Income, Psychological Well-being, and the Dynamics of Poverty*

M. Alloush†

January 24, 2023

Abstract

Evidence across disciplines suggests a bi-directional relationship between psychological and economic well-being indicating a possible feedback loop that can reinforce poverty. However, estimating these causal links is difficult due to this simultaneity. I use a panel GMM approach and a large-scale dataset from South Africa to estimate a system of dynamic equations where income and psychological well-being are simultaneously determined. I find evidence of heterogeneous effects in both directions highlighting the vulnerability of the poor with low levels of psychological well-being. Simulations suggest this relationship can double the overall impact of shocks and explain prolonged poverty spells.

JEL codes: I32, I12, I15, O12, C33

*I am grateful for the support and guidance of Michael Carter, Dalia Ghanem, Travis Lybbert, and Ashish Shenoy throughout this project. This work has benefited from discussions with many economists and psychologists through one-on-one discussions and in seminars and presentations. I thank two anonymous referees and the editor—Marcel Fafchamps—for valuable feedback helping me push the work forward. Finally, I thank Murray Leibbrandt and the team at the South Africa Labor and Development Research Unit for their helpful comments and for their commitment to keeping the psychological well-being module in the National Income Dynamics Study of South Africa. All errors are my own.

†Hamilton College, Economics Department. 198 College Hill Rd, Clinton, NY 13323. Telephone: +1(202)413-8390. Email: mismaila@hamilton.edu.

1 Introduction

One in five adults around the world suffers from a common psychological disorder every year; psychological disorders account for nearly 13% of the overall global disease burden (Steel et al., 2014; Collins et al., 2011). With a lifetime prevalence of nearly 20%, depression is the most common psychological disorder, estimated to affect 4.4% of adults worldwide at any given point in time (Friedrich, 2017). Psychological well-being is pertinent to economists as an important end in itself but also, in part, because the *lack of it* likely plays a role in shaping economic outcomes such as employment or poverty. However, despite the ubiquity of psychological disorders, empirical evidence on their role in determining economic outcomes and economic decision-making is limited, especially in developing countries.

This dearth of empirical evidence is, at least partially, due to the difficulty of empirically untangling the causal relationship between psychological well-being and economic outcomes. While a change in an individual's psychological well-being *can* influence their earnings, at the same time, their level of economic well-being likely plays a role in determining their state of mental health. In addition to other potential sources of endogeneity, this simultaneity makes it difficult to pin down estimates of causal links using observational data. While experimental and quasi-experimental studies have estimated the effect of both positive and negative income or wealth shocks on psychological well-being, estimating the opposite relationship experimentally is challenging among more representative samples. In this paper, I use a dynamic panel data approach and a large dataset from South Africa to estimate the simultaneous relationship between economic and psychological well-being and explore its implications on the dynamics of poverty.

This bi-directionality between psychological and economic well-being and its potential to push some individuals into a vicious cycle is well established in the psychology literature. The *social drift* hypothesis posits that individuals with psychological disorders are more likely to enter into or remain in poverty due to reduced productivity, loss of earnings, and wasteful spending. At the same time, the *social causation* hypothesis states that conditions of poverty increase the risk of mental illness, and affect psychological well-being through malnutrition, violence, and social exclusion (Lund et al., 2011). Recent work in economics suggests that such a feedback loop could lead to a psychological poverty trap (Ridley et al., 2020; Haushofer, 2019). Given the prevalence of psychological disorders around the world, the question as to whether the relationship between economic and psychological well-being can pose as an impediment for some to achieve their full economic potential should be of interest to economists and policymakers alike.

In this paper, I extend panel data methods popularized by Arellano and Bond (1991) and use a generalized method of moments (GMM) approach to estimate a system of dynamic simultaneous equations.[1] With data from the National Income Dynamics Study of South

[1] Starting with methods by Anderson & Hsiao (1981), Holtz-Eakin, Newey and Rosen (1988), and Arellano

Africa I first answer two main questions: Do changes in depressive symptoms affect an individual's own income? And, does economic well-being (proxied by household income) play a significant role in determining an individual's level of depressive symptoms? The answer I find to both is yes, on average, with significant heterogeneity.

I show, through further analysis, that changes at the low end of the Center for Epidemiologic Studies Depression (CES-D) scale (few depressive symptoms) do not seem to affect income; however, for changes closer to the threshold used by psychologists to screen for depression at the higher end of the scale, I find large effects on individual income. The estimates predict that for an average working-age individual, a one standard deviation (SD) increase in depressive symptoms decreases their income by nearly 16%. The results suggest that one possible avenue through which this occurs is a decreased likelihood of being economically active. Turning to the opposite direction of causality, I find that a 20% decrease in household income per capita increases an individual's CES-D score by 0.4 points (0.13 SD) on average. I also find similar statistically significant estimates when using other measures of economic well-being specifically food expenditure per capita and a household wealth index. By investigating the heterogeneity by baseline poverty status, I find that the effect of changes in economic well-being is larger among the poor.

The relationship between psychological and economic well-being is difficult to untangle with observational data. I use an approach that controls for important individual fixed effects and time-varying individual and household characteristics, but requires assumptions on the dynamic process and correlation of shocks over time. The spacing of the waves and the recall periods of the questions make these assumptions more plausible: nonetheless, they are strong assumptions. I show results of the effect of income on psychological well-being using an alternative estimation method: a local randomization regression discontinuity approach requiring a different set of assumptions yet I find similar point estimates. The uniformity of the estimates across these methods provides support to the core results.

The results indicate that income and depressive symptoms are intertwined; still, it is clear that not every poor person suffers from depression and that not every person suffering from depression experiences significant declines in their income. While the average impacts in either direction are both statistically and economically significant, the estimated dynamics do not, on average, suggest the existence of multiple-equilibrium poverty traps. The results do highlight that the poor with high levels of depressive symptoms are particularly vulnerable and may be disproportionately affected by shocks. Simulations using the estimated system of dynamic equations show that this feedback loop can explain prolonged poverty spells and reduced resilience. The simulations show that the dynamic bi-directional relationship increases vulnerability to long-term poverty up to 30 percentage points among

& Bond (1991), I show that a system of dynamic simultaneous equations can be estimated with at least four rounds of data (proofs and Monte Carlo simulations are available in Appendix D). This approach requires assumptions on the dynamic process and the correlation of shocks over time which I discuss in detail in Section 3.

those with low levels of psychological well-being.

The literature on mental health (and specifically depression) in economics is growing. Depressive disorders are associated with diminished quality of life and higher risk of mortality (Hays et al., 1995; Spijker et al., 2004). Moreover, reduced functioning in occupational and social roles is pervasive among those suffering from depression and this is consistent across contexts and cultures (Ormel et al., 1993). De Quidt and Haushofer (2016) summarize the seminal work of Aaron Beck (1967) on the symptoms and behavioral changes associated with depression and highlight ways in which several aspects of depression could be of interest to economists. It is possible and likely that depression can have a substantial impact on one's economic decision-making, productivity, and consequent outcomes.

Experimental evidence shows that decreasing depressive symptoms through therapy and/or antidepressants among those suffering from depression and seeking treatment significantly improves several economic outcomes including investment in children and employment at the intensive and extensive margins (Bolton et al., 2003; Patel et al., 2017; Angelucci and Bennett, 2021; Lund et al., 2020). Other work shows productivity increases with improved psychological well-being (Berndt et al., 1998; Oswald, Proto and Sgroi, 2015; Shreekumar and Vautrey, 2022). Only a handful of studies look at the effects of mental health on employment and income among representative populations. The existing evidence suggests that psychological distress significantly reduces earnings and the likelihood of employment (Biasi, Dahl and Moser, 2021; Chatterji, Alegria and Takeuchi, 2011; Frijters, Johnston and Shields, 2014; Bryan et al., 2020). However, recent studies show mixed results for programs aimed at improving psychological well-being (Baranov et al., 2019; Haushofer, Mudida and Shapiro, 2020; Angelucci and Bennett, 2021).

Interest in economics has long focused on the reverse causal link, the effect of income (or other measures of economic well-being) on mental health. Several studies use exogenous shocks to income to show that income does affect mental health. Early work by Gardner and Oswald (2007) compares British lottery winners to a control group of other lottery players and find that those who win the lottery show higher levels of psychological well-being. A recent study by Christian, Hensel and Roth (2019) shows that income shocks lower suicide rates in Indonesia with suggestive evidence that the mechanisms is through a reduction in depression rates. Experimental evidence shows that, among poor households in Kenya, increased economic well-being reduced stress and decreased depressive symptoms measured by a reduction in the CES-D 20 scale (Haushofer and Shapiro, 2016).[2]

[2] A non-exhaustive list of additional studies that includes experimental and quasi-experimental variation in income is McGuire, Kaiser and Bach-Mortensen (2022); Frijters, Haisken-DeNew and Shields (2004, 2005); Macours, Schady and Vakis (2012); Baird, De Hoop and Özler (2013); McInerney, Mellor and Nicholas (2013). Moreover, studies differentiate between psychological well-being and life satisfaction/happiness which is the subject of many different studies over the years (see for example Stevenson and Wolfers (2013); Graham and Pettinato (2002); Kahneman and Deaton (2010)). The CES-D scale used in this paper measures depressive symptoms as a proxy for psychological well-being which is associated with but different from life satisfaction (Headey, Kelley and Wearing, 1993; Das et al., 2009; Lindqvist, Östling and Cesarini, 2020).

More broadly, this paper contributes to a growing eld aimed at understanding the multitude of stresses faced in poverty. The psychological consequences of poverty are gaining increased attention among economists investigating the mechanisms through which poverty can a ect economic productivity and decision-making (Schilbach, Scho eld and Mullainathan, 2016). One avenue through which these e ects may occur is lower levels of mental health (Haushofer and Fehr, 2014; Ridley et al., 2020).

Overall, the evidence in the literature suggest that economic and psychological well-being are intertwined. I add to the di erent strands of the literature with three main contributions: 1) I estimate both e ects in a dynamic and simultaneous system of equations with a large, more broadly representative sample in a middle income country; this allows me to to use simulations to better understand the e ect of the feedback loop in the long-run and estimate the overall e ect of shocks; 2) I show signi cant heterogeneity in both directions whereby changes on the higher end of the scale have a bigger e ect on an individual's income and, on the other hand, the poor are more a ected by changes in income when it comes to their depressive symptoms; 3) I extend panel data methods and show that the approach leads to similar point estimates to that estimated using a regression discontinuity approach.

While I mainly stress the potential negative consequences of the relationship between poverty and psychological well-being, there is a positive story to tell. Poverty-alleviation programs may have an added bene t of positive impacts on psychological well-being an important goal in itself which may also enhance an individual's capability to further improve their economic well-being. In this sense, psychological well-being is both a constitutive freedom and an instrumental one (Sen, 1999).

The rest of this paper is structured as follows. In Section 2, I introduce the data, discuss the measure of psychological well-being, and highlight relevant descriptive statistics. Section 3 outlines the main empirical strategy and the key assumptions required for consistency of the econometric approach, and Section 4 presents the results. Section 5 shows some implications for poverty dynamics using simulations. Finally, Section 6 concludes.

2 Data, Measurement, and Descriptive Statistics

In this section I introduce the data used in this analysis, the Center for Epidimiologic Studies Depression Scale which is the main measure of depressive symptoms I use in this paper, and discuss some key motivating descriptive statistics.

2.1 Data

The panel data used in this analysis comes from the National Income Dynamics Study (NIDS) of South Africa.[3] The rst survey wave was conducted in 2008 and households

[3]This is a panel study conducted by the South Africa Labor and Development Research Unit at the University of Cape Town. An analysis of mental health and socioeconomic status using the rst round of

were interviewed every two years until 2014 (Wave 1-4) and once again in 2017 (Wave 5). The study began with a nationally representative sample of nearly 27,000 individuals (16,758 completing the adult individual-level questionnaire) in 6,598 households. Rich data was collected on socio-economic at the household and individual levels. Most important to this analysis, NIDS contains a psychological well-being module comprising of the 10-item Center for the Epidemiological Studies Depression Scale (CES-D) for adults (at least 16 years old) in all waves. This is unprecedented in a nationally representative panel survey in a developing country.

For the main analysis in this paper, I use data from the rst four relatively equally-spaced waves of the NIDS. I use the fth wave to conduct robustness checks. In this study, I use a balanced sample which includes only those who were at least 16 years of age in the rst wave and responded to the all relevant questions in the individual-level questionnaire in all four waves. I trim the top and bottom 0.25% of individuals based their household income per capita to remove outliers.[4] The resulting sample size is 6,281 individuals. Table 1 presents Wave 4 descriptive statistics of the this sample. This study sample is poorer on average than the full representative samples from each wave (See Table A2 in the Appendix). However, those that complete the CES-D in all four waves are similar to the balanced sample. The median income and expenditure variables are more similar across these samples suggesting that wealthier households and individuals are more likely to attrit from the panel. The results in the paper therefore are not necessarily representative of wealthier South Africans.

2.2 Measurement of Depressive Symptoms

In this analysis, I use the 10-item Center for Epidemiologic Studies Depression scale to measure depressive symptoms and proxy psychological well-being. The CES-D scale was developed to assess depressive symptoms and screen for depression in the general population (Radlo , 1977). It is a widely-used measure of depressive symptoms (Santor, Gregus and Welch, 2006) and is comprised of questions that ask individuals how often in the last week they felt certain emotions related to depression (See Table A1 in the Appendix for the list of questions and more information on CES-D). The scores for all questions are summed for an overall score between 0 and 30. A higher overall CES-D score indicates more depressive symptoms. The distribution of CES-D scores across all four waves in the study sample is shown in Figure 1.

The CES-D is used to screen for depression and scores above certain thresholds indicate that an individual is increasingly likely to be su ering from what clinical evaluation would diagnose as depression. In the 10-item CES-D, a threshold score of 10 is most commonly used for depression screening; however, Baron, Davies and Lund (2017) suggest that a

data of this study can be found in Ardington and Case (2010).

[4]The results in the paper are robust to the trimming of the top and bottom 0.5,1, and 5%, however, the results are not statistically signi cant without any trimming (see Table A3 in the Appendix).

Table 1: Study Sample Characteristics

VARIABLES - Wave 4	Mean	(SD)
Household Income Per Capita (ZAR)	1,934	(1,963)
Household Food Expenditure Per Capita (ZAR)	363	(365)
Individual Income (ZAR)	2,267	(2,912)
CES-D score	7.22	(4.22)
Household Size	5.05	(3.33)
Female	0.63	(0.48)
Age	42.85	(15.88)
Economically Active	0.62	(0.49)
Disabled/Chronically Ill	0.08	(0.28)
Observations	6,281	

Notes: This table provides Wave 4 descriptive statistics for the study sample used in this paper. Table A2 in the Appendix shows characteristics of the representative samples from all four waves versus the balanced sample. The sample in this study include individuals who completed the individual section of the survey including the CES-D section for the rst four rounds of NIDS. Table A2 shows that this sample is on average poorer, however, in most other characteristics, the study sample is similar.

threshold score of 11 is appropriate for screening for depression among most populations in South Africa. The CES-D 10 is commonly used in South Africa and is shown to be internally consistent and veri ed as an e ective screening tool for depression (Hamad et al., 2008; Johnes and Johnes, 2004; Myer et al., 2008; Baron, Davies and Lund, 2017). An important characteristic of CES-D is that its questions do not explicitly mention psychiatric illnesses. This helps mitigate the e ect of stigma on the quality of data as mental illnesses are highly stigmatized in South African communities (Hugo et al., 2003).[5]

In the study sample, the mean CES-D score for all four waves is 7.73 (4.49) and shows a decreasing trend where the average score is 8.39 (4.64) in 2008 and 7.22 in 2014. Within person standard deviation in the CES-D score is 3.6. The incidence of scores above 11 show a similar pattern and decrease from about 20.67% in 2008 to 15.06% in 2014. Nearly 52% of the panel sample record a CES-D score of 11 or above at least once in all four waves. Figure B1 in the Appendix shows the predictive value of lagged CES-D scores on current CES-D scores and the likelihood of having a CES-D score greater than or equal to the threshold of 11 where individuals are at high risk of depression. This illustrates within-person correlation over time.

[5] This is evident in the data as the rate of response about speci c mental illnesses is very low. However, the response rate on the CES-D module is high; on average, 94.6% of individuals who completed the individual-level questionnaire completed the CES-D questionnaire.

Figure 1: Distribution of CES-D scores: Histogram of the CES-D scores shows that a signi cant portion of the population have scores above the threshold of 10 used by psychologists to screen for depression. Figure B1 illustrates the correlation of CES-D scores over time.

2.3 Descriptive Statistics

South Africa is a middle-income country with one of the highest levels of income inequality in the world. The mean monthly household income per capita (standard deviation in brackets) in the study sample in 2014 was 1,934 ZAR (1,963)[6]. This hides signi cant inequality as recent reports estimate that nearly 54% of the population is living in poverty and about 20% live in extreme poverty (Leibbrandt, Finn and Woolard, 2012). In the study sample, nearly 84% of individuals report food expenditure levels that are considered poor in at least one of the four waves. 45% are poor in at least three out of the four waves and 21% are poor in all four waves of the panel.

Figure 2 shows the share of individuals with scores above the thresholds of 10, 11, and 12 by wealth decile. High scores like these indicate more depressive symptoms and a higher likelihood of depression. The share of individuals with scores above the threshold decreases with wealth whereby the share among the highest wealth decile is nearly half that of the lowest. Figure B2 in the Appendix graphs a histogram of CES-D scores by poverty status in all four waves. These gures illustrate a clear correlation between psychological and economic well-being that we observe in our data and in other contexts around the world (Lund et al., 2011; Ridley et al., 2020). The next section outlines the empirical strategy to estimate the causal relationships between the two.

[6] This corresponds to 190 US Dollars or approximately $340 PPP adjusted. The GDP per capita in South Africa in 2014 was $6,434 corresponding to a monthly income per capita of $536. The distribution of income is skewed and the trimming of the top and bottom extremes in income for my study sample brings down the mean re ecting the high levels of inequality in the country. Income and expenditure numbers are adjusted for in ation and are in November 2014 prices.

Figure 2: Economic Well-being and CES-D: The share of individuals with CES-D scores above 10, 11, and 12 by wealth decile. High scores indicate an increasing likelihood of clinical depression and it is clear that as wealth increases, the share of individuals reporting scores higher than these thresholds is decreasing. Figure B2 in the Appendix shows overlapping histograms by poverty status. Among the poor, the distribution is shifted to the right where they are more likely to have scores above the depression threshold of 11. A Kolmogorov-Smirnov test for the di erence in the two distributions shows that they are statistically di erent from each other ($p = 0.000$).

3 Econometric Approach

In this analysis, I exploit the panel nature of the data to estimate the relationship between economic and psychological well-being as a system of two simultaneous dynamic equations that captures both causal links at the same time. I provide details on this econometric approach in Appendix D; this includes proofs and simulations that show the consistency of the GMM estimators under di erent assumptions. In Section 3.1, I brie y describe the estimation approach and the assumptions on the dynamic process it requires. In Section 3.2 I discuss the main threats to identi cation and outline how I show robustness of the main results.

3.1 Core Econometric Approach

First, I represent the relationship between income and psychological well-being with the following system of linear, dynamic, and simultaneous equations:

$$y_{i,t} = {}_1 d_{i,t} + {}_1 y_{i,t \; 1} + G_1 x_{i,t} + {}_i + e_{i,t}$$

$$d_{i,t} = {}_2 h_{i,t} + {}_2 d_{i,t \; 1} + G_2 x_{i,t} + {}_i + u_{i,t}$$

where $y_{i,t}$ is individual income and $d_{i,t}$ is a measure of psychological well-being for individual i in time t, and $h_{i,t}$ is a measure of economic well-being. In this paper, I will consider mainly household income per capita. Using household income per capita instead of individual income is not an identifying assumption. Individual income is part of household income and the simultaneity remains. The intuition behind this choice is that psychological well-being likely affects individual income directly and while this will then affect household income, there can be potentially some compensatory behavior by other household members that would also affect the household income. On the other hand, household income is likely a better proxy for the level of economic well-being one experiences.[7] $_i$ and $_i$ are individual fixed effects; and $e_{i,t}$ and $u_{i,t}$ are the unobserved error terms for their respective equations. $x_{i,t}$ is a vector of time varying individual characteristics for individual i at time t. While the focus is not on the dynamics of income and psychological well-being, I allow for state dependence in the underlying process by having lagged income ($y_{i,t-1}$) and psychological well-being ($d_{i,t-1}$) as explanatory variables in their respective equations.

The individual fixed effects v_i and $_i$ are likely important determinants of both income and psychological well-being. I control for these individual fixed effects by first-differencing both equations to get the following:

$$Dy_{i,t} = {}_1Dd_{i,t} + {}_1Dy_{i,t-1} + G_1Dx_{i,t} + De_{i,t} \qquad (1)$$

$$Dd_{i,t} = {}_2Dh_{i,t} + {}_2Dd_{i,t-1} + G_2Dx_{i,t} + Du_{i,t} \qquad (2)$$

In this system of equations, I am interested in estimating the coefficients of four endogenous variables, namely $_1$, $_1$, $_2$, and $_2$.[8] If considering each single equation separately and abstracting away from bi-directionality, dynamic panel data methods suggests that, assuming sequential exogeneity and that the error terms $e_{i,t}$ and $u_{i,t}$ are serially uncorrelated, the lagged levels $y_{i,t-2}, y_{i,t-3}, \ldots$ and $d_{i,t-2}, d_{i,t-3}, \ldots$ may be used as instruments to consistently estimate the parameters of the equation (1); the same set of instruments may be used to estimate equation (2) as well (Anderson and Hsiao, 1982; Arellano and Bond, 1991; Holtz-Eakin, Newey and Rosen, 1988). However, the bi-directionality inherently implies that assumption of sequential exogoneity is violated.

If the number of observations per individual is at least four ($T \geq 4$), I can relax the assumption of sequential exogeneity and extend this dynamic panel approach to estimate a system of simultaneous dynamic equations. I discuss this in greater detail in the Appendix

[7] Using the same income variable (either individual income or household income) gives results that are weaker than the ones I present below. I show these results in Appendix tables A6 and A7. These results are less statistically significant and more noisy however broadly follow the same pattern as the main results in the paper.

[8] Two of these variables are assumed to be endogenous ($Dd_{i,t}$ in equation (1) and $Dy_{i,t-1}$ in equation (2)) and the dependent variable lags ($Dy_{i,t-1}$ and $Dd_{i,t-1}$) are endogenous due to the first differencing. Right hand side variable $Dy_{i,t}$ and its regressor $Dy_{i,t-1}$ share a common variable $y_{i,t-1}$ making the regressor correlated with the error term $De_{i,t}$. Similarly for $Dd_{i,t}$ and its regressor $Dd_{i,t-1}$ and the error term $Du_{i,t}$.

Figure 3: Dynamic process allowed by Assumption A rules out direct effects from $y_{i,t-1}$ ($d_{i,t-1}$) on $d_{i,t}$ ($y_{i,t}$).

D. With four rounds of data in the panel, I can estimate the system of equations under two slightly different assumptions. A visual representation of the dynamic and simultaneous relationship allowed under the first assumption (Assumption A) can be found in Figure 3. Assumption A implies six moments that identify the four coefficients of the system of equations (see Appendix D.2.1). A two-step GMM estimator is consistent for estimating the coefficients $_1$, $_1$, $_2$, and $_2$.

For Assumption A to hold, the $e_{i,t}$ may not be correlated with $y_{i,t-2}$ and $d_{i,t-1}$ and further lags of these variables, and $u_{i,t}$ may not be correlated with $y_{i,t-1}$ and $d_{i,t-2}$ and further lags of these variables. This is a weaker condition than sequential exogeneity, but estimation requires a larger minimum T. The simultaneity of the equations also implies that both error terms cannot be serially correlated. After controlling for state dependence (through the lagged dependent variable), individual fixed effects, and observable time varying characteristics, the remaining unobserved errors may not be correlated across t. Effectively, this assumption means that a shock to income in one period can affect income next period through state dependence, but it cannot affect the likelihood of shocks in the next period. Similarly for shocks to psychological well-being.

The plausibility of this assumption is tied to the time between observations for each individual and the time-frame of the variable. It would be difficult to impose such an assumption if the panel was short and consisted of monthly observations. For example, if an individual loses a job, it is very likely that this shock to income is still manifesting itself next month. However, assuming a constant T, more time between observations makes a lack of serial correlation more plausible. In this study, income is reported for the last month. Yearly observations of income in the past month can be viewed as observations of monthly income 12 time periods apart. In the NIDS data, income in the last month is reported approximately two years apart. Thus, in this context, shocks to monthly income can be correlated for up to 23 months but not more under Assumption A. Similarly for depressive symptoms which are reported for the last week.

Throughout the rest of this paper, the main results are based on Assumption A. As a robustness check, I also present results that I estimate with moment conditions implied by a less restrictive assumption (Assumption B, Appendix D.2.2). This assumption allows for a more flexible dynamic simultaneous process illustrated visually in Figure 4. The implied

Figure 4: Dynamic process allowed by Assumption B. Assumptions A and B imply slightly different dynamic processes. Under the less restrictive assumption B, the assumptions do not allow for direct effects across 2 time periods. Income and psychological well-being in time $t-2$ can only affect income and psychological well-being in time t through income and psychological well-being in time $t-1$.

moment condition identifies the coefficients of the system of equations and is also implied under Assumption A. With fewer lagged levels being used as instruments, the system is now just identified.

Under Assumption B, unlike under Assumption A, the error terms may be first-order moving-average serially correlated. Assuming first-order serial correlation in shocks is common in the literature on income dynamics and state dependence of income and employment that uses yearly income data (Guvenen, 2007; Meghir and Pistaferri, 2004; Magnac, 2000). The time between each wave in the NIDS dataset is two years, and income is reported for the past month; this allows for moving-average serial correlation in the unobserved error terms that lasts no more than 47 months. Moreover, under Assumption B, the error terms $e_{i,t}$ and $u_{i,t}$ may be correlated with $u_{i,t-1}$ and $e_{i,t-1}$, respectively, allowing for common shocks that could persist up to one period.

Throughout Section 4, I show, where appropriate, estimates based on both assumptions A and B. While Assumption B is weaker, the estimator under Assumption A is more efficient. I present both sets of results with the trade-off between consistency and efficiency in mind. The estimates do not differ significantly throughout indicating that the main results are robust to first-order moving-average serial correlation.[9]

The bi-directional relationship between health (including mental health) and employment has been modeled and estimated previously in the literature (Hamilton, Merrigan and Dufresne, 1997; Haan and Myck, 2009; Bubonya, Cobb-Clark and Ribar, 2019; Steele, French and Bartley, 2013). The structural endogeneity created by the bi-directional relationship is addressed in two main ways: using instrumental variables for each of employment which comes with validity concerns. Other approaches attempt to control for reverse causality by modelling lagged employment and health outcomes when determining health and employment, respectively and using correlated random-effects to address other sources of endogeneity (Bubonya, Cobb-Clark and Ribar, 2019; Steele, French and Bartley, 2013). In this paper, I model the simultaneity directly within the same time period while taking

[9] Appendix section D.2.2 provides a discussion on how I test for serial correlation in my data.

into account state-dependence and individual fixed effects and require a different set of assumptions than those in previous studies.

3.2 Potential Identification Threats and Robustness Checks

Panel data methods are not very common in empirical microeconomics analyses such as this one. This approach comes with strong assumptions regarding the serial correlation of shocks within and across equations. The two-year spacing of the waves and the recall time of the main variables make the assumptions discussed above more plausible. Below, I briefly discuss threats to identification in the context of this study and outline an alternative estimation approach I use as a robustness check. I provide a more thorough discussion in Appendix D.3.

Common shocks that increase the likelihood of future shocks for more than four years would violate the weaker of the two Assumptions. For example, if a shock to income in Waves 1 or 2 makes divorce more likely in later waves, divorce could lead to income and psychological well-being shocks in the future (Charles and Stephens, 2004). If the income shock increases the likelihood of divorce for more than four years post shock, then past income can affect current income and psychological well-being in a way that is not captured by the model thus violating Assumption B the weaker of the two assumptions.

Moreover, it may be the case that the dynamics are mis-specified: it could well be that another $t-2$ lagged term is important in directly determining income and psychological well-being in time t. This would be a violation of the weaker of the two assumptions. The panel, however, is not long enough to test this assumption directly.

While we cannot entirely rule out long persistence of shocks and mis-specification, I further test the robustness of the one side of results by using a local randomization regression discontinuity approach to estimate the effect of income on psychological well-being (see Appendix C). I leverage the discontinuity of eligibility for the Older Person's Grant and restrict the sample to individuals in households with economically inactive members in narrow windows around the age of 60 and using the age-eligibility for the Older Person's Grant as an instrument for household income.[10] The required assumptions for consistent estimators here are different, yet, I find that the estimated effect of changes in household income are very similar to those estimated using the GMM approach. I am unable to conduct a similar exercise for the effect of psychological distress on income.[11] However, the uniformity of the estimates on income gives credence to the core results.

[10] This is the same empirical method used in Alloush and Wu (2023) to estimate the effect of income on life satisfaction and is discussed in detail there. More discussion on local randomization approaches in regression discontinuity designs can be found in Cattaneo, Idrobo and Titiunik (forthcoming).

[11] As is evident from the literature, it is difficult to find a valid instrument for psychological well-being; I discuss this in more detail in Appendix C.

4 Results

In this section, I first show results for both equations. Second, I focus on the working age population when specifically looking at the effect of psychological well-being on individual income, its potential heterogeneity, and the potential mechanisms. Third, I show results linking alternative measures of economic well-being to psychological well-being and show heterogeneity by initial wealth.

4.1 Simultaneous Equations

In Section 3, I consider a simple linear version of the system of equations to illustrate the estimation strategy. The results presented in the rest of the paper are mainly estimates of the following system of equations:

$$Dy_{i,t} = {}_1Dd_{i,t} + {}_1Dy_{i,t-1} + GDx_{i,t} + De_{i,t} \tag{3}$$

$$Dd_{i,t} = a_1Dh_{i,t} + a_2Dh^2_{i,t} + b_1Dd_{i,t-1} + QDx_{i,t} + Du_{i,t} \tag{4}$$

With this system of equations I add a quadratic term of household income per capita $h_{i,t}$.[12] Table 2 also shows results without this quadratic term in column 1.

The two-step GMM results for the study sample are shown in Table 2. The results in columns 1-3 of the upper half of the table show that changes in CES-D have a significant effect on individual income on average. In column 1, I show the basic dependent and in column 2 I add the quadratic term on household income per capita as shown in equation (4) above. Lastly, I add controls that include household size, marital status, disability/chronic illness, and number of children in the household in column 3. The point estimates do not change significantly. These controls may well be endogenous, however, a similar specification that treats all these variables as endogenous does not change the main results that suggest strong causal links in both directions.

Results in column 4 show the estimates under the less restrictive Assumptions B that allow for first-order serial correlation in shocks. The results show similar patterns and suggest that the results are robust to less restrictive assumptions. A Hausman-type test shows that the differences in the estimates in columns 3 and 4 are not statistically significant.[13] In addition, testing for overidentifying restrictions provides Hansen J-test statistics that do not reject the validity of the instruments (p value = 0.315).[14]

[12] Intuitively, changes in income may affect psychological well-being at a decreasing rate. The assumptions required for validity do not change. I add quadratic terms of the lagged levels of household income the instrumental variable used to estimate equation (4). I explore the non-linearity in the effect of CES-D on individual income in Section 4.2.

[13] This Hausman-type test is done under the premise that the estimator using Assumptions B is consistent and that under assumptions A it is efficient; the result p value = 0.52.

[14] A standard test for weak instruments in dynamic panel GMM does not currently exist and diagnostics

Table 2: System of Simultaneous Equations: GMM Estimates

	$Z^A_{i,t}$			$Z^B_{i,t}$
	(1)	(2)	(3)	(4)
Dependent Variable: Individual Income				
CES-D_t	-2.260	-2.048	-2.067	-4.178
	(1.232)	(1.016)	(0.730)	(1.554)
Individual $Income_{t-1}$	1.241	0.916	0.269	0.636
	(0.609)	(0.501)	(0.344)	(0.280)
Dependent Variable: CES-D				
HH Income Per $Capita_t$	-0.146	-0.338	-0.323	-0.342
	(0.051)	(0.083)	(0.152)	(0.102)
HH Income Per $Capita_t^2$		0.004	0.004	0.003
		(0.001)	(0.001)	(0.002)
CES-D_{t-1}	0.050	0.051	0.036	0.090
	(0.030)	(0.030)	(0.029)	(0.032)
Controls			Yes	Yes
Observations	6,281	6,281	6,281	6,281

Notes: Cluster robust standard errors in parentheses. All standard errors shown in the tables are clustered at the PSU level. PSUs are de ned geographic areas based on the 2001 census in South Africa based on which the sampling for NIDS took place. Income numbers are in 100 South African Rands. Controls include household size, marital status, disability, and number of children in the household. Two-stage GMM for the study sample using two approaches with di erent instrument matrices that are consistent under two slightly di erent assumptions (A and B) show similar statistically signi cant results for the simultaneous e ects. The results that include a individual/household income in both equations are shown in Appendix tables A6 and A7.

The results show statistically signi cant results in both directions. The estimates in Table 2 suggest that a 2 point increase in the CES-D score (0.45 SD) decreases individual income by ZAR 470. This causal link is explored further in Section 4.2. The results also show signi cant e ects of income on depressive symptoms whereby a ZAR 200 increase in household income per capita decreases CES-D scores (decreases depressive symptoms) by about 0.9 points on average (0.25 SD). The quadratic term is statistically signi cant suggesting that increases in income decrease depressive symptoms at a decreasing rate.

from linear IV regressions do not carry over to this more general dynamic panel GMM setting (Stock and Wright, 2000; Bazzi and Clemens, 2013). However, I show results in the appendix (Table A4 and A5) where I apply a systems GMM approach to each side of the equation and get similar estimates. This approach is shown to be more robust to conditions which may imply that lagged levels are weak instruments, however it requires additional assumptions on the stationarity of the dynamic process (Blundell and Bond, 1998); I use this approach to conduct direct tests on serial correlation using Wave 5 data and I nd no evidence of rst-order moving-average serial correlation in either equation.

This side of the simultaneous equations, the effect of income on psychological well-being, is analyzed in more detail in Section 4.3.

4.2 The Impact of CES-D on Individual Income

To further study the effect of depressive symptoms on individual income, I focus specifically on equation (3) from the system of equations above:

$$\Delta y_{i,t} = \beta_1 \Delta d_{i,t} + \alpha_1 \Delta y_{i,t-1} + \Gamma \Delta x_{i,t} + \Delta e_{i,t} \tag{3}$$

I first show the heterogeneity that might exists based on the baseline level of CES-D after which I investigate potential mechanisms through which changes in psychological well-being affect income.

4.2.1 Non-linear Effects: Clinical Depression Threshold

The psychology literature on the CES-D scale indicates that high scores suggests that a person is increasingly likely to be suffering from what would be clinically diagnosed as depression. 10, 11, or 12 are common thresholds used to screen for depression (Baron, Davies and Lund, 2017). Given that clinical depression is increasingly likely at higher CES-D levels, it may be that that the effects of changes in CES-D scores are non-linear.

To capture this non-linearity, I would like to estimate the marginal effect of changes in CES-D at each baseline CES-D score. To do so, an ideal dataset would have a very large number of observations at each baseline (Wave 3) CES-D score and all individuals would experience a change of 1 or -1 in their CES-D score between Waves 3 and 4. Applying the same econometric specification as above using sub-samples of individuals who report the given baseline CES-D score would estimate the marginal effect of changes in CES-D on individual income at each CES-D score. I apply this approach to estimating the marginal effects to this dataset, however, in order to attain adequate sample sizes, I increase the bandwidth to 1 in local linear regression terms, and I restrict the sample to individuals who experience changes less than or equal to the absolute value of 4 in their CES-D score between Wave 3 and 4 (instead of 1).[15] I view this approach as a type of non-parametric estimation that is achieved by focusing on small changes in short intervals of a discrete variable.

The results of this estimation method are shown in Figure 5. While psychologists offer a clear hypothesis that changes in depressive symptoms may matter more when clinical depression is more likely, since these marginal effects were estimated using 13 regressions, I

[15] In this discrete variable case, when estimating the marginal effect at CES-D = 5 in 3, I would include individuals who report a CES-D score of 4,5, or 6. Observations where baseline $CES-D_3 = j \pm 1$ were weighted at 1/2 that of j. The results are robust to a range of different weighting specifications. Results in which the sample is restricted to individuals who experience smaller and larger changes (2,3 and 5) exhibit a similar pattern, but have slightly different point estimates.

Figure 5: Marginal e ects of CES-D on individual income: The impact of CES-D on individual income based on baseline CES-D score in Wave 3. This gure was based on 13 linear regressions: For each baselines CES-D score, the sample was restricted to those with CES-D scores 1 and who experienced changes with absolute value 4. Dashed con dence intervals are Bonferroni-Holm corrected CIs to control the family-wise error rate. The vertical line at 10 indicates the threshold commonly used by psychologists to screen for depression. This gure was estimated with instruments matrix $Z^{A}_{i,t}$. Income is in 100 South African Rands.

present a conservative Bonferroni-Holm corrected con dence interval (dashed grey line) to control for the family-wise error rate (Holm, 1979).[16] The estimates suggest that when an individual is at the threshold of 10, a 1-point increase in their CES-D score decreases income by nearly ZAR 500. This estimate is signi cant at the 1% level even after a Bonferroni-Holm adjustment to the p-value. The model estimates slightly smaller marginal e ects at CES-D scores 11, 12, and 13 that are statistically signi cant at the 5% level. Moreover, the largest estimate is at the CES-D score of 8. If I consider an individual with a median CES-D score of 6 in Wave 3, a 1 SD increase in their CES-D (approximately 4 points) is estimated to decrease their individual income by nearly ZAR 1,200 or about 0.3 SD on average. The average income of an employed individual with CES-D equal to 6 is ZAR 4,250: the estimates predict that a 4-point increase in CES-D score would decrease the individual's income by over 20%.

While overall changes in CES-D do seem to a ect an individual's income on average in a statistically signi cant way, the results presented in this section suggest that there is

[16] This step-down Bonferroni procedure (also known as Holm-Ryan) controls for family-wise error rate by sequentially adjusting the standard errors after ranking by the p-values from the largest to the smallest. Thus, coe cients with the smallest p-values get the largest corrections to their standard errors. This correction makes the probability of Type I error for each test / rank where = 0.05.

Table 3: Mechanisms Labor Supply

	Economically Active (1)	Employed [y] (2)	Hours Worked [z] (3)
CES-D$_t$	-.052	-.038	-0.263
	(0.028)	(0.037)	(1.93)
Controls	Yes	Yes	Yes
Observations	4,891	2,451	1,455

Notes: Cluster robust standard errors in parentheses. Samples restricted to working age adults. [y]Conditional on being economically active in Wave 3 and 4. [z]Conditional on being employed in Wave 3 and 4. Controls include lagged dependent, household size, number of children per household, disability, and marital status. Moment conditions requiring adapted assumptions similar to assumption A are used to estimate these results.

signi cant heterogeneity/non-linearity. Depression is increasingly likely among individuals with higher CES-D scores. The results show that for those who are in the upper half of the CES-D distribution, changes in depressive symptoms have large impacts on their income. This paper adds to literature by clearly showing these non-linearities.

4.2.2 Mechanisms and Other E ects

In Table 3, I present results that investigate some of the possible mechanisms through which changes in CES-D might a ect individual income speci cally labor supply at the extensive and intensive margins. All the results in Table 3 are estimated using a systems GMM speci cation for the variable of interest, $m_{i,t}$, that is similar to the system speci cation above where I use an equivalent vector of instruments. The equation I focus on is the following:

$$Dm_{i,t} = \ _1Dd_{i,t} + \ _2Dm_{i,t\ 1} + \ _3Dx_{i,t} + D\ _{i,t}$$

The estimates in column 1 of Table 3 suggest that one of the likely mechanisms through which an increase in CES-D decreases income is through decreased labor force participation. The results predict that a 1-point increase in an individual's CES-D score results in a 5.2 percentage-point decrease in the likelihood of labor force participation. The point estimate for employment (given in the labor force) and hours worked (given employment) in columns 2 and 3 show no e ects of increases in CES-D.

4.3 The Impact of Income on Psychological Well-Being

The results in Section 4.1 show that changes in income a ect psychological well-being for the average individual in the sample. To further explore this e ect, I focus on equation (4)

Table 4: Impact of Other Measures of Economic Well-being

	$Z^A_{i,t}$		$Z^B_{i,t}$	
Dependent Variable: CES-D	(1)	(2)	(3)	(4)
Food Exp Per Capita$_t$	-2.61		-2.39	
	(0.79)		(0.81)	
Food Exp Per Capita^2_t	0.086		0.074	
	(0.031)		(0.030)	
Wealth Index$_t$		-6.12		-6.44
		(2.58)		(3.55)
Controls	Yes	Yes	Yes	Yes
Observations	6,281	6,281	6,281	6,281

Notes: Cluster robust standard errors in parentheses. Income and expenditure numbers are in 100 South African Rands. Controls include household size, marital status, disability, and number of children in the household.

in the system of equations above:

$$Dd_{i,t} = a_1 Dh_{i,t} + a_1 Dh^2_{i,t} + b_1 Dd_{i,t\ \ 1} + QDx_{i,t} + Du_{i,t} \tag{4}$$

When estimating the impact of changes in food expenditure per capita, and wealth, I use the system estimation strategy used in Section 4.1.

Table 4 shows the coe cient estimates for di erent variations of equation (4) using two-step GMM estimation. Columns 1-2 present results for the approach that uses the vector of instruments consistent under assumption A. The results show that for two di erent measures of economic well-being food expenditure per capita and a household wealth index a change in economic well-being a ects the CES-D score in a statistically signi cant way. The estimates in Table 2 suggests a decreasing marginal e ect of household income per capita due.

To test the robustness of these results, I replace household income with other measures of economic well-being, namely food expenditure per capita and wealth. The results are similar in sign and statistical signi cance for both alternative measures of economics well-being. The estimates in Table 4 predict that a ZAR 50 (mean food expenditure per capita is nearly ZAR 330) decrease in food expenditure increases CES-D score by over 1 point. Also a 0.1-SD increase in wealth measured by the wealth index is predicted to decrease CES-D scores by near 0.6 points.

Columns 3-4 show estimates for the econometric speci cation that requires less restrictive assumptions. The results are similar to those in columns 1-2 again showing robustness.

The magnitude of the estimated e ects of changes in household income on psychological well-being are in line with other experimentally estimated impacts. In Haushofer and Shapiro (2016), the unconditional cash transfer averaging out at nearly PPP $45 per capita per year targeting the poor led to an additional increase in revenue of nearly $16 on average. The treated households showed an overall decrease in nearly 1.2 in their CES-D20 score. Back-of-the-envelope calculation and an equivalent PPP adjustment shows that a similar increase in household income among the poorest 20% in South Africa would lead to nearly 0.64 reduction in CES-D. Noting that Haushofer and Shapiro use CES-D20 in their analysis and abstracting away from the complexity of predicting CES-D 20 scores with CES-D 10, the estimates in this analysis on the impact of household income per capita on CES-D are similar in size.

4.3.1 Heterogeneity by Poverty Status

Intuitively, the impact of changes in income on psychological well-being may be larger for the poor. While a 100 Rand change in household income per capita is large for individuals in poor households, this is a small change for those in wealthier households. I use the natural log of household income per capita to facilitate a percent change interpretation in this section. In Appendix Table A8, the estimated coe cient on log of income per capita for the whole sample (Column 1) suggests that a 20% increase in household income leads to a 0.4 point decline in CES-D scores on average, however, this coe cient is not statistically signi cant. I then restrict my sample to the poorest 54% and 20% which Leibbrandt, Finn and Oosthuizen (2016) suggest is the poverty and extreme poverty head count percentages in South Africa. The results in Table A8 show larger statistically signi cant point estimates for the poor especially the extremely poor.

I further analyze these potentially heterogeneous e ects with a gure analogous to Figure 5 whereby I estimate the relationship between log of income and CES-D by baseline wealth deciles. Figure 6 shows the results. While the overall point estimates are negative for all and not statistically di erent from each other, the estimated e ects are larger among the poorer half of the sample and more consistently statistically signi cant.

Even in percentage changes, this heterogeneity makes sense: a 20% increase in household income per capita for a wealthy family may not have the same e ect on psychological well-being as it would for individuals in a poor family barely able to meet their basic needs.

4.3.2 Alternative Approach

In Appendix C, I use an alternative approach to estimate the e ect of income on psychological well-being. Namely, I leverage the discontinuity in eligibility for the Older Person's Grant and restrict the sample to individuals with economically inactive household members around the age eligibility threshold. I then employ a fuzzy local randomization approach using having a household member eligible for the grant as an instrument for household in-

Figure 6: Heterogeneous e ects of the Log of Income on CES-D: The impact of changes in income on CES-D based on baseline wealth decile in Wave 3. This gure was based on 8 linear regressions: For baseline wealth deciles 2-9, the sample was restricted to those were in deciles 1. Dashed con dence intervals are Bonferroni-Holm corrected CIs to control the family-wise error rate. This gure was estimated with instruments matrix $Z^{A}_{i,t}$.

come in narrow windows around the age eligibility threshold.[17] The results estimating the e ect of income on CES-D scores, even when restricted to indirect recipients, show point estimates similar to those shown in the average e ects shown in Figure 6 and Table A8.

While the estimation approach requires strong assumptions on the dynamic process and a lack of long persistence in common shocks, I show similar estimates of the e ect of income on psychological well-being using a di erent identi cation approach. Under the assumptions outlined in Section 3, the results demonstrate that economic and psychological well-being are intertwined. There are signi cant e ects in both directions with important heterogeneity. The estimated e ect of changes in depressive symptoms are stronger at the higher end of the scale. On the other hand, the e ect of changes in economic well-being are especially pronounced for the poorer part of the sample.

5 Poverty Dynamics: Simulations

In Section 4, I estimate a system of dynamic simultaneous equations that show the extent to which psychological well-being is intertwined with income and poverty. The estimated coe cients do not, on average, suggest a poverty trap in the strict sense of the term.

[17] Detailed discussion on local randomization approaches in regression discontinuity designs can be found in Cattaneo, Idrobo and Titiunik (forthcoming).

In this section, I show that psychological well-being can still play an important role in the dynamics of income and the persistence of poverty. I show how the bi-directional relationship exacerbates the impacts of shocks to either variable over time. I then use simulations to illustrate the impact this relationship can have on the persistence of poverty.

First, I borrow from the structural vector auto-regression literature and conduct an impulse response function analysis on the estimated dynamic simultaneous relationship between income and psychological well-being. As expected, this simultaneity increases the impact of shocks in a certain time period on either variable in the future. Compared to an AR(1) income process (current income depends only only on lagged income and an error term) where psychological well-being plays no role, the estimated bi-directional relationship exacerbates the effect of the initial shock but also has an added impact over time.[18] The estimated system of equations predicts that the overall impact of an income shock (includes current and future loss) is nearly double that estimated through an AR(1) process. The results in Section 4 show significant heterogeneity/non-linearity. Particularly, among the poor in the upper half of the CES-D distribution, the estimated effects are larger the impulse response function analysis shows the added initial and long-run impacts that this group might experience. This suggest that an across-the-board shock to either income or psychological well-being affects some individuals the poor with low levels of psychological well-being (approximately 18% of the NIDS sample) disproportionately.

This heterogeneity in the overall impact of shocks can help explain low levels of resilience among some. But can the relationship between income and psychological well-being with its key non-linearities also help explain the persistence of poverty? To illustrate the implications on poverty dynamics and allow for repeated shocks, I use the estimated coefficients of the system of dynamic equations to simulate income and CES-D over time. In these simulation, I make simplifying assumptions that the path of income over time does not change, and I extrapolate what happens over a longer period of time using estimates from one time period. In addition, I assume that there are no intra-household responses to changes in individual income and that households face the same income path. I independently and randomly draw income and CES-D values at time 0 with means, variances, and zero-centered normally distributed shocks every time period calibrated by the NIDS data. At time zero, CES-D score is independent of income and poverty status.

Figure 7 shows the probability of being poor after 10 years based on initial income (y) and CES-D score. Figure 7(a) shows probabilities of poverty based on income dynamics that do not include psychological well-being. It is clear that the poverty in the future depends solely on initial income. Figure 7(b) displays this probability when I simulate income over time using the estimated system of equations. The observed pattern suggests that those who initially start with low levels of psychological well-being are more likely to be poor after 10 years even with those who start with income levels above the poverty line. Figure

[18] Figure E1 in Appendix E shows this clearly. More on this analysis and calculations in the Appendix E.

(a) No simultaneous causality

(b) Estimated system of dynamic simultaneous equations

Figure 7: Heatmap of Poverty Probabilities by initial Income and CES-D. Simulations results show the probability of being poor after 10 years based on the independently and randomly drawn income (y) and CES-D in time $t = 0$. The estimated system of equations clearly increases the probabilities especially for those with higher CES-D scores.

8 shows the di erence in the probability of poverty that the estimated equations predict. While increasing overall vulnerability to long-term poverty, this increase is especially large for those with low levels of psychological well-being. In addition, those near the poverty line become more vulnerable; a negative shock could put them on a vicious cycle that is di cult to get out of.

These gures illustrate the e ect of psychological well-being on poverty dynamics in a setting where there can be successive negative shocks. An individual who starts above the poverty line and under the depression threshold may experience an income or psychological well-being shock that lowers both their income and level of psychological well-being. If the shock is strong enough, it could potentially put them in the red zone (Figure 8 poor with

Figure 8: Di erence in Poverty Probabilities: A Heatmap. The di erence in probabilities of Figure 7(b) and 7(a). The estimated relationship increases the likelihood of poverty after 10 years for most. However, this di erence is especially pronounced for the moderately poor with low levels of psychological well-being. In addition, those near the poverty have an added risk of shocks potentially pushing them into the zone where the feedback loop is strong.

low levels of psychological well-being) where the causal links of this relationship (in both directions are stronger) and where another negative shock could have even bigger long-term impacts. While the estimated relationship does not itself constitute a multi-equilibrium poverty trap on average, it does make it increasingly di cult to get out of poverty by exacerbating the initial and long-run impact of shocks.

6 Conclusion

This paper explores the bi-directional relationship between income and psychological well-being. Despite its importance, this relationship is understudied among general populations partly because of the di culty in establishing causality with limited data. Experimentally identifying the e ects of improvements in psychological well-being can be achieved among very speci c samples those su ering from psychological distress who seek treatment. However, sample selection hampers the generalizability of such estimates. The recent availability of large-scale and high-quality panel datasets that track mental health allow for the use of econometric methods that answer important policy-relevant questions.

With the caveats regarding inferring causality from observational data in mind, the goal of this paper is to shed light on the relationship between mental health and poverty in a general population. I nd signi cant impacts in both directions with important heterogeneity. While the system of equations do not explicitly suggest a poverty trap, the magnitude of the estimated e ects suggests larger overall e ects of shocks and longer lasting poverty spells. The results highlight an especially vulnerable group the poor with low levels of psychological well-being who could be disproportionately a ected by shocks.

The results of this paper also add to the discussion on unexpectedly large impacts of some poverty alleviation programs. A stable income through aid likely improves levels of psychological well-being which allows individuals to realize their capabilities and further improve their economic well-being in a way that exceeds initial expectations. This suggests that aside from being a constitutive and important outcome in itself, psychological well-being is also an instrumental one (Sen, 1999). The results rea rm the conclusion of Haushofer and Fehr (2014) stressing the importance of considering psychological variables as avenues for novelty in poverty-alleviation programs.

To estimate the relationship between income and psychological well-being as a system of dynamic simultaneous equations, I use a dynamic panel GMM approach and a large dataset from South Africa. This approach requires strong assumptions on the dynamic process which are di cult to test with short panels. However, I provide alternative supporting results using a regression discontinuity local randomization approach which generates closely comparable estimates for the e ect of income on psychological well-being. Future work with longer panels could add to this analysis by directly testing for serial correlation and further relaxing the assumptions on sequential exogeneity. In addition, larger datasets could allow for investigating heterogeneity by sex, type of work an individual is involved in, or by key household characteristics.

The results from this paper should encourage future research on this topic including focusing on other components of psychological well-being. In the future, di erentiating the e ects of negative versus positive economic shocks on psychological well-being is a fruitful endeavor. The econometric approach here does not allow me to restrict to positive or negative shocks and thus shows the average of both. In addition, shedding light on the mechanisms through which changes in psychological well-being a ect income is important. In this paper and in De Quidt and Haushofer (2016), labor supply seems to be an important mechanism. However, researchers in psychology have shown various ways di erent mental disorders a ect preferences and even cognitive ability. Investigating these mechanisms is germane to the design of e ective poverty alleviation policy.

References

Alloush, Mo, and Jeffrey Bloem. 2022. Neighborhood Violence, Poverty, and Psychological Well-being. Journal of Development Economics, 154(102756).

Alloush, Mo, and Stephen Wu. 2023. Income Improves Subjective Well-being: Evidence from South Africa. Economic Development and Cultural Change.

Anderson, T.W., and Cheng Hsiao. 1982. Formulation and estimation of dynamic models using panel data. Journal of Econometrics.

Angelucci, Manuela, and Daniel Bennett. 2021. The Economic Impact of Depression Treatment in India. IZA Discussion Paper.

Ardington, C., and A. Case. 2010. Interactions between mental health and socioeconomic status in the South African National Income Dynamics Study. Journal for Studies in Economics and Econometrics, 34(3): 69 85.

Arellano, Manuel. 2003. Panel Data Econometrics. Oxford, United Kingdom:Oxford University Press.

Arellano, Manuel, and Stephen Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. The Review of Economic Studies, 58(2): 277.

Baird, Sarah, Jacobus De Hoop, and Berk Özler. 2013. Income shocks and adolescent mental health. Journal of Human Resources, 48(2): 370 403.

Baranov, Victoria, Sonia Bhalotra, Pietro Biroli, and Joanna Maselko. 2019. Maternal Depression, Women's Empowerment, and Parental Investment: Evidence from a Randomized Control Trial. American Economic Review.

Baron, Emily Claire, Thandi Davies, and Crick Lund. 2017. Validation of the 10-item Centre for Epidemiological Studies Depression Scale (CES-D-10) in Zulu, Xhosa and Afrikaans populations in South Africa. BMC Psychiatry, 17(1): 6.

Bazzi, Samuel, and Michael A. Clemens. 2013. Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. American Economic Journal: Macroeconomics.

Beck, Aaron T. 1967. Depression: Causes and Treatment.. 1 ed., Philadelphia:University of Pennsylvania Press.

Berndt, Ernst R, Stan N Finkelstein, Paul E Greenberg, Robert H Howland, Alison Keith, A John Rush, James Russell, and Martin B Keller. 1998. Workplace performance effects from chronic depression and its treatment. Journal of health Economics, 17(5): 511 535.

Biasi, Barbara, Michael S Dahl, and Petra Moser. 2021. Career effects of mental health. National Bureau of Economic Research.

Blundell, Richard, and Stephen Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics.

Bolton, Paul, Judith Bass, Richard Neugebauer, Helen Verdeli, Kathleen F Clougherty, Priya Wickramaratne, Liesbeth Speelman, Lincoln Ndogoni, and Myrna Weissman. 2003. Group Interpersonal Psychotherapy for Depression in Rural Uganda: A Randomized Controlled Trial. Jama, 289(23): 3117 3124.

Bryan, Mark L, Nigel Rice, Jennifer Roberts, and Cristina Sechel. 2020. Mental health and employment: a bounding approach using panel data. The She eld Economic Research Paper Series (SERPS), 2020006(2020006).

Bubonya, Melisa, Deborah A Cobb-Clark, and David C Ribar. 2019. The reciprocal relationship between depressive symptoms and employment status. Economics & Human Biology, 35: 96 106.

Cattaneo, Matias D, Nicolás Idrobo, and Rocío Titiunik. forthcoming. A practical introduction to regression discontinuity designs: Extensions. Cambridge University Press.

Charles, Kerwin Ko , and Melvin Stephens. 2004. Disability, job displacement and divorce. Journal of Labor Economics, 22(2): 489 522.

Chatterji, Pinka, Margarita Alegria, and David Takeuchi. 2011. Psychiatric disorders and labor market outcomes: Evidence from the National Comorbidity Survey-Replication. Journal of Health Economics, 30(5): 858 868.

Christian, Cornelius, Lukas Hensel, and Christopher Roth. 2019. Income Shocks and Suicides: Causal Evidence From Indonesia. The Review of Economics and Statistics, 101(5): 905 920.

Collins, Pamela Y, Vikram Patel, Sarah Joestl, Dana March, Thomas R. Insel, and Abdallah S. Daar. 2011. Grand challenges in global mental health. Nature, 475(N.A): 37 30.

Currie, Janet, and Erdal Tekin. 2012. Understanding the cycle childhood maltreatment and future crime. Journal of Human Resources, 47(2): 509 549.

Das, Jishnu, Quy-Toan Do, Jed Friedman, and David McKenzie. 2009. Mental Health Patterns and Consequences: Results from Survey Data in Five Developing Countries. World Bank Economic Review, 23(1): 31 55.

De Quidt, Jonathan, and Johannes Haushofer. 2016. Depression for Economists. NBER Working Paper, 1 33.

Friedrich, M.J. 2017. Depression Is the Leading Cause of Disability Around the World. Journal of the American Medical Association, 317(15): 1517.

Frijters, Paul, David W Johnston, and Michael A Shields. 2014. The E ect of Mental Health on Employment: Evidence from Australian Panel Data. Health economics, 23: 1058 1071.

Frijters, Paul, John P Haisken-DeNew, and Michael A Shields. 2004. Money Does Matter ! Evidence from Increasing Real Income and Life Satisfaction in East Germany Following Reuni cation. The American Economic Review, 94(3): 730 740.

Frijters, Paul, John P Haisken-DeNew, and Michael A Shields. 2005. The causal e ect of income on health: Evidence from German reuni cation. Journal of Health Economics, 24(5): 997 1017.

Gardner, Jonathan, and Andrew J. Oswald. 2007. Money and mental wellbeing: A longitudinal study of medium-sized lottery wins. Journal of Health Economics, 26(1): 49 60.

González, Patricia, Alicia Nuñez, Erin Merz, Carrie Brintz, Orit Weitzman, Elena L. Navas, Alvaro Camacho, Christina Buelna, Frank J. Penedo, Sylvia Wassertheil-Smoller, Krista Perreira, Carmen R. Isasi, James Choca, Gregory A. Talavera, and Linda C. Gallo. 2017. Measurement properties of the Center for Epidemiologic Studies Depression Scale (CES-D 10): Findings from HCHS/SOL. Psychological Assessment, 29(4): 372 381.

Graham, Carol, and Stefano Pettinato. 2002. Frustrated Achievers: Winners, Losers and Subjective Well-Being in New Market Economies. The Journal of Development Studies, 38(4): 100 140.

Guvenen, Fatih. 2007. Learning your earning: Are labor income shocks really very persistent? American Economic Review, 97(3): 687 712.

Haan, Peter, and Michal Myck. 2009. Dynamics of health and labor market risks. Journal of health economics, 28(6): 1116 1125.

Hamad, R, L C H Fernald, D S Karlan, and J Zinman. 2008. Social and economic correlates of depressive symptoms and perceived stress in South African adults. Journal of epidemiology and community health, 62(6): 538 544.

Hamilton, Vivian H, Philip Merrigan, and Éric Dufresne. 1997. Down and out: estimating the relationship between mental health and unemployment. Health economics, 6(4): 397 406.

Haushofer, Johannes. 2019. Is there a Psychological Poverty Trap? Working Paper.

Haushofer, Johannes, and Ernst Fehr. 2014. On the psychology of poverty. Science, 344(6186): 862 7.

Haushofer, Johannes, and Jeremy Shapiro. 2016. The short-term impact of unconditional cash transfers to the poor: Experimental evidence from kenya. Quarterly Journal of Economics, 131(4): 1973 2042.

Haushofer, Johannes, Robert Mudida, and Jeremy P Shapiro. 2020. The comparative impact of cash transfers and a psychotherapy program on psychological and economic well-being. National Bureau of Economic Research.

Hays, Ron D., K B Wells, C D Sherbourne, W Rogers, and K Spritzer. 1995. Functioning and well-being outcomes of patients with depression compared with chronic general medical illnesses. Archives of General Psychiatry, 52(1): 11 9.

Headey, Bruce, Jonathan Kelley, and Alex Wearing. 1993. Dimensions of mental health: Life satisfaction, positive a ect, anxiety and depression. Social Indicators Research, 29(1): 63 82.

Holm, Sture. 1979. A simple sequentially rejective multiple test procedure. Scandinavian journal of statistics, 65 70.

Holtz-Eakin, Douglas, Whitney Newey, and Harvey S. Rosen. 1988. Estimating Vector Autoregressions with Panel Data. Econometrica, 56(6): 1371 1395.

Hsiao, Cheng, and Qiankun Zhou. 2015. Statistical inference for panel dynamic simultaneous equations models. Journal of Econometrics, 189(2): 383 396.

Hugo, Charmaine J., Dorothy E L Bosho , Annelene Traut, Nompumelelo Zungu-Dirwayi, and Dan J. Stein. 2003. Community attitudes toward and knowledge of mental illness in South Africa. Social Psychiatry and Psychiatric Epidemiology, 38(12): 715 719.

Johnes, G, and J Johnes. 2004. International Handbook on the Economics of Education. Knowledge Creation Di usion Utilization , 28(6): 1578 1580.

Kahneman, Daniel, and Angus Deaton. 2010. High income improves evaluation of life but not emotional well-being. Proceedings of the National Academy of Sciences, 107(38): 16489 93.

Leibbrandt, Murray, Arden Finn, and Ingrid Woolard. 2012. Describing and decomposing post-apartheid income inequality in South Africa. Development Southern Africa, 29(1): 19 34.

Leibbrandt, Murray, Arden Finn, and Morné Oosthuizen. 2016. Poverty, inequality, and prices in post-apartheid South Africa. Growth and poverty in sub-Saharan Africa, 393 417.

Lindqvist, Erik, Robert Östling, and David Cesarini. 2020. Long-run e ects of lottery wealth on psychological well-being. The Review of Economic Studies, 87(6): 2703 2726.

Lund, Crick, Kate Orkin, Marc Witte, Thandi Davies, Johannes Haushofer, Judy Bass, and V Patel. 2020. Economic impacts of mental health interventions in low and middle-income countries: a systematic review and meta-analysis. Working paper.

Lund, Crick, Mary De Silva, Sophie Plagerson, Sara Cooper, Dan Chisholm, Jishnu Das, Martin Knapp, and Vikram Patel. 2011. Poverty and mental disorders: Breaking the cycle in low-income and middle-income countries. The Lancet, 378(9801): 1502 1514.

Macours, Karen, Norbert Schady, and Renos Vakis. 2012. Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment. American Economic Journal: Applied Economics, 4(2): 247 273.

Magnac, Thierry. 2000. Subsidised Training and Youth Employment : Distinguishing Unobserved Heterogeneity from State Dependence in Labour Market Histories. The Economic Journal, 110(466): 805 837.

McGuire, Joel, Caspar Kaiser, and Anders M Bach-Mortensen. 2022. A systematic review and meta-analysis of the impact of cash transfers on subjective well-being and mental health in low-and middle-income countries. Nature Human Behaviour, 1 12.

McInerney, Melissa, Jennifer M. Mellor, and Lauren Hersch Nicholas. 2013. Recession depression: Mental health e ects of the 2008 stock market crash. Journal of Health Economics, 32(6): 1090 1104.

Meghir, Costas, and Luigi Pistaferri. 2004. Income Variance Dynamics and Heterogeneity. Econometrica, 72(1): 1 32.

Myer, Landon, Joalida Smit, Liezel Le Roux, Siraaj Parker, Dan J Stein, and Soraya Seedat. 2008. Common mental disorders among HIV-infected individuals in South Africa: prevalence, predictors, and validation of brief psychiatric rating scales. AIDS patient care and STDs, 22(2): 147 158.

Ormel, Johan, Tineke Oldehinkel, Els Brilman, and Wim v Brink. 1993. Outcome of Depression and Anxiety in Primary Care. Archives of General Psychiatry, 50(10): 759.

Oswald, Andrew J, Eugenio Proto, and Daniel Sgroi. 2015. Happiness and productivity. Journal of labor economics, 33(4): 789 822.

Patel, Vikram, Benedict Weobong, Helen A Weiss, Arpita Anand, Bhargav Bhat, Basavraj Katti, Sona Dimidjian, Ricardo Araya, Steve D Hollon, Michael King, et al. 2017. The Healthy Activity Program (HAP), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomised controlled trial. The Lancet, 389(10065): 176 185.

Pilkonis, Paul A., Lan Yu, Nathan E. Dodds, Kelly L. Johnston, Catherine C. Maihoefer, and Suzanne M. Lawrence. 2014. Validation of the depression item bank from the Patient-Reported Outcomes Measurement Information System (PROMIS®) in a three-month observational study. Journal of Psychiatric Research, 56: 112 119.

Radlo , Lenore Sawyer. 1977. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. Applied Psychological Measurement, 1(3): 385 401.

Ridley, Matthew, Gautam Rao, Frank Schilbach, and Vikram Patel. 2020. Poverty, depression, and anxiety: Causal evidence and mechanisms. Science, 370(6522): eaay0214.

Samuels, Fiona, and Maria Stavropoulou. 2016. `Being Able to Breathe Again': The E ects of Cash Transfer Programmes on Psychosocial Wellbeing. The Journal of Development Studies, 52(8): 1099 1114.

Santor, Darcy A., Michelle Gregus, and Andrew Welch. 2006. Eight Decades of Measurement in Depression. Measurement: Interdisciplinary Research & Perspective, 4(3): 135 155.

Saylor, C F, G L Edwards, and J A McIntosh. 1987. Children's Depression Inventory: Reliability over repeated administrations. Journal of Clinical Child . . . , 16(4): 339 341.

Schilbach, Frank, Heather Scho eld, and Sendhil Mullainathan. 2016. The psychological lives of the poor. American Economic Review Paper & Proceedings, 106(5): 435 440.

Sen, Amartya. 1999. Development as Freedom. Oxford Press, 1 50.

Shreekumar, Advik, and Pierre-Luc Vautrey. 2022. Managing Emotions: The E ects of Online Mindfulness Meditation on Mental Health and Economic Behavior. Working Paper.

Spijker, J, R Graaf, R V Bijl, a T F Beekman, J Ormel, and W a Nolen. 2004. Functional disability and depression in the general population. Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). Acta psychiatrica Scandinavica, 110(3): 208 14.

Steele, Fiona, Robert French, and Mel Bartley. 2013. Adjusting for selection bias in longitudinal analyses using simultaneous equations modeling: the relationship between employment transitions and mental health. Epidemiology, 703 711.

Steel, Zachary, Claire Marnane, Changiz Iranpour, Tien Chey, John W. Jackson, Vikram Patel, and Derrick Silove. 2014. The global prevalence of common mental disorders: A systematic review and meta-analysis 1980-2013. International Journal of Epidemiology.

Stevenson, Betsey, and Justin Wolfers. 2013. Subjective Well-Being and Income: Is There Any Evidence of Satiation? American Economic Review: Papers & Proceedings, 103(1033): 598 604.

Stock, James H., and Jonathan H. Wright. 2000. GMM with weak identi cation. Econometrica, 68: 1055 1096.

Zhang, Wendy, Nadia O'Brien, Jamie I. Forrest, Kate A. Salters, Thomas L. Patterson, Julio S G Montaner, Robert S. Hogg, and Viviane D. Lima. 2012. Validating a shortened depression scale (10 item CES-D) among HIV-Positive people in British Columbia, Canada. PLoS ONE, 7(7): 1 5.

Appendix

A. Tables

Table A1: CES-D 10 Questionnaire

	In the past week...	Rarely or none of the time (Less than 1 day)	Some or little of the time (1-2 days)	Occasionally or a moderate amount of the time (3-4 days)	Most or all of the time (5-7 days)
1	I was bothered by things that usually don't bother me	0	1	2	3
2	I felt depressed	0	1	2	3
3	I felt lonely	0	1	2	3
4	I had trouble keeping my mind on what I was doing	0	1	2	3
5	I felt that everything I did was an e ort	0	1	2	3
6	I felt hopeful about the future	3	2	1	0
7	I felt fearful	0	1	2	3
8	My sleep was restless	0	1	2	3
9	I was happy	3	2	1	0
10	I could not get going	0	1	2	3

Notes: The scores for all questions are summed for an overall score between 0 and 30. For negative feelings such as how often an individual felt loneliness or an inability to get going, the respondent gets a 0 score if they respond with Not at all or rarely, 1 for Some or little of the time, 2 for Occasionally, and 3 for All the time. For positive statements such as feeling hopeful, the scores are reversed. The numbers corresponding to the answers are then added for all questions. The shortened 10-item CES-D scale used here attains satisfactory prediction accuracy and reliability in assessing depressive symptoms and correlates very highly with the full 20-item questionnaire (Zhang et al., 2012) and is shown to be stable over time (González et al., 2017; Saylor, Edwards and McIntosh, 1987). Language and culture likely a ect the way questions are understood and answered so it is important to consider whether the CES-D is valid in South Africa (Samuels and Stavropoulou, 2016). The CES-D 10 is commonly used in South Africa and is shown to be internally consistent and veri ed as an e ective screening tool for depression (Hamad et al., 2008; Johnes and Johnes, 2004; Myer et al., 2008; Baron, Davies and Lund, 2017). In the 10-item CES-D, a threshold score of 10 is most commonly used for depression screening; however, Baron, Davies and Lund (2017) suggest that a threshold score of 11 is appropriate for screening for depression among most populations in South Africa. For more on how these thresholds are determined see Baron, Davies and Lund (2017). A threshold of 11 on average correctly classi ed 84% of cases (as depressed or not depressed) in a follow up on the NIDS sample in South Africa. This threshold score is determined based on a trade-o between sensitivity, speci city, and positive predictive value for clinical depression. The CES-D Scale has slightly better positive predictive value than the PHQ-9, another commonly used depression screening tool. The CES-D scale and PHQ-9 are shown to be highly correlated [0.8-0.88] (Pilkonis et al., 2014).

Table A2: Comparison of Wave Samples vs Balanced Sample

	Sample:				
Wave 4 Mean (SD)	Wave 1	Wave 2	Wave 3	Wave 4	Balanced
Household Income Per Capita	2,582	2,576	2,621	2,624	1,997
	(2,985)	(2,953)	(3,022)	(3,023)	(2,022)
Food Expenditure Per Capita	422	419	426	426	373
	(437)	(421)	(439)	(437)	(386)
Individual Income	3,141	3,155	3,209	3,220	2,363
	(4,614)	(4,589)	(4,662)	(4,672)	(2,992)
CES-D score	7.22	7.20	7.22	7.21	7.24
	(4.30)	(4.31)	(4.33)	(4.31)	(4.28)
Household Size	4.78	4.80	4.77	4.77	4.99
	(3.21)	(3.20)	(3.19)	(3.18)	(3.28)
Female	0.58	0.57	0.56	0.56	0.62
	(0.49)	(0.50)	(0.50)	(0.50)	(0.48)
Age	42.86	42.58	42.60	42.54	42.47
	(15.28)	(15.13)	(15.15)	(15.14)	(15.63)
Observations	9,150	9,425	9,563	9,662	7,967

Notes: This table provides some descriptive statistics samples from each of the NIDS Waves and the balanced sample without the CES-D restriction. The sample in this study include individuals who completed the individual section of the survey including the CES-D section for all four rounds of NIDS (Table 1). Here I show the descriptive statistics for those in in all four waves but may have not responded to the CES-D questions in one of those waves. Given that the balanced sample requires the individual to have responded to the adult individual questionnaire in Wave 1 (and after), I restrict the samples to individuals to those who are age 22 and above in Wave 4 to be comparable to the balanced sample used in the study. The sample used in the study appears to be, on average, poorer than representative samples from each wave. However, there does not seem to be selection when it comes to CES-D response.

Table A3: GMM Estimates are robust to di erent outlier trimming.

	5%		0.5%		None	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable: Individual Income						
CES-D_t	-1.81	-2.66	-2.22	-4.50	-4.95	-2.41
	(0.57)	(0.82)	(0.81)	(1.59)	(3.47)	(3.48)
Dependent Variable: CES-D						
HH Income Per $Capita_t$	-0.36	-0.46	-0.48	-0.36	-0.14	-0.099
	(0.099)	(0.13)	(0.098)	(0.082)	(0.038)	(0.050)
HH Income Per $Capita_t^2$	0.005	0.0044	0.003	0.0021	0.00	0.00
	(0.002)	(0.002)	(0.001)	(0.000)	(0.000)	(0.000)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,645	5,645	6,860	6,860	7,103	7,103

Notes: Cluster robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Income numbers are in 100 South African Rands. Controls include lagged dependent variables, household size, marital status, and number of children in the household. Each pair of regression results use two-stage GMM for the study sample using two approaches with di erent instrument matrices that are consistent under two slightly di erent assumptions (A and B). The results are consistent to di erent levels of trimming for outliers. The main paper trims to top and bottom 2.5% of changes in household income per capita and individual income.

Table A4: Systems GMM Results (Blundell & Bond (1998)): E ect of CES-D on Individual Income

Dependent Variable: Individual Income	Study Sample (1)	Including Wave 5 (2)
CES-D_t	-1.972	-2.98
	(0.782)	(1.75)
Controls	Yes	Yes
AR(2) p-value		0.38
Sargan test p-value		1.00
Hansen test p-value		0.33

Notes: Cluster robust standard errors in parentheses. Income numbers are in 100 South African Rands. Controls include household size, number of children per household, marital status, and wave xed e ects. Similar results using methods that are more robust to the lagged levels being weak instruments for rst di erences. In addition, using data from Wave 5 allows me to test for serial correlation: I nd no evidence of serial correlation and similar results.

Table A5: Systems GMM Results (Blundell & Bond (1998)): E ect of Economic Well-being on CES-D

Dependent Variable:	Study Sample			Including Wave 5
CES-D	(1)	(2)	(3)	(4)
HH Income Per Capita$_t$	-0.18			-0.47
	(0.07)			(0.0014)
Food Exp Per Capita$_t$		-0.80		
		(0.37)		
Wealth Index$_t$			-2.54	
			(1.04)	
Controls	Yes	Yes	Yes	Yes
AR(2) p-value				0.32
Sargan test p-value				1.00
Hansen test p-value				0.02

Notes: Cluster robust standard errors in parentheses. Income numbers are in 100 South African Rands. Controls include household size, number of children per household, marital status, and wave xed e ects. Similar results using methods that are more robust to the lagged levels being weak instruments for rst di erences. In addition, using data from Wave 5 allows me to test for serial correlation: I nd no evidence of serial correlation and similar results.

Table A6: GMM Estimates when using household income per capita in both equations.

	$Z_{i,t}^{A}$	$Z_{i,t}^{B}$
	(1)	(2)
Dependent Variable: Household Income per Capita		
CES-D_t	-0.38	-1.58
	(0.72)	(2.11)
Dependent Variable: CES-D		
HH Income Per Capita_t	-0.27	-0.28
	(0.099)	(0.09)
HH Income Per Capita_t^{2}	0.004	0.004
	(0.001)	(0.001)
Controls	Yes	Yes
Observations	6,281	6,281

Notes: Cluster robust standard errors in parentheses. Income numbers are in 100 South African Rands. Controls include lagged dependent variables, household size, marital status, and number of children in the household. Under either assumption, using household income per capita in both equations shows estimates that are consistent with results in the main paper; however, one's own psychological well-being does not a ect overall household income as it does to one's individual income.

Table A7: GMM Estimates when using individual income in both equations.

	$Z^A_{i,t}$	$Z^B_{i,t}$
	(1)	(2)
Dependent Variable: Individual Income		
CES-D_t	-1.77	-6.23
	(1.03)	(1.40)
Dependent Variable: CES-D		
Individual Income_t	-0.05	-0.10
	(0.04)	(0.036)
Controls	Yes	Yes
Observations	6,281	6,281

Notes: Cluster robust standard errors in parentheses. Income numbers are in 100 South African Rands. Controls include lagged dependent variables, household size, marital status, and number of children in the household. Under either assumption, using household income per capita in both equations shows estimates that are consistent with results in the main paper.

Table A8: Log Transformations and Estimates for the Poor.

Dependent Variable: CES-D	Full Sample	Poorest 54%	Poorest 20%
	(1)	(2)	(3)
Log(HH Income Per Capita_t)	-2.10	-2.20	-4.34
	(1.36)	(1.13)	(1.96)
Controls	Yes	Yes	Yes
Observations	6,281	3,375	1,383

Notes: Cluster robust standard errors in parentheses. Controls include lagged dependent, household size, marital status, disability, and number of children in the household. Poverty level is determined by the wealth index in Wave 3. Instruments requiring Assumptions A used. Instruments $z^B_{i,t}$ that require less restrictive assumptions show similar results

B. Figures

(a) Relationship between Lagged and Current CES-D

Figure B1: Current and Future CES-D: Past CES-D scores are predictive of current scores.

Figure B2: CES-D and Economic Well-being: A histogram of CES-D scores by poverty status.

(a) No simultaneous causality

(b) Estimated system of dynamic simultaneous equations

Figure B3: Simulations: Income CDFs after 10 years. Top (a) shows income at time 0 and after 5 periods; if CES-D plays no role then initial levels of CES-D will not a ect the distribution of income over time. In the lower part of the Figure (b), the full system of equations estimated above shows that when psychological well-being plays a role, those who randomly begin with lower levels of psychological well-being will have higher rates of poverty in the future.

C. Alternative Identification Approach

In this appendix, I use an alternative approach to estimate the effect of income on CES-D. The results are similar to those estimated using the GMM panel approach. This approach requires different assumptions and shows very similar results. For important questions that are unlikely to be answered with experiments (natural or otherwise) applying a multitude of different approaches requiring different assumptions with qualitatively similar results is important and a second-best approach (Currie and Tekin, 2012). Uniformity of the estimates across methods suggests confidence in the core results.

While I am able to do this for income, it is difficult to find clearly relevant and exogenous instrumental variables for psychological well-being. Frijters, Johnston and Shields (2014) use the reported death of friends outside of the household with no financial ties as an instrumental variables for psychological well-being while controlling for individual fixed effects with an Australian panel. This instrument is not necessarily valid as there are potential avenues through which the death of a friend could affect an individual's income such as have to take time off to attend a funeral or loss of important social ties. The data in NIDS is limited regarding reporting deaths of friends; in Wave 1 and 3, the questionnaire differentiates between deaths of friends and relatives with financial ties to the household and those without. In other waves, they questionnaire only asks about friends and relatives specifically with financial ties to the household. In Alloush and Bloem (2022), we show using the NIDS data that neighborhood violence affects CES-D scores. While violent crime is clearly relevant, changes in neighborhood violence can potentially affect income directly and may be a mechanism through which psychological well-being is affected thus it is unlikely exogenous to my model.

6.1 Alternative Instrument for Income

To check for the robustness of the estimates of the impact of income on psychological well-being using the panel GMM approach, I use an alternative approach to estimate the effect of household income on CES-D scores. Specifically, I use eligibility for the South Africa's Older Person's Grant a cash transfer program that individuals become eligible for when they turn 60. I do so by restricting the sample to households with economically inactive individuals in narrow windows around the age 60. This approach is a local randomization regression discontinuity approach used as the main identification strategy in Alloush and Wu (2023) to estimate the effect of income on life satisfaction.[19] It is essentially a fuzzy regression discontinuity design with a continuous treatment (household income per capita);

[19] While life satisfaction and mental health are related, studies have shown different effects of income on either (Lindqvist, Östling and Cesarini, 2020).

Table C1: Restricted Samples all and non-recipient Results

		All Members			Indirect Recipients		
		(1)	(2)	(3)	(4)	(5)	(6)
Age Range 54-65 N=19,093	Log(HH Income Per$_{i,t}$)	-1.53 (0.59)	-1.99 (0.82)	-2.40 (0.64)	-1.21 (0.49)	-1.26 (0.55)	-1.05 (0.83)
Age Range 55-64 N=16,462	Log(HH Income Per$_{i,t}$)	-1.71 (0.68)	-2.34 (0.91)	-2.88 (0.74)	-0.99 (0.52)	-1.15 (0.59)	-1.56 (0.94)
Age Range 56-63 N=13,691	Log(HH Income Per$_{i,t}$)	-1.71 (0.80)	-1.98 (0.96)	-2.40 (0.81)	-0.88 (0.59)	-0.81 (0.64)	-1.07 (0.97)
Age Range 57-62 N=10,724	Log(HH Income Per$_{i,t}$)	-1.91 (0.94)	-2.20 (1.05)	-2.60 (1.06)	-1.41 (0.76)	-1.72 (0.74)	-1.35 (1.24)
Age Range 58-61 N=7,352	Log(HH Income Per$_{i,t}$)	-2.57 (1.35)	-2.67 (1.57)	-2.89 (1.72)	-2.28 (1.18)	-1.93 (1.05)	-2.59 (2.37)
Controls			X	X		X	X
Individual Fixed Effects				X			X

Notes: Cluster robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. In these regressions, I control for wave and district fixed effects independently.

however, the running variable (age of eligible household member) is discrete, resulting in a small number of mass points around the threshold. Thus I use an approach akin to fuzzy local randomization (Cattaneo, Idrobo and Titiunik, forthcoming).[20] While households with elderly living in it are different from those with no people over 60, I posit that to an individual having a economically inactive 58-year old member in the household is similar to having an economically inactive 61-year old who was also inactive at 58.[21] After restricting the sample in this way in narrow windows around age 60, I use the number of relatives eligible for the grant (due to their age) as an instrument for household income and estimate the following equation:

$$d_{i,t} = {}_0 + {}_1 h_{i,t} + {}_t + Qx_{i,t} + {}_{i,t}$$

As before, $d_{i,t}$ is the CES-D score, $h_{i,t}$ is household income per capita, ${}_t$ is a wave fixed effect, $x_{i,t}$ is a vector of time varying individual and household characteristics including household size, number of children in the household, marital status, age (cubic), gender, race, and education.

[20] See discussion in Cattaneo, Idrobo and Titiunik (forthcoming): with a small number of mass points around the cutoff, the sample size in continuity-based approaches is essentially the number of mass points, which in this case is very small. Cattaneo, Idrobo and Titiunik (forthcoming) suggest local randomization approaches as more appropriate for this type of data.

[21] I use the panel to impose this restriction.

The estimated coefficients are shown in Table C1. I show results for all members of the household and specifically for indirect recipients (other members living in the household who are not near the age of 60). In addition, I show results for different window sizes around the age of 60.[22] The results again suggest that an increase in household income per capita decreases an individual's CES-D score reflecting a decrease in depressive symptoms. Moreover, the point estimates are very similar in magnitude to the point estimates (Table 2 and specifically Table A8) calculated using the dynamic panel GMM approach above. Results from the first stage of the IV regressions producing the estimates in Column 3 suggest that having a grant eligible (above age 60) individual in the household increases household income per capita by 17% (a very strong first stage). A 17% change in household income per capita is estimated to decrease CES-D scores by approximately 0.3 points.[23]

The results in Columns (4)-(6) show the estimated effects on the non-recipient sample. The results are not as large or consistently statistically significant, but show an overall pattern of decrease in depressive symptoms due to increase in household income from the Older Person's Grant.

In this section, I provide an alternative method of estimating the effect of income on depressive symptoms. This approach requires assumptions and has limitations in terms of interpretation and generalizability. However, the point estimates are similar to those estimated using the GMM approach in the main paper. While I am unable to similarly estimate the effect of changes in CES-D on individual income, the uniformity in the estimates for the other equation gives confidence in the GMM approach.

[22] It is important to note that this analysis is restricted to households with individuals who are around the age of 60 and are economically inactive. In addition, for those above 60, I drop those who were in the labor force in previous waves when they were just below 60.

[23] Details on this method and the increase in economic well-being due to the old age grant can be found in Alloush and Wu (2023). In addition, more detailed discussion on local randomization approaches in regression discontinuity designs can be found in Cattaneo, Idrobo and Titiunik (forthcoming).

D. Econometric Approach: Proofs and Illustration (for online publication only)

The main di culty in empirically unpacking the relationship between income and poverty on one hand and psychological well-being and mental health on the other is the bi-directionality of this relationship: there are potentially causal impacts in both directions. Many relationships that we are interested in studying as economists may potentially exhibit this bi-directionality. Examples include income and health at the individual level, attitudes and laws at the community level, electricity access and poverty at the regional level, and institutions and GDP growth at the country level to name a few. While this bi-directional causation makes empirical identi cation di cult, it also can make the implications of this relationship all the more important. Simultaneous bi-directionality can imply a feedback loop that if strong enough can put the unit of interest (individuals/households/communities/regions /countries) on vicious or virtuous cycles.

The increasing availability of panel data especially at the individual and household levels allows us to use the added information that repeated observations give us to answer important economic questions where it is di cult to nd cross-sectional instrumental variables or natural experiments and that cannot (or in a lot of cases should not) be answered experimentally. In this extended appendix, I adapt panel data methods (Holtz-Eakin, Newey and Rosen, 1988; Arellano and Bond, 1991; Anderson and Hsiao, 1982) where lagged levels are e ectively used as instrumental variables to estimate a system of simultaneous and dynamic equations. This extension to panel data methods requires assumptions on the dynamic process that includes at most rst-order moving average serial correlation in shocks and a minimum of four rounds of data. I discuss this method using income and psychological well-being as the two variables of interest, however, the method can be generalized to any number of variables that are persistent over time and are potentially simultaneous.

D.1 System of Equations

As discussed, the main source of endogeneity when studying the relationship between mental health and income is simultaneity. Psychological well-being may have an impact on an individual's own earnings, but at the same time, income or the level of economic well-being can a ect their psychological well-being. This type of potentially simultaneous bi-directional causality is common in economics as well as in other disciplines (psychology, evolutionary biology, etc...). Conceptually, this bi-directional relationship can be described using a system of simultaneous equations as follows:

$$y_{i,t} = f(d_{i,t}) + e_{i,t}$$

$$d_{i,t} = g(y_{i,t}) + u_{i,t}$$

where $y_{i,t}$ is income and $d_{i,t}$ is a measure of psychological well-being for individual i in time t.[24] $e_{i,t}$ and $u_{i,t}$ are the unobserved error terms for their respective equations. Importantly, I assume that both variables are state dependent they exhibit some level of persistence and are not completely determined independently in each period. I add a lag of the dependent variable to both equations to take into account this state dependence resulting in the following dynamic simultaneous system of equations:

$$y_{i,t} = f(d_{i,t}, y_{i,t\ \ 1}) + e_{i,t}$$

$$d_{i,t} = g(y_{i,t}, d_{i,t\ \ 1}) + u_{i,t}$$

I also want to allow for individual xed e ects which can be a source of omitted variable bias if not taken into account. I insert $\ _i$ and $\ _i$ as additive individual xed e ects in their respective equations. When estimating the system of equations, I can easily add $x_{i,t}$: a vector of exogenous time varying individual characteristics for individual i at time t, however, to illustrate the estimation approach in a simple way, I will ignore $x_{i,t}$ beyond the equations below:

$$y_{i,t} = f(d_{i,t}, y_{i,t\ \ 1}, x_{i,t}) + \ _i + e_{i,t}$$

$$d_{i,t} = g(y_{i,t}, d_{i,t\ \ 1}, x_{i,t}) + \ _i + u_{i,t}$$

I consider only parametric speci cations for $f(.)$ and $g(.)$. To outline and justify the proposed estimation approach that extends panel data methods to estimate a system of dynamic simultaneous equations, I present a simple linear form of the above system of equations:

$$y_{i,t} = \ _1 d_{i,t} + \ _1 y_{i,t\ \ 1} + \ _i + e_{i,t}$$

$$d_{i,t} = \ _2 y_{i,t} + \ _2 d_{i,t\ \ 1} + \ _i + u_{i,t}$$

The individual xed e ects $\ _i$ and $\ _i$ are likely important determinants of both income and psychological well-being. I can control for these individual xed e ects by rst-di erencing both equations as such:

$$Dy_{i,t} = \ _1 Dd_{i,t} + \ _1 Dy_{i,t\ \ 1} + De_{i,t} \quad (5)$$

$$Dd_{i,t} = \ _2 Dy_{i,t} + \ _2 Dd_{i,t\ \ 1} + Du_{i,t} \quad (6)$$

In this rst-di erences system of simultaneous dynamic equations, I am interested in estimating the coe cients on four variables, namely $\ _1$, $\ _2$, $\ _1$, and $\ _2$. Two of these vari-

[24] I use income and psychological well-being throughout as they are the two variables I am interested in in this paper, however the method I outline can be applied to any two simultaneous and persistent variables.

ables are assumed to be endogenous ($\Delta d_{i,t}$ in equation 6 and $\Delta y_{i,t}$ in equation 7) and the dependent variable lags ($\Delta y_{i,t-1}$ and $\Delta d_{i,t-1}$) are endogenous due to the first differencing.[25] By considering each single equation separately, dynamic panel data methods (commonly referred to as Arellano-Bond methods)[26] suggests that, assuming sequential exogeneity and that the error terms $\varepsilon_{i,t}$ and $u_{i,t}$ are serially uncorrelated, the lagged levels $y_{i,t-2}, y_{i,t-3}, ...$ and $d_{i,t-2}, d_{i,t-3}, ...$ may be used as instruments to estimate the parameters of the equation (2.1); the same set of instruments may be used to estimate equation (7) as well (Anderson and Hsiao, 1982; Arellano and Bond, 1991; Holtz-Eakin, Newey and Rosen, 1988). With a minimum of three observations per unit ($T \geq 3$), the resulting estimator is consistent with large N. Despite not always achieving asymptotic efficiency, the estimates are consistent under very general conditions such as conditional and time series heteroskedasticity, correlation between the individual fixed effect and the error terms, and predetermined initial conditions.[27]

The intuition behind this estimation strategy is as follows: after controlling for individual fixed effects, lagged income, psychological well-being, and other time varying individual characteristics, what remains in the error terms is random assuming no serial correlation. The lagged levels are correlated with the first differences thus relevant, and are uncorrelated with the remaining error terms and thus valid instrumental variables.

This method has been refined to increase its efficiency and robustness.[28] It has also been used in numerous studies with panel datasets over the years especially with income. Some persistence in income and the assumption that lagged income is a sufficient statistic that summarizes earlier states of income make it ideal for this kind of estimation.

D.2 Extending Arellano-Bond

Below, I derive the conditions necessary to extend this dynamic panel data method to estimate a system of dynamic simultaneous equations shown above.[29] The assumption of

[25] Right hand side variable $\Delta y_{i,t}$ and its regressor $\Delta y_{i,t-1}$ in equation (6) share a common variable $y_{i,t-1}$ making the regressor correlated with the error term $\Delta\varepsilon_{i,t}$. Similarly for $\Delta d_{i,t}$ and its regressor $\Delta d_{i,t-1}$ and the error term $\Delta u_{i,t}$ in equation (2.2).

[26] The first to propose this type of approach where lagged level values would be used as instruments for first-differenced dynamic equations were Anderson and Hsiao (1982). This work was further developed by Holtz-Eakin, Newey and Rosen (1988) and later Arellano and Bond (1991). A number of different refinements were made over the years that suggested that additional moment conditions can be imposed in a GMM framework (Arellano & Bover, 1990; 1995; Ahn & Schmidt, 1995; Blundell & Bond, 1998).

[27] For more see Chapter 6 in Arellano (2003).

[28] Additional moment conditions are available under more restrictive assumptions on initial conditions, heteroskedasticity, stationarity, and correlation between the error terms and the individual fixed effect. These additional assumptions can make the estimation robust to models with autoregressive coefficients outside the unit root, for example.

[29] For more on dynamic systems of equations and consistency of the estimations using GMM approaches, see Hsiao and Zhou (2015): A GMM approach is consistent for a fixed T and large N.

sequential exogeneity that is typical when using Arellano-Bond methods to estimate the coefficients of a single dynamic equation is inherently violated due to simultaneous bi-directionality and state-dependence. Sequential exogoneity assumes the following on $e_{i,t}$ and $u_{i,t}$:

$$E\,[e_{i,t} j\; y_{i,t-1}, y_{i,t-2}, y_{i,t-3}, :::; d_{i,t}, d_{i,t-1}, d_{i,t-2}, :::] = 0$$

and

$$E\,[u_{i,t} j\; y_{i,t}, y_{i,t-1}, y_{i,t-2}, :::; d_{i,t-1}, d_{i,t-2}, d_{i,t-3}, :::] = 0$$

Because of bi-directionality, $d_{i,t}$ cannot be uncorrelated with $e_{i,t}$: a shock to $y_{i,t}$ will affect $d_{i,t}$ because of the simultaneity. The same is true for $y_{i,t}$ and $u_{i,t}$. I show below that with a larger minimum number of observations per unit (T 4), sequential exogeneity can be relaxed and the coefficients of the following system of dynamic simultaneous equations:

$$Dy_{i,t} = {}_1 Dd_{i,t} + {}_1 Dy_{i,t-1} + De_{i,t}$$

$$Dd_{i,t} = {}_2 Dy_{i,t} + {}_2 Dd_{i,t-1} + Du_{i,t}$$

can be estimated consistently with the moment conditions implied by lagged levels of $y_{i,t}$ and $d_{i,t}$ under certain assumptions.

With $T = 4$, I can estimate the system of equations above under two different assumptions on the error terms $e_{i,t}$ and $u_{i,t}$. The first, I refer to as Assumption A, implies more moment conditions which is preferred with limited data, the other Assumption B is less restrictive and the assumptions required for consistent estimation allow for more flexible relationships between the two variables across time. With $T > 4$, the assumptions can be relaxed even further; I discuss this briefly in appendix Section D.2.3.

D.2.1 Assumption A

The system of simultaneous dynamic equations can be estimated under the following somewhat restrictive assumptions on the error terms:

$$E\,[e_{i,t}\; j\; y_{i,t-2}, y_{i,t-3}, :::; d_{i,t-1}, d_{i,t-2}, :::] = 0$$

and

$$E\,[u_{i,t}\; j\; y_{i,t-1}, y_{i,t-2}, :::; d_{i,t-2}, d_{i,t-3}, :::] = 0$$

Proposition . If Assumption A holds, then it implies the following moment condition that identifies the system of equations (2.1) and (2.2) above:

$$E \quad Z^{A\,0}_{i,t} DU_{i,t} \quad = 0$$

Figure D1: Dynamic process allowed by Assumption A rules out direct effects from $y_{i,t-1}$ ($d_{i,t-1}$) on $d_{i,t}$ ($y_{i,t}$).

where

$$Z^{A}_{i,t} = \begin{pmatrix} d_{i,t-2} & d_{i,t-3} & y_{i,t-3} & 0 & 0 & 0 \\ 0 & 0 & 0 & d_{i,t-3} & y_{i,t-2} & y_{i,t-3} \end{pmatrix}$$

and $U_{i,t}$ is a vector of the unobserved error terms $e_{i,t}$ and $u_{i,t}$.

Under Assumption A and with $T = 4$, the lags provide six moment conditions to identify four coefficients in the system of equations. For Assumption A to hold, the $e_{i,t}$ may not be correlated with $y_{i,t-2}$ and $d_{i,t-1}$ and further lags of these variables, and $u_{i,t}$ may not be correlated with $y_{i,t-1}$ and $d_{i,t-2}$ and further lags of these variables. This is a weaker condition than sequential exogeneity, but estimation requires a larger minimum T. The simultaneity of the equations also inherently implies that both error terms cannot be serially correlated.[30] After controlling for state dependence through the lagged dependent variable, individual fixed effects, and observable time varying characteristics, the remaining unobserved errors may not be correlated across time. Effectively, this assumption means that a shock to income in one period can affect income next period through state dependence, but it cannot affect the likelihood of shocks in the next period. Similarly for shocks to psychological well-being.

What does Assumption A imply on the relationship between y_i, d_i, and their lags? Figure A1 below shows the relationships that are allowed under assumption A. It rules out direct effects from $y_{i,t-1}$ ($d_{i,t-1}$) on $d_{i,t}$ ($y_{i,t}$) the lag of the simultaneous variable can only affect the variable of interest through the lagged dependent and the current simultaneous variable.

The time between waves and the way data is collected may make this assumption more plausible. Monthly income data for an individual is unlikely to satisfy these conditions. For one, serial correlations in shocks to income will violate these assumptions. However, data on income in the last month for an individual collected in yearly intervals more plausibly

[30] The error terms may be correlated with each other within the same time period.

satis es these assumptions.

To show that Assumption A implies the moment condition, assume $z^1_{i,t} = \begin{bmatrix} d_{i,t-2} & d_{i,t-3} & y_{i,t-3} \end{bmatrix}$ and $z^2_{i,t} = \begin{bmatrix} d_{i,t-3} & y_{i,t-2} & y_{i,t-3} \end{bmatrix}$. I expand the left hand side of the equation below:

$$E\left[\begin{pmatrix} z^{1\prime}_{i,t} & 0 \\ 0 & z^{2\prime}_{i,t} \end{pmatrix}\begin{pmatrix} \Delta e_{i,t} \\ \Delta u_{i,t} \end{pmatrix}\right] = E\left[\begin{pmatrix} z^{1\prime}_{i,t}\Delta e_{i,t} \\ z^{2\prime}_{i,t}\Delta u_{i,t} \end{pmatrix}\right] = \left[\begin{pmatrix} E[z^{1\prime}_{i,t}e_{i,t} - z^{1\prime}_{i,t}e_{i,t-1}] \\ E[z^{2\prime}_{i,t}u_{i,t} - z^{2\prime}_{i,t}u_{i,t-1}] \end{pmatrix}\right]$$

Distributing further and applying the law of iterated expectations gives:

$$E\left[\begin{pmatrix} d_{i,t-2}E[e_{i,t}\mid d_{i,t-2}] - d_{i,t-2}E[e_{i,t-1}\mid d_{i,t-2}] \\ d_{i,t-3}E[e_{i,t}\mid d_{i,t-3}] - d_{i,t-3}E[e_{i,t-1}\mid d_{i,t-3}] \\ y_{i,t-3}E[e_{i,t}\mid y_{i,t-3}] - y_{i,t-3}E[e_{i,t-1}\mid y_{i,t-3}] \\ d_{i,t-3}E[u_{i,t}\mid d_{i,t-3}] - d_{i,t-3}E[u_{i,t-1}\mid d_{i,t-3}] \\ y_{i,t-2}E[u_{i,t}\mid y_{i,t-2}] - y_{i,t-2}E[u_{i,t-1}\mid y_{i,t-2}] \\ y_{i,t-3}E[u_{i,t}\mid y_{i,t-3}] - y_{i,t-3}E[u_{i,t-1}\mid y_{i,t-3}] \end{pmatrix}\right] = 0$$

It is clear that under Assumption A, each term in the vector above would be equal to zero. Furthermore, with six moment conditions and four coe cients in the system of equations we are overidenti ed. A test of overidentifying restrictions can provide information on the validity of these lagged levels as instruments for the rst di erences. Simulation results (shown in Section D.2.3) verify that under an error structure and dynamic process that satis es assumption A, using a two-step GMM and instruments matrix $Z^A_{i,t}$ leads to consistent estimates of the coe cients ${}_1$, ${}_2$, ${}_1$, and ${}_2$.

D.2.2 Assumption B

With $T = 4$, the coe cients of the system of dynamic simultaneous equations can also be identi ed under the following less restrictive assumptions:

$$E[e_{i,t} \mid y_{i,t-2}, y_{i,t-3}, \ldots; d_{i,t-2}, d_{i,t-3}, \ldots] = 0$$

and

$$E[u_{i,t} \mid y_{i,t-2}, y_{i,t-3}, \ldots; d_{i,t-2}, d_{i,t-3}, \ldots] = 0$$

Proposition . If Assumption B holds, then it implies the following moment condition that identi es the system of equations (2.1) and (2.2) above:

$$E\left[Z^{B\prime}_{i,t}\Delta U_{i,t}\right] = 0$$

where

$$Z^B_{i,t} = \begin{pmatrix} d_{i,t-3} & y_{i,t-3} & 0 & 0 \\ 0 & 0 & d_{i,t-3} & y_{i,t-3} \end{pmatrix}$$

In addition, $Z^B_{i,t}$ also provides moment conditions to identify the coefficients under the more restrictive set of Assumptions A.

The proof follows the same logic as the proof for $Z^A_{i,t}$. Expanding $E\left[Z^{B\prime}_{i,t}\Delta U_{i,t}\right]$ and applying the law of iterated expectations gives the following:

$$E\left[\begin{pmatrix} d_{i,t-3}E[e_{i,t}\,|\,d_{i,t-3}] - d_{i,t-3}E[e_{i,t-1}|d_{i,t-3}] \\ y_{i,t-3}E[e_{i,t}\,|\,y_{i,t-3}] - y_{i,t-3}E[e_{i,t-1}|y_{i,t-3}] \\ d_{i,t-3}E[u_{i,t}|d_{i,t-3}] - d_{i,t-3}E[u_{i,t-1}|d_{i,t-3}] \\ y_{i,t-3}E[u_{i,t}|y_{i,t-3}] - y_{i,t-3}E[u_{i,t-1}|y_{i,t-3}] \end{pmatrix}\right] = 0$$

which holds under both Assumption A and B. Four moments allow for the estimation of the four coefficients of interest and we are just identified.[31]

Under Assumption B, the error terms may be first-order moving-average serially correlated.[32] Assuming this type of first-order serial correlation is common in the literature on yearly income dynamics and state dependence of income and employment (Guvenen, 2007; Meghir and Pistaferri, 2004; Magnac, 2000).[33] Moreover, the error terms $e_{i,t}$ and $u_{i,t}$ may be correlated with $u_{i,t-1}$ and $e_{i,t-1}$, respectively. Under assumption B, twice lagged levels (and further) can only affect y_t and d_t through y_{t-1} and d_{t-1}. A visual representation of the implied dynamic processes under Assumptions B is shown in Figure A2.

Throughout the paper, I show, where appropriate, estimates based on both Assumptions A and B. If the results are not different under Assumption B, this would indicate that they are robust to first-order serial correlation. To test if the results under the two sets of assumptions are different, a Hausman-type test should be conducted. Neither regression gives results that are efficient, thus, when testing for the statistical significance of the difference of the estimates, the variance of the difference can be estimated using a bootstrap. If the difference is not statistically significantly different from zero, this would suggest, albeit indirectly, that the error terms are not strongly serially correlated. This is under the assumption that the dataset has four waves of data ($T = 4$); after taking the first difference and using lagged levels $t-2$ and $t-3$ as instruments, this effectively means that

[31] Removing the lagged $t-2$ level variables from the matrix of instruments allows for less restrictive assumptions on the error terms.

[32] In a moving-average serial correlation, $e_{i,t}$ is correlated with $e_{i,t-1}$ but this correlation breaks down with $e_{i,t-2}$. An autoregressive serial correlation implies a decaying correlation such that $e_{i,t}$ and $e_{i,t-2}$ are still correlated. This would violate the assumptions needed to for this approach (and typical Arellano-Bond approaches) to be consistent and no lags would be proper instruments.

[33] The time periods considered are often of higher frequency. The time between each wave in this paper is two years — and income is reported for the past month — making serial correlation such as this less likely.

Figure D2: Dynamic process allowed by Assumption B. Assumptions A and B imply slightly different dynamic processes. Under the less restrictive assumption B, the assumptions do not allow for direct effects across 2 time periods. Income and psychological well-being in time $t-2$ can only affect income and psychological well-being in time t through income and psychological well-being in time $t-1$.

we have one observation per individual. This means that we cannot directly test for serial correlation which is an important and obvious violation of Assumption A. When using $Z^A_{i,t}$, if a test of overidentifying restrictions rejects the validity of the instruments, it would be evidence against the dynamic process assumed and potential serial correlation; however, the validity of the instruments is not rejected in any of the results presented in the rest of the paper. Using additional data from Wave 5 and methods from Blundell and Bond (1998) to estimate the two equations independently, I conduct direct tests of serial correlation of the error terms. I do not find evidence for first-order serial correlation in either equation.[34]

D.2.3 Assumptions when $T > 4$

With additional waves ($T > 4$), we can test directly for serial correlation, the weaker Assumption B would be over-identified, we have more power, and would be able to identify the system under even weaker assumptions that allow for additional orders of moving average serial correlation. I illustrate some these properties using simulations in the next section.

Suppose $T > 4$. The weakest assumption we can impose and be over-identified in estimated a system of dynamic simultaneous equations is:

$$E\left[e_{i,t} \mid y_{i,t-(T-2)}, y_{i,t-(T-1)}, \ldots; d_{i,t-(T-2)}, d_{i,t-(T-1)}, \ldots\right] = 0$$

and

$$E\left[u_{i,t} \mid y_{i,t-(T-2)}, y_{i,t-(T-1)}, \ldots; d_{i,t-(T-2)}, d_{i,t-(T-1)}, \ldots\right] = 0$$

and just identified:

$$E\left[e_{i,t} \mid y_{i,t-(T-1)}, \ldots; d_{i,t-(T-1)}, \ldots\right] = 0$$

[34] See Appendix Tables A4 and A5.

and

$$E\left[u_{i,t} \mid y_{i,t-(T-1)},\ldots;d_{i,t-(T-1)},\ldots\right] = 0$$

However, in most circumstances, it is unnecessary to relax the assumptions to this level. With increasing T, these lags case will likely suffer from weak instruments problems. With assumptions similar to A or B, we can estimate the system with more observations (added observations per individual).

With additional assumptions on initial conditions and stationarity, conditional and time series heteroskedasticity, and correlation between the error terms and the individual fixed effects (or lack thereof), additional moment conditions would improve the efficiency of this estimation approach.

D.3 Potential Identification Threats and Robustness Checks

While panel data methods are used extensively to estimate the state-dependence of income, they are not very common in empirical microeconomics analyses such as this one. This approach comes with strong assumptions regarding the serial correlation of shocks within and across equations. The two-year spacing of the waves and the recall time-frame of the main variables (one month for income and one week for CES-D) make the assumptions discussed above more plausible. Below, I discuss threats to identification in the context of this study. While we cannot entirely rule out long persistence of shocks, I outline an alternative approach (regression discontinuity) that I use to show the robustness of some of the results.

The panel GMM approach outlined above is an instrumental variable approach and thus deals with most omitted variable bias under the assumptions specified above. For example, if a person becomes chronically ill, this will likely affect both psychological well-being and income (a common shock). However, this will not lead to biased results unless the chronic illness was the result of a lagged shock to income or psychological well-being. Suppose a shock to income in past waves affects an omitted variable which then, in turn, affects income and psychological well-being independently in the future in a way that is not captured by the lagged terms; this would violate the assumptions.[35] For example, if a shock to income in Waves 1 or 2 makes divorce more likely in later waves, divorce could lead to income and psychological well-being shocks in the future (Charles and Stephens, 2004). If the income shock increases the likelihood of divorce for more than four years post shock, then past income can affect current income and psychological well-being in a way that is not captured by the model thus violating Assumption B the weaker of the two assumptions.

[35] Shocks such as this one that happen prior to the first wave that make shocks to income and psychological well-being more likely would be captured in the individual fixed effects which are controlled for in this method.

Under Assumption B, shocks can be correlated within each variable and across the two for up to one time period but not more. Examples of such variables that could change due to a past shock that could lead to future common shocks include but are not restricted to having children, a disability/chronic illness, a death in the family, etc... While in the main regressions, I do control for changes in some concurrent time-varying individual and household characteristics, I do not account for changes in these variables in past waves. To see if the results are robust to such changes within the panel data approach, I am able to do two things: first, control for changes in these variables several lags into the past. Second, I can restrict the sample to those who did not experience changes in these variables throughout the four waves. Neither of these two approaches leads to noteworthy changes in the main results.

There are examples of other channels that I cannot control for. If a past income shock (in Wave 1 or 2) creates financial worries that increase the likelihood of accidents or bad financial decision-making because of inattention or lowered levels of cognitive and executive function and this is not captured by lagged dependents and persists for four years, this would again violate the conditions required for consistent estimators.

Lastly, it may be the case that the dynamics are mis-specified: it could well be that another $t-2$ lagged term is important in directly determining income and psychological well-being in time t. This would be a violation of the weaker of the two assumptions. Given the spacing of the waves in this setting (2 years), this is not likely an important issue. However, the panel is not long enough to test this assumption directly.

I further test the robustness of the one side of results by using a local randomization regression discontinuity approach to estimate the effect of income on psychological well-being—this is discussed and the results are shown in Appendix C.[36] I leverage the discontinuity of eligibility for the Older Person's Grant by restricting the sample to individuals in households with economically inactive members in narrow windows around the age of 60 and using the eligibility for Older Person's Grant (age 60 or above) as an instrument for household income.[37] The required assumptions for consistent estimators in this approach are different from those required using the panel data method. However, I find that the estimated effects of changes in household income are very similar to those estimated using the GMM approach.[38] As is evident from the literature, it is difficult to find a valid instrument

[36]The panel data approach and this approach are both instrumental variable approaches and thus remove concern related to measurement error of our variables of interest (Björn, 2000).

[37]This is the same empirical method used in Alloush and Wu (2023) to estimate the effect of income on life satisfaction and is discussed in detail there. More discussion on local randomization approaches in regression discontinuity designs can be found in Cattaneo, Idrobo and Titiunik (forthcoming).

[38]This is similar in spirit to Currie and Tekin (2012) who use several approaches to study the effect of childhood maltreatment—a link that is unlikely to be studied experimentally. They argue that approaches requiring different assumptions lead to qualitatively similar results give confidence in the core results.

for psychological well-being; several instruments have been proposed however their validity is debated. I discuss this more thoroughly in Appendix C. Thus I am unable to conduct a similar exercise for the effect of psychological distress on income. However, the uniformity of the estimates on income gives credence to the core results.

D.4 Monte Carlo Simulations

In this section, I use the following set of simultaneous equations:

$$y_{i,t} = -0.5d_{i,t} + 0.7y_{i,t-1} + \nu_i + e_{i,t}$$

$$d_{i,t} = -0.45y_{i,t} - 0.35d_{i,t-1} + \rho_i + u_{i,t}$$

to conduct a Monte Carlo simulation to illustrate the consistency of this econometric approach. $y_{i,0}$ and $d_{i,0}$ are independently drawn from a $N(0,1)$ distribution. ν_i and ρ_i are individual fixed effects randomly and independently drawn from a uniform[-1,1] distribution. From these initial conditions, I use the dynamic equations above to generate four rounds of data where the error terms are constructed as such:

$$e_{i,t} = \lambda_e z^e_{i,t-1} + \zeta z^u_{i,t-1} + z^e_{i,t}$$

$$u_{i,t} = \lambda_u z^u_{i,t-1} + \zeta z^e_{i,t-1} + z^u_{i,t}$$

where $z^e_{i,t}$ and $z^u_{i,t}$ are randomly drawn from an $N(0,1)$ distribution and with a correlation of -0.25. This defines the within-period covariance between $e_{i,t}$ and $u_{i,t}$. λ contributes to the moving-average serial correlation in error terms and ζ is the correlation between $e_{i,t}$ and $u_{i,t\pm1}$.

I use the data ($T = 4$) created under the above data generating process to estimate the coefficients of the following equations:

$$\mathrm{D}y_{i,t} = \beta_1 \mathrm{D}d_{i,t} + \gamma_1 \mathrm{D}y_{i,t-1} + \mathrm{D}e_{i,t}$$

$$\mathrm{D}d_{i,t} = \beta_2 \mathrm{D}y_{i,t} + \gamma_2 \mathrm{D}d_{i,t-1} + \mathrm{D}u_{i,t}$$

using a two step GMM and two different instrument matrices $Z^A_{i,t}$ and $Z^B_{i,t}$ where

$$Z^A_{i,t} = \begin{pmatrix} d_{i,t-2} & d_{i,t-3} & y_{i,t-3} & 0 & 0 & 0 \\ 0 & 0 & 0 & d_{i,t-3} & y_{i,t-2} & y_{i,t-3} \end{pmatrix}$$

and

$$Z_{i,t}^{B} = \begin{pmatrix} d_{i,t-3} & y_{i,t-3} & 0 & 0 \\ 0 & 0 & d_{i,t-3} & y_{i,t-3} \end{pmatrix}$$

Below, I compare the consistency of using these two instrument matrices under different conditions.

D.4.1 No Serial Correlation

Correlation between error terms in the same t (α) does not affect consistency and in the throughout all the simulation results below I assume that $\alpha = -0.4$. Table D1 shows summary results when there is no serial correlation in the error terms ($\lambda_e = 0$ and $\lambda_u = 0$), and $\zeta = 0$. Using either instruments matrix $Z_{i,t}^{A}$ or $Z_{i,t}^{B}$ is consistent and the bias and standard deviation of the Monte Carlo estimates of the four coefficients of interest are decreasing with N. It is evident (in the top half of the table) that as N increases, bias and variance of the estimated coefficients is decreasing and going to 0—bias and ssd in the table of each coefficient is decreasing with N.

We have two measures of the standard error of the estimates: by taking the standard deviation of the simulated coefficients and by averaging the standard error of the estimate. We see in column 4 that the se/ssd ratio is inching towards 1 with larger sample sizes, suggesting that we have a consistent estimator of the finite-sample variance.[39] The simulation results suggest that the test is consistent against fixed alternatives as the probability of rejecting the null of 0 with a test of size 0.05 and 0.1 goes to 1 as N increases. Under these conditions, using $Z_{i,t}^{A}$ strictly dominates in terms of power rejecting the null of zero with smaller N.

In the bottom half of Table D1, I show results of using this estimation approach when there is no simultaneity: $\beta_1 = 0$ and $\beta_2 = 0$. The estimation approach is consistent under these conditions with either instruments matrix. Again, using matrix $Z_{i,t}^{A}$ that has more instruments has more power in rejecting the null for the lagged dependent terms. Importantly, for the coefficients that are equal to zero, this approach rejects the null of zero at about the size of the test (0.05 and 0.10).

Table D2 shows results for similar conditions with respect to serial correlation, only the random variables and the unobserved error/shock terms are calibrated to look more like the data for income and psychological well-being in South Africa. We can see that the estimation approach under the condition of no serial correlation is consistent and the tests are well-behaved.

[39] In these simulations, I do not assume that the process is stationary; the observations of y and d are not necessarily in their long-run equilibrium means and standard deviations. With stationarity, these ratios are close to 1 even in small sample sizes.

D.4.2 First-Order Serial Correlation

In Table D3, I add first-order moving-average serial correlation to both error terms. I do so by setting $\lambda = 0.3$. Under these conditions, only estimation using $Z^B_{i,t}$ will be consistent. In Table D3, we can see that the bias when using $Z^A_{i,t}$ does is not decreasing to zero with N. On the other hand, using $Z^B_{i,t}$ is consistent and the bias and the standard deviation of the Monte Carlo estimates of the coefficients of the system of equations is decreasing to zero with increasing N.

In the lower half of Table D3, we can see that under conditions of first-order serial correlation, even if the true data generating process has no simultaneity (β_1 and β_2 are zero) using $Z^A_{i,t}$ will be not be consistent. This is due to the within period correlation between the two error terms.[40] Importantly, when using $Z^B_{i,t}$ the test is well-behaved and this approach rejects the null at about the size of the test.

D.4.3 Cross-Error Serial Correlation

The results in Table D4 show that if, in addition to first-order serial correlation, ζ which effectively introduces correlation across the two error terms over time is not equal to zero, then using $Z^A_{i,t}$ will not be consistent even when there is no simultaneous causality (β_1 and β_2 are zero).

Using matrix $Z^B_{i,t}$, however, is consistent and the estimated coefficients have bias and standard deviation of the estimates that go to zero as N increases.

D.4.4 Higher Order Serial Correlation and $T > 4$

Moving-average serial correlation of an order more than one would create bias with the two approaches suggested here when $T = 4$. This is not shown in the simulations but is obvious from both assumption A and B. In this case, having $T > 4$ would allow us to estimate the system. Table D5 shows results when the moving average serial correlation is of order 2, there is cross correlation in error terms both concurrently and over 1 time period. Having further lags allows for consistent estimators by starting with further lags.

D.5 Conclusion

This appendix provides an extension of dynamic panel data methods which shows that with more observations per individual—at least four, a system of simultaneous dynamic equations can be estimated using lagged levels as instruments for first differences. I show that with $T =$

[40] If $\alpha = 0$ and there is first-order moving-average serial correlation, using $Z^A_{i,t}$ will be consistent under conditions of no simultaneity. Under these conditions, despite the bias of $Z^A_{i,t}$, it can still be useful in showing that there is simultaneity.

4, two approaches can be consistent depending on the data generating process. Through proofs and simulations, I show that, overall, using $Z^B_{i,t}$ is more robust to different conditions, however, under conditions of no serial correlation in shocks, the approach that uses $Z^A_{i,t}$ is consistent and more powerful. Even with if first-order moving-average serial correlation is suspected, the simulations show that $Z^A_{i,t}$ can be useful in informing us about the existence of simultaneous bi-directional causality.

Table D1: Simulation Results 1 – No Serial Correlation

$\lambda_e = 0, \lambda_u = 0, \alpha = -0.4, \zeta = 0$											
		$Z^A_{i,t}$					$Z^B_{i,t}$				
	N	**bias**	**ssd**	$\frac{sse}{ssd}$	**p5**	**p10**	**bias**	**ssd**	$\frac{sse}{ssd}$	**p5**	**p10**
Main Equation											
$\beta_1 = -0.5$	500	0.010	0.173	0.952	0.829	0.877	0.114	4.140	1.666	0.681	0.722
	1,000	0.006	0.122	0.949	0.967	0.980	0.236	6.761	0.753	0.755	0.800
	2,000	0.002	0.080	1.017	0.999	1.000	0.002	0.134	0.954	0.910	0.935
$\gamma_1 = 0.7$	500	-0.019	0.214	0.970	0.921	0.956	0.158	7.055	1.652	0.364	0.536
	1,000	-0.009	0.147	0.992	0.998	1.000	0.348	10.241	0.774	0.776	0.865
	2,000	-0.004	0.101	1.019	1.000	1.000	0.000	0.225	0.896	0.988	0.995
$\beta_2 = -0.45$	500	-0.014	0.470	0.625	0.584	0.648	-0.034	0.720	0.884	0.430	0.495
	1,000	-0.013	0.200	0.938	0.718	0.784	-0.001	0.369	0.858	0.516	0.595
	2,000	-0.001	0.138	0.960	0.859	0.897	0.009	0.216	0.949	0.630	0.700
$\gamma_2 = -0.35$	500	0.000	0.050	1.020	0.989	0.993	0.017	0.204	0.930	0.621	0.720
	1,000	-0.001	0.035	0.970	1.000	1.000	0.002	0.127	0.848	0.907	0.938
	2,000	-0.001	0.024	0.995	1.000	1.000	-0.001	0.074	0.954	0.997	0.997
No Simultaneity											
$\beta_1 = 0$	500	0.008	0.131	0.986	0.047	0.093	-0.004	0.860	0.906	0.045	0.075
	1,000	0.006	0.090	1.013	0.040	0.090	0.037	0.376	0.853	0.040	0.083
	2,000	0.002	0.064	0.992	0.051	0.101	0.004	0.210	0.948	0.047	0.089
$\gamma_1 = 0.7$	500	-0.040	0.247	1.027	0.871	0.943	-0.021	0.396	1.051	0.613	0.742
	1,000	-0.015	0.179	1.003	0.999	1.000	-0.005	0.214	1.022	0.954	0.978
	2,000	-0.008	0.127	0.999	1.000	1.000	-0.004	0.141	0.987	1.000	1.000
$\beta_2 = 0$	500	-0.001	0.259	0.954	0.056	0.114	-0.008	0.485	0.867	0.060	0.107
	1,000	0.002	0.177	0.976	0.047	0.103	0.005	0.303	0.906	0.055	0.099
	2,000	0.004	0.122	0.996	0.040	0.095	0.009	0.198	0.970	0.053	0.096
$\gamma_2 = -0.35$	500	-0.006	0.120	0.986	0.809	0.869	0.014	0.151	0.918	0.660	0.743
	1,000	-0.004	0.084	0.994	0.962	0.976	0.005	0.100	0.947	0.909	0.946
	2,000	-0.001	0.062	0.953	0.999	1.000	0.002	0.070	0.952	0.992	0.996

Results are based on simulations using 1,000 replications.
N: Number of observations
bias: Bias of the estimated coefficient
ssd: Standard Deviation of the estimated coefficients
$\frac{sse}{ssd}$: Mean of the ratio of the estimated standard error to **ssd**
p5,p10: Probability of rejecting null where coefficient equals 0 with test of size 5, 10.

Table D2: Simulation Results 2 – Calibrated Data

$\lambda_e = 0, \lambda_u = 0, \alpha = -0.4, \zeta = 0$											
		$Z^A_{i,t}$					$Z^B_{i,t}$				
	N	**bias**	**ssd**	$\frac{sse}{ssd}$	**p5**	**p10**	**bias**	**ssd**	$\frac{sse}{ssd}$	**p5**	**p10**
Main Equation											
$\beta_1 = -0.5$	500	-0.029	0.257	0.997	0.607	0.688	-0.033	0.539	0.938	0.474	0.541
	1,000	-0.015	0.185	0.978	0.758	0.833	0.020	0.480	0.761	0.506	0.573
	2,000	-0.007	0.127	1.006	0.944	0.968	-0.005	0.231	0.965	0.669	0.738
$\gamma_1 = 0.7$	500	-0.009	0.117	0.989	1.000	1.000	-0.016	0.211	0.976	0.945	0.964
	1,000	-0.005	0.080	1.013	1.000	1.000	0.005	0.170	0.864	0.988	0.993
	2,000	-0.003	0.057	1.011	1.000	1.000	-0.006	0.094	0.968	1.000	1.000
$\beta_2 = -0.45$	500	-0.003	0.055	0.916	1.000	1.000	-0.002	0.055	0.951	0.998	0.999
	1,000	-0.002	0.035	1.012	1.000	1.000	-0.001	0.035	1.032	1.000	1.000
	2,000	0.000	0.026	0.981	1.000	1.000	0.000	0.026	0.989	1.000	1.000
$\gamma_2 = -0.35$	500	-0.003	0.036	1.022	1.000	1.000	0.000	0.063	0.923	0.992	0.994
	1,000	-0.001	0.026	1.009	1.000	1.000	0.001	0.042	0.963	0.999	0.999
	2,000	0.000	0.018	1.000	1.000	1.000	0.001	0.029	0.994	1.000	1.000
No Simultaneity											
$\beta_1 = 0$	500	0.013	0.415	0.979	0.048	0.095	-0.054	2.082	0.832	0.040	0.067
	1,000	0.013	0.285	1.003	0.047	0.091	0.075	1.067	0.867	0.038	0.080
	2,000	0.005	0.204	0.986	0.051	0.105	0.001	0.625	0.956	0.043	0.089
$\gamma_1 = 0.7$	500	-0.010	0.143	0.992	1.000	1.000	-0.009	0.197	0.989	0.951	0.962
	1,000	-0.004	0.100	0.995	1.000	1.000	-0.001	0.109	1.033	0.996	0.997
	2,000	-0.005	0.069	1.007	1.000	1.000	-0.001	0.072	1.019	1.000	1.000
$\beta_2 = 0$	500	-0.001	0.057	0.946	0.055	0.101	-0.003	0.070	0.928	0.059	0.112
	1,000	0.000	0.038	0.987	0.037	0.086	0.000	0.044	1.009	0.040	0.072
	2,000	0.001	0.027	1.002	0.042	0.090	0.001	0.032	0.992	0.044	0.090
$\gamma_2 = -0.35$	500	-0.007	0.116	0.985	0.829	0.883	0.008	0.135	0.918	0.751	0.811
	1,000	-0.003	0.083	0.975	0.963	0.979	0.003	0.091	0.959	0.944	0.968
	2,000	0.001	0.061	0.945	1.000	1.000	0.003	0.066	0.947	0.997	1.000

Results are based on simulations using 1,000 replications.
N: Number of observations
bias: Bias of the estimated coefficient
ssd: Standard Deviation of the estimated coefficients
$\frac{sse}{ssd}$: Mean of the ratio of the estimated standard error to **ssd**
p5,p10: Probability of rejecting null where coefficient equals 0 with test of size 5, 10.

Table D3: Simulation Results 3 – First-order Serial Correlation

	N	bias	ssd	$\frac{sse}{ssd}$	p5	p10	bias	ssd	$\frac{sse}{ssd}$	p5	p10
$\lambda_e = 0.3, \lambda_u = 0.3, \alpha = -0.4, \zeta = 0$											
		$Z_{i,t}^A$					$Z_{i,t}^B$				
Main Equation											
$\beta_1 = -0.5$	500	0.142	0.976	0.599	0.349	0.393	0.009	1.556	0.592	0.451	0.508
	1,000	0.341	0.891	0.532	0.275	0.319	0.100	2.680	0.430	0.483	0.560
	2,000	0.364	0.462	0.645	0.256	0.297	-0.007	0.274	0.917	0.603	0.686
$\gamma_1 = 0.7$	500	0.010	0.347	0.608	0.927	0.944	-0.033	0.847	0.845	0.941	0.958
	1,000	0.088	0.313	0.550	0.992	0.994	0.027	0.751	0.478	0.985	0.989
	2,000	0.093	0.154	0.694	1.000	1.000	-0.005	0.097	0.929	0.998	1.000
$\beta_2 = -0.45$	500	-0.006	0.060	0.824	1.000	1.000	-0.002	0.052	0.960	1.000	1.000
	1,000	-0.002	0.038	0.898	1.000	1.000	0.000	0.033	1.050	1.000	1.000
	2,000	0.002	0.028	0.851	1.000	1.000	0.000	0.025	0.981	1.000	1.000
$\gamma_2 = -0.35$	500	0.045	0.087	0.594	0.947	0.954	0.002	0.072	0.929	0.979	0.986
	1,000	0.029	0.048	0.761	0.995	0.995	0.001	0.048	0.977	0.998	0.999
	2,000	0.024	0.028	0.939	1.000	1.000	0.001	0.032	1.004	1.000	1.000
No Simultaneity											
$\beta_1 = 0$	500	0.475	0.708	0.952	0.066	0.159	0.058	4.088	0.805	0.035	0.074
	1,000	0.478	0.476	0.984	0.134	0.239	-0.023	4.796	0.736	0.025	0.077
	2,000	0.475	0.338	0.969	0.277	0.410	-0.024	1.067	0.863	0.033	0.076
$\gamma_1 = 0.7$	500	-0.001	0.125	1.019	1.000	1.000	-0.013	0.199	1.067	0.913	0.936
	1,000	0.006	0.087	1.010	1.000	1.000	0.003	0.237	0.863	0.985	0.990
	2,000	0.005	0.061	1.005	1.000	1.000	-0.003	0.064	1.034	0.999	0.999
$\beta_2 = 0$	500	-0.018	0.048	0.956	0.085	0.146	-0.003	0.061	0.950	0.057	0.105
	1,000	-0.017	0.033	0.974	0.092	0.152	0.000	0.039	1.026	0.031	0.072
	2,000	-0.016	0.023	0.992	0.128	0.214	0.001	0.028	0.980	0.040	0.101
$\gamma_2 = -0.35$	500	0.032	0.170	0.954	0.551	0.633	0.015	0.194	0.914	0.510	0.615
	1,000	0.035	0.125	0.924	0.752	0.828	0.007	0.138	0.904	0.749	0.833
	2,000	0.044	0.089	0.917	0.916	0.953	0.005	0.095	0.937	0.928	0.951

Results are based on simulations using 1,000 replications.
N: Number of observations
bias: Bias of the estimated coefficient
ssd: Standard Deviation of the estimated coefficients
$\frac{sse}{ssd}$: Mean of the ratio of the estimated standard error to **ssd**
p5,p10: Probability of rejecting null where coefficient equals 0 with test of size 5, 10.

Table D4: Simulation Results 4 – First Order Serial Correlation and Cross Correlation

	$\lambda_e = 0.3, \lambda_u = 0.3, \alpha = -0.4, \zeta = -0.2$										
		$Z^A_{i,t}$					$Z^B_{i,t}$				
	N	**bias**	**ssd**	$\frac{sse}{ssd}$	**p5**	**p10**	**bias**	**ssd**	$\frac{sse}{ssd}$	**p5**	**p10**
		Main Equation									
$\beta_1 = -0.5$	500	-1.563	0.434	0.590	0.993	0.996	-0.049	1.884	0.734	0.422	0.490
	1,000	-1.643	0.324	0.565	0.999	0.999	0.015	0.640	0.848	0.412	0.488
	2,000	-1.705	0.222	0.595	1.000	1.000	-0.007	0.370	0.892	0.518	0.595
$\gamma_1 = 0.7$	500	-0.550	0.164	0.451	0.529	0.609	-0.007	0.599	0.683	0.913	0.936
	1,000	-0.572	0.117	0.442	0.613	0.672	0.002	0.167	0.890	0.977	0.982
	2,000	-0.588	0.082	0.443	0.681	0.737	-0.004	0.103	0.879	0.998	0.999
$\beta_2 = -0.45$	500	-0.048	0.102	0.671	0.989	0.994	0.001	0.076	0.935	0.983	0.990
	1,000	-0.035	0.076	0.642	0.999	0.999	-0.001	0.047	1.026	1.000	1.000
	2,000	-0.024	0.054	0.645	1.000	1.000	0.001	0.034	0.992	1.000	1.000
$\gamma_2 = -0.35$	500	0.327	0.112	0.506	0.328	0.415	0.009	0.102	0.915	0.901	0.920
	1,000	0.346	0.079	0.459	0.363	0.451	0.001	0.067	0.976	0.993	0.997
	2,000	0.361	0.055	0.434	0.429	0.507	0.001	0.046	1.000	1.000	1.000
		No Simultaneity									
$\beta_1 = 0$	500	0.841	0.768	1.025	0.104	0.231	-1.445	31.081	1.891	0.042	0.085
	1,000	0.886	0.389	1.308	0.380	0.587	0.002	3.108	0.624	0.034	0.085
	2,000	0.906	0.274	1.312	0.834	0.928	-0.027	1.313	0.752	0.030	0.079
$\gamma_1 = 0.7$	500	-0.129	0.172	0.860	0.959	0.974	0.080	1.543	1.677	0.840	0.878
	1,000	-0.133	0.095	1.004	0.998	0.999	-0.003	0.223	0.816	0.946	0.956
	2,000	-0.138	0.066	1.012	1.000	1.000	0.000	0.110	0.883	0.991	0.993
$\beta_2 = 0$	500	-0.300	0.064	0.886	1.000	1.000	0.001	0.095	0.965	0.085	0.133
	1,000	-0.300	0.046	0.858	1.000	1.000	-0.001	0.063	0.953	0.059	0.102
	2,000	-0.301	0.031	0.898	1.000	1.000	0.000	0.045	0.949	0.058	0.116
$\gamma_2 = -0.35$	500	-0.553	0.205	0.803	0.995	0.997	0.015	0.218	1.009	0.471	0.570
	1,000	-0.577	0.141	0.816	1.000	1.000	0.003	0.150	0.931	0.691	0.759
	2,000	-0.585	0.101	0.812	1.000	1.000	0.001	0.110	0.918	0.875	0.911

Results are based on simulations using 1,000 replications.
N: Number of observations
bias: Bias of the estimated coefficient
ssd: Standard Deviation of the estimated coefficients
$\frac{sse}{ssd}$: Mean of the ratio of the estimated standard error to **ssd**
p5,p10: Probability of rejecting null where coefficient equals 0 with test of size 5, 10.

Table D5: Simulation Results 5 – Second-order Serial Correlation

$\lambda_e = 0.3, \lambda_u = 0.3, \alpha = -0.4, \zeta = -0.2$											
		$Z^A_{i,t}$					$Z^B_{i,t}$				
	N	**bias**	**ssd**	$\frac{sse}{ssd}$	**p5**	**p10**	**bias**	**ssd**	$\frac{sse}{ssd}$	**p5**	**p10**
Main Equation											
$\beta_1 = -0.5$	500	-0.609	1.098	0.676	0.654	0.708	-0.488	7.068	1.675	0.606	0.667
	1,000	-0.279	1.154	0.689	0.572	0.624	-0.685	1.672	1.363	0.554	0.607
	2,000	-0.136	0.612	0.759	0.551	0.610	0.036	14.521	1.326	0.528	0.573
$\gamma_1 = 0.7$	500	-0.149	0.260	0.714	0.890	0.911	-0.165	1.514	1.506	0.551	0.616
	1,000	-0.070	0.274	0.732	0.937	0.947	-0.208	0.490	1.348	0.605	0.676
	2,000	-0.033	0.153	0.749	0.991	0.993	0.053	5.461	1.299	0.690	0.744
$\beta_2 = -0.45$	500	-0.007	0.105	0.933	0.934	0.950	-0.004	0.458	0.804	0.867	0.897
	1,000	0.002	0.078	0.882	0.972	0.979	-0.002	0.204	0.915	0.922	0.935
	2,000	-0.002	0.052	0.893	0.996	0.996	0.002	0.066	0.959	0.978	0.981
$\gamma_2 = -0.35$	500	0.008	0.179	0.823	0.730	0.779	0.009	1.103	0.827	0.442	0.506
	1,000	0.014	0.136	0.806	0.814	0.848	0.012	0.604	0.823	0.522	0.584
	2,000	-0.001	0.084	0.872	0.959	0.972	0.003	0.177	0.899	0.668	0.722
No Simultaneity											
$\beta_1 = 0$	500	-0.218	2.640	0.759	0.035	0.082	-1.089	18.930	0.810	0.034	0.063
	1,000	0.081	1.665	0.773	0.043	0.090	-0.249	5.834	1.093	0.036	0.072
	2,000	0.031	1.000	0.845	0.033	0.088	-0.093	6.188	1.101	0.039	0.076
$\gamma_1 = 0.7$	500	-0.043	0.313	1.032	0.776	0.853	0.000	1.326	1.009	0.471	0.575
	1,000	-0.027	0.166	1.106	0.965	0.977	-0.018	0.485	0.930	0.731	0.797
	2,000	-0.011	0.115	1.023	0.999	0.999	-0.021	0.255	1.264	0.872	0.896
$\beta_2 = 0$	500	-0.007	0.100	0.951	0.045	0.090	0.003	0.158	1.438	0.037	0.069
	1,000	-0.002	0.088	0.716	0.040	0.080	0.002	0.108	1.125	0.029	0.069
	2,000	-0.002	0.038	1.008	0.048	0.086	0.000	0.058	1.025	0.037	0.085
γ_2=-0.35	500	-0.167	0.593	0.927	0.354	0.436	-0.155	1.120	1.313	0.222	0.298
	1,000	-0.062	0.466	0.823	0.386	0.476	-0.006	0.669	1.037	0.275	0.356
	2,000	-0.034	0.259	0.964	0.490	0.568	0.004	0.340	0.927	0.369	0.475

Results are based on simulations using 1,000 replications.
N: Number of observations
bias: Bias of the estimated coefficient
ssd: Standard Deviation of the estimated coefficients
$\frac{sse}{ssd}$: Mean of the ratio of the estimated standard error to **ssd**
p5,p10: Probability of rejecting null where coefficient equals 0 with test of size 5, 10.

E. Impulse Response Function (for online publication only)

While in the system of equations used in this paper, I distinguish between household income and individual income, in this section, I abstract away from this distinction and I treat them as the same variable: however, I temper the effect of changes in CES-D on household income and reduce it to 0.54 times the estimated effect on individual income.[41] Moreover, since I am considering small marginal changes or shocks, I ignore the estimated quadratic term. The simplified system of equations is the following:

$$Dh_{i,t} = \alpha_1 DD_{i,t} + \beta_1 Dh_{i,t-1} + De_{i,t}$$

$$DD_{i,t} = a_1 Dh_{i,t} + b_1 DD_{i,t-1} + Du_{i,t}$$

I can represent these equations in the following matrix form:

$$AY_{i,t} = BY_{i,t-1} + \ _{i,t}$$

where $Y_{i,t} = \begin{pmatrix} Dh_{i,t} & DD_{i,t} \end{pmatrix}$, $\ _{i,t} = \begin{pmatrix} De_{i,t} & Du_{i,t} \end{pmatrix}$, $A = \begin{pmatrix} 1 & -\alpha_1 \\ -a_1 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} \beta_1 & 0 \\ 0 & b_1 \end{pmatrix}$, which can be rewritten as:

$$Y_{i,t} = A^{-1}BY_{i,t-1} + A^{-1} \ _{i,t}$$

A Wold decomposition of the above equation gives the following:

$$Y_{i,t} = \sum_{j=0}^{\infty} (A^{-1}B)^j A^{-1} \ _{i,t-j}$$

This decomposition allows me to look at the effects of shocks (in) on $Y_{i,t}$ over time. For example, an income shock of size 1 in time $t-j$ has the following effect on $Y_{i,t}$:

$$\frac{\delta Y_{i,t}}{\delta e_{i,t-j}} = (A^{-1}B)^j A^{-1} e_1$$

where $e_1 = \begin{pmatrix} 1 & 0 \end{pmatrix}$. Figure E1 shows the plot of the impulse response function of a negative shock to income over time. The dotted black line shows the impulse response for an AR(1) process that ignores psychological well-being.[42]. The solid blue line shows the

[41]This is the lower bound of the 95% confidence interval of the estimated effect of changes in individual income on household income per capita (this is calculated by estimating a system of equations that includes the effect of individual income on household income per capita.)

[42]The coefficient on lagged individual income is estimated from the data to be 0.71.

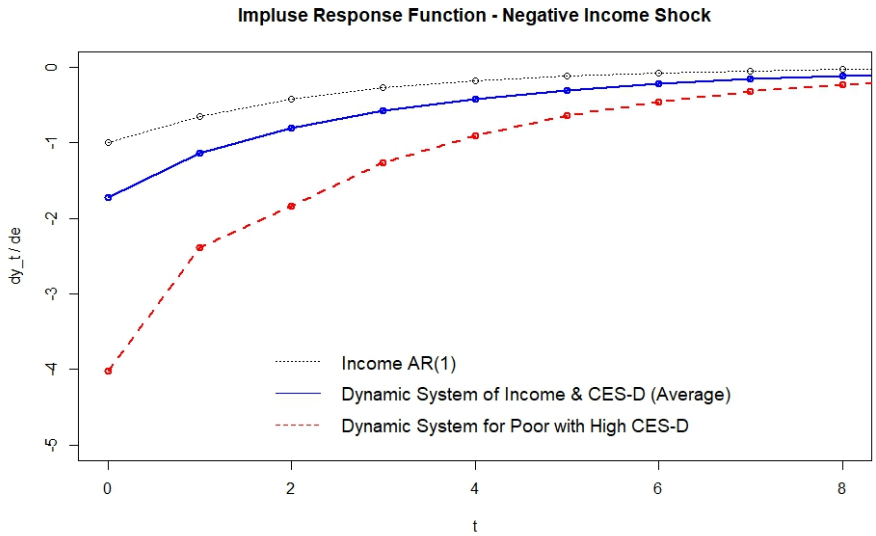

FIGURE E1: Impulse response function shows the effect of a shock to income in time 0 on future income.

long-term impact of a negative income shock for the average estimated system of dynamic simultaneous equations. This relationship accentuates the effect of the initial shock but also has an added impact over time. The estimated system of equations predicts that the overall impact of an income shock (current and future loss) is nearly double that estimated through an AR(1) process.[43] The results estimated above show significant heterogeneity. Particularly, among the poor near the depression threshold, the estimated effects are larger which suggests that shocks could affect this group disproportionately. The dashed (red) line shows the estimated impulse response for this group. The overall effect of shocks is nearly four times that of the AR(1) process. In the NIDS sample, over 30% of individuals in the lower half of the household income per capita distribution have a CES-D score of nine or above. The results reiterates that an across-the-board shock to either income or psychological well-being affects some individuals – the poor with low levels of psychological well-being (approximately 18% of the NIDS sample) – disproportionately.

The dynamics of income and psychological well-being are similarly affected by shocks to either variable: Figure E2 illustrates how shocks to income or psychological well-being affect both variables by amplifying the effect initially and over time.

[43] The overall impact of a -1 shock to income is -5.71 vs -3.45. This can be calculated by adding the infinite sum of $\frac{\delta Y_{i,t+j}}{\delta e_{i,t}}$. $\sum_{j=0}^{\infty}(A^{-1}B)^j A^{-1}e_1$ which is a geometric series and converges to $(I - A^{-1}B)^{-1}A^{-1}e_1$. For an AR(1) process it is simply $1-(1-\beta)$.