Microeconometrics
Supervision 2

Samuel Lee

Question 1

(a)

The coefficient estimates will generally differ if taking first differences removes a latent variable that is correlated with some of the observed covariates. For instance, suppose the true data-generating process is

$$weight_{it} = \beta_0 + \beta_1 school_{it} + \beta_2 MARRIED_{it} + \beta_3 height_{it} + \beta_4 age_{it} + \beta_5 BLACK_{it} + \alpha_i + \varepsilon_{it}$$

where $\text{Cov}[MARRIED_{it}, \alpha_i] \neq 0$, for example if α_i represents some genetic predisposition to weight gain which also reduces one's chances of getting married. Then the OLS coefficient estimates will all be biased and inconsistent unless $MARRIED_{it}$ is orthogonal to all other observables. On the other hand, taking first differences gives us

$$\Delta weight_{it} = \beta_1 \Delta school_{it} + \beta_2 \Delta MARRIED_{it} + \beta_3 \Delta height_{it} + \beta_4 \Delta age_{it} + \beta_5 \Delta BLACK_{it} + \Delta\varepsilon_{it}$$

in which case the coefficient estimates are unbiased and consistent. Therefore $\mathbb{E}[\boldsymbol{\beta}_{OLS} - \boldsymbol{\beta}_{FD}] \neq \mathbf{0}$ and $\boldsymbol{\beta}_{OLS} - \boldsymbol{\beta}_{FD} \overset{P}{\nrightarrow} \mathbf{0}$.

On the other hand, the estimates of $\text{Var}[\varepsilon_{it}]$ and $\text{Cov}[\varepsilon_{it}, \varepsilon_{it-1}]$ (which are also parameters of the model) will typically differ. For example, if ε_{it} is i.i.d. white noise, then $\text{Var}[\Delta\varepsilon_{it}] = 2\text{Var}[\varepsilon_{it}]$, and $\hat{\sigma}^2_{FD}/\hat{\sigma}^2_{OLS} \xrightarrow{P} 2$.

(b)

This might be due to behavioural factors: as mentioned, if α_i is some genetic condition which is associated with weight gain and also happens to reduce one's chances of getting married, then we would observe a negative correlation between marriage and body weight. Likewise, the genetic condition might hamper one's ability to study. This means that the OLS regression picks up this negative correlation between schooling and marriage status with the unobserved α_i.

There could also be practical reasons for this. It might just be that there is not enough time variation in years of schooling and marriage status, which makes the standard errors of the coefficient estimates large. If there is very little true variation in these two variables, and the two variables are measured imperfectly, it is possible that most of the variation observed in first differences are due to measurement error, which creates a severe attenuation bias.

Height and BLACK are excluded since one would expect little to no variation in these two variables, and including them would lead to multicollinearity; their coefficients are unidentifiable.

(c)

This would lead to multicollinearity if we did not make any adjustments: the intercept (which is really equivalent to a dummy variable equal to 1 for all observations) is a multiple of the change in age. But an intercept was excluded from the FD regression, so there is no issue here and age acts like a time trend.

(d)

There might be reasons to prefer the FD regression (for example, if our concerns regarding individual fixed effects were legitimate), but the R^2 can be high simply because we've differenced out most of the interesting variation in the data. Only the estimated coefficient on age is statistically significant, it is difficult to say if this is because the other variables truly have negligible causal effects, or because there is too little variation in observables. We would usually need more contextual knowledge before deciding which specification is better.

(e)

Instead of an FD regression, we could include a dummy variable for each individual and test if they are jointly significant, for example with an F-test. This is manageable if the number of individuals is small; it may not be feasible with large n since the number of estimated parameters grows linearly in n, and the operations required to recover the coefficient estimates could be computationally difficult.

Question 2

(a)

Using the command `reg lnhr lnwg`, we estimate the equation using pooled OLS without specifying individual-specific intercepts. We get

	Dependent variable $\ln hr_{it}$
$\ln wg_{it}$	0.0827^{***}
	(0.00913)
constant	7.442^{***}
	(0.0241)
N	5320

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Even if α_i is uncorrelated with all the explanatory variables, estimating the equation using pooled OLS with common intercept will lead to serial correlation in the composite errors $v_{it} = \alpha_i + \varepsilon_{it}$, since

$$\mathrm{Cov}[v_{it}, v_{it-1}] = \mathrm{Cov}[\alpha_i + \varepsilon_{it}, \alpha_i + \varepsilon_{it-1}] = \sigma_\alpha^2$$

which leads to biased and inconsistent standard errors.

(b)

The results are reported below:

	Pooled OLS	Fixed effects	Random effects
$\ln wg_{it}$	0.0827^{***}	0.168^{***}	0.119^{***}
	(0.00913)	(0.0189)	(0.0136)
constant	7.442^{***}	7.220^{***}	7.346^{***}
	(0.0241)	(0.0493)	(0.0364)
N	5320	5320	5320

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

All three estimates are statistically significant. The fixed effects and random effects estimates are larger than the pooled OLS estimate, and also have larger standard errors. The constant reported for the fixed effects model is the average value of the estimated fixed effects.

(c)

Again, the composite error in the equation is $\nu_{it} = \alpha_i + \varepsilon_{it}$. The $T \times T$ covariance matrix of ν_{it} is

$$\Omega_{OLS} = \begin{pmatrix} \sigma_\alpha^2 + \sigma_\varepsilon^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\varepsilon^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 + \sigma_\varepsilon^2 \end{pmatrix}$$

We require spherical errors for the estimated standard errors to be consistent, so we must have $\sigma_\alpha^2 = 0$. In that case α_i is constant across all i, and acts just like a normal intercept.

From the Frisch–Waugh–Lovell theorem, the fixed effects estimates can be recovered by running a regression of the demeaned dependent variable on the demeaned regressors, and the standard errors are equal to that with a dummy variable regression after making a degrees-of-freedom correction. The demeaned errors are $\nu_{it} - \bar{\nu} = \varepsilon_{it} - \bar{\varepsilon} = \varepsilon_{it} - \frac{1}{T}\sum_{\tau=1}^{T} \varepsilon_{i\tau}$, with the following covariance matrix

$$\Omega_{FE} = \begin{pmatrix} \left(1 - \frac{1}{T}\right)\sigma_\varepsilon^2 & -\frac{1}{T}\sigma_\varepsilon^2 & \cdots & -\frac{1}{T}\sigma_\varepsilon^2 \\ -\frac{1}{T}\sigma_\varepsilon^2 & \left(1 - \frac{1}{T}\right)\sigma_\varepsilon^2 & \cdots & -\frac{1}{T}\sigma_\varepsilon^2 \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{T}\sigma_\varepsilon^2 & -\frac{1}{T}\sigma_\varepsilon^2 & \cdots & \left(1 - \frac{1}{T}\right)\sigma_\varepsilon^2 \end{pmatrix}$$

which tends to $\sigma_\varepsilon^2 \boldsymbol{I}_T$ as T goes to infinity. Therefore, the standard errors for the fixed effects estimators are consistent in T, but the above assumes ε_{it} is homoscedastic and serially uncorrelated.

We can get the GLS estimators by pre-multiplying the equation by $\Omega_{OLS}^{-\frac{1}{2}}$ (or an estimate of $\Omega_{OLS}^{-\frac{1}{2}}$ if σ_α^2 and/or σ_ε^2 are unknown) and estimating the new equation by OLS. The covariance matrix of the resultant errors is

$$\mathrm{Cov}[\Omega_{OLS}^{-\frac{1}{2}} \boldsymbol{\nu}_i] = \mathbb{E}\left[\Omega_{OLS}^{-\frac{1}{2}} \boldsymbol{\nu}_i \left(\Omega_{OLS}^{-\frac{1}{2}} \boldsymbol{\nu}_i\right)'\right] = \Omega_{OLS}^{-\frac{1}{2}} \mathbb{E}[\boldsymbol{\nu}_i \boldsymbol{\nu}_i'] \Omega_{OLS}^{-\frac{1}{2}} = \boldsymbol{I}_T$$

using the fact that Ω_{OLS} is symmetric. Therefore, the standard errors for the GLS estimators are consistent, but also only if ε_{it} is homoscedastic and serially uncorrelated.

(d)

In this context, panel-robust or clustered standard errors allow for cross-sectional heteroscedasticity in α_i as well as within-panel heteroscedasticity in ε_{it}. The usual non-clustered robust standard errors are applicable for cross-sectional data; they are not available in Stata when estimating a fixed effects model, where the robust option defaults to the panel-robust covariance matrix. We might use panel robust standard errors if we suspect α_i or ε_{it} are not independently and identically distributed, since those ensure consistency of the standard errors as explained in (c).

Question 3

(a)

The specification is that of a two-way fixed effects model with state-specific time trends. Applying the Frisch–Waugh–Lovell theorem, we get that the coefficient estimates are equivalent to that from a regression of the state-detrended and state and time-demeaned variables. This means we are trying to identify β through within-state, within-time variation above or below state trends.

Estimated over the first period or the first two periods, the estimates for β are small, positive, and not statistically significant. It is only when the model is estimated over the entire time period that the estimated coefficient is negative and statistically significant. Taken at face value, it might suggest that increases in the minimum wage have had greater disemployment effects over time. However, much of the increase in statistical significance is driven by the standard errors getting smaller as the sample gets larger. The estimate for the first period has a larger magnitude than the estimate for the entire sample, even though the former is not statistically significant. Therefore, the economic significance of the point estimate is actually greater for the first sample period.

The fact that the standard errors decreases as the sample period expands might suggest that the statistical significance of the last estimate is merely driven by a larger sample size making the estimate more precise. The standard deviation graph also provides another useful insight: over the first ten years, there was little variation in minimum wages across states. Over the next ten years, there was more variation in minimum wages across states, but the standard deviation hovered around the same level over the period, and the minimum wage may well have (though not necessarily) varied little over time within states. In the last ten years, not only was the standard deviation of minimum wages positive, the change in the standard deviation also stayed positive, and this means there must have been variation in minimum wages within states across time, on top of variation across states. This would mean that the last period contains more of the type of variation that persists after demeaning the variables both by time and state, and this could be the reason the coefficient estimate only becomes significant when the last period is included.

We might think that the first two estimates are not as reliable, since there might have been very little variation to work with once time and state fixed effects have been partialled out. However, it would be ideal to also see what the estimates are when estimated over the three periods separately, rather than cumulatively expanding the sample period. This way we can separate the effect of having greater variation in the data from the effect of simply having a larger sample size. We would also be able to see if the coefficient estimates still went from positive to negative as we shifted the sample period, which might hint at time-heterogeneous effects of the minimum wage.

(b)

We interpret the coefficients assuming the estimated model is true, and that minimum wages are zero to set aside differences in minimum wages. Differences in the individual fixed effects between two states would represent the expected difference in state-detrended youth employment between those two states within a given time period:

$$\mathbb{E}\left\{\left[\ln(empy_{it}) - \gamma_i t\right] - \left[\ln(empy_{jt}) - \gamma_j t\right]\right\} = \alpha_i - \alpha_j$$

Differences in the time fixed effects between two time periods would represent the expected difference in state-detrended youth employment between those two time periods within a given state:

$$\mathbb{E}\left\{\left[\ln(empy_{it}) - \gamma_i t\right] - \left[\ln(empy_{is}) - \gamma_i s\right]\right\} = f_t - f_s$$

Differences in the state-specific time trend coefficients between two adjacent time periods represent the expected difference-in-differences of youth employment between the two time periods and any two states:

$$\mathbb{E}\left\{\left[\ln(empy_{it}) - \ln(empy_{it-1})\right] - \left[\ln(empy_{jt}) - \ln(empy_{jt-1})\right]\right\} = \gamma_i - \gamma_j$$

(c)

The price index is just another time fixed effect, so the coefficient on $\ln(mw_{it})$ will not be affected.

(d)

There will be a simultaneity bias here, and the strict exogeneity assumption is no longer satisfied. For example, if minimum wages are set according to the following rule:

$$\ln(mw_{it}) = \delta \ln(empy_{it}) + \nu_{it}$$

Then we have

$$\begin{aligned}\ln(mw_{it}) &= \delta\left[\alpha_i + f_t + \gamma_i t + \beta \ln(mw_{it}) + \varepsilon_{it}\right] + \nu_{it} \\ &= \frac{\delta}{1-\beta\delta}\left[\alpha_i + f_t + \gamma_i t + \varepsilon_{it}\right] + \frac{1}{1-\beta\delta}\nu_{it}\end{aligned}$$

which means

$$\mathrm{Cov}[\ln(mw_{it}), \varepsilon_{it}] = \frac{\delta}{1-\beta\delta}\sigma_\varepsilon^2 \neq 0$$

Even weak exogeneity is not satisfied, and the coefficient estimates will be biased and inconsistent. We could overcome this using an instrument variable, but we would need to find one with sufficient variation even after both fixed effects and the state-specific time trends have been taken out; finding a strong enough instrument that meets this criteria could be challenging.

(e)

We can add an interaction term between the indicator variable and $\ln(mw_{it})$, and test if its estimated coefficient is statistically significant. We reject the null hypothesis that the effect of minimum wages is the same in a recession if the estimate is found to be statistically significant.

Question 4

(a)

With a linear probability model, the average partial effect is equal to the partial effect at the mean since expectations preserve linearity. In this case there are no interaction terms, so the partial effects are also independent of covariate values, and both are equal to

$$APE = PEA = \frac{\partial \hat{\mathbb{E}}[workedm_i|\bar{\boldsymbol{x}}_i]}{\partial morekids_i} = \hat{\beta}_{morekids} \approx -0.16056$$

(b)

Both model estimates are reported below:

	LPM (s.e.)		Probit (s.e.)	
morekids	-0.161***	(0.00204)	-0.422***	(0.00549)
agem1	0.0227***	(0.000314)	0.0600***	(0.000848)
agefstm	-0.0380***	(0.000396)	-0.101***	(0.00111)
boy1st	0.000524	(0.00192)	0.00139	(0.00507)
boy2nd	-0.00525**	(0.00192)	-0.0139**	(0.00507)
blackm	0.188***	(0.00401)	0.521***	(0.0121)
hispm	0.0477***	(0.00421)	0.127***	(0.0112)
othracem	0.0602***	(0.00465)	0.158***	(0.0124)
educm	0.0272***	(0.000456)	0.0726***	(0.00127)
constant	0.339***	(0.00990)	-0.430***	(0.0260)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The only things for which direct comparisons across models make sense are the signs of the coefficients. These tell us the direction of change in the predicted probability given a change in any of the dependent variables. As shown, all of the signs the same across the two models except for the constant. The magnitudes are not directly comparable since the covariates enter linearly into a non-linear function in the probit model.

We can use the command `margins, dydx(morekids)` to compute the average partial effect $\mathbb{E}\left[\Delta\Phi(\boldsymbol{x}_i'\boldsymbol{\beta})\right]$ and `margins, dydx(morekids) atmeans` to compute the partial effect at the mean values of the covariates, $\Delta\Phi(\mathbb{E}[\boldsymbol{x}_i']\boldsymbol{\beta})$, where Δ returns the difference when the operand is evaluated at $morekids_i = 1$ and $morekids_i = 0$. We must also take care to declare the indicator variables as factor variables so that Stata calculates $\Delta\Phi(\boldsymbol{x}_i'\boldsymbol{\beta})$ rather than $\frac{\partial}{\partial morekids_i}\Phi(\boldsymbol{x}_i'\boldsymbol{\beta})$. We get that the APE is -0.1601583 and the PEA is -0.1672401, both of which are very similar to that from the LPM.

(c)

Couples may choose to have less kids because the mother is working, leading to simultaneity bias. We can use $samesex_i$ as an instrument because having two children of the same sex is likely to be exogenous (the possibility of abortion might introduce some endogeneity; not too much, one should hope). Some couples might prefer to have diversity in sex among their children, and might keep trying until they have at least one of each. This provides a source of plausibly exogenous variation in the number of children in a family.

Using the command `ivregress 2sls workedm agem1 agefstm boy1st boy2nd blackm hispm othracem educm (morekids=samesex), robust`, we use *samesex* along with all the other covariates to instrument for *morekids* via two-stage least squares. The results are reported below:

	LPM (s.e.)		Probit (s.e.)		LPM 2SLS (s.e.)	
morekids	-0.161***	(0.00204)	-0.422***	(0.00549)	-0.118***	(0.0277)
agem1	0.0227***	(0.000314)	0.0600***	(0.000848)	0.0214***	(0.000900)
agefstm	-0.0380***	(0.000396)	-0.101***	(0.00111)	-0.0363***	(0.00116)
boy1st	0.000524	(0.00192)	0.00139	(0.00507)	0.000922	(0.00194)
boy2nd	-0.00525**	(0.00192)	-0.0139**	(0.00507)	-0.00493*	(0.00193)
blackm	0.188***	(0.00401)	0.521***	(0.0121)	0.185***	(0.00444)
hispm	0.0477***	(0.00421)	0.127***	(0.0112)	0.0434***	(0.00508)
othracem	0.0602***	(0.00465)	0.158***	(0.0124)	0.0584***	(0.00479)
educm	0.0272***	(0.000456)	0.0726***	(0.00127)	0.0276***	(0.000548)
constant	0.339***	(0.00990)	-0.430***	(0.0260)	0.322***	(0.0150)
N	254654		254654		254654	

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The estimated coefficients are still of the same sign, but the coefficient on *morekids* (which is also the average partial effect and partial effect at the average) has a smaller estimated magnitude, and it is possible that the non-instrumented linear probability model overstated the effect of having children on mothers' workforce participation at the extensive margin.

(d)

The quote seems to be getting at the fact that when all the covariates are sparse and discrete, the model is close to saturated; it would be fully saturated if all possible interaction terms were included. With a saturated model, there is a parameter for every possible combination of values the covariates can have, and the predicted values at any given combination of covariate values are exactly equal to the empirical mean conditional on that combination of covariate values. This is true whether or not we use a linear model or one with a limited dependent variable, since the maximum likelihood estimators in the usual limited dependent variable models happen to be the conditional empirical mean when we have as many degrees of freedom as there are combinations of covariate values. Without interaction terms, the model is only saturated when there is at most one covariate, but if the covariates are sparse and discrete a similar logic follows and the predictions and marginal effects are close to equal whatever our model choice.