178724

Applied Econometrics Methods

Investigating the Impact of Psychological Distress on Wealth

Luis Vieira
23012096

June, 2024

Index

1 Introduction

This study investigates the relationship between psychological distress (PD) and wealth, with particular interest on how this relationship varies across different socioeconomic status (SES) levels. Using longitudinal data inspired by the Panel Study of Income Dynamics (PSID)[1], the analysis covers the period from 2003 to 2019, allowing for a robust exploration of wealth accumulation dynamics among a diverse sample of individuals.

The study is divided into two main sections. The first part compares Ordinary Least Squares (OLS) and Entity and Time Fixed Effects (FE) models to illustrate the differences in estimated coefficients and the explainability of the models. This comparison highlights the limitations of OLS due to issues like serial correlation, omitted variable bias, and endogeneity, and demonstrates how FE models can provide more reliable estimates by accounting for unobserved heterogeneity but not without caveats.

The second part of the study aims to confirm the hypothesis that PD significantly affects wealth and explores whether this effect differs based on socioeconomic status (SES). Using entity FE models, the analysis controls for time-invariant characteristics and includes interactions to capture time-variant factors, such as negative life events. The study takes into consideration the potential for reverse causality and omitted variable bias, ensuring a comprehensive examination of the relationship between PD and wealth.

The findings from this analysis have important implications for addressing the economic consequences of psychological distress, in particular how PD interacts with SES and negative life events to influence wealth outcomes.

[1] Source: https://psidonline.isr.umich.edu/

2 Descriptive Statistics and EDA

2.1 Dataset Description

The dataset used for this analysis of wealth accumulation dynamics is simulated data inspired by the Panel Study of Income Dynamics (PSID) from the research by Balloch et al. (2022). It covers the period from 2003 to 2019 and includes a sample of 210,388 observations across 26 variables (Table 1). The data is already organised in a longitudinal format, allowing to investigate and analyse individuals over time and gain a deeper understanding of the factors influencing their wealth accumulation.

Table 1: Variables description[2]

Variables	Description	Data Type
	Dependent Variables	
Financial distress (financial_distress)	A dummy variable that is equal to 1 if net wealth is negative and is equal to 0 if net wealth is positive or zero.	Dummy
Net worth (wealth)	Captures the values of assets minus debts, where positive net worth means surplus wealth and negative net worth means deficit wealth.	Continuous
	Key Independent Variable	
Psychological distress (pd)	Captures the psychological distress score of the respondents, where higher score means higher psychological distress (maximum score is 24 and minimum score is 0).	Continuous
	Control Variables	
Education (education)	Captures the respondents' years of schooling.	Continuous
Income (income)	Captures the combined labour income of all household members (in logs).	Continuous
Age (age)	This variable is equal to the respondents' age in years.	Continuous
Male (male)	It takes the value of one if the respondent is male, and zero otherwise.	Dummy
Employed (employed)	Equal to one if the respondent is employed, and zero otherwise.	Dummy
Divorce (divorce)	Equal to one if the respondent recently experienced a divorce, and zero otherwise.	Dummy
Marriage (marriage)	Equal to one if the respondent recently got (re)married, and zero otherwise.	Dummy
Birth of child (childbirth)	Equal to one if a household member recently gave birth, and zero otherwise.	Dummy
Death of family member (familydeath)	Equal to one if a household member recently died, and zero otherwise.	Dummy
Lay off (laidoff)	Equal to one if the respondent was recently laid off from work, and zero otherwise.	Dummy
Missed work with illness (missedwork)	Captures the total number of weeks of work missed due to illness.	Continuous
White (white)	Equal to one for "White" ethnicity, and zero otherwise.	Dummy
Black (black)	Equal to one for "Black" ethnicity, and zero otherwise.	Dummy
Hispanic (hispanic)	Equal to one for "Hispanic" ethnicity, and zero otherwise.	Dummy
Other ethnicity (otherethnicity)	Equal to one for reports of ethnicity other than "Black", "Hispanic" or "White", and zero otherwise.	Dummy
College degree (collegedegree)	Equal to one for respondents with a college degree. (Other options: No college degree = 5; Did not study in college = 0; No answer = 9)	Dummy
Student loan (studentloan)	Captures the respondents' student loan outstanding. This variable is only captured for those who took a student loan.	Continuous

[2] Table source: *in Appendix*, Assessment 2 Guidelines for course Applied Econometrics Methods 178724

Socio-economic status (socioeconomic)	Captures the respondents' socio-economic status, where higher score means higher socio-economic status. Socio-economic status variable is measured as: Total income – Poverty threshold.	Continuous
Year (year)	Captures the year of the biannual (data collected once in two years) surveys. Also, the year variable has been encoded as a categorical number, where: 0=2003, 1=2005, 2=2007, 3=2009, 4=2011, 5=2013, 6=2015, 7=2017, 8=2019.	Categorical
ID (id)	Captures the unique identifier of the respondents.	Categorical
Head (head)	This variable is equal to one if the respondent is the head of the household.	Dummy
Not moved (notmoved)	This variable is equal to one if the head did not change within the family.	Dummy
No family change (nofamichange)	This variable is equal to one if the respondent did not move in or out of the family.	Dummy

2.2 Descriptive Statistics & EDA

Table 2: Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
year	210388	3.615	2.546	0	8
financial distress	210388	.153	.36	0	1
wealth	210388	220330.3	1058502.3	-3197000	100555000
pd	201659	3.366	3.976	0	24
education	197122	12.928	2.743	0	17
income	203764	10.742	1.033	0	16
age	204856	43.458	14.264	16	101
male	204905	.729	.445	0	1
employed	204905	.751	.433	0	1
divorce	210388	.071	.257	0	1
marriage	210388	.103	.304	0	1
childbirth	210388	.19	.392	0	1
familydeath	210388	.026	.159	0	1
laidoff	210381	.059	.236	0	1
missedwork	204905	1.079	3.583	0	75
white	210367	.538	.499	0	1
black	210367	.358	.479	0	1
hispanic	210367	.017	.131	0	1
otherethnicity	210367	.087	.281	0	1
collegedegree	204832	1.278	1.939	0	9
studentloan	79823	9664.373	29444.26	0	700000
socioeconomic	210302	37318.865	102922.48	-96352	6278577

Table 2 presents the descriptive statistics for key variables in the study. This study's main dependent variable Wealth shows a mean of 220,330.3 with a substantial standard deviation of 1,058,502.3, showing a large variability and the presence of extreme values and right skewness, with a range from -3,197,000 to 100,555,000. PD has a mean of 3.37 and a standard deviation of 3.98, indicating lower levels of distress for most individuals, with scores ranging from 0 to 24. Socioeconomic status displays a mean of 37,318.87 and a standard deviation of 102,922.48, reflecting like wealth skewness and considerable differences in socioeconomic conditions, ranging from -96,352 to 6,278,577.

Income, which is log-transformed, has a mean of 10.74 and a standard deviation of 1.03, showing a range from 0 to 16. Education averages above the completed secondary school and with a standard deviation of 2.74 years, indicates diverse educational backgrounds. The average age is considerably high in the dataset (just under 44 years), with ages reaching up to 101 years! The dataset also includes demographic variables such as gender, with 72.9% of the sample being male, and employment status, with 75.1% being employed.

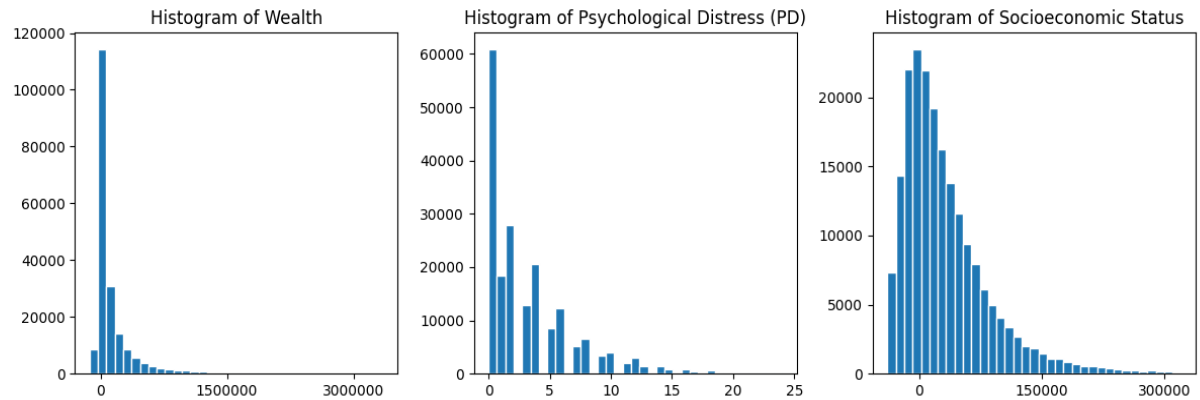
It is important to note the presence of missing values in some variables, such as PD (8,729 missing), education (13,266 missing), income (6,624 missing), and age (5,532 missing). However, the variable for student loans has over half of the data missing (130,565), but will not impact the study as I do not plan to use it. Despite these missing values, the dataset still allows for a robust analysis, focusing on the relationships between wealth, psychological distress, and socioeconomic status.

2.3 Outliers & Transformations

The summary descriptive statistics (Table 1) and the histograms in Figure 1 reveal that both wealth and socioeconomic status (SES) variables are right skewed, with extreme outliers likely having a significant impact. To deal with outliers, I chose trimming over the standard deviation method or Winsorising. Trimming does not artificially alter the data, unlike reducing the weight of outliers, which is similar to imputing values. Although trimming removes some data, it effectively excludes extreme values that do not represent the majority, reducing bias and keeping the analysis accurate (Tukey, 1962).

Thus, I trimmed values below the 1st percentile and above the 99th percentile (wealth between -128,000 and 3,355,000, and SES between -45,713 and 342,892).

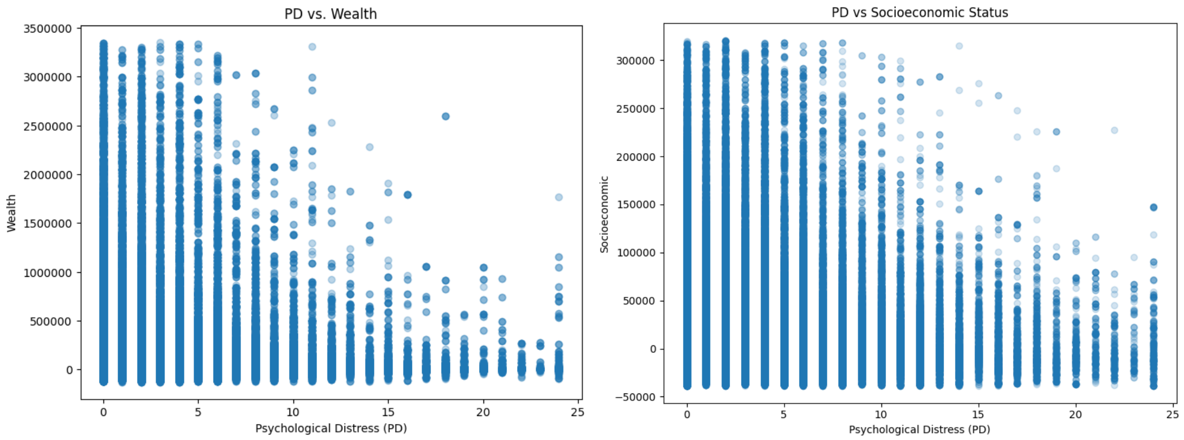
Figure 1: Histograms of Wealth, PD and Socioeconomic Status (Trimmed Outliers)



Additionally, I used both variables inverse hyperbolic arcsine (arsinh) transformations for improving normalisation of their distributions, in a similar way to what was already done for the income variable (log).

The arsinh transformation is widely used and in econometrics, with the added advantage of handling zero and negative values without requiring adjustments and thanks to its similarity to the logarithmic transformation (Burbidge et al., 1988). This approach allows for a more accurate analysis by mitigating the distortions caused by extreme values.

Figure 2: PD vs Wealth and PD vs Socioeconomic Status (Trimmed Outliers)

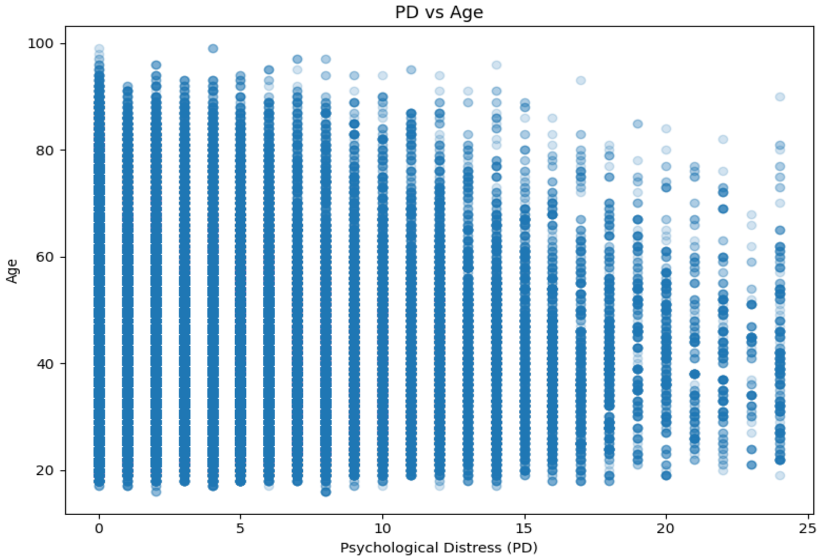


Visualising the data in Figure 2 above, the two-way graphs of Wealth vs PD and SES vs PD display a similar pattern, that wealthier and high SES individuals do not tend to reach critical PD levels, levels of PD of 12 and above are mostly dominated by lower wealth and low SES.

Figure 3: Individuals Network (Wealth) by Age

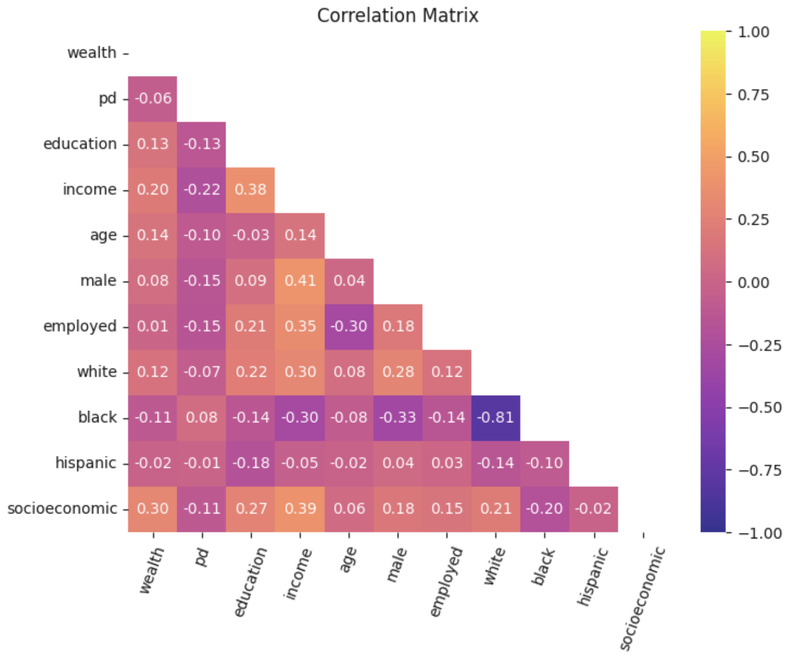
Considering the relationship between wealth and age, displaying an inverted U-shape, I include a polynomial term age*age to address non-linearity. The dynamic of the individuals after retirement and range extending beyond 100 years old, in my modelling I later focus on working age of individuals thus truncated between 18 and the USA retirement age of 67 (18 < age < 68), also as it Figure 4 hereafter shows, it appears that individuals after retirement are less likely to have very high levels of PD.

Figure 4: Psychological Distress by Age



The correlation matrix in Figure 5 shows the relationships between key variables in this study, with Wealth having a positive correlation with income (20%) and socioeconomic status (30%). This indicates that higher income and SES levels are associated with greater wealth. PD is negatively correlated with Wealth (-6%) and income (-22%), suggesting that higher distress levels are linked to limited income levels. Education correlates positively with income (38%) but negatively with pd (-13%), implying that higher education is associated with higher income and lower PD.

Figure 5: Pearson Correlation Matrix of Main Variables



The descriptive statistics and EDA provided a detailed overview of the dataset, the key relationships and identifying potential issues such as skewness and outliers. With the dataset now appropriately transformed, we can move forward to the econometric analysis.

3 Comparing OLS with Fixed Effects Models

In the previous analysis using OLS, the panel data was treated as a cross-sectional dataset, ignoring individual specific background effects and time dynamics. This approach led to several limitations, such as serial correlation, omitted variable bias, and endogeneity due to unobserved factors influencing both the predictors and the dependent variable. Additionally, there were issues with heteroskedasticity due to variance changes across different individuals over time. I suggested to mitigate these problems the use of robust standard errors (White, 1980) or clustered standard errors (Wooldridge, 2010), but they could not address all biases, especially endogeneity (Gujarati & Porter, 2010).
As we now transition from OLS to the using Fixed Effects modelling, and reverting back to our OLS assumptions failures, it was also suggested to use FE models to overcome most of these issues. These take into account unobserved heterogeneity by including individual specific and can overcome these prior issues. FE models control for time-invariant background, reducing bias from omitted variables correlated with explanatory variables. This approach is often preferred due to the ability to provide more reliable estimates whilst focusing on within-entity variations and eliminating background information.
Conversely to OLS, coefficients in FE models are not considered true betas because they reflect the relationship within entities over time, considering time-invariant characteristics rather than the overall average effect across all entities. Moreover, we must note that there are also limitations to FE models. Time-invariant characteristics for each individual that FE models cannot capture as they are ignored, such as expected shocks in wealth that individuals can and often anticipate and start affecting their PD even before the actual event happens. Using lag variables can provide some improvements, but a better more robust method would be using Difference-in-Differences (DiD). This technique can assess and compare the differential changes in outcomes around the event, controlling for both time-invariant and time-varying unobserved factors.
This time we focus only on comparing OLS and FE models, where I try to explain the differences in estimated coefficients (betas) and demonstrate how FE models address potential biases in OLS estimates, leading to more accurate results (Angrist & Pischke, 2009).

The base model used for comparing OLS and FE models is based on the one used previously (removed "white" and did not truncate nor removed outliers) and specified as:

$$\text{arsinh(wealth)} = \beta_0 + \beta_1 * \text{pd} + \beta_2 * \text{education} + \beta_3 * \text{income} + \beta_4 * \text{age} + \beta_5 * \text{male} + \beta_6 * \text{black} + \beta_7 * \text{employed} + \beta_8 * \text{arsinh(socioeconomic)} + \epsilon$$

When we compare the OLS, Entity-FE and Time-FE models (Table 3), we can clearly notice differences in the estimated coefficients, and the reduced R-squared for Entity-FE. Models shows interesting results, PD has a strong negative effect on wealth in both the OLS and Time-FE models (-0.149), but smaller effect in the Entity-FE model (-0.054). This means some of the impact in OLS is due to unchanging personal abilities, like a person's long term mental health. Socioeconomic status has a positive effect on wealth, with coefficients of 0.081 (OLS) and 0.077 (Time-FE), but only 0.018 (Entity-FE), showing OLS might overestimate this due to, for example, someone from a wealthy family might always have higher wealth, independently of changes in income or education. Education's negative impact on wealth is stronger in the Entity-FE model (-0.134) compared to OLS (-0.032), likely due to student loans. The race variable black is significant in OLS (-1.560) and Time-FE (-1.526) but not in Entity-FE (0.031), suggesting racial wealth differences in OLS are influenced by personal background factors.

Table 3: OLS vs Entity Fixed Effects

VARIABLES	(1) OLS	(2) Entity-FE	(3) Time-FE
pd	-0.149***	-0.054***	-0.149***
	(0.005)	(0.005)	(0.005)
education	-0.032***	-0.134***	-0.008
	(0.007)	(0.013)	(0.007)
income	1.053***	0.373***	1.129***
	(0.023)	(0.025)	(0.023)
age	0.125***	0.065***	0.126***
	(0.001)	(0.002)	(0.001)
male	1.234***	0.608***	1.200***
	(0.045)	(0.062)	(0.045)
black	-1.560***	0.031	-1.526***
	(0.040)	(0.173)	(0.040)
employed	0.549***	0.535***	0.415***
	(0.047)	(0.050)	(0.047)
ars_ses	0.081***	0.018***	0.077***
	(0.002)	(0.002)	(0.002)
Constant	-9.355***	1.785***	-10.372***
	(0.225)	(0.301)	(0.226)
Observations	187,611	187,611	187,611
R-squared	0.163	0.013	0.166
Number of ids		29,841	
Number of years			9

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Including entity effects helps account and control for individual backgrounds isolating their lifestyle, abilities, social (like family) and financial backgrounds' impact on wealth, such as individual capacity to manage their finances. Whereas time fixed effects, capture the macroeconomic conditions and changes which are likely to affect all individuals. Using FE approaches ensures that our estimates are not biased by factors that vary across individuals or over time.

When using Hausman test against the Random Effects model the Entity-FE rejected the null hypothesis (chi^2= 3010.96), thus indicating that Entity-FE model is appropriate.

In summary, comparing OLS, Entity-FE and Time-FE models reveals significant differences in the estimated coefficients. The OLS model tends to overestimate or underestimate relationships due to unobserved heterogeneity. Fixed effects models, both entity and time, give more reliable estimates by controlling for these unobserved factors. An FE approach can provide a more accurate understanding of the true relationship between psychological distress and wealth.

4 Does PD Significantly Affects Wealth?

In this section, I aim to confirm the hypothesis that PD significantly affects wealth and explore how this relationship may differ based on SES. The analysis uses entity fixed effects models to control for time-invariant characteristics of individuals and includes interactions to capture time-variant factors. Also accounting for potential conflicting and distorting factors that might affect both PD and wealth.

4.1 Investigating PD Effects Wealth interacting with SES Levels

To investigate the impact of PD on wealth, I started by removing missing values after excluding the student loan variable to retain the maximum amount of observations possible (187,611). Initially removing outliers, I simply applied an arsinh transformation to the variables wealth and socioeconomic to normalise their distributions.
I then ran a baseline entity FE model (Model 1 in the table 4) and a model (2) including an interaction term for PD and high SES. The SES groups were split into quartiles with high SES considered as those in the top 25% percentile.

Table 4: Entity Fixed Effects (Baseline and PD interaction with High SES)

VARIABLES	(1) Base-FE	(2) InteractionSES-FE
pd	-0.053***	-0.045***
	(0.005)	(0.006)
pd_high_ses		-0.056***
		(0.015)
age	0.064***	0.064***
	(0.002)	(0.002)
male	0.613***	0.613***
	(0.063)	(0.063)
education	-0.124***	-0.124***
	(0.013)	(0.013)
income	0.399***	0.398***
	(0.025)	(0.025)
ars_ses	0.017***	0.017***
	(0.002)	(0.002)
black	-0.073	-0.069
	(0.182)	(0.182)
hispanic	1.100***	1.097***
	(0.141)	(0.141)
otherethnicity	-0.231*	-0.230*
	(0.139)	(0.139)
employed	0.523***	0.526***
	(0.050)	(0.050)
divorce	-1.039***	-1.038***
	(0.067)	(0.067)
marriage	0.265***	0.265***
	(0.059)	(0.059)
childbirth	-0.132***	-0.132***
	(0.047)	(0.047)
familydeath	-0.223**	-0.219*
	(0.112)	(0.112)
laidoff	-0.090	-0.090
	(0.077)	(0.077)
missedwork	-0.019***	-0.019***
	(0.005)	(0.005)
Constant	1.549***	1.567***
	(0.306)	(0.306)
Observations	187,611	187,611
R-squared	0.015	0.016
Number of ids	29,841	29,841

The baseline model shows a significant negative direct effect of PD on wealth (about 5.3% decrease). The interaction model also reveals that high SES individuals experience an additional negative effect of PD (4.5% plus about 5.6% decrease) on wealth when compared with other SES groups. This suggests that while PD negatively statistically impacts wealth across all SES groups, the effect is more pronounced for those in the high SES group but there could be something else affecting both PD and wealth.

4.2 Investigating PD Effect on Wealth interacting with SES and Life Events

Further investigation into whether the relationship between PD and wealth varies across different SES levels, I used a binary variable created to capture negative life events, such as divorce, missed work, or being laid off, which could impact financial status. Family death was excluded because it might lead to increased wealth due to inheritance, especially for high SES individuals. Similarly, having children and getting married were excluded as they are generally planned events, more likely to involve savings or synergies that reduce costs and potentially lead to higher wealth. I also included the age polynomial term age^2. The objective in here was to determine if the observed effect of PD on wealth is not just due to SES or other confounding factors.

Table 5: Entity FE – Trimmed Outliers (Baseline vs PD interaction with High SES and Neg. Life Events)

	(1)	(2)	(3)
VARIABLES	Baseline-FE	Inter_SES-FE	Inter_SESNegEvents-FE
pd	-0.051***	-0.049***	-0.030***
	(0.006)	(0.006)	(0.007)
pd_high_ses		-0.033**	-0.034**
		(0.016)	(0.016)
pd_neg_life_events			-0.060***
			(0.008)
age	0.067***	0.337***	0.336***
	(0.003)	(0.013)	(0.013)
age^2		-0.003***	-0.003***
		(0.000)	(0.000)
education	-0.123***	-0.121***	-0.121***
	(0.014)	(0.014)	(0.014)
income	0.483***	0.415***	0.418***
	(0.027)	(0.027)	(0.027)
ars_ses	0.020***	0.018***	0.018***
	(0.002)	(0.002)	(0.002)
black	-0.022	-0.024	-0.029
	(0.191)	(0.191)	(0.191)
hispanic	0.874***	0.742***	0.755***
	(0.148)	(0.148)	(0.148)
otherethnicity	-0.033	-0.012	-0.012
	(0.147)	(0.147)	(0.147)
employed	0.557***	0.401***	0.413***
	(0.054)	(0.055)	(0.055)
divorce	-1.060***	-1.078***	-0.893***
	(0.071)	(0.070)	(0.075)
marriage	0.202***	0.304***	0.299***
	(0.061)	(0.062)	(0.062)

childbirth	-0.067	-0.026	-0.029
	(0.049)	(0.049)	(0.049)
familydeath	-0.067	0.002	-0.007
	(0.136)	(0.136)	(0.136)
laidoff	-0.119	-0.141*	0.026
	(0.080)	(0.080)	(0.083)
missedwork	-0.022***	-0.023***	-0.024***
	(0.005)	(0.005)	(0.005)
Constant	0.928***	-3.255***	-3.301***
	(0.332)	(0.387)	(0.387)
Observations	167,214	167,214	167,214
R-squared	0.016	0.019	0.019
Number of ids	28,473	28,473	28,473

Due to the trimming of outliers and truncation of age to consider only working force between 18 and 67 led to a slight drop in observations to 167,214 and loss of about 1500 ids. The R-squared improved but very slightly.

The baseline model now shows that unit increases in PD lead to decrease in wealth of approximately 5.1%, indicating a significant negative effect although lower than before removing outliers.

In the interaction model with high SES now shows a negative effect of PD on wealth and remains significant at 4.9%, an increase when compared to before removing outliers. And there is an additional 3.3% decrease in wealth due to PD for high SES individuals.

Comparing with the added interaction of PD with negative life events, both significant but PD's negative effect on wealth is more dominated by the negative life events with an increased 6% burden on the wealth. Negative events significantly exacerbate the negative effect of PD on wealth, suggesting that negative life shocks compound the financial impact of psychological distress.

The initial hypothesis that psychological distress (PD) affects wealth needs to account for the possibility that the observed relationship might be driven by other factors, such as socioeconomic status (SES) and negative life events. By including interaction terms for high SES and negative life events, we can better understand whether the relationship between PD and wealth is direct or if it is influenced by other underlying factors. The results show that while PD negatively affects wealth, the impact is more severe for high SES individuals and those experiencing negative life events, likely meaning these individuals have more to lose. This suggests that SES and life events play a crucial role in the relationship between PD and wealth.

In short, the analysis confirms that psychological distress negatively affects wealth, with the effect being more severe for high SES individuals and those experiencing negative life events. This confirms the importance of considering SES and life events when examining the impact of PD on wealth, ensuring that the relationship is not confounded by these factors.

Looking into the residuals QQ-Plot (Figure 6), we can see a clear departure from normality particularly at the tails. By using clustered standard errors (Table 6) we see there is little change in the model 3, nonetheless the standard errors in the clustered model are slightly larger, especially for PD and its interactions (pd: 0.007 to 0.008, pd_high_ses: 0.016 to 0.017, pd_neg_life_events: 0.008 to 0.010).

This increase in standard errors suggests that accounting for within-cluster correlation provides a more conservative estimate of the variability in the coefficients, leading to slightly larger confidence intervals and potentially affecting the statistical significance of some variables.

In conclusion, the clustered standard errors indicate a more robust approach to estimating the variability of the coefficients, addressing potential issues with heteroscedasticity and within-group correlation observed in the residuals.

Figure 6: Model (3) Residuals QQ-Plot and Histogram

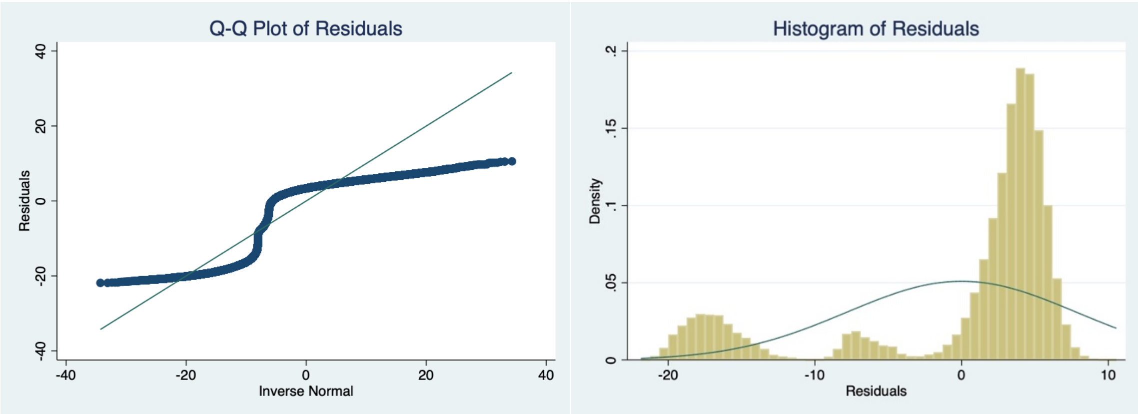


Table 6: Clustered Std. Errors Entity-FE with PD Interactions with SES and Neg. Life Events

VARIABLES	Inter_SESNegEvents-ClustErr
pd	-0.030***
	(0.008)
pd_high_ses	-0.034**
	(0.017)
pd_neg_life_events	-0.060***
	(0.010)
age	0.336***
	(0.016)
age2	-0.003***
	(0.000)
education	-0.121***
	(0.020)
income	0.418***
	(0.030)
ars_ses	0.018***
	(0.003)
black	-0.029
	(0.262)
hispanic	0.755***
	(0.158)
otherethnicity	-0.012
	(0.176)
employed	0.413***
	(0.061)
divorce	-0.893***
	(0.091)
marriage	0.299***
	(0.074)
childbirth	-0.029

	(0.057)
familydeath	-0.007
	(0.138)
laidoff	0.026
	(0.099)
missedwork	-0.024***
	(0.006)
Constant	-3.301***
	(0.478)
Observations	167,214
Number of ids	28,473
R-squared	0.019

4.3 Investigating PD Effect on Wealth with Fixed Head of Household

Finally, I narrowed down and focused on households where the head remained unchanged to control for potential biases introduced by changes in household composition. The dataset was filtered to include only observations where the head of the household did not change and there were no moves into or out of the family. The data as before was also trimmed to focus on working age individuals. Two models were used, a baseline FE model and an interaction model for fixed head of households with clustered standard errors.

Table 7: Clustered Std. Errors Entity-FE with PD Interactions with SES and Neg. Life Events for Fixed Head of Households

VARIABLES	(1) Baseline-FE	(2) Interac_FixHead-FE
pd	-0.027**	-0.018
	(0.012)	(0.014)
pd_high_ses		0.006
		(0.029)
pd_neg_life_events		-0.035**
		(0.017)
age	0.068***	0.067***
	(0.008)	(0.008)
education	-0.033	-0.034
	(0.060)	(0.060)
income	0.422***	0.424***
	(0.052)	(0.052)
ars_ses	0.002	0.002
	(0.005)	(0.005)
black	-0.172	-0.183
	(0.711)	(0.711)
hispanic	0.666*	0.670*
	(0.382)	(0.382)
otherethnicity	-0.023	-0.024
	(0.456)	(0.456)
employed	0.485***	0.493***
	(0.106)	(0.106)
divorce	-0.751***	-0.641***
	(0.180)	(0.185)
marriage	0.251	0.248

	(0.153)	(0.153)
childbirth	-0.018	-0.021
	(0.122)	(0.122)
familydeath	0.023	0.020
	(0.238)	(0.238)
laidoff	0.005	0.103
	(0.174)	(0.179)
missedwork	-0.017*	-0.018*
	(0.010)	(0.010)
Constant	0.479	0.476
	(1.008)	(1.007)
Observations	50,859	50,859
R-squared	0.008	0.008
Number of ids	11,280	11,280

Immediately, we can notice the large drop in observations and ids, and obviously that both PD and its interaction with high SES lost their significance, thus the direct negative effect of PD on wealth is not robust. The interaction term for PD and high SES is also not significant, suggesting that high SES does not significantly change the impact of PD on wealth in this model. Whereas, the interaction term of PD with negative life events is significant (around -3.5%), indicating that these increment the negative effect of PD on wealth. The significance and direction of other variables remain consistent with the baseline model.

From these results, we observe a substantial drop in observations and IDs, which may contribute to the loss of significance for PD and its interaction with high SES. There are a wide number of reasons this could be happening, such as the reduction in sample size decreasing the statistical power, making it harder to detect significant effects. Additionally, the specific subset of households with fixed head might have less variability in PD and wealth missing important factors affecting these, further contributing to the loss of significance for the interaction terms. This could potentially mean that the pd effect on wealth for different SES groups might not be as robust as estimated before, there might be potentially omitted variables and reverse causality not being fully addressed.

Interestingly, the interaction term for PD and negative life events remains significant, which suggests that negative life events play a crucial role in explaining the impact of PD on wealth. This indicates that much of the negative effect of PD on wealth is explained by the occurrence of negative life events, rather than PD alone. This highlights the importance of considering life events when analysing the relationship between psychological distress and wealth.

5 Conclusions

This study investigated the relationship between psychological distress and wealth, with a particular focus on how this relationship varies across different socioeconomic status levels. Using longitudinal data inspired by the PSID over the period of 2003 to 2019, providing an insightful exploration on wealth dynamics among a diverse sample of individuals.

The comparison between OLS and FE models revealed significant differences in the estimated coefficients. OLS models, which do not account for individual-specific effects and time dynamics, tend to overestimate or underestimate relationships due to unobserved heterogeneity. In contrast, FE models account for time-invariant characteristics and provide more accurate within-entity estimates. However, the coefficients from FE models are not true betas as they reflect the relationship within entities over time rather than the overall average effect across all entities. Despite the improvements, FE models still miss time-variant background effects that could influence both PD and wealth, such as anticipated shocks. Using a DiD approach could provide a more robust analysis by capturing these time-variant factors and better isolating the causal impact of PD on wealth.

The investigation into the effect of PD on wealth, accounting for interactions with SES and negative life events, demonstrated that PD has a significant negative impact on wealth. High SES individuals and those experiencing negative life events face a slightly more pronounced impact of PD on wealth. The results showed that much of the negative effect of PD on wealth is explained due to the negative life events when these occur, reminding us of the importance of considering life events when analysing the relationship between PD and wealth. The findings emphasise the complex effects between psychological distress, socioeconomic status, and life events in understanding financial outcomes.

Future analyses could use DiD and focus on the effects of PD on wealth during and around the Global Financial Crisis of 2008, identifying which individuals were more affected and had a higher propensity for increased effects of PD on wealth. This could provide valuable insights into the impact of economic downturns on psychological distress and financial stability.

References

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Balloch, A., Engels, C., & Philip, D. (2022). When It Rains It Drains: Psychological Distress and Household Net Worth. *Journal of Banking and Finance*, Forthcoming. http://dx.doi.org/10.2139/ssrn.3521323

Burbidge, J.B., Magee, L., & Robb, A.L. (1988). Alternative Transformations to Handle Extreme Values of the Dependent Variable. *Journal of the American Statistical Association*, 83(401), 123-127.

Gujarati, D. N., & Porter, D. C. (2010). *Essentials of econometrics* (4th ed.). McGraw-Hill/Irwin.

Tukey, J. W. (1962). The Future of Data Analysis. *Annals of Mathematical Statistics*. 33 (1), 1–67 [p. 17-18]. https://doi.org/10.1214/aoms/1177704711

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48(4), 817-838.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). The MIT Press.

Appendix

Acknowledgment

Given the several issues of using SAS via Massey's virtual machine (sudden crashes, loss of work and impossibility at times to open the software in the VM), I appreciate the opportunity to conduct this assessment in Stata. Due to the steep learning curve, I took the liberty of using a Python notebook in parallel with Stata. All analysis was conducted in Stata (complete code hereafter), however EDA graphs are from the python notebook because of difficulties formatting some of Stata's graphs. The python notebook file can be found in the submission box of this assessment.

STATA Code:

1 Load and Order Data

```
* Import the data
import excel "/Users/lvmacpro/Desktop/724_Applied_Econm/A2/A2_data.xlsx", sheet("Sheet1") firstrow clear

* Renaming variables
* rename financial_distress fd
rename otherrace otherethnicity

* Reorder variables
order id year financial_distress wealth pd education income age male employed divorce marriage childbirth familydeath laidoff missedwork white black hispanic otherethnicity collegedegree studentloan socioeconomic head notmoved nofamichange

* Generate numeric id from string id
egen numeric_id = group(id)

* Sort by id and year
sort numeric_id year
```

2 Descriptive Statistics and EDA

```
* Missing Values
misstable summarize

* Descriptive Statistics:
summarize financial_distress wealth pd education income age male socioeconomic

* OR for all variables simply:
summarize
```

EDA and Correlation Matrix

```
* Histogram of Wealth, PD, and SES
histogram wealth, normal title("Wealth Histogram")
histogram pd, normal title("Psychological Distress (PD) Histogram")
histogram socioeconomic, normal title("Socioeconomic Status Histogram")

* Scatter plot of PD vs. Wealth
twoway (scatter wealth pd) (lfit wealth pd), title("PD vs. Wealth")

* Wealth over Time
graph bar (mean) wealth, over(year) title("Average Wealth over Time")

* PD vs Age / SES
twoway (scatter age pd), title("PD vs Age")
twoway (scatter socioeconomic pd), title("PD vs Socioeconomic Status")

* Correlation Matrix
pwcorr wealth pd education income age male white black hispanic otherethnicity socioeconomic, star(.05)
```

Handle Missing Values and Transformations

```
* Keep relevant variables only (no student loan)
keep id year financial_distress wealth pd education income age male employed divorce marriage childbirth familydeath laidoff missedwork white black hispanic otherethnicity collegedegree socioeconomic head notmoved nofamichange numeric_id

* Exclude rows with missing values in key variables
drop if missing(id, year, financial_distress, wealth, pd, education, income, age, male, employed, divorce, marriage, childbirth, familydeath, laidoff, missedwork, white, black, hispanic, otherethnicity, collegedegree, socioeconomic, head, notmoved, nofamichange, numeric_id)
```

3 Comparing Base Model: OLS and Fixed-effects Models

```
* Applying arcsinh transformation WITHOUT removing outliers
gen ars_wealth = asinh(wealth)
gen ars_ses = asinh(socioeconomic)

* OLS Model
regress ars_wealth pd education income age male black employed ars_ses

* Entity Fixed-effects Model (entity-specific effects)
xtset numeric_id year
xtreg ars_wealth pd education income age male black employed ars_ses, fe
estimates store fe

* Time Fixed-effects Model (time-specific effects only)
xtset year
xtreg ars_wealth pd education income age male black employed ars_ses, fe

* 2-Way Fixed-effects Model (entity and time-specific effects)
xtset numeric_id year
xtreg ars_wealth pd education income age male black employed ars_ses i.year, fe

* Random Effects Model
xtreg ars_wealth pd education income age male black employed ars_ses, re
estimates store re

* Hausman Test to decide between FE and RE
hausman fe re
```

4.1 Hypothesis: PD affects Wealth

[Load data and do steps up to dropping missing values, skip the Stats/EDA and OLS vs FE]

```
* If needed, Drop transformations/interactions used before (didn't exclude outliers)
drop ars_ses ars_wealth ses_group pd_high_ses

* Applying arcsinh transformation WITHOUT removing outliers
gen ars_wealth = asinh(wealth)
gen ars_ses = asinh(socioeconomic)

* Classify SES based on the initial year
by id (year), sort: gen initial_socioeconomic = socioeconomic[1]

* Create SES groups based on initial SES and fix them over the period
xtile initial_ses_quartile = initial_socioeconomic, nq(4)

* Create SES groups based on initial_ses_quartile
gen ses_group = .
replace ses_group = 1 if initial_ses_quartile == 1
replace ses_group = 2 if initial_ses_quartile == 2 | initial_ses_quartile == 3
replace ses_group = 3 if initial_ses_quartile == 4

 * Drop temporary variables
drop initial_socioeconomic initial_ses_quartile
* Creating interactions for PD with SES
gen pd_high_ses = pd * (ses_group == 3)
```

```
gen pd_low_ses = pd * (ses_group == 1)

* Regression (Entity Fixed Effects)
xtset numeric_id year
xtreg ars_wealth pd age male education income ars_ses black hispanic otherethnicity employed divorce marriage childbirth familydeath laidoff missedwork, fe
estimates store fe1

* Regression w/ Interaction (Entity Fixed Effects)
xtreg ars_wealth pd pd_high_ses age male education income ars_ses black hispanic otherethnicity employed divorce marriage childbirth familydeath laidoff missedwork, fe
estimates store fe_interac
```

4.2 Hypothesis: PD affects Wealth w/out Outliers

[Load data and do steps up to dropping missing values, skip the Stats/EDA and OLS vs FE]

```
* If Needed! Drop transformations/interactions if used before (didn't exclude outliers)
drop ars_ses ars_wealth ses_group pd_high_ses

* Wealth by Age (removed wealth outliers)
keep if wealth > -101700 & wealth < 3175000
twoway (scatter wealth age), title("Wealth by Age (Wealth outliers removed)")

* Filter dataset when studying only head, ensure head didn't change nor moved to different household, & trim age and outliers
keep if wealth > -101700 & wealth < 3175000 & age > 18 & age < 68 & socioeconomic > -38538 & socioeconomic < 320530

* Classify SES based on the initial year
by id (year), sort: gen initial_socioeconomic = socioeconomic[1]
 * Create SES groups using xtile
xtile initial_ses_quartile = initial_socioeconomic, nq(4)
 * Create SES groups based on initial_ses_quartile
gen ses_group = .
replace ses_group = 1 if initial_ses_quartile == 1
replace ses_group = 2 if initial_ses_quartile == 2 | initial_ses_quartile == 3
replace ses_group = 3 if initial_ses_quartile == 4
 * Drop temporary variables
drop initial_socioeconomic initial_ses_quartile
 * Creating interactions for pd
gen pd_high_ses = pd * (ses_group == 3)
gen pd_low_ses = pd * (ses_group == 1)
gen neg_life_events = (divorce == 1 | laidoff == 1 | missedwork == 1)
gen pd_neg_life_events = pd * neg_life_events

* Including Age^2
gen age2 = age * age

* Apply arcsinh transformation AFTER removing outliers
gen ars_wealth = asinh(wealth)
gen ars_ses = asinh(socioeconomic)

* Regression (Entity Fixed Effects)
xtset numeric_id year
xtreg ars_wealth pd age education income ars_ses black hispanic otherethnicity employed divorce marriage childbirth familydeath laidoff missedwork, fe

* Regression w/ Interaction SES (Entity FE)
xtreg ars_wealth pd pd_high_ses age age2 education income ars_ses black hispanic otherethnicity employed divorce marriage childbirth familydeath laidoff missedwork, fe

* Regression w/ Interaction SES & Neg Events (Entity FE)
xtreg ars_wealth pd pd_high_ses pd_neg_life_events age age2 education income ars_ses black hispanic otherethnicity employed divorce marriage childbirth familydeath laidoff missedwork, fe

* Predict fitted values and residuals for the interaction model
predict yhat, xb
predict residuals, resid
```

```
* Residual plot, Histogram and Q-Q plot of residuals
scatter residuals yhat, title("Residuals vs Fitted Values")
histogram residuals, normal title("Histogram of Residuals")
qnorm residuals, title("Q-Q Plot of Residuals")

* Clustered Standard Errors for the Entity-FE w/ Interactions
xtreg ars_wealth pd pd_high_ses pd_neg_life_events age age2 education income ars_ses black hispanic otherethnicity employed divorce marriage childbirth familydeath laidoff missedwork, fe vce(cluster numeric_id)
```

4.3 Hypothesis: PD affects Wealth w/out Outliers and w/ Fixed Head

[Load data and do steps up to dropping missing values, skip the Stats/EDA and OLS vs FE]

```
* Filter dataset when studying only head, ensure head didn't change nor moved to different household, & trim age and outliers
keep if head == 1 & nofamichange == 1 & notmoved == 1 & wealth > -101700 & wealth < 3175000 & age > 18 & age < 68 & socioeconomic > -38538 & socioeconomic < 320530

* Classify SES based on the initial year
by id (year), sort: gen initial_socioeconomic = socioeconomic[1]

* Create SES groups using xtile
xtile initial_ses_quartile = initial_socioeconomic, nq(4)

* Create SES groups based on initial_ses_quartile
gen ses_group = .
replace ses_group = 1 if initial_ses_quartile == 1
replace ses_group = 2 if initial_ses_quartile == 2 | initial_ses_quartile == 3
replace ses_group = 3 if initial_ses_quartile == 4

* Drop temporary variables
drop initial_socioeconomic initial_ses_quartile

* Creating interactions for pd
gen pd_high_ses = pd * (ses_group == 3)
gen pd_low_ses = pd * (ses_group == 1)
gen neg_life_events = (divorce == 1 | laidoff == 1 | missedwork == 1)
gen pd_neg_life_events = pd * neg_life_events

* Including Age^2
gen age2 = age * age

* Apply arcsinh transformation AFTER removing outliers
gen ars_wealth = asinh(wealth)
gen ars_ses = asinh(socioeconomic)

* Regression w/ Interaction & Fix head (Entity Fixed Effects w/ clustered errors)
xtset numeric_id year
xtreg ars_wealth pd age education income ars_ses black hispanic otherethnicity employed divorce marriage childbirth familydeath laidoff missedwork, fe vce(cluster numeric_id)

xtreg ars_wealth pd pd_high_ses  pd_neg_life_events age education income ars_ses black hispanic otherethnicity employed divorce marriage childbirth familydeath laidoff missedwork, fe vce(cluster numeric_id)
```

*** FOR SAVING RESULTS/OUTPUTS** (add after model code, append or replace)

```
outreg2 using results.doc, word append ctitle("Entity-FE") dec(3)
asdoc summarize, save(summarize.doc) replace
```