	0.00800	0.01359	0.01470	0.00181
7	0.06162	10.32813	0.00254	0.00002997	0.25955	0.02921	0.57704	0.06259	0.00180	0.00616	0.00046780
8	0.01676	19.80372	0.07402	0.00156	0.68496	0.06557	0.12045	0.06888	0.02772	0.00979	0.00524
9	0.00034803	137.42746	0.92211	0.99809	0.00446	0.00035825	0.00001619	0.00166	0.00138	0.00212	0.98824

3. Coefficients for the same model from 1.2 with FixOne, FixTwo, RanOne, RanTwo (significant variables highlighted in bold)

Variable/Model	FixOne	Fixtwo	RanOne	RanTwo
Intercept	-0.87079	**4.655149**	0.392766	0.392882
age	**0.129034**	0	**0.069502**	**0.069495**
edu	0	0	**0.072589**	**0.072588**
hr	**-0.00031**	**-0.00031**	**-0.00026**	**-0.00026**
married	0.125309	0.126931	**0.14445**	**0.144458**
salaried	**0.116997**	**0.118684**	**0.197964**	**0.197988**
selfempl	**-0.21552**	**-0.21658**	**-0.26199**	**-0.262**
kid1	**-0.13127**	**-0.13123**	**-0.12255**	**-0.12255**
agesq	**-0.00135**	**-0.00134**	**-0.00069**	**-0.00069**

4. As per our findings, when running a simple linear regression model, an overall of 24% is the R-squared value, which is a very small variation in the independent variables to estimate the wage and as mentioned earlier, there could be unobserved heterogeneity in the model. A case of unobserved heterogeneity could be ability, which is not clearly a quantified variable. The age as well as the education level has an overall positive effect in the wage and is significant. This aligns with the economic theory. The age causes the wage to increase by approximately 1%, whereas the education level increases the wage by 6.7%. The hours of work put in has an effect close to zero. A person who is married is probable to have a wage 18% higher than someone who is not. A person who is salaried, is probable to have a 30% higher salary than one who is not. A person who is self-employed, would have a 34% lower salary than one who is not. A person with one kid would have a 9% decrease in salary than one who does not. All the variables in the model are significant in the case of the simple linear regression model. By running the fixed effects model, we remove the case of the unobserved heterogeneity in the model and we observe an R-squared value of 87%, which is a significant increase from that of the simple linear regression model. Education has no effect because it is a time invariant variable. Being married is highly insignificant in this case because it is a slow changing variable over time and slow changing variables are not very significant in estimating dependent variables in a fixed effects model. The age causes the wage to increase by a value of 13% and is highly significant. Being salaried causes the wage to increase by 11%, which does not make sense as again it could be a slow changing variable over time. Considering the random effects in this case does not make much sense as the data is a balanced panel data model and random effects does not work very well for a balanced panel data model and the Hausmann test shows us a very low P-value which clearly shows that the fixed effects model is the best in this case and makes the most sense.

In the case of multicollinearity, VIF and COLLIN results suggest that there are no signs of multicollinearity in the data. From the White test and Breusch-Pagan test, it is evident that heteroscedasticity exists in the data.

5.The observation after running the panel data models gives a better insight as to how the parameters are changing, after considering the cross-sectional as well as the time-series heterogeneity.

When comparing the fixed effect one and linear regression model, we can see that fixed effects model has low t-values than linear model. Also, fixed effects model explains 87% of the variation in wages in comparison to 25% in linear model. Hours worked, selfempl, kid1 and agesq variables seem to be having a negative effect on both the models. The estimates of age and kid1 variable are bigger in fixed effects model than linear model and all other variables have estimates smaller in the case of fixed effects.

Number of hours worked has a very negligible effect on the wage and this is not very sensible because from a economic standpoint, there should be a positive effect. The effect of being married causes the wage to increase by a value of 12.5% and it's not significant in fixed effects model.

An approximate 6% change is observed between fixed and random effects model for age variable. In the case of education is having no effect because it's a time invariant variable and this effect is captured in the random effects model, where we see a 7% change from the fixed effects model. Observed effect of marriage between fixed and random effects model is approximately 2%. Salaried employee effect is approximately 8% between panel data models. Self-employed effect is approximately 5%, having one kid effect is approximately 1% between panel data models.

6. The education variable can only be estimated by the case of the linear regression model because in the case of the fixed effects model, as discussed earlier, there will be no change as it is a time invariant variable and the effect is seen in the random effects model, but the random effects model is not the best model here. In the simple linear regression model, we see that the education causes the wage to increase by 6.7% and in the case of the random effects model, it increases the wage by 6.9%. There is hardly a change of 0.2%.

Q2.1) As per the model, it is observed that being a pioneer in the market causes a positive effect (7.17) on the market share and it is very significant. The effect of market share across different models has been explored based on the 9 hypotheses below:

(i) H1: Higher product quality increases market share and market pioneers tend to have a higher product quality.
As per our finding from the running the regression values, it is observed that product quality has a significant effect on Market share (0.51) and it is seen that the being a market pioneer doesn't necessarily increase the product quality, but the variable is not that significant. Thus, there is some evidence to support H1, where we see a Market share increase of 0.51 percentage points.

(ii) H2: Broader product lines increase market share and market pioneers tend to have broader product lines.
The model suggests that the product line breadth has a negative impact on the market share which is ludicrous and at the same time, the effect of being a market pioneer on product line breadth is highly insignificant and does not really answer the question at hand.

(iii) H3: Pioneers have higher market share in industries where product has a high purchase frequency because of distribution advantages.
The model produces a very insignificant effect of the pioneer's high purchase frequency on the Market Share and hence, H3 cannot be supported.

(iv) H4: Pioneers tends to have higher market share with intensive and higher marketing expenditure.
The parameter for the relative market expenditure is negative and at the same time, not significant at a 5% level and again, there is weak support for H4.

(v) H5: Lower prices increase market share and market pioneers tend to have lower prices
The price parameter in the model is insignificant and hence there is not much support for H5.

(vi) H6: Considering absolute cost advantages, pioneers have lower relative direct costs leading to a stronger relative marketing mix and higher market share
The effect of being a pioneer on relative direct costs has an insignificant effect. Moving to the case of the marketing mix, the relative direct cost does not have the intended effect on the product quality, where it is showing a positive effect. The relative direct cost does have the intended negative effect on the product line breadth and is statistically significant and hence, there is not enough proof to support H6.

(vii) H7: Due to scale economy advantages, pioneers have lower relative direct costs leading to stronger marketing mix and a higher market share.
The market share parameter on direct cost is insignificant and, the H7 hypotheses is not supported by enough evidence.

(viii) H8: Owing to consumer information advantages, pioneers have higher share in industries where product is purchased infrequently and lower shares in industries, where products are changed in seasonal, annual or periodic basis.

The parameter of being a pioneer and having a product with a low purchase frequency has a very insignificant effect on the market share and similarly, the effect of having a product that changes seasonally or annually is also insignificant. As a whole, there is weak evidence to support H8.

(ix) H9: A deterioration of lower product quality, product line breadth, price absolute cost advantages over time leads to lower pioneer shares.
The effect of being a 20-year pioneer on the product quality in insignificant and the similar insignificant effect can be seen for the product line breadth as well. The effect of being a 20-year pioneer also has an insignificant effect on the price as observed from the model and the similar effect is observed for the relative direct cost effect as well. On a holistic level, there is weak evidence to support H9 as well.

It can be concluded from the above hypotheses and results that the overall effect of being a pioneer is not clear and makes no sense from the above dataset provided. Further research and analysis would need to be conducted to reach a concrete conclusion for the effect of pioneering on market share for industrial goods.

Q2.2) Based on the model, the effect of being a pioneer for the industrial goods industry has an overall positive effect and is highly significant. Across the different models(2SLS and simple linear regression model), there isn't much change in the effect as most of the parameters have yielded a very insignificant result and hence, the overall effect of being a pioneer is not clearly captured and estimated.