Ivan Hung

Shattered Stability: Investigation into COVID-19's impact on Wellbeing on individuals residing in London and Southeast.

Background & Methodology

Wellbeing is a crucial 'indicator' for policymakers to see what factors influence an individual's wellbeing as they can utilise this information to help improve the life quality of citizens, similarly suggested by Grajek et al. (2021). It is particularly important to focus on how individuals' wellbeing level (referred to as utility units throughout our analysis) changes before and after the pandemic to see the impact it had on mental health. It has been speculated by O'Connor et al. (2020) that there are likely to be profound and long-lasting changes to our society; we already see changes in individual behaviours through increased usage of social media post-pandemic (Liu H et al, 2021). Despite the belief that there are a plethora of reasons which can generate a change in ones wellbeing, interpretation of coefficients will focus on what we believe are the main drivers that generally determine an individual's mental health to find a best-fit model, shown as follows: net income ($linc$), education ($educ$), and job-status ($jobstatus$) [details of variables found in Appendix A].

The focus is on this triad as qualitative evidence suggests they generally have the largest impact on an individual's wellbeing. It is without doubt that income has an impact on an individual's wellbeing; therefore, following Guardiola et al. (2014), we proceed to use net income as it provides us with a more accurate representation of an individual's expenditure. Vieira K.M et al. (2021) found that wellbeing was more optimistic with individuals who had job and income stability as it helped to mitigate the ambiguity of their future, thus segregating our job-status variable similar to Dockery (2006). Sandner et al. (2022) identified education as a significant factor during the pandemic, noting that declines in mental health were strongly linked to educational and career challenges - with many students looking towards post-graduate education. Taking this into account, we have aligned $educ$ with how Smith (1982) categorised their education variable.

This paper ultimately aims to find how wellbeing changes in London and Southeast pre and post-pandemic by utilising panel data from the Understand Standing Society Survey (UKHLS), across waves 7 to 11. It is a national-based survey of respondents how are based in the UK. Findings will be segregated into two sections. We will start by estimating 3 models (Pooled Ordinary Least Squares, POLS; Random Effects, RE; and Fixed Effects, FE) to identify which model best fits as we try to unveil whether there are unobserved factors that correlate with our present explanatory variables (known as unobserved heterogeneity).

Each of the models have their strengths and limitations. However, POLS is typically eliminated when dealing with panel data due to omitted variable bias occurring as it assumes that there are no differences between time-variant and invariant characteristics: the most common invariant characteristics are race or gender, whilst variant characteristics are age or income.

This leaves us with RE and FE: RE is considered to be more efficient than the FE model as unobserved heterogeneity is uncorrelated with the explanatory variables included in the model [$Cov(X_{it}, u_i) = 0$], whereas FE controls for unobserved heterogeneity by observing how each individual changes overtime [$Cov(X_{it}, u_i) \neq 0$]. To distinguish which one to use, we utilise the Regression-based Hausman test to identify whether the time-varying variable group means are jointly significantly or not through estimating a Correlated Random Effects model (using Mundlak's approach) and using the Wald test. All models used throughout the paper are shown in Fig. 1.

Ivan Hung

Figure 1 – Proposed Models for Analysis

<u>POLS and RE Model:</u>

$$ghqrev_{it} = \alpha + \beta_1 linc_{it} + \beta_2 ug_{it} + \beta_3 pg_{it} + \beta_4 gcse_{it} + \beta_5 other_{it} + \beta_6 unemployed_{it} + \beta_7 NILF_{it} + \eta_n \omega_{it} + \eta_m z_i + \beta_{14} time_t + u_i + \epsilon_{it}$$

Notes: Where t =1, 2, 3 (for part a) and t = 4,5 (for part b); $\eta_{n/m}$ is a vector representing estimated coefficients of other included variables in model; ω_{it} represents other included time-variant variables; z_i represents other included time-invariant variables; u_i represents the error term for unobserved heterogeneity (time-invariant variables); and ϵ_{it} represents the error term for idiosyncratic (time-variant) variables.
Check Appendix Bi for full notes.

<u>FE Model:</u>

$$ghqrev_{it} = \alpha + \beta_1 linc_{it} + \beta_2 ug_{it} + \beta_3 pg_{it} + \beta_4 gcse_{it} + \beta_5 other_{it} + \beta_6 unemployed_{it} + \beta_7 NILF_{it} + \eta_n \omega_{it} + \eta_m z_i + \beta_{14} time_t + \epsilon_{it}$$

Notes: Where t =1, 2, 3 (for part a) and t = 4,5 (for part b). Check Appendix Bii for full notes.

<u>Correlated Random Effects Model (using Mundlak's approach):</u>

$$\begin{aligned} ghqrev_{it} = {} & \alpha + \beta_1 linc_{it} + \gamma_1 \overline{linc_i} + \beta_2 ug_{it} + \gamma_2 \overline{ug_{it}} + \beta_3 pg_{it} + \gamma_3 \overline{pg_i} + \beta_4 gcse_{it} + \gamma_4 \overline{gcse_{it}} \\ & + \beta_5 other_{it} + \gamma_5 \overline{other_i} + \beta_6 unemployed_{it} + \gamma_6 \overline{unemployed_i} + \beta_7 NILF_{it} + \gamma_7 \overline{NILF_i} \\ & + \eta_1 \omega_{it} + \gamma_8 \overline{\omega_i} + (\eta_2 + \gamma_8) z_i + \beta_{14} time_t + \psi_i + \epsilon_{it} \end{aligned}$$

Notes: Where t =1, 2, 3 (for part a) and t = 4,5 (for part b); and $u_t = \psi_i + w_{it}$.
Check Appendix Biii for full notes.

<u>Treatment Effect Models:</u>

Restricted Model (x)

$$ghqrev_{it} = \alpha + \beta_{12} child_{it} + \beta_{14} time_t + \Gamma child_i + \kappa(time_t \times child_i) + u_i + \epsilon_{it}$$

Unrestricted Model (y)

$$ghqrev_{it} = \alpha + \eta_n \omega_{it} + \eta_m z_i + \beta_{12} child_{it} + \beta_{14} \mathrm{time}_t + \Gamma child_i + \kappa(time_t \times child_i) + u_t + \epsilon_{it}$$

Notes: Where t = 4,5; $\eta_{n/m}$; u_i represents the error term for unobserved heterogeneity (time-invariant variables); $u_i = 0$ when estimating for FE Model; Γ is our estimated coefficient for our treatment dummy; κ is our estimated coefficient for the average treatment effect on the treated; and ϵ_{it} represents the error term for idiosyncratic (time-variant) variables.
Check Appendix Biv for full notes.

Further analysis examines the impact of COVID on individuals' wellbeing on those with and without children, using variable '*child*', within the London and Southeast regions. This portion of analysis will hone into individuals which had interviews both before and after the pandemic. By identifying the difference in wellbeing between individual who are parents, or not, can help policymakers make informed decisions on prioritising policies to support various demographics and improve wellbeing post-pandemic.

Determining the Most Efficient Model (Part a)

To determine which model is most optimal for our dataset, we will compare POLS against RE. If RE is preferred, compare RE against FE using the Regression-based Hausman Test (Appendix C illustrates the process). Table 1 represents the estimated coefficient from our POLS and our respective RE models. We can see that both models have coefficients that differ significantly, particularly PG and GCSE, where POLS (RE) model suggests, when holding everything else constant, individuals with a PG are more likely to have 0.0116 (0.108) more utility points, conversely, individuals with GCSEs are more likely to have 0.103 (0.0204) fewer utility points compared to those with an A-level qualification. Since estimates differ significantly, it indicates that unobserved heterogeneity is both present and relevant. Hence, we prefer the Random Effects model in this case.

Table 1 - POLS Clustered vs RE vs RE Clustered

	POLS (w/Clustered SEs)	RE	RE (w/Clustered SEs)
Degree	0.146	0.267^{***}	0.267^{***}
	(0.103)	(0.102)	(0.103)
Postgraduate Degree [PG]	0.0116	0.108	0.108
	(0.129)	(0.126)	(0.129)
GCSE	-0.103	-0.0204	-0.0204
	(0.119)	(0.114)	(0.119)
No Qualification	-0.329^{**}	-0.294^{**}	-0.294^{**}
	(0.132)	(0.125)	(0.132)
R^2	0.044		
N	41767	41767	41767

Notes: Coefficients represent marginal effects. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Full regression results can be found in Appendix (D)

We can confirm this by performing a Breusch-Pagan Lagrangian Multiplier (BPLM) test. Through estimating a RE model, we can utilise the estimated variance of the random effects and idiosyncratic error (ϵ_{it}) term to determine the Lagrangian Multiplier (refer to Appendix E). We obtain the p-value is less than 1%, showing there is strong evidence to reject the null hypothesis (in favour of the alternative hypothesis), suggesting that our RE model is most appropriate for our dataset, bringing us to compare the RE and FE models.

Typically, we would prefer the FE model if our primary focus were on time-varying variables, whilst the RE model would be used for highlighting time-invariant variables. Thus, we expect that a FE model would estimate more accurate coefficients. Informally, and based off coefficients from Table 2, we can see that most variables in the RE model have been underestimated, suggesting that the variables have some degree of correlation with the individual-specific heterogeneity error term (u_i). The FE model addresses this by focussing on within individual-variation overtime. Although, if there is little variation in the time-varying variables (e.g. $educ$), it becomes harder to estimate coefficients with high accuracy, hence explaining the larger standard error values. Check Appendix Fi for further information.

Table 2 - RE vs "Standard FE" vs "FE without t" vs "FE without age"

	RE (w/Clustered SEs)	FE [1]	FE [2]	FE [3]
Net Income (log)	0.110***	0.0490	0.0490	0.0490
	(0.0339)	(0.0405)	(0.0405)	(0.0405)
Degree	0.267***	1.396**	1.396**	1.396**
	(0.103)	(0.551)	(0.551)	(0.551)
Postgraduate Degree	0.108	1.431**	1.431**	1.431**
	(0.129)	(0.643)	(0.643)	(0.643)
GCSE	-0.0204	1.305**	1.305**	1.305**
	(0.119)	(0.548)	(0.548)	(0.548)
No Qualification	-0.294**	1.042	1.042	1.042
	(0.132)	(1.005)	(1.005)	(1.005)
Non-Contracted Employment	-1.080***	-0.676***	-0.676***	-0.676***
	(0.101)	(0.147)	(0.147)	(0.147)
NILF	0.429***	0.558***	0.558***	0.558***
	(0.107)	(0.171)	(0.171)	(0.171)
R^2		0.006	0.006	0.006
N	41767	41767	41767	41767

Notes: Coefficients represent marginal effects. Clustered Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. For FE models, time-invariant (unobserved heterogeneity) variables have been omitted.
Full regression results can be found in Appendix (Fii)

Formally, the Regression-based Hausman test can be performed to statistically distinguish which model is most optimal for our analysis. Utilising Mundlak's approach to estimate a CRE model, we obtain the results on Table 4. This approach enables us to estimate means of time-varying variables, allowing us to assess the strength of the correlation between the error term for unobserved heterogeneity and explanatory variables.

Usually, estimates from the CRE are very close to FE estimates, although this is not the case here for some variables due to several and aforementioned reasons. One could be the way the models treat individual effects. The FE model fully eliminates them (i.e. explicitly neglects time-invariant variables from model), whereas the CRE model is more flexible, allowing individual-specific effects through estimated individual means ($\gamma\bar{X}_\iota$, where X is a vector of time-dependent variables). As a result, the difference is caused as the CRE model emulates the FE model by altering a RE model, allowing the error term for unobserved heterogeneity to be correlated with explanatory variables.

Table 4 - CRE vs "FE without age"

	CRE	FE [3]
Net Income (log)	0.0647*	0.0490
	(0.0392)	(0.0405)
Degree	1.023**	1.396**
	(0.511)	(0.551)
Postgraduate Degree [PG]	1.559**	1.431**
	(0.632)	(0.643)
GCSE	1.308**	1.305**
	(0.529)	(0.548)
No Qualification	1.036	1.042
	(0.936)	(1.005)
Non-Contracted Employment	-0.587***	-0.676***
	(0.141)	(0.147)
NILF	0.544***	0.558***
	(0.163)	(0.171)
mlinc (Net income, log)	0.205***	-
	(0.0711)	
meduc1 (Degree	-0.927*	-
	(0.521)	
meduc2 (Postgraduate Degree)	-1.601**	-
	(0.646)	
meduc4 (GCSE)	-1.422***	-
	(0.541)	
meduc5 (No Qualification)	-1.337	-
	(0.948)	
mjobstat1 (Non-Contracted Employment)	-0.978***	-
	(0.200)	
mjobstat3 (NILF)	-0.169	-
	(0.219)	
R^2		0.006
N	41767	41767

Notes: Coefficients represent marginal effects. Clustered Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. m'variable' coefficients link group averages to well-being.
Full regression results can be found in Appendix (G)

After estimating our coefficients for the individual means, through the CRE model, we can complete the Hausman test by performing a Wald test. This joint test statistically ensures whether our individual mean coefficients can be accepted. Table 5 shows our test outcomes for when we allow and don't allow (resp.) for the time-invariant error term to be correlated to the explanatory variables. Regardless of which model we use, we can conclude, at the 99% confidence level, that our individual means are not equal to 0: rejecting the null hypothesis. Therefore, we should use the Fixed Effects model.

Table 5 – Wald test after estimating CRE models

	With heteroskedasticity and Serial Correlation (Autocorrelation) Robust SEs	Without heteroskedasticity and Serial Correlation (Autocorrelation) Robust SEs
χ^2	128.45	164.23
p-value	0.000	0.000

Notes: LHS represents the Wald Statistic when the test uses robust standard errors that are adjusted to account for heterogeneity and autocorrelation; the test follows a chi-squared distribution, hence the estimated χ-statistic; $Cov(X_{it}, u_t) = 0$ (vice versa for RHS).
Hypotheses can be found in Appendix H.

Well-Being Levels Before and After the Pandemic (Part b)

The next part of the analysis will be looking at how SARS-CoV-2 impacted wellbeing. As previously mentioned, the variable $child$ (a dummy variable segregating observations into if they have a child or not) will be used as our treatment effect. In this case, the treated group will be those who have a child, whereas the control group will be those without a child. I have segregated the time periods to before and after the first lockdown was announced in the UK, from 24th March 2020 according to Dropkin (2020), so that we can analyse the impact of lockdown on wellbeing.

To start, we should figure out the difference in utility points between the two categories and see whether there is a difference in wellbeing, regardless of if we differentiate the categories by treatment/control group. Running a simple regression between our scalable wellbeing variable ($ghqrev$) and our treatment term (child), we find that, on average, those who have children have 0.287 less utility points compared to those who don't have children (LHS of Appendix I). There is no surprise here as there are many factors that can contribute to a parent's wellbeing such as stress and pressure, as found by Townshend et al. (2016). Nevertheless, we can further investigate if any biases occurred when estimating for our above coefficient. We can summarise the statistics of variables which could have an impact on a parent's wellbeing (found in Appendix J).

Doing this, we find a key point that we should take note of: notice that despite the averages of the variables in both categories being similar, there may be a bias generated with those who have a child because there is only half the number of observations compared to those who don't have a child. Therefore, it is ambiguous to determine whether our coefficient is positively or negatively biased. However, we can run a regression with these additional variables to see how our coefficient changes. We find that the coefficient, keeping all things constant, increases and suggests that those who have children have 0.279 less utility points compared to those who don't have children (shown on RHS of Appendix I). As a result, we find that our initial coefficient was likely to be positively biased.

We can now find our difference-in-difference (κ) coefficient, to help us see the impact of COVID-19's exposure on those who are/aren't parents. We start off with the most basic model: by finding the difference in the means between both the treated and control group (where μ is the mean):

$$\kappa = (\mu_{treated,t=2} - \mu_{treated,t=1}) - (\mu_{control,t=2} - \mu_{control,t=1}) \approx -0.65477$$

The coefficient is found to be -0.655, suggesting that COVID further reduced the utility points by 0.655 points to those in the treatment group. We can confirm this coefficients plausibility by estimating regressions to identify the average wellbeing for our control, treated and actually treated (ATT) groups. We estimate both a POLS and FE model (found on Table 6). Comparing both is useful to see the robustness of each as solely relying on FE makes it impossible to estimate for our individual heterogeneity variables, making it impossible to estimate our treatment coefficient (Γ). As a result, we can use the POLS model to make plausible estimation for Γ and assess the sensitivity of the coefficients across different models.

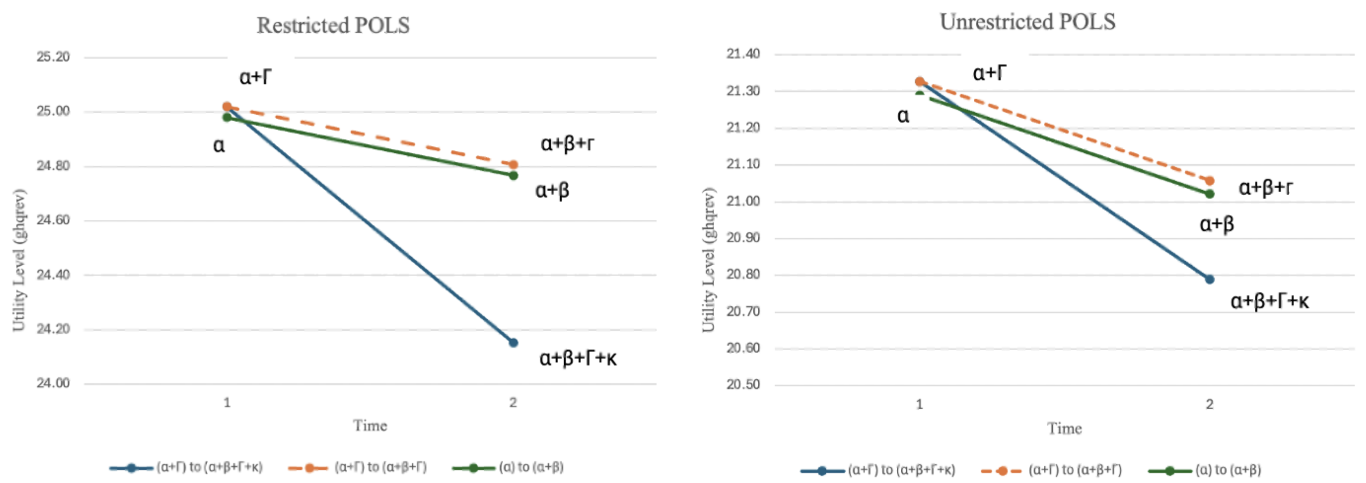
Table 6 - Identifying the Difference-in-Difference (κ) Coefficient

	Models (x)		Models (y)	
	Restricted FE	Restricted POLS	Unrestricted FE	Unrestricted POLS
Have Children (Γ)	-	0.0399	-	0.0371
		(0.0935)		(0.0922)
Time [β]	-0.212***	-0.212***	-0.327**	-0.269***
	(0.0496)	(0.0496)	(0.164)	(0.0497)
Interaction term between time and child [κ]	**-0.655*****	**-0.655*****	**-0.651*****	**-0.640*****
	(0.0901)	(0.0901)	(0.0902)	(0.0902)
_cons (α)	25.00***	24.98***	21.15***	21.29***
	(0.0208)	(0.0548)	(4.397)	(0.344)
R^2	0.012	0.003	0.012	0.039
N	27948	27948	27948	27948

Notes: Coefficients represent marginal effects. Clustered Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Full regression results (for unrestricted model) can be found in Appendix (K)

Our estimated κ has strong statistical evidence and stays relatively the same throughout, with it increasing marginally with the unrestricted models. With the clarification of our models, we can bring our focus to the impact the pandemic on our treated and control groups. Looking at Fig. 2, we see that the Average Treated Effect on the Treated (ATT) group ($\alpha + \beta + \Gamma + \kappa$) were impacted more significantly by COVID compared to the control. This is expected as we previously anticipated that this would be the case, since parents would have increased stress and psychological burden (Chu et al. [2021]) from children which takes a toll on their wellbeing.

Figure 2 – Estimated Treatment Effects Coefficients Visualised



Appendix L for further detail

On the contrary, an interesting highlight to point out is the impact of COVID on those who are in the treatment group but weren't 'treated' ($\alpha + \beta + \Gamma$). We see that their level of wellbeing is not only better than those in the treated group, but better than the control group too. This begs to differ the question; can children bring both positive and negative effects to a parent's wellbeing? Sander et al. (2023) found that those with happier children would also have a higher level of utility – suggesting a direct relationship between a child's and parent's wellbeing. Although the data doesn't explicitly tell us, we can infer that having a child didn't necessarily decrease a parent's wellbeing.

<u>Conclusion</u>

This study has found that since our highlighted focus on the dataset leaned towards time-variant variables, our series of tests found that the most preferred estimation method was the FE model. We further proceeded to look at the wellbeing of parents and non-parents before and after COVID. Generally, we found that those with children were severely impacted by SARS-CoV-2, facing external pressures of looking after others, regardless of job security and other personal factors (Mathur et al. [2023]).

A noteworthy mention from this study is that there are an abundant of studies, like Woodland et al. (2023) and Patrick et al. (2020), that hone into the wellbeing of children during this time period. However, there are little to none which focus on the wellbeing parents, as backed by Dawes et al. (2021). Research has missed out a crucial part of the analysis: optimistic families or policies supporting them can trigger a domino effect, as parents with higher 'utility' are better able to care for their children, boosting morale during tough times like the pandemic, aligning with our findings.

(2200 Words)

Bibliography

Chu, K.A., Schwartz, C., Towner, E., Kasparian, N.A. and Callaghan, B. (2021). Parenting under pressure: A mixed-methods investigation of the impact of COVID-19 on family life. *Journal of Affective Disorders Reports*, 5, p.100161. doi:https://doi.org/10.1016/j.jadr.2021.100161.

Dawes, J., May, T., McKinlay, A., Fancourt, D. and Burton, A. (2021). Impact of the COVID-19 pandemic on the mental health and wellbeing of parents with young children: a qualitative interview study. *BMC psychology*, [online] 9(1), p.194. doi:https://doi.org/10.1186/s40359-021-00701-8.

Dockery, A. (2006). *MENTAL HEALTH AND LABOUR FORCE STATUS: PANEL ESTIMATES WITH FOUR WAVES OF HILDA #*. [online] Available at: https://melbourneinstitute.unimelb.edu.au/assets/documents/hilda-bibliography/working-discussion-research-papers/2006/Dockery_Mental_Health_Four_Waves_HILDA.pdf [Accessed 16 Dec. 2024].

Dropkin, G. (2020). COVID-19 UK Lockdown Forecasts and R0. *Frontiers in Public Health*, 8. doi:https://doi.org/10.3389/fpubh.2020.00256.

Grajek, M. and Sobczyk, K. (2021). Original Article Well-being levels among students during the COVID-19 pandemic. *Journal of Physical Education and Sport ® (JPES)*, [online] 21(4), pp.1682–1687. doi:https://doi.org/10.7752/jpes.2021.04213.

Guardiola, J. and Guillen-Royo, M. (2014). Income, Unemployment, Higher Education and Wellbeing in Times of Economic Crisis: Evidence from Granada (Spain). *Social Indicators Research*, 120(2), pp.395–409. doi:https://doi.org/10.1007/s11205-014-0598-6.

Liu, H., Liu, W., Yoganathan, V. and Osburg, V.-S. (2021). COVID-19 information overload and generation Z's social media discontinuance intention during the pandemic lockdown. *Technological Forecasting and Social Change*, [online] 166(120600). doi:https://doi.org/10.1016/j.techfore.2021.120600.

Mathur, M., Robiolio, H., Ebert, L. and Kerr, B. (2023). Family stress during the COVID-19 pandemic: a qualitative analysis. *BMJ Open*, [online] 13(5), p.e061396. doi:https://doi.org/10.1136/bmjopen-2022-061396.

Mundlak (1978). Models with Variable Coefficients: Integration and Extension. *Annales de l'inséé*, (30/31), p.483. doi:https://doi.org/10.2307/20075301.

O'Connor, R.C., Wetherall, K., Cleare, S., McClelland, H., Melson, A.J., Niedzwiedz, C.L., O'Carroll, R.E., O'Connor, D.B., Platt, S., Scowcroft, E., Watson, B., Zortea, T., Ferguson, E. and Robb, K.A. (2020). Mental Health and Wellbeing during the COVID-19 pandemic: Longitudinal Analyses of Adults in the UK COVID-19 Mental Health & Wellbeing Study. *The British Journal of Psychiatry*, [online] 218(6), pp.1–17. doi:https://doi.org/10.1192/bjp.2020.212.

Patrick, S.W., Henkhaus, L.E., Zickafoose, J.S., Lovell, K., Halvorson, A., Loch, S., Letterie, M. and Davis, M.M. (2020). Well-being of Parents and Children During the COVID-19 Pandemic: A National Survey. *Pediatrics*, [online] 146(4), p.e2020016824. doi:https://doi.org/10.1542/peds.2020-016824.

Sandner, M., Patzina, A., Anger, S., Bernhard, S. and Dietrich, H. (2023). The COVID-19 pandemic, well-being, and transitions to post-secondary education. *Review of Economics of the Household*, 21. doi:https://doi.org/10.1007/s11150-022-09623-9.

Skripkauskaite, S., Creswell, C., Shum, A., Pearcey, S., Lawrence, P., Dodd, H. and Waite, P. (2023). Changes in UK parental mental health symptoms over 10 months of the COVID-19 pandemic. *JCPP Advances*, 3(2). doi:https://doi.org/10.1002/jcv2.12139.

Smith, T.W. (1982). College Dropouts: An Analysis of the Psychological Well-Being and Attitudes of Various Educational Groups. *Social Psychology Quarterly*, 45(1), p.50. doi:https://doi.org/10.2307/3033674.

Townshend, K., Jordan, Z., Stephenson, M. and Tsey, K. (2016). The effectiveness of mindful parenting programs in promoting parents' and children's wellbeing. *JBI Database of Systematic Reviews and Implementation Reports*, [online] 14(3), pp.139–180. doi:https://doi.org/10.11124/jbisrir-2016-2314.

Vieira, K.M., Potrich, A.C.G., Bressan, A.A. and Klein, L.L. (2021). Loss of financial well-being in the COVID-19 pandemic: Does job stability make a difference? *Journal of Behavioral and Experimental Finance*, 31, p.100554. doi:https://doi.org/10.1016/j.jbef.2021.100554.

Woodland, L., Smith, L.E., Brooks, S.K., Webster, R.K., Amlôt, R., Rubin, A. and G. James Rubin (2023). Parent-reported child's close contact with non-household family members and their well-being during the COVID-19 pandemic: A cross-sectional survey. *PLOS ONE*, [online] 18(10), pp.e0292344–e0292344. doi:https://doi.org/10.1371/journal.pone.0292344.

<u>Appendix</u>

Appendix A – Definition table for variables, showing percentage distribution of categorical variables and base categories (**base categories shown in bold for part a**):

Variable Name	Description	Coding (for categorised var)
pidp	Individual identifier	
t (time)	Time (wave) identifier	1 = Wave 7 2 = Wave 8 3 = Wave 9 4 = Wave 10 5 = Wave 11
region	Region of England	**1 = London** 2 = Southeast
istrtday	Interview date (year)	Ranges between 2015 and 2021
istrtdatm	Interview date (month)	1 = January 2 = February, etc.
istrtdatd	Interview date (day)	Ranges from 1 to 31
ghq	Subjective well-being (GHQ)	Dependent Variable (unreversed).
age	Age of respondent (in years)	
male	Male	**0 = female**, 1 = male
jobstatus (jbstat)	Current labour force status	1 = Not paid employment (17.35%) **2 = Paid employment** (51.59%) 3 = NILF (Not In Labour Force) (31.06%)
gincome	*Monthly gross income*	
netinc (nincome)	Monthly net income	
linc	Logarithmic form of monthly net income	
marryst	Marital Status	**0 = Not Married (other)** 1 = Married
children	Number of children in the household	
educ (educat)	Highest qualification	1 = Degree (32.53%) 2 = Postgraduate Degree (13.40%) **3 = A-Level** (18.95%) 4 = GCSE (18.85%) 5 = Other or None (16.27%
race	Ethnicity	0 = Not White **1 = White**

ghqrev	Subjective well-being (GHQ)	Dependent Variable (0-36 point scale).
T	Represents the number of observations who responded in 1, 2, 3, 4 or 5 waves	
intdate	Interview Date (Date-Month-Year)	

Appendix Bi – Model for estimating the POLS and RE Model:

$$\begin{aligned} ghqrev_{it} = \alpha &+ \beta_1 linc_{it} + \beta_2 ug_{it} + \beta_3 pg_{it} + \beta_4 gcse_{it} + \beta_5 other_{it} + \beta_6 unemployed_{it} + \beta_7 NILF_{it} \\ &+ \beta_8 age_{it} + \beta_9 age_{it}^2 + \beta_{10} male_i + \beta_{11} married_{it} + \beta_{12} child_{it} + \beta_{13} notwhite_i \\ &+ \beta_{14} time_t + \epsilon_{it} \end{aligned}$$

Notes:

Given that time = 1,2,3 (for part a) and time = 4,5 (for part b).

$\eta_{n/m}$ is a vector representing the estimated coefficients of the other included variables in the model.

u_i represents the error term for unobserved heterogeneity (time-invariant variables).

ϵ_{it} represents the error term for idiosyncratic (time-variant) variables.

Appendix Bii – Model for estimating the FE Model:

$$\begin{aligned} ghqrev_{it} = \alpha &+ \beta_1 linc + \beta_2 ug_{it} + \beta_3 pg_{it} + \beta_4 gcse_{it} + \beta_5 other_{it} + \beta_6 unemployed_{it} + \beta_7 NILF_{it} \\ &+ \beta_8 age_{it} + \beta_9 age_{it}^2 + \beta_{10} male_i + \beta_{11} married_{it} + \beta_{12} child_{it} + \beta_{13} notwhite_i \\ &+ \beta_{14} time_t + u_t + \epsilon_{it} \end{aligned}$$

Given that time = 1,2,3 (for part a) and time = 4,5 (for part b).

u_i represents the error term for unobserved heterogeneity (time-invariant variables).

ϵ_{it} represents the error term for idiosyncratic (time-variant) variables.

Appendix Biii – Estimated Correlated Random Effects Model (using Mundlak's approach):

$$\begin{aligned} ghqrev_{it} = \alpha &+ \beta_1 age_{it} + \gamma_1 \overline{age_i} + \beta_2 age_{it}^2 + \gamma_2 \overline{age_i^2} + (\beta_3 + \gamma_3) male_i + \beta_4 unemployed_{it} \\ &+ \gamma_4 \overline{unemployed_i} + \beta_5 NILF_{it} + \gamma_5 \overline{NILF_i} + \beta_6 nincome_{it} + \gamma_6 \overline{nincome_i} + \beta_7 married_{it} \\ &+ \gamma_7 \overline{married_i} + \beta_8 child_{it} + \gamma_8 \overline{child_i} + \beta_9 ug_{it} + \gamma_9 \overline{ug_i} + \beta_{10} pg_{it} + \gamma_{10} \overline{pg_i} + \beta_{11} gcse_{it} \\ &+ \gamma_{11} \overline{gcse_i} + \beta_{12} other_{it} + \gamma_{12} \overline{other_i} + (\beta_{13} + \gamma_{13}) notwhite_{it} + \beta_{14} time_t + \psi_i + \epsilon_{it} \end{aligned}$$

Given that time = 1,2,3 (for part a) and time = 4,5 (for part b).

u_i represents the error term for unobserved heterogeneity (time-invariant variables).

ϵ_{it} represents the error term for idiosyncratic (time-variant) variables.

$Where\ u_t = \psi_i + w_{it}$, proposed by Mundlak [1978].

ψ_i is an error term which is uncorrelated with all explanatory variables.

w_{it} is the individual-specific means of time-varying explanatory variables.

γ_i are the estimated coefficients of the individual-specific means of time-varying explanatory variables.

Appendix Biv - Treatment Effect Model (using our Fixed Effects Model from part a):

Restricted Model (x)

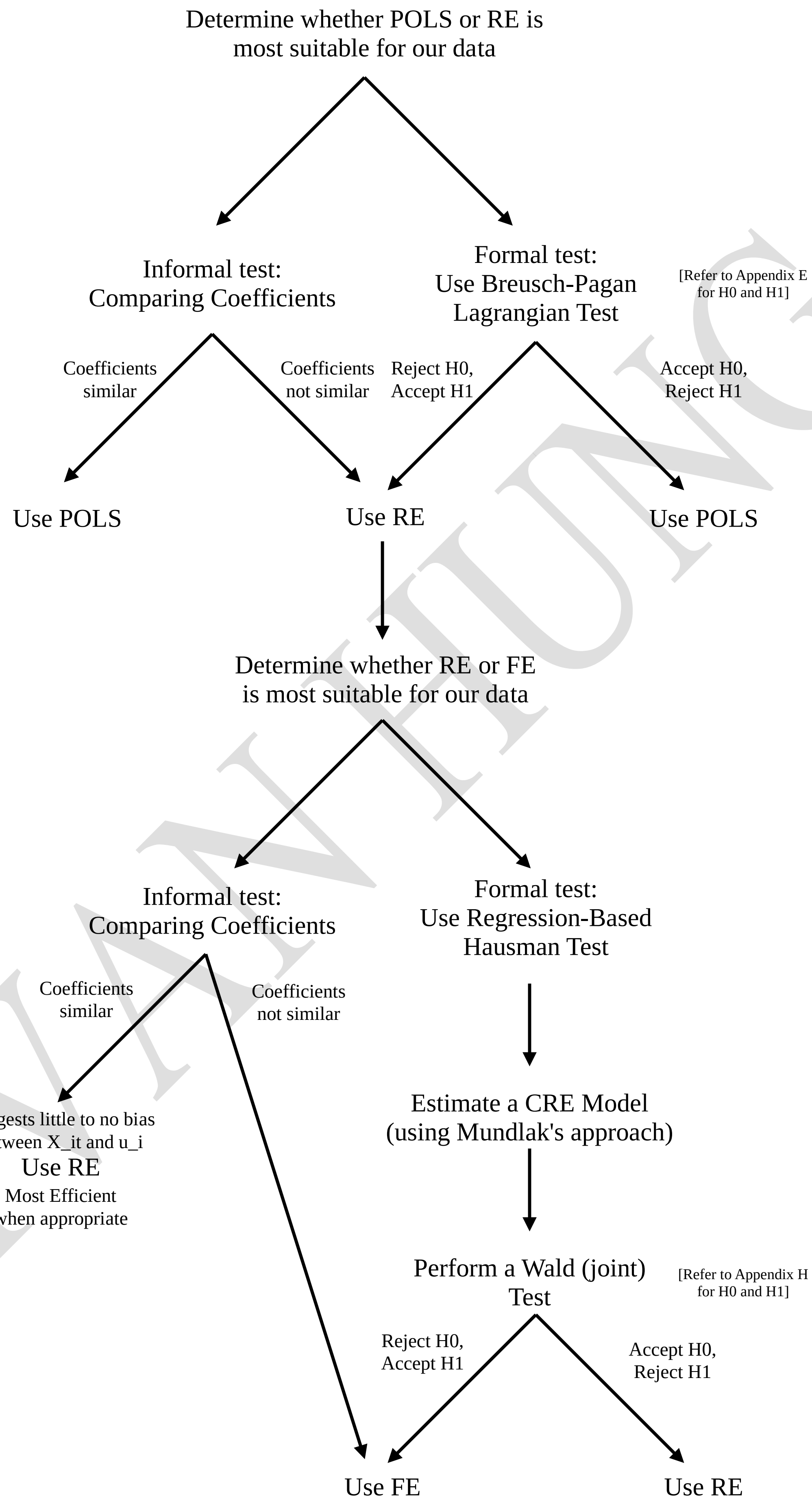
$$ghqrev_{it} = \alpha + \boldsymbol{\beta_{12} child_{it}} + \beta_{14} time_t + \Gamma child_i + \kappa(time_t \times child_i) + u_i + \epsilon_{it}$$

Unrestricted Model (y)

$$ghqrev_{it} = \alpha + \beta_1 linc_{it} + \beta_2 ug_{it} + \beta_3 pg_{it} + \beta_4 gcse_{it} + \beta_5 other_{it} + \beta_6 employed_{it} + \beta_7 NILF_{it} + \beta_8 age_{it} + \beta_9 age_{it}^2 + \beta_{10} male_i + \beta_{11} married_{it} + \boldsymbol{\beta_{12} child_{it}} + \beta_{13} notwhite_i + \beta_{14} time_t + \Gamma child_i + \kappa(time_t \times child_i) + u_r + \epsilon_{it}$$

Notes: Where t = 4,5; $\eta_{n/m}$
u_i represents the error term for unobserved heterogeneity (time-invariant variables). $u_i = 0$ when estimating for FE Model
Γ is our estimated coefficient for our treatment dummy.
κ is our estimated coefficient for the average treatment effect on the treated (ATT); difference-in-difference estimator.
$time_t \times child_i$ is an interaction term between treatment group and time.
ϵ_{it} represents the error term for idiosyncratic (time-variant) variables.

Appendix C – Flowchart of process in determining best-fit model for our dataset


Determine whether POLS or RE is most suitable for our data
Informal test: Comparing Coefficients
Formal test: Use Breusch-Pagan Lagrangian Test
[Refer to Appendix E for H0 and H1]
Coefficients similar
Coefficients not similar
Reject H0, Accept H1
Accept H0, Reject H1
Use POLS
Use RE
Use POLS
Determine whether RE or FE is most suitable for our data
Informal test: Comparing Coefficients
Formal test: Use Regression-Based Hausman Test
Coefficients similar
Coefficients not similar
Suggests little to no bias between X_it and u_i
Use RE
Most Efficient when appropriate
Estimate a CRE Model (using Mundlak's approach)
Perform a Wald (joint) Test
[Refer to Appendix H for H0 and H1]
Reject H0, Accept H1
Accept H0, Reject H1
Use FE
Use RE

Ivan Hung

Appendix D - POLS Clustered vs RE vs RE Clustered

	POLS (w/Clustered SEs)	RE	RE (w/Clustered SEs)
Net Income (log)	0.190***	0.110***	0.110***
	(0.0410)	(0.0303)	(0.0339)
Degree	0.146	0.267***	0.267***
	(0.103)	(0.102)	(0.103)
Postgraduate Degree	0.0116	0.108	0.108
	(0.129)	(0.126)	(0.129)
GCSE	-0.103	-0.0204	-0.0204
	(0.119)	(0.114)	(0.119)
No Qualification	-0.329**	-0.294**	-0.294**
	(0.132)	(0.125)	(0.132)
Non-Contracted Employment	-1.294***	-1.080***	-1.080***
	(0.114)	(0.0807)	(0.101)
NILF	0.394***	0.429***	0.429***
	(0.120)	(0.110)	(0.107)
Age	-0.101***	-0.101***	-0.101***
	(0.0148)	(0.0140)	(0.0145)
Age^2	0.00112***	0.00114***	0.00114***
	(0.000145)	(0.000140)	(0.000141)
Male	0.871***	0.917***	0.917***
	(0.0739)	(0.0737)	(0.0732)
Married	0.896***	0.811***	0.811***
	(0.0808)	(0.0737)	(0.0788)
Has Children	-0.0594	0.0339	0.0339
	(0.0942)	(0.0865)	(0.0884)
Caucasian	0.0113	0.0385	0.0385
	(0.111)	(0.102)	(0.111)
t	-0.213***	-0.212***	-0.212***
	(0.0226)	(0.0215)	(0.0226)
_cons	25.40***	25.79***	25.79***
	(0.442)	(0.389)	(0.415)
R^2	0.044		
N	41767	41767	41767

Notes: Coefficients represent marginal effects. (Clustered) Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix E – Breusch-Pagan Lagrangian Multiplier test statistic

$$LM = \frac{N}{2} \times \frac{\widehat{\sigma_u^2}}{\widehat{\boldsymbol{\sigma_u}}^2 + \widehat{\sigma_\epsilon^2}}$$

Where N is the number of individuals, $\widehat{\sigma_u^2}$ is the estimated variance of RE, $\widehat{\sigma_\epsilon^2}$ is the estimated variance of the idiosyncratic error term

Where our hypothesis test is the following:
$H_0: \sigma_u^2 = 0$
$H_1: \sigma_u^2 > 0$

H_0 in this case this suggests that heteroskedasticity isn't present, so in our context, using the Pooled Ordinary Least Squares (POLS) would be preferred since it is more efficient compared our Random Effects (RE) model.

H_1 on the other hand suggests that heteroskedasticity is present, so in our context, using the Pooled Ordinary Least Squares (POLS) wouldn't be preferred since $Cov(X_{it}, u_t) = 0$. Therefore, we should prefer to utilise the Random Effects (RE) model.

Appendix Fi – Further Explanation on larger standard errors when estimating a Fixed Effects Model

educ	Overall		Between		Within Percent
	Frequency	Percent (%)	Frequency	Percent (%)	
Degree	22728	33.53	4658	33.33	97.59
Postgraduate	9365	13.40	1938	13.87	96.65
A-level	13239	18.95	2786	19.94	95.04
GCSE	13169	18.85	2713	19.41	97.08
No Qualification	11369	16.27	2303	16.48	98.73
Total (∑)	69870	100	14398	**103.03**	97.06

jobstatus	Overall		Between		Within Percent
	Frequency	Percent (%)	Frequency	Percent (%)	
N-C Employment	7379	17.67	3376	24.16	73.51
Employed	22010	52.70	8213	58.79	89.55
NILF	12378	29.64	4498	32.20	91.92
Total (∑)	41767	100	16087	**115.15**	86.85

Tabulating these tables, we can see how many individuals moved to a different category: we find that only 3.03% of people actually continued to do the next level of education. This backs up our argument in the main body of why standard errors increase due to the little variation in individuals across the time periods. Moreover, we find that the mean age of this dataset is ~51 years old, which could explain why there is little variation in individuals' education level; if we had a dataset with younger people, the proportion of individuals between the levels of education could increase.

Conversely, we see that there is larger variation with individuals moving between the job status categories. Due to this higher variation, we are able to get more statistically clearer results that provides us with more confidence when interpreting estimated coefficients (we can see that all of the categories for $jobstatus$ are statistically significant at the 99% confidence level in Appendix Fii).

Appendix Fii - RE Clustered vs Standard FE vs FE without t vs FE without age

	RE (w/Clustered SEs)	FE [1]	FE [2]	FE [3]
Net Income (log)	0.110***	0.0490	0.0490	0.0490
	(0.0339)	(0.0405)	(0.0405)	(0.0405)
Degree	0.267***	1.396**	1.396**	1.396**
	(0.103)	(0.551)	(0.551)	(0.551)
Postgraduate Degree	0.108	1.431**	1.431**	1.431**
	(0.129)	(0.643)	(0.643)	(0.643)
GCSE	-0.0204	1.305**	1.305**	1.305**
	(0.119)	(0.548)	(0.548)	(0.548)
No Qualification	-0.294**	1.042	1.042	1.042
	(0.132)	(1.005)	(1.005)	(1.005)
Non-Contracted Employment	-1.080***	-0.676***	-0.676***	-0.676***
	(0.101)	(0.147)	(0.147)	(0.147)
NILF	0.429***	0.558***	0.558***	0.558***
	(0.107)	(0.171)	(0.171)	(0.171)
Age	-0.101***	-0.541***	-0.541***	-
	(0.0145)	(0.0827)	(0.0827)	
Age^2	0.00114***	0.00335***	0.00335***	0.00335***
	(0.000141)	(0.000697)	(0.000697)	(0.000697)
Male	0.917***	-	-	-
	(0.0732)			
Married	0.811***	0.344*	0.344*	0.344*
	(0.0788)	(0.207)	(0.207)	(0.207)
Has Children	0.0339	0.280	0.280	0.280
	(0.0884)	(0.186)	(0.186)	(0.186)
Caucasian	0.0385	-	-	-
	(0.111)			
t	-0.212***	-	-	-0.541***
	(0.0226)			(0.0827)
_cons	25.79***	41.81***	41.81***	14.35***
	(0.415)	(2.392)	(2.392)	(2.040)
R^2		0.006	0.006	0.006
N	41767	41767	41767	41767

Notes: Coefficients represent marginal effects. Clustered Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. For FE models, time-invariant (unobserved heterogeneity) variables have been omitted.

Appendix G - CRE vs FE without 'age'

	CRE	FE[3]
Net Income (log)	0.0647*	0.0490
	(0.0392)	(0.0405)
Degree	1.023**	1.396**
	(0.511)	(0.551)
Postgraduate Degree	1.559**	1.431**
	(0.632)	(0.643)
Degree	1.308**	1.305**
	(0.529)	(0.548)
No Qualification	1.036	1.042
	(0.936)	(1.005)
Non-Contracted Employment	-0.587***	-0.676***
	(0.141)	(0.147)
NILF	0.544***	0.558***
	(0.163)	(0.171)
Age^2 {c.age#c.age}	0.00316***	{0.00335***}
	(0.000696)	(0.000697)
Male	0.829***	-
	(0.0765)	
Married	0.0903	0.344*
	(0.193)	(0.207)
Has Children	0.128	0.280
	(0.170)	(0.186)
Caucasian	-0.00406	-
	(0.111)	
mlinc (Net income, log)	0.205***	-
	(0.0711)	
meduc1 (Degree)	-0.927*	-
	(0.521)	
meduc2 (Postgraduate Degree)	-1.601**	-
	(0.646)	
meduc4 (GCSE)	-1.422***	-
	(0.541)	
meduc5 (No Qualification)	-1.337	-
	(0.948)	
mjobstat1(Non-Contracted Employment)	-0.978***	-
	(0.200)	

mjobstat3 (NILF)	-0.169	-
	(0.219)	
mage (age)	-0.0992***	-
	(0.0153)	
mage2 (age^2)	-0.00204***	-
	(0.000707)	
mmarry (Married)	0.901***	-
	(0.209)	
mchildren (Has Children)	-0.214	-
	(0.202)	
t	-0.519***	-0.541***
	(0.0825)	(0.0827)
_cons	25.68***	14.35***
	(0.569)	(2.040)
R^2		0.006
N	41767	41767

Notes: Coefficients represent marginal effects. Clustered Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Appendix H – Wald test after estimating CRE models

$H_0: \gamma_i = 0$
$H_1: \gamma_i \neq 0$

H_0 in this case means that the individual mean is equal to zero. This suggests that it doesn't contribute significantly to the dependent variable.

H_1 in this case means that the individual mean does not equal zero. This suggests that there is enough statistical evidence to show that the individual mean contributes significantly to the dependent variable.

	With heteroskedasticity and Serial Correlation (Autocorrelation) Robust SEs	Without heteroskedasticity and Serial Correlation (Autocorrelation) Robust SEs
χ^2	128.45	164.23
p-value	0.000	0.000

Notes: LHS represents the Wald Statistic when the test uses robust standard errors that are adjusted to account for heterogeneity and autocorrelation, $Cov(X_{it}, u_t) = 0$ (vice versa for RHS).

Appendix I - Difference in utility points of those with/without children

	Restricted Model for Child Coefficient	Unrestricted Model for Child Coefficient
Have Children	-0.287***	-0.279***
	(0.0676)	(0.0669)
Age	-	0.0411***
		(0.00203)
Male	-	0.892***
		(0.0650)
Net Income (log)	-	0.202***
		(0.0369)
Married	-	0.756***
		(0.0674)
Race	-	0.0168
		(0.0961)
Education Level	-	-0.0584**
		(0.0232)
_cons	24.88***	20.43***
	(0.0393)	(0.314)
R^2	0.001	0.034
N	27948	27948

Notes: Coefficients represent marginal effects. Robust Standard errors in parentheses* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix J – Summary Statistics, under the condition of whether individual has child or not

Variable – Has child	Observations (N)	Mean	Standard Deviation	Min	Max
Age	9752	55.1452	16.17994	19	96
Gender	9752	0.4312961	0.4952827	0	1
Net Income (log)	9752	7.275988	0.9219138	-2.52729	10.7522
Marriage Status	9752	0.6062346	0.4866089	0	1
Race	9752	0.7965546	0.4025817	0	1
Education	9752	2.63587	1.478315	1	5

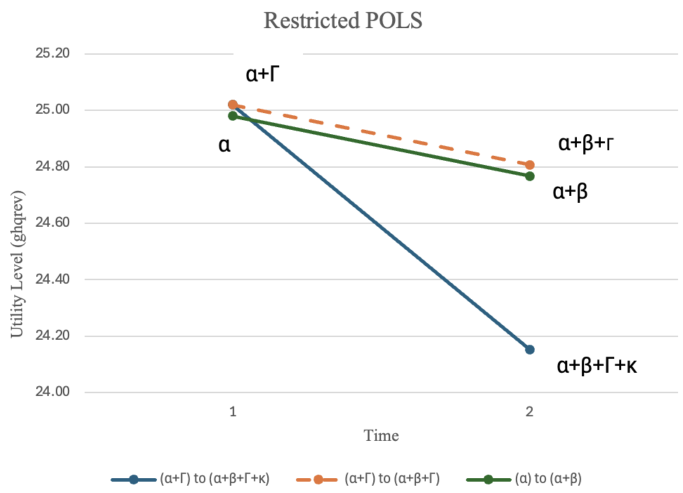
Variable – No child	Observations (N)	Mean	Standard Deviation	Min	Max
Age	18196	55.22082	16.11336	19	99
Gender	18196	0.4406463	0.4964783	0	1
Net Income (log)	18196	7.248737	0.9375728	-3.506558	11.24584
Marriage Status	18196	0.6165641	0.4862364	0	1
Race	18196	0.8777753	0.3275441	0	1
Education	18196	2.752638	1.490349	1	5

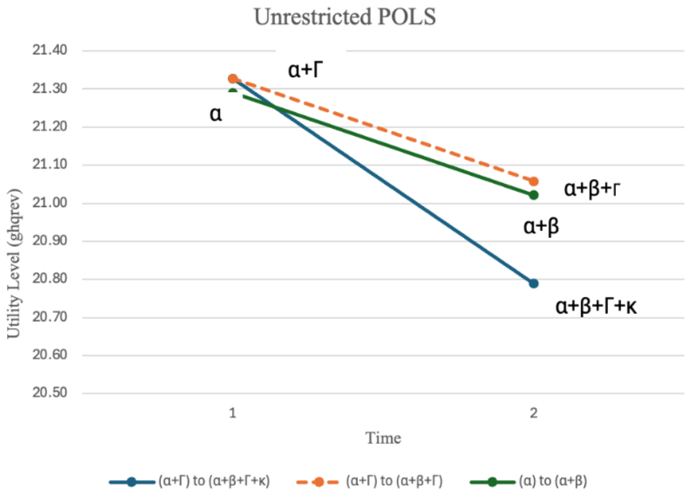
Appendix K - Identifying the Difference-in-Difference (κ) Coefficient

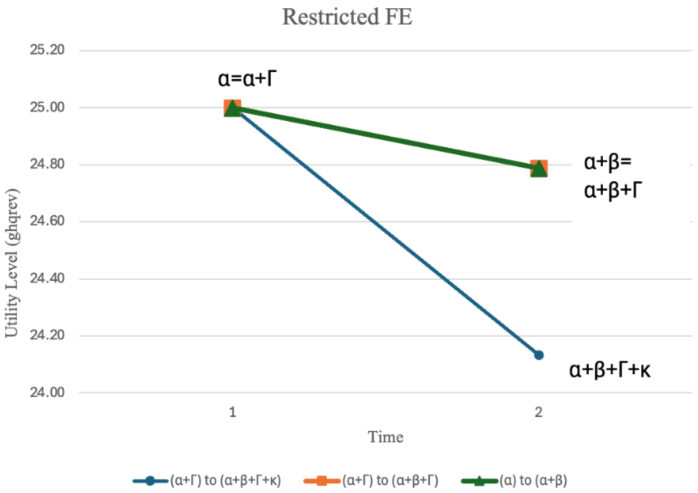
	Models (x)		Models (y)	
	Restricted FE	Restricted POLS	Unrestricted FE	Unrestricted POLS
Have Children (Γ)	-	0.0399	-	0.0371
		(0.0935)		(0.0922)
Time [β]	-0.212***	-0.212***	-0.327**	-0.269***
	(0.0496)	(0.0496)	(0.164)	(0.0497)
Interaction term between time and child [κ]	-0.655***	-0.655***	-0.651***	-0.640***
	(0.0901)	(0.0901)	(0.0902)	(0.0902)
Net Income (log)	-	-	-0.0232	0.211***
			(0.0510)	(0.0419)
Degree	-	-	0.0387	0.0884
			(0.966)	(0.115)
Postgraduate Degree	-	-	0.915	-0.0120
			(1.129)	(0.143)
GCSE	-	-	0.519	-0.0198
			(1.218)	(0.132)
No Qualification	-	-	1.060	-0.291**
			(1.133)	(0.147)
Age^2	-	-	0.00103	0.000410***
			(0.00132)	(0.0000229)
Male	-	-	0	0.879***
			(.)	(0.0812)
Married	-	-	0.422	0.795***
			(0.326)	(0.0832)
Caucasian	-	-	-	-0.0197
				(0.120)
_cons (α)	25.00***	24.98***	21.15***	21.29***
	(0.0208)	(0.0548)	(4.397)	(0.344)
R^2	0.012	0.003	0.012	0.039
N	27948	27948	27948	27948

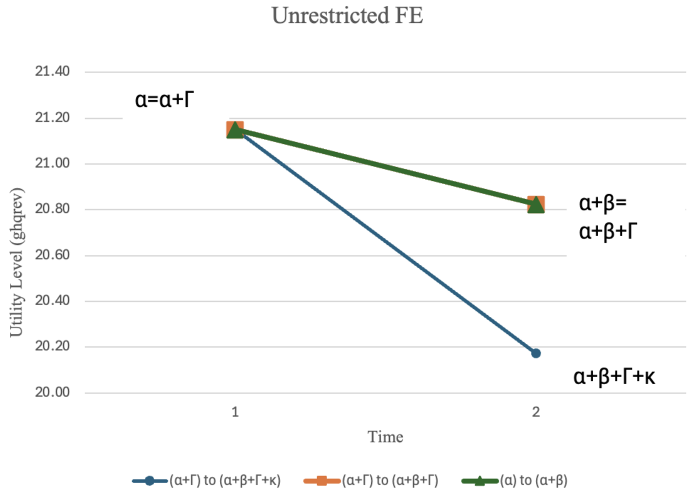
Notes: Coefficients represent marginal effects. Clustered Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Refer to Appendix J for estimated coefficient visualised.

Appendix L – Treatment Effect Values Visualised









To view models in excel, please click the following: [Models in Excel]

Values from each of the respective models:

Restricted FE

α	25.0000	α+β	24.7880	T	1	2
β	-0.2120	α+β+Γ	24.7880	(α+Γ) to (α+β+Γ+κ)	25.00	24.1330
Γ	0.0000	α+β+Γ+κ	24.1330	(α+Γ) to (α+β+Γ)	25.00	24.79
κ	-0.6550	α+Γ	25.0000	(α) to (α+β)	25.00	24.79

Restricted POLS

α	24.9800	α+β	24.7680	T	1	2
β	-0.2120	α+β+Γ	24.8079	(α+Γ) to (α+β+Γ+κ)	25.02	24.1529
Γ	0.0399	α+β+Γ+κ	24.1529	(α+Γ) to (α+β+Γ)	25.02	24.81

κ	-0.6550	α+Γ	25.0199	(α) to (α+β)	24.98	24.77

Unrestricted FE

α	21.1500	α+β	20.8230	T	1	2
β	-0.3270	α+β+Γ	20.8230	(α+Γ) to (α+β+Γ+κ)	21.15	20.1720
Γ	0.0000	α+β+Γ+κ	20.1720	(α+Γ) to (α+β+Γ)	21.15	20.82
κ	-0.6510	α+Γ	21.1500	(α) to (α+β)	21.15	20.82

Unrestricted POLS

α	21.2900	α+β	21.0210	T	1	2
β	-0.2690	α+β+Γ	21.0581	(α+Γ) to (α+β+Γ+κ)	21.33	20.7891
Γ	0.0371	α+β+Γ+κ	20.7891	(α+Γ) to (α+β+Γ)	21.33	21.06
κ	-0.2690	α+Γ	21.3271	(α) to (α+β)	21.29	21.02

Appendix M – Stata Inputs

```
/*Restrict sample to South East and London*/
/*Note that South East and London are coded as 7 and 8*/

codebook region
replace region =. if inlist(region, 1,2,3,4,5,6,9)
replace region = 1 if region == 7
replace region = 2 if region == 8
codebook region

label var region "1=London, 2=SE"
label define region 1 "London" 2 "South East"
label values region region

/*Balancing the data*/
xtdes
bysort pidp: generate T=_N
keep if T==5
xtset pidp time

/*Reversing the General Health Questionnaire Score, so that the higher the number the better the health
heatlh*/
gen ghqrev = 36-ghq

/*Formatting Interview Date*/
gen intdate = dmy(istrtdatd, istrtdatm, istrtdaty)
format intdat %td

*----*

/*Cleaning the data, categorisation*/
label define male 0 "female" 1 "male"
label values male male

replace children = 1 if inlist(children,1,2,3,4,5,6,7,8)
tab children
label var children "0 = No Children, 1 = Has Children"
```

Ivan Hung

```
label define child 0 "no child" 1 "child"
label values children child

rename educat educ
replace educ = 5 if inlist(educ,5,9)
tab educ
label var educ "1 = Degree, 2 = Post-Grad, 3 = A-Level, 4 = GCSE, 5 = Other or none"
label define educ 1 "ug" 2 "pg" 3 "a-level" 4 "gcse" 5 "n/a"
label values educ educ

//Proportion of people who move into the 'next' category in each time period (from T=1 to T=5)
xttrans educ
xttab educ

gen jobstatus = jbstat
replace jobstatus=1 if inlist(jobstatus,1,3,5,6,7,8,9,10,11,12)
replace jobstatus=. if inlist(jobstatus,13,97)
replace jobstatus=3 if jobstatus==4
tab jobstatus
label var jobstatus "1 = Not paid employment, 2 = Paid employment, 3 = NILF"
label define jobstatus 1 "unemployed" 2 "employed" 3 "NILF"
label values jobstatus jobstatus

tab marryst
replace marryst = 0 if inlist(marryst,1,3,4,5,6,7,8,9,10)
replace marryst = 1 if marryst==2
label var marryst "0 = Not Married (other), 1 = Married"
label define marryst 0 "not married" 1 "married"
label values marryst marryst

replace race = 0 if inlist(race,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,97)
tab race
label var race "0 = Not White, 1 = White"
label define race 0 "not white" 1 "white"
label values race race

gen linc = log(nincome)

rename time t
/*Don't forget to mention how the split between the categories are disproportionate (~85:15 split, so any interpretations made here may not be reflected accurately*/

*----*


/*Carrying out Analysis:*/

//If coefficients estimated between RE and FE are very close, just go with the RE model
//If coefficients aren't close to each other, go with the FE model
//When we interpret a coefficient from a FE model, you are holding all other variables and individual heterogenetiy constant
//Within percent indicates how time persistant a variable is
```

Ivan Hung

```
*POLS Model
reg ghqrev linc ib(3).educ ib(2).jobstatus c.age##c.age i.male i.marryst i.children i.race t if inlist(t,1,2,3)
estimates store POLS

reg ghqrev linc ib(3).educ ib(2).jobstatus c.age##c.age i.male i.marryst i.children i.race t if inlist(t,1,2,3), vce(cluster pidp)
estimates store POLS_cluster

esttab POLS POLS_cluster using POLS.rtf, replace se r2 star(* 0.1 ** 0.05 *** 0.01) title ("POLS vs POLS Clustered") obslast mtitles compress

*RE Model
xtreg ghqrev linc ib(3).educ ib(2).jobstatus c.age##c.age i.male i.marryst i.children i.race t if inlist(t,1,2,3), re theta
estimates store RE

xtreg ghqrev linc ib(3).educ ib(2).jobstatus c.age##c.age i.male i.marryst i.children i.race t if inlist(t,1,2,3), re vce(cluster pidp) theta
estimates store RE_cluster

esttab RE RE_cluster using RE.rtf, replace se r2 star(* 0.1 ** 0.05 *** 0.01) title ("RE vs RE Clustered") obslast mtitles compress

*FE Model 1 - Standard
xtreg ghqrev linc ib(3).educ ib(2).jobstatus c.age##c.age i.male i.marryst i.children i.race t if inlist(t,1,2,3), fe
estimates store FE

xtreg ghqrev linc ib(3).educ ib(2).jobstatus c.age##c.age i.male i.marryst i.children i.race t if inlist(t,1,2,3), fe vce(cluster pidp)
estimates store FE_cluster

esttab FE FE_cluster using FE.rtf, replace se r2 star(* 0.1 ** 0.05 *** 0.01) title ("FE vs FE Clustered") obslast mtitles compress

*FE Model 2 - Without time variable
xtreg ghqrev linc ib(3).educ ib(2).jobstatus c.age##c.age i.male i.marryst i.children i.race if inlist(t,1,2,3), fe
estimates store FE2

xtreg ghqrev linc ib(3).educ ib(2).jobstatus c.age##c.age i.male i.marryst i.children i.race if inlist(t,1,2,3), fe vce(cluster pidp)
estimates store FE2_cluster

*FE Model 3 - Without age variable (but age^2 still included)
xtreg ghqrev linc ib(3).educ ib(2).jobstatus c.age#c.age i.male i.marryst i.children i.race t if inlist(t,1,2,3), fe
estimates store FE3

xtreg ghqrev linc ib(3).educ ib(2).jobstatus c.age#c.age i.male i.marryst i.children i.race t if inlist(t,1,2,3), fe vce(cluster pidp)
estimates store FE3_cluster

*---*
/*Comparison Tables used in paper*/
```

Ivan Hung

```
*POLS Clustered vs RE vs RE Clustered:
esttab POLS_cluster RE RE_cluster using POLS_RE.rtf,replace se r2 star(* 0.1 ** 0.05 *** 0.01) title
("POLS Clustered vs RE vs RE Clustered") obslast mtitles compress

*RE Clustered vs Standard FE vs FE without t vs FE without age:
esttab RE_cluster FE_cluster FE2_cluster FE3_cluster using RE_FE.rtf, replace se r2 star(* 0.1 ** 0.05 ***
0.01) title ("RE Clustered vs Standard FE vs FE without t vs FE without age") obslast mtitles compress

*Standard FE vs FE without t vs FE without age:
esttab FE_cluster FE2_cluster FE3_cluster using FEs.rtf, replace se r2 star(* 0.1 ** 0.05 *** 0.01) title
("Standard FE vs FE without t vs FE without age") obslast mtitles compress

*---*
/*Bruesch Pagan LM Test - POLS vs RE*/
*Informal by comparing the coefficients between both models:
reg ghqrev linc ib(3).educ ib(2).jobstatus c.age##c.age i.male i.marryst i.children i.race t if inlist(t,1,2,3),
vce(cluster pidp)
estimates store POLS_cluster

xtreg ghqrev linc ib(3).educ ib(2).jobstatus c.age##c.age i.male i.marryst i.children i.race t if inlist(t,1,2,3),
re theta
estimates store RE

xtreg ghqrev linc ib(3).educ ib(2).jobstatus c.age##c.age i.male i.marryst i.children i.race t if inlist(t,1,2,3),
re vce(cluster pidp) theta
estimates store RE_cluster

esttab POLS_cluster RE RE_cluster using POLSvRE, replace se r2 star(* 0.1 ** 0.05 *** 0.01) title ("POLS
Clustered vs RE") obslast mtitles compress

*More formal (Using the BPLM Test):
xtreg ghqrev linc ib(3).educ ib(2).jobstatus c.age##c.age i.male i.marryst i.children i.race t if inlist(t,1,2,3),
re vce(cluster pidp) theta
xttest0


*---*

/*Regression-based Hausman Test - RE vs FE*/

*Generating Individual-specific means*
gen age2 = age^2
tab educ, gen(education)
tab jobstatus, gen(jobstat)

bysort pidp: egen mage=mean(age)
bysort pidp: egen mage2=mean(age^2)

xttab jobstatus if inlist(t,1,2,3)
bysort pidp: egen mjobstat1=mean(jobstat1)
bysort pidp: egen mjobstat2=mean(jobstat2)
bysort pidp: egen mjobstat3=mean(jobstat3)
```

Ivan Hung

```
xttab education if inlist(t,1,2,3)
bysort pidp: egen meduc1=mean(education1)
bysort pidp: egen meduc2=mean(education2)
bysort pidp: egen meduc3=mean(education3)
bysort pidp: egen meduc4=mean(education4)
bysort pidp: egen meduc5=mean(education5)

bysort pidp: egen mlinc=mean(linc)

xttab marryst if inlist(t,1,2,3)
bysort pidp: egen mmarry=mean(marryst)

xttab children if inlist(t,1,2,3)
bysort pidp: egen mchildren=mean(children)


/*Estmating a Correlated Random Effects Model*/
xtreg ghqrev linc ib(3).educ ib(2).jobstatus age2 i.male i.marryst i.children i.race mlinc meduc1 meduc2
meduc4 meduc5 mjobstat1 mjobstat3 mage mage2 mmarry mchildren t if inlist(t,1,2,3), re vce(cluster pidp)

estimates store CRE

*The CRE Model should be equivalent to a FE model:
xtreg ghqrev linc ib(3).educ ib(2).jobstatus c.age#c.age i.male i.marryst i.children i.race t if inlist(t,1,2,3), fe
vce(cluster pidp)'

estimates store CREFE

esttab CRE CREFE using CRE_FE.rtf, replace se r2 star(* 0.1 ** 0.05 *** 0.01) title ("CRE vs FE
Clustered") obslast mtitles compress

/*Hausman Test*/
*Without heteroskedasticity and Serial Correlation Robust SEs
xtreg ghqrev age2 male ib(2).jobstatus linc marryst children ib(3).educ race mage mage2 mjobstat1
mjobstat3 mlinc mmarry mchildren meduc1 meduc2 meduc4 meduc5 t if inlist(t,1,2,3), re

test mage mage2 mjobstat1 mjobstat3 mlinc mmarry mchildren meduc1 meduc2 meduc4 meduc5
*p-value < 0.01. Therefore, we prefer to use the FE model

*With heteroskedasticity and Serial Correlation Robust SEs

xtreg ghqrev age2 male ib(2).jobstatus linc marryst children ib(3).educ race mage mage2 mjobstat1
mjobstat3 mlinc mmarry mchildren meduc1 meduc2 meduc4 meduc5 t if inlist(t,1,2,3), re vce(cluster pidp)

test mage mage2 mjobstat1 mjobstat3 mlinc mmarry mchildren meduc1 meduc2 meduc4 meduc5
*p-value < 0.01. Therefore, we prefer to use the FE model

*----------*

xtreg ghqrev linc ib(3).educ ib(2).jobstatus c.age#c.age i.male i.marryst i.children i.race t, fe vce(cluster
pidp)

drop if t==1
```

Ivan Hung

```
drop if t==2
drop if t==3

replace t=1 if t==4
replace t=2 if t==5

/*Identifing the treated vs control group*/
bysort pidp (intdate): replace children = intdate>td(23/03/2020) if _n==2
bysort pidp (intdate): replace children = children[2] if _n==1

tab children t

reg ghqrev i.children##i.t, vce(cluster pidp)

tab children

ttest ghqrev, by(children) unequal reverse
/*on average, those who have children have -0.2874496 utility points compared to those who don't have
children
Note: we include the option 'unequal' to control for possible unequal variance of student performance
between treatment and control groups (heteroskedasticity)
*/

*Gives us the same result as above
reg ghqrev children, vce(robust)

estimates store childaveragereg
esttab childaveragereg using avg.rtf, replace se r2 star(* 0.1 ** 0.05 *** 0.01) title ("Difference in utility
points of those with/without children") obslast mtitles compress

sum age male linc marryst race educ if children==1
sum age male linc marryst race educ if children==0

*Notice from these summaries that the averages of both categories are very similar. Despite not have
generating a positive/negative bias from either category, we must be careful when interpretating the impacts
of individuals who have chilren as we only have half the number of observations for them compared to those
without children. We should be concious of having sampling bias when interpretating coefficients later.

reg ghqrev children c.age male linc marryst race educ, vce(robust)
estimates store childaverageregfull
esttab childaveragereg childaverageregfull using avgfull.rtf, replace se r2 star(* 0.1 ** 0.05 *** 0.01) title
("Difference in utility points of those with/without children") obslast mtitles compress
*-0.2791538


*Treated Group
mean ghqrev if t==1 & children==1
mean ghqrev if t==2 & children==1
di 24.1571 - 25.024
// = -0.8669

gen D = 0
replace D = 1 if children==1 & t==2
```

Ivan Hung

```
*FE Model
xtreg ghqrev D if children==1, fe vce(cluster pidp)
*-0.8668991


*Control Group
mean ghqrev if t==1 & children==0
mean ghqrev if t==2 & children==0
di 24.77193 - 24.98406
// = -0.21213


*Diff-in-Diff
di (24.1571 - 25.024) - (24.77193 - 24.98406)
* Diff-in-Diff = -0.65477

*FE Model
xtreg ghqrev children##i.t, fe vce(cluster pidp)
estimates store DiDFE
*Diff-in-Diff = -0.6547646

*POLS Model
reg ghqrev children##i.t, vce(cluster pidp)
estimates store DiDPOLS
*Diff-in-Diff = -0.6547646


*FE Model included with explanatory variables
xtreg ghqrev i.children##i.t linc ib(3).educ c.age#c.age i.male i.marryst i.children i.race, fe vce(cluster pidp)
estimates store DiDFEfull
*Diff-in-Diff increases to -0.65117

*POLS Model included with explanatory variables
reg ghqrev i.children##i.t linc ib(3).educ c.age#c.age i.male i.marryst i.children i.race, vce(cluster pidp)
estimates store DiDPOLSfull
*Diff-in-Diff increases to -0.6395484

esttab DiDFE DiDPOLS DiDFEfull DiDPOLSfull using DiD.rtf, replace se r2 star(* 0.1 ** 0.05 *** 0.01)
title ("Identifying the Difference-in-Difference Coefficient") obslast mtitles compress
```