Department of Economics Discussion Papers

ISSN 1473-3307

Time-stability of risk preferences: A new approach with evidence from developed and developing countries

Nicolás Salamanca, Buly A. Cardak, Edwin Ip, and Joe Vecci

Paper number 23/05

Time-stability of risk preferences: A new approach with evidence from developed and developing countries.*

Nicolás Salamanca† Buly A. Cardak‡ Edwin Ip§ Joe Vecci¶

April 2023

Abstract

Time-stability of preferences is a crucial assumption in economics. We develop a novel test-retest method to examine the stability of risk preferences over time, while quantifying the importance of both idiosyncratic shocks and measurement error. Using eight large, representative datasets from developing and developed countries, we find risk preferences to be unstable in developing countries. In contrast, they are very stable in developed countries, except for low-income individuals in the U.S.. We discuss the important implications of these findings for policies and research.

Key Words: Risk preferences; Stability; Economic development

JEL Codes: D01, D81, O10, C18

*This version supersedes a previous version titled "The Dynamic Properties of Economic Preferences". Cardak acknowledges that this project was completed with the assistance of a La Trobe University Social Research Platform Grant. Salamanca acknowledges the Australian Research Council Centre of Excellence for Children and Families over the Life Course (project numbers CE140100027 and CE200100025). Vecci acknowledges support from the Swedish Research Council (No. 348-2014-4030). This paper uses data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project is funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research (MIAESR). The findings and views reported in this paper are those of the authors and should not be attributed to FaHCSIA or the MIAESR. We acknowledge helpful comments from Deborah Cobb-Clark, Denis de Brombrugghe, Julian Dyer, Samuel Engle, Jan Feld, Alex de Gendre, Bart Golsteyn, Dan Hamermesh, Simon Georges-Kot, Remi Levin, Mahreen Mahmud, Stefi Schurer, Benedikt Vogt, Jongsay Yong, Ulf Zölitz, and participants in various seminars and conferences. Any errors in this paper are our own. The HILDA data used in this paper was extracted using the Add-On package PanelWhiz for Stata® (see Hahn and Haisken-DeNew (2013) and Haisken-DeNew and Hahn (2010)). This paper uses data from the DNB Household Survey.

†n.salamanca@unimelb.edu.au, Melbourne Institute: Applied Economic & Social Research, University of Melbourne, Australia, ARC Centre of Excellence for Children and Families over the Life Course and IZA.

‡b.cardak@latrobe.edu.au, Department of Accounting, Data Analytics, Economics and Finance, La Trobe Business School, La Trobe University, Australia.

§e.ip@exeter.ac.uk, Department of Economics, University of Exeter, United Kingdom.

¶joseph.vecci@economics.gu.se, Department of Economics, University of Gothenburg, Sweden and IZA.

1 Introduction

A fundamental assumption in economics is that model primitives such as preferences do not change over time. Leaning on this assumption, economic theory typically seeks to explain changes in individual behavior through changes in prices and constraints rather than changes in preferences (Golsteyn and Schildberg-Hörisch, 2017; Schildberg-Hörisch, 2018; Stigler and Becker, 1977). In empirical work, preference stability also has wide ranging implications; e.g., it is implicitly assumed in many empirical models of dynamic choice (Barsky et al., 1997), allows for identification in collective household models (Browning and Chiappori, 1998) and rules out simultaneity in empirical studies (Dohmen et al., 2011). In finance, modern portfolio theory (Markowitz, 1952; Merton, 1973) is founded on the assumption of stable preferences, leading to the mutual fund separation theorem and the capital asset pricing model (CAPM), the most widely used model in asset pricing.[1]

While preference stability is an important and widespread assumption in economics, there is a growing literature exploring the possibility that preferences change over time (Dohmen et al., 2017; Grubb et al., 2016; Schildberg-Hörisch, 2018; Meier, 2022; Akesaka et al., Forthcoming), respond to shocks from natural disasters (Cameron and Shah, 2015; Hanaoka et al., 2018) and financial markets (Dohmen et al., 2016; Guiso et al., 2018; Weber et al., 2012), vary with macroeconomic conditions (Sahm, 2012; Bucciol and Miniaci, 2018) or appear to change due to measurement error (Chuang and Schechter, 2015; Meier and Sprenger, 2015).

One of the most common methods used to study stability is the test-retest approach (see reviews in e.g., Roberts and DelVecchio, 2000; Golsteyn and Schildberg-Hörisch, 2017). This entails estimating the correlation between a person's observed preference at two different points in time. Generally, a preference that has a high correlation over time is considered to be stable (Dohmen et al., 2016; Bland and Altman, 1986; Schildberg-Hörisch, 2018).[2] While this method provides a guide to the reliability of a measure, there are a number of limitations. First, judging a test-retest correlation as sufficiently high is often based on subjective interpretations of what constitutes a high correlation, and ignores factors such as sample size and time elapsed between

[1]Recent evidence of the growing interest in the stability of economic preferences can be seen by the excellent review of the stability of risk literature by Schildberg-Hörisch (2018); Golsteyn and Schildberg-Hörisch (2017) who summarise the challenges in analysing the stability of preferences in economics and psychology, and highlight areas where further progress can be made; Hardardottir (2017) analyze the long–term stability and determinants of time preferences, using DHS data, and Duersch et al. (2017) investigate the intertemporal stability of ambiguity preferences; while Frey et al. (2017) study the stability of risk using multiple measures with 109 adults. None of these papers, however, propose a way to test the time-stability of the preferences analyzed, nor do they quantify measurement error and idiosyncratic shocks. Further, most of these studies use non representative samples.

[2]Cicchetti (1994) argues that a correlation of 0.40 to 0.59 as fair, 0.60 to 0.74 as good, and above 0.75 as excellent.

observations (Watson, 2004; Berchtold, 2016). Second, as shown by Bland and Altman (1986), a high positive correlation between observations is not a sufficient condition to establish stability. Third, the correlation between observations at two points in time provided by a test-retest does not distinguish between changes due to measurement error and actual changes in preferences. In the presence of measurement error, the test-retest correlation will be biased downward because previously observed preferences will be correlated with the error term, resulting in endogeneity (Roberts and DelVecchio, 2000; Fraley and Roberts, 2005; Gillen et al., 2018; Schildberg-Hörisch, 2018; Salamanca et al., 2020). Despite these issues, an overwhelming number of studies simply compute the correlation between observations at two points in time and conclude good stability properties if the correlation is sufficiently high.

In the first part of this paper, we develop a model for testing the stability of preferences that addresses many of these shortcomings. Our model, which we refer to as the "enhanced test-retest" builds on the conventional approach to provide estimates of the correlation between preferences observed from one point in time to the next. However, our model builds in corrections for endogenity, in particular measurement error, using an instrumental variable (IV) approach, with a previously observed preference serving as an instrument. While difficult to deal with and often ignored in the literature, measurement error in observed preferences can bias parameter estimates of the effect of preferences on outcomes (Gillen et al., 2018; Schildberg-Hörisch, 2018; Salamanca et al., 2020), a problem that also applies to estimates of the stability of preferences. The estimate from this enhanced approach can be treated as an adjusted test-retest correlation, unattenuated by measurement error and net of predictable changes in preferences based on observable characteristics.

In addition to correcting for measurement error, our enhanced test-retest model makes three other innovations. First, the model provides estimates of the variance of measurement error in stated preferences. This quantifies the relative importance of measurement error in the measurement of preferences and allows for the comparison of measurement error between different survey instruments. Second, the model provides estimates of the variance of idiosyncratic shocks to latent preferences. These shocks can be understood as the variation in latent preferences that cannot be predicted by observable characteristics or by the natural change in preferences measured by the coefficient on past preferences. The variance of these shocks measures the degree of unobserved heterogeneity in latent preferences and is a crucial component of preference stability. Third, by combining parameter estimates from our model, we can formally test the null hypothesis that preferences are stable, a common assumption which is central to most economic analyses but that is never directly tested in the existing literature (Golsteyn and Schildberg-Hörisch, 2017). This enables us to go beyond using an arbitrarily determined "high correlation" between preferences in two periods as an indicator of stability, or to measure the rate of change of preferences in response to only a few observable events (e.g., a natural disaster or sharp changes in personal circumstances). Our method provides researchers with a measure of the rate of change in preferences attributable to gradual changes over time, changes due to observed factors, changes due to unobserved shocks, and measurement error. This can help researchers assess the seriousness of any preference instability problems in their setting and potential causes. Finally, for applied researchers, the method can be easily implemented in terms of both data and technical requirements. The companion Stata code to this paper, `dynreg`, is available upon request.

In the second part of this paper, we use our model to investigate the stability of risk preferences over a short time period by analyzing large representative panel data sets from four developed countries (Australia, Germany, the Netherlands and United States) and four developing countries (Kyrgyzstan, Malawi, Thailand and Vietnam). We thus provide by far the largest and most comprehensive analysis of the stability of risk preferences and are the first to compare risk preference stability in developed and developing countries.[3]

We estimate the stability of risk preferences separately for each country. Remarkably, despite differences in time span, country and method of preference elicitation, we find consistent results within the set of developed and developing countries. In three of the four developed economies the correlation between preferences is very close to one. The null hypothesis of stability cannot be rejected for Germany and, while it is rejected for Australia and the Netherlands, the respective estimates of 0.962 and 0.964 imply risk preferences change only at a very slow rate and thus behave as if stable. In the United States the correlation is 0.88, suggesting greater instability in risk preferences. We investigate potential reasons for this in a separate section. In all developed countries except for the United States, idiosyncratic shocks account for only a small part (up to 8 percent of the standard deviation) of the year to year variation in measured risk preferences, while measurement error accounts for a much larger fraction of the variation (between 18 and 38 percent of the standard deviation). We also find that included observed characteristics have little influence on the parameter estimates of the model or on the overall assessment of the stability of risk preferences.

Risk preferences are markedly less stable in the four developing countries considered. The null hypothesis of stability is rejected in all four cases with the estimated correlation between preferences found to be 0.85 (Kyrgyzstan), 0.30 (Malawi), 0.72

[3]The various panel data sets used here provide self reported measures of willingness to take risk that are widely referred to in the literature as risk preferences (Falk et al., 2018; Schildberg-Hörisch, 2018; Dohmen et al., 2017; Chuang and Schechter, 2015; Becker et al., 2012; Cheung, 2015; Hanaoka et al., 2018; Einav et al., 2012).

(Thailand) and 0.49 (Vietnam). Idiosyncratic shocks play an important role in the evolution of risk preferences in these countries, accounting for at least 21 percent of a standard deviation of the variation in risk preferences; far higher than in developed countries. Measurement error is similarly important, accounting for up to 0.77 standard deviations of the variation.

The high level of measurement error present in all countries illustrates that failure to account for it–for example, by using the traditional test-retest method–would result in biased estimates of preference stability, with estimated coefficients on past preferences well below one (Chuang and Schechter, 2015; Meier and Sprenger, 2015). A similar pattern is observed in our data when we estimate the standard test-retest model. These estimates range between 0.51 and 0.68 in developed countries and 0.19 to 0.45 in developing countries, all markedly lower than those of our enhanced test-retest model. While these correlations are slightly higher than those found in existing studies of developed countries (which range from 0.13 to 0.68; see e.g. Meier and Sprenger, 2015; Chuang and Schechter, 2015; Schildberg-Hörisch, 2018), they would all be misleadingly considered as fair or moderate.[4] This highlights the fact that standard test-retest correlations as low as 0.50 can, unintuitively, be consistent with risk preferences that do not change over time.

An important empirical contribution of our paper is the comparison of preference stability between developed and developing countries, for which there is limited evidence. In a review of the literature on the stability of risk, time and social preferences, Chuang and Schechter (2015) identify a small number of studies in developing countries which test preference stability (Carlsson et al., 2014; Sakha, 2019; Kirby et al., 2002; Dean and Sautmann, 2020; Dasgupta et al., 2017). Unlike our analysis, these studies rely on small and non-representative samples. We provide the largest and most comprehensive study of the stability of risk preferences undertaken thus far, analysing representative data from eight countries, and including over 180,000 observations that span multiple years.[5] We therefore provide systematic and generalizable evidence on the stability of risk preferences around the globe and over a wide range of institutional settings.

To understand differences between developed and developing countries, including the estimated instability in the US, we turn to differences in observable characteristics such as income levels and availability of health and social security provisions. We argue that lower levels of financial and social security can affect stability of risk preferences

[4]The only other study we could identify that reports a correlation coefficient using developing country data on risk preferences is Chuang and Schechter (2015), who find a correlation between 0.04-0.07 with only around 140 observations.

[5]A recent study by Liebenehm (2018) examines the impact of changes in consumption on changes in risk preferences in Thailand and Vietnam but not explicitly the stability of risk preferences over time. We use the same data in this study, taking advantage of two additional waves available since their work was undertaken.

through two channels. Socially and financially insecure people may be more susceptible to shocks that can affect their risk preferences. The United States has a wider distribution of incomes than the other developed countries and a larger vulnerable population due to its relatively low levels of social insurance (OECD, 2020; Jusko, 2016; Faricy, 2015; Garfinkel and Smeeding, 2010). This may explain the lower stability estimates we observe for the United States relative to the other developed countries studied. To examine the influence of income on the relatively low stability estimates observed for the United States and developing countries, we estimate separate models of risk preference stability for different segments of the income distribution.

We find for nearly all countries studied, people with low incomes have less stable risk preferences than those with high incomes, and this is especially true in the United States. In the United States we find stark differences in the estimated correlation on risk preferences between low- and high-income people (0.70 versus 0.96) when comparing the top and bottom 30%. High-income people have risk preference stability similar to those in other developed countries, whereas people at the bottom of the income distribution have markedly less stable risk preferences, more similar to people in developing countries. Our estimates across countries also show that idiosyncratic shocks generally explain a greater proportion of the variation in risk preferences for people with low income, and this is especially true for the US. These patterns of stability and the role of idiosyncratic shocks over the income distribution suggest the lower stability estimates observed in the US and developing countries are tied to those with lower income levels, consistent with recent evidence on the instability of preferences to income shocks for vulnerable people (Akesaka et al., Forthcoming).

Our results on the stability of risk preferences across countries have at least two important implications. First, people in developing countries have been found to be risk averse (Haushofer and Fehr, 2014), which has been shown to, at times, perpetuate poverty by inducing suboptimal risk avoiding behaviors (Liu, 2013; Liu and Huang, 2013).[6] This has led to the development of aid programs that embolden people to take greater risks by, for example, encouraging entrepreneurship or changing the agricultural practices of farmers. However, our results suggest aid programs targeting risk alone may not be sufficient if the instability of risk preferences is not accounted for and incorporated into the policy framework, especially for low-income people. Policies aimed solely at encouraging people to increase their risk taking may fail to deliver the desired response when risk preferences change over time. Further, estimates of such programs may lack internal validity and lead to incorrect inferences. Our results also suggest the

[6]For instance in China, Liu (2013) finds that farmers who are more risk or loss averse take longer to adopt better cotton varieties. Risk aversion is found to influence agricultural decision making and the effectiveness of a cash grant program in Ghana (Karlan et al., 2014), while Dercon and Christiaensen (2011) explicitly show that Ethiopian farmers limit their technology adoption due to risk aversion.

provision of a social safety net may serve a further purpose by enhancing the stability of risk preferences, which can potentially generate large complementarities between these and other development programs and increase their joint efficiency. Finally, our findings highlight the importance of frequently measuring and monitoring risk preferences in the implementation of programs in developing countries. Since risk preferences are less stable in these contexts, repeated measurement can help designers implement interventions that are incentive-compatible and welfare-increasing in these populations.

The paper proceeds as follows. In Section 2 we discuss the standard test-retest model and then our enhanced test-retest model, providing details on the estimation approach. We detail our data sources and summarize key features of the data in Section 3. Our main estimation results are presented in Section 4 along with multiple robustness tests, with the decomposition of results by income presented in Section 5. Section 6 provides conclusions.

2 Enhanced test-retest: A model of time-varying preferences

2.1 The Standard Test-Retest Model

Test-retest models used in the literature are generally based on simple correlations between preferences measured in two periods (Schildberg-Hörisch, 2018; Dohmen et al., 2016; Cobb-Clark and Schurer, 2013; Shou et al., 2022), occasionally controlling for observable characteristics (Chuang and Schechter, 2015; Fraley and Roberts, 2005). Allowing person i's measured level of preference at time t to be denoted by $P_{i,t}$, the standard test-retest model is:

$$P_{i,t} = \rho P_{i,t-1} + \phi' X_{i,t} + u_{i,t}, \tag{1}$$

where ρ is the parameter of interest, reflecting the correlation between current and past preference. The model may include controls for observable characteristics $X_{i,t}$, with $u_{i,t}$ being the error term, assumed to have mean zero and constant variance.

There are several limitations with this approach. First, preferences are virtually always observed with error. Measurement error in $P_{i,t}$ is more or less innocuous, resulting in less precise estimates of ρ. More importantly, however, measurement error in $P_{i,t-1}$ results in attenuation bias (i.e., lower estimates of ρ than would be found in its absence, often referred to as the error in variable problem). It is therefore easy to misinterpret low test-retest correlations as evidence of preference instability when preference measures are noisy (Bland and Altman, 1986; Watson, 2004; Berchtold, 2016). Despite this significant concern, measurement error is typically not addressed in the preference stability literature, implying estimates of the correlation between observed preference at two points in time are likely to be biased. Second, related to the first point, the disturbance term in equation (1) confounds measurement error with unobservable idiosyncratic factors, limiting the ability of researchers to quantify the effect of such shocks on preferences and their stability.

By quantifying the magnitude of measurement error and separating it from the variance of idiosyncratic shocks, researchers would be able to understand the impacts of each on the stability of observed preferences, their relative importance to observable factors, and make a better assessment of whether preferences are stable according to data. In the next section, we propose an enhanced test-retest approach that introduces these innovations.

2.2 The Enhanced Test-Retest Model

Let $P^*_{i,t}$ be the corresponding latent (i.e., unobserved) preference for individual i at time t. We model the dynamics of preference P via the following two equations:

$$P_{i,t} = P^*_{i,t} + \varepsilon_{i,t} \tag{2}$$

$$P^*_{i,t} = \delta P^*_{i,t-1} + \beta' X_{i,t} + \eta_{i,t} \tag{3}$$

Equation (2) defines the observed preference $P_{i,t}$ as the sum of the latent preference $P^*_{i,t}$ and a measurement error term $\varepsilon_{i,t}$. This measurement error is assumed to be classical — i.e., homoscedastic, independent of $P^*_{i,t}$, serially uncorrelated, and with well-defined first and second moments.[7]

Equation (3) defines the evolution of the latent preference $P^*_{i,t}$ as a function of past preferences with a level consistent with observable characteristics given by $\beta' X_{i,t}$.[8] For simplicity, we assume that the linear projection of $X_{i,t}$ on $P^*_{i,t}$ is a good approximation of the impact of individual characteristics on preferences, making β a sufficient parameter to describe the effect of observable characteristics on preference P. The $\beta' X_{i,t}$ term allows preferences to tend towards a conditional mean level determined by individual characteristics, effectively letting preferences converge towards different values for observably different people. The error term $\eta_{i,t}$ represents idiosyncratic shocks to the latent preference $P^*_{i,t}$, assumed to be independent and homoscedastic, with finite second moments.

Four parameters in this model provide useful information about the dynamics of preference P. The correlation coefficient on past preferences, δ, identifies the time-stability of $P^*_{i,t}$ — that is, how similar the current preferences are relative to past

[7]Autocorrelated measurement error can also be accommodated (see Section 2.3).

[8]More complicated preference structures, such as using the past two periods, can also be accommodated. However, the interpretation of this type of structure is not clear. Adding further past periods also seems empirically unimportant in the sense that they only result in small changes in the serial correlation test statistic of Wooldridge (2010).

preferences. The parameter β identifies how individual characteristics determine the level that preferences naturally tend to. The variance of the idiosyncratic shocks to preferences, σ_η^2, identifies the severity of variations in latent preferences once individual characteristics have been accounted for. While the variance of the measurement error, σ_ε^2, identifies the noise in the measurement of preference P.

To estimate the model's δ and β parameters, begin by rearranging and substituting (2) into (3) to obtain:

$$P_{i,t} = \delta P_{i,t-1} + \beta' X_{i,t} + \{\eta_{i,t} + \varepsilon_{i,t} - \delta\varepsilon_{i,t-1}\} = \beta' X_{i,t} + \delta P_{i,t-1} + \upsilon_{i,t}, \tag{4}$$

noting that preferences in this equation are all observable, though the composite error term $\upsilon_{i,t}$ is correlated with $P_{i,t-1}$. We explain in section 2.3 how we solve this endogeneity issue to obtain consistent estimates of δ and β.

To derive expressions for the variance of $\eta_{i,t}$ and $\varepsilon_{i,t}$ we define transformed past preferences as $\tilde{P}^*_{i,t} = P^*_{i,t} - \beta'(X_{i,t})$ and $\tilde{P}_{i,t} = \tilde{P}^*_{i,t} + \varepsilon_{i,t}$. Taking the k^{th} difference of $\tilde{P}_{i,t}$ and replacing recursively for $\tilde{P}^*_{i,t+k} = \delta\tilde{P}^*_{i,t+k-1} + \delta\beta'(X_{i,t+k-1}) + \eta_{i,t+k}$ using (3) yields:

$$\begin{aligned}
\tilde{P}_{i,t+k} - \tilde{P}_{i,t} &= \tilde{P}^*_{i,t+k} + \varepsilon_{i,t+k} - \tilde{P}^*_{i,t} - \varepsilon_{i,t} \\
&= \delta\tilde{P}^*_{i,t+k-1} + \delta\beta' X_{i,t+k-1} + \eta_{i,t+k} + \varepsilon_{i,t+k} - \tilde{P}^*_{i,t} - \varepsilon_{i,t} \\
&= \delta\{\delta\tilde{P}^*_{i,t+k-2} + \delta\beta' X_{i,t+k-2} + \eta_{i,t+k-1}\} + \delta\beta' X_{i,t+k-1} + \eta_{i,t+k} + \varepsilon_{i,t+k} - \tilde{P}^*_{i,t} - \varepsilon_{i,t} \\
&\vdots \\
&= (\delta^k - 1)\tilde{P}^*_{i,t} + \sum_{j=1}^{k} \delta^j \beta' X_{i,t+k-j} + \sum_{j=0}^{k-1} \delta^j \eta_{i,t+k-j} + \varepsilon_{i,t+k} - \varepsilon_{i,t}.
\end{aligned}$$

Based on equation (2) and the definitions of $\tilde{P}^*_{i,t}$ and $\tilde{P}_{i,t}$, we can substitute $(\tilde{P}_{i,t} - \varepsilon_{i,t})$ for $\tilde{P}^*_{i,t}$ which provides:

$$\tilde{P}_{i,t+k} - (\delta^k)\tilde{P}_{i,t} = \sum_{j=1}^{k} \delta^j \beta' X_{i,t+k-j} + \sum_{j=0}^{k-1} \delta^j \eta_{i,t+k-j} - \delta^k \varepsilon_{i,t} + \varepsilon_{i,t+k}. \tag{5}$$

The conditional variance of Equation (5) provides an expression that can be used to estimate the remaining parameters in the model. Note that the left-hand side of (5) depends on parameters δ and β for which consistent estimates can be obtained and observables $X_{i,t}$ and $X_{i,t+k}$. We can therefore obtain estimates of the variance of (5), $Var[\tilde{P}_{i,t+k} - (\delta^k)\tilde{P}_{i,t}]$, for any k as long as we have data for both periods t and $t+k$. The first term on the right-hand side of (5) shows that the variation due to individual characteristics can be accounted for by conditioning on $\bar{X}_{i,t+k} = (k-1)^{-1}\sum_{j=1}^{k} X_{i,t+k-j}$, the averages of individual characteristics up to time $t+k-1$.[9] Taking the variance of (5) conditional on $\bar{X}_{i,t+k}$ yields:

[9] If we treat k as the number of years between observations of preferences, then in cases where data on X_i are not observed in periods (years) between t and $t+k$, i.e. where survey waves are more than one year apart, $\bar{X}_{i,t+k}$ is evaluated as the mean of observed values at t and $t+k$.

$$Var[\tilde{P}_{i,t+k} - (\delta^k)\tilde{P}_{i,t}|\bar{X}_{i,t+k}] = \sigma_\eta^2 \sum_{j=1}^{k} \delta^{2j} + \sigma_\varepsilon^2(\delta^{2k}+1) \text{ for } k = 1, ..., K, \tag{6}$$

where K is the maximum number of lags of the observed preference P available.

Through Equation (6) we can identify the variances of shocks to latent preferences and preference measurement error. The left-hand side of (6) is all identified. The right-hand side of equation (6) includes the terms $\sum_{j=1}^{k} \delta^{2j}$ and $(\delta^{2k}+1)$, for which we have consistent estimators, and the variances σ_η^2 and σ_ε^2, which are the parameters of interest. These parameters can be obtained as the solutions to a linear system of equations with two unknowns, an extension of the technique used to identify the variance of permanent income by Carroll and Samwick (1997). For example, replacing δ by its consistent estimator, $\hat{\delta}$, and taking data for two lags, $k = 1, 2$ we can solve the system

$$\begin{aligned} Var[\tilde{P}_{i,t+1} - \hat{\delta}\tilde{P}_{i,t}|\bar{X}_{i,t+1}] &= \sigma_\eta^2\hat{\delta}^2 + \sigma_\varepsilon^2(\hat{\delta}^2+1) \\ Var[\tilde{P}_{i,t+2} - (\hat{\delta}^2)\tilde{P}_{i,t}|\bar{X}_{i,t+2}] &= \sigma_\eta^2(\hat{\delta}^2+\hat{\delta}^4) + \sigma_\varepsilon^2(\hat{\delta}^4+1) \end{aligned} \tag{7}$$

for σ_η^2 and σ_ε^2.[10] This yields estimators for $\hat{\sigma}_\eta^2$ and $\hat{\sigma}_\varepsilon^2$ which, together with $\hat{\delta}$ and $\hat{\beta}$ complete the model's parameter estimates.

2.3 Estimation

Estimates of δ and β can be obtained directly from (4). However, ordinary least squares (i.e., the standard test-retest) will yield inconsistent estimates since $P_{i,t-1}$ and $\varepsilon_{i,t-1}$ are correlated. To obtain consistent estimates, we exploit the fact that $P_{i,j}$ is uncorrelated with the composite error term in (4) for $j \leq t-2$, or equivalently that the moment conditions $E[P_{ij} \cdot v_{i,t}] = 0$, $\forall\, j = 0, \ldots, t-2$ identify the parameters of interest. These are straightforward moment choices given the time structure of the problem, and the same moment conditions implied in an IV approach, where $P_{i,t-1}$ is instrumented by further lags of P. Therefore, consistent estimates of the parameters in equation (4) can be obtained through IV estimation.

Different moment conditions can be used to balance data requirements and the strictness of the underlying assumptions for the validity of the instruments. For example, serial correlation in $\varepsilon_{i,t}$ does not make the method infeasible. It merely restricts the lag structure of the valid instruments available for estimation. If $\varepsilon_{i,t-1}$ follows an

[10]In the empirical analysis below we treat k as the number of years between surveys as most of the data sets we use are collected annually. To ensure comparability of results between data sets, in cases where surveys are conducted more than one year apart, for example every three years as in the cases of Thailand and Vietnam - detailed below, the equations in (7) are derived by setting $k = 3$ and $k = 6$ in equation (6) to reflect the number of years between surveys.

autoregressive process of order 3, the IV estimation will have to either (i) exploit lags of $P_{i,j}$ starting at $j = t-3$, or (ii) exploit other moment conditions that hold given this structure in the error term. Other data restrictions and model extensions will imply different valid instruments, yet many possible configurations will be admitted in the general IV estimation.

The variance of the idiosyncratic shocks (σ^2_η) and the measurement error (σ^2_ε) can be estimated by solving a two-equation two-unknown system based on Equation (6) for two arbitrary periods. However, a more efficient estimator combines the information of all valid k–lengths via a non-linear regression, obtaining the variances as non-linear least squares parameters constrained to be positive. Finally, since $P_{i,t}$ can easily be standardized to have unit variance, the noise-to-signal ratio, a comparable metric of the amount of measurement error in preferences across models, can be estimated as $s = \sigma^2_\varepsilon/(1-\sigma^2_\varepsilon)$ (Cameron and Trivedi, 2005, p. 903).

2.4 Interpretation and inference

One of our main goals is to enhance the standard test-retest by providing a statistical test of the time-stability of preference P. In the proposed model the test is of the null hypothesis $H_0 : \delta = 1$. That is, whether the correlation coefficient of past preferences indicates that preferences remain the same as in the previous period. Interpretation of $\delta = 1$, however, is not straightforward. If viewed as a test-retest correlation, $\delta = 1$ would imply a perfect correlation over time and thus complete stability.

An issue if one considers a **long time** horizon is that $\delta = 1$ implies that any idiosyncratic shock to preferences will be infinitely-lived and thus preferences would be, in that very specific way, unstable. This highlights the importance of considering estimates of σ^2_η alongside estimates of δ for interpreting whether the data is consistent with stable preferences. An estimate of δ close to one and a small estimate of σ^2_η is consistent with stable preferences where idiosyncratic shocks are largely unimportant. Larger estimates of σ^2_η would instead imply that preferences might not be as stable over time, yet they would be "stable" in the sense that past measures of preferences would still be good approximations of present measures of preferences. Either way, an estimate of δ that is close to one is taken as evidence of stable preferences.

Some further related points are worth noting about statistical inference for δ. As δ approaches unity, testing for $\hat{\delta} = 1$ would seem to become increasingly similar to a panel data unit root test. The similarity suggests that $\hat{\delta}$ could have a distribution that degenerates too fast as the number of individuals, time periods, or observations tends to infinity. This phenomenon, called super-consistency, makes t-statistics unsuitable for inference. The estimator may have a distribution with asymptotic behavior that depends on the rate at which the number of individuals and the time periods tend to infinity.

However, when testing the stability of preferences in this setting, information accumulates by increasing the number of individuals, not each individual's number of observed periods. This means that the relevant asymptotics for inference in this model are large N and fixed T. Harris and Tzavalis (1999) show that with large N and fixed T asymptotics, the limiting distribution is $\sqrt{N}(\hat{\delta}_{OLS} - 1) \xrightarrow{d} N(0, 2\{T(T-1)\}^{-1})$. As a robustness check, we use their test for inference on $\hat{\delta}$ noting that their derivation is only valid for balanced panels, implying a need to accommodate these tests to this restriction.[11]

Another consideration is the potential existence of confounding unobserved individual heterogeneity. In this setting, confounding heterogeneity would take the form of individual unobserved differences in preferences that, if correlated with past preferences or individual characteristics, could bias the estimates. The standard way to account for this heterogeneity is to eliminate its time-invariant individual-specific component via fixed or random effects (e.g., Sahm, 2012). In this model, however, this heterogeneity is the identifying variation of the estimates, and eliminating it would not be reasonable. To see this more clearly, consider the scenario where preferences are time-stable, and thus $P^*_{i,t} = P^*_i$. In this model, past preferences would be fully captured by time-invariant individual heterogeneity, and in any panel model that eliminates it, such as Arellano and Bond (1991), the correlation coefficient of past preferences δ will be unidentified under the null hypothesis. Therefore, we do not consider such models as providing feasible estimators for the purposes of this paper. Mixed effect models, on the other hand, can accommodate time-invariant individual unobserved heterogeneity uncorrelated to past preferences but partially correlated to individual characteristics (see Mundlak (1978) for their development and Sahm (2012) for a related application). However, since the coefficients on controls (β) are not the focus of our analysis, and as such models would not aid in the identification of the stability parameter δ, we do not consider them in this paper. Instead, we focus on estimating $\hat{\delta}$, $\hat{\sigma}^2_\eta$ and $\hat{\sigma}^2_\varepsilon$ in the empirical exercise below.

As we discuss below, most of our measures of risk preferences are ordinal with categories ranging between 0-10. Given the categorical nature of this data, a possible empirical approach is to estimate an ordered probit model. However, this would make the interpretation of the coefficients more challenging, and it is not commonly used in the analysis of this type of data (Falk et al., 2018; Hanaoka et al., 2018; Schildberg-Hörisch, 2018; Dohmen et al., 2017; Cheung, 2015; Chuang and Schechter, 2015; Becker et al., 2012; Einav et al., 2012; Frey et al., 2017). Further, ordered probit models place similar restrictions, as our preferred model, on the coefficients and the variance of the latent error term. Therefore, for ease of interpretation and to be consistent with the

[11] A minor issue is that the efficiency loss in IV compared to OLS implies that standard errors may be larger for our estimates, making the Harris and Tzavalis tests of robustness conservative.

literature we estimate our models using an IV GMM approach rather than an ordered probit specification.[12]

Finally, heterodscedasticity in measurement error (σ^2_ε) and idiosyncratic shocks (σ^2_η) may be an issue. However, we believe if heterodscedasticity exists it is likely to be minimal for two reasons. First, when we estimate σ^2_η with and without controls, we find very similar parameter estimates. This suggests that observable factors have little influence on our estimates and we would therefore be surprised if unobservables did. Second, the R^2 in our models are relatively high, which means there is little room for model error variance, especially for the countries with high δ, which implies there is minimal potential for heteroscedasticity and unobservables to matter.

2.5 Comparison with related models

An overview of the different approaches to preference stability and key characteristics of each method are presented in Table 1. We focus on five important characteristics to understand the dynamics of stability. The first is whether the model allows the user to formally test for preference stability. The second is whether the model mitigates endogeneity, in particular from measurement error. The third and fourth characteristics are respectively whether the model can estimate the impact of measurement error or idiosyncratic shocks. As discussed above, both play a critical role in establishing the stability of empirically observed preferences. Finally, we consider the ability of each method to be used to understand observable factors influencing changes in preferences.

We begin with the standard test-retest approach which, as described above, computes the correlation of the same preference measured at two points in time. However, as discussed in Section 2.1 it has various shortcomings. It relies on a biased estimator and is silent on the role of measurement error and idiosyncratic shocks, which are both integral to interpreting any observed (in)stability of preferences.

Unlike the standard test-retest approach, Gillen et al. (2018) offers a method to study measurement error. They show measurement error is important in understanding behavior and can often explain significant variation in behaviour. However, their method cannot directly examine stability nor measure idiosyncratic shocks.

Another approach is to estimate a model that regresses changes in preferences against a host of control variables such as age or marital status (see e.g., Cobb-Clark and Schurer, 2013, 2012). These models are generally useful to understand the correlates of changes in preferences but do not explicitly test for stability nor do they examine

[12] One could also estimate σ^2_ε and σ^2_η using an ordered probit specification. While this may be more efficient, it will not change results nor will it reduce any bias. Improvements in efficiency of the estimate is beyond the scope of this paper. Similarly, for ease of interpretation we avoid using an ordered probit specification to estimate these parameters.

Table 1: Overview of Methods

	Test-Retest	Gillen et. al (2019)	Cobb-Clark & Schurer (2012,2013)	Sahm (2012)	Enhanced Test-Retest
Stability Test	x			x	x
Correct for ME.					x
Estimate ME.		x			x
Estimate Idiosyn. Shocks				x	x
Determinants	x		x		

Notes: ME. denotes measurement error. Examples of Test Retest include Roberts and DelVecchio (2000), Fraley and Roberts (2005), Chuang and Schechter (2015), Dohmen et al. (2016) and Beauchamp et al. (2017).

measurement error.[13] In related work, Sahm (2012) adopts an innovative approach to investigate the impacts of time varying characteristics such as wealth, income, age and macroeconomic variables on risk preferences as well as cross sectional variation in preferences. However, the analysis does not directly address the issue of time stability of risk preferences and while it includes data on a number of life events (observed idiosyncratic shocks), it is unable to address measurement error.

Our Enhanced Test-Retest model overcomes many of the limitations of existing approaches. It is designed to provide a direct statistical test of the stability of preferences, it explicitly quantifies and accounts for measurement error, and also quantifies the variation of idiosyncratic shocks to preferences. Despite these benefits the approach has limitations. Unlike the determinants model, it is not designed to understand the correlates of risk preferences and in its most basic version it has stronger data requirements –it requires preferences to be observed in at least 3 time periods. However, it represents a significant improvement to existing methods used to study the stability of preferences.

3 Data

The empirical strategy requires repeated observations of stated risk preferences. The analysis utilizes household survey data from four developed countries (Australia, Germany, The Netherlands and the United States) and four developing countries (Kyrgyzstan, Malawi, Thailand and Vietnam), estimating the model outlined in Section 2 **separately** for each country. The data for each country is described below. These data sets were selected as they were the only data sets we could identify that met the minimal criteria of: i) being nationally representative and ii) where risk preferences are measured for the same individuals at least three times.[14]

[13] These studies estimate the correlation between individual characteristics, such as age, and changes in observed risk preferences. This allows the researcher to test if individual characteristics are related to changes in risk preferences. In contrast, our approach explicitly estimates the correlation between risk preferences from one observation to the next in order to test for the stability of risk preferences, seeking to identify whether there are any statistically significant changes in risk preferences.

[14] To identify relevant data we searched the common individual/household panels as well as databases such as the World Bank DataVerse.

3.1 Developed countries

Australia

The data on Australian household risk preferences and demographic characteristics are taken from the Household, Income and Labour Dynamics in Australia (HILDA) survey, a nationally representative household survey collected annually since 2001. The survey involves more than 9,500 households with over 17,000 people responding in each wave; see Summerfield et al. (2017) for full details of the HILDA survey.

Over the 16 available waves, the HILDA survey provides repeated observations on financial risk preferences assessed with the question "Which of the following statements comes closest to describing the amount of financial risk that you are willing to take with your spare cash? That is, cash used for savings or investment", with ordinal responses ranging on a 4-point scale from "I am not willing to take any financial risks" through to "I take substantial financial risks expecting to earn substantial returns". This question is asked in 2001-2004, 2006, 2008 and 2010-2016, providing a sample of 67,378 observations.[15] We select the standard demographics as controls and to the best of our ability we try and ensure controls are comparable across countries. Our controls for the HILDA data, $X_{i,t}$, include household income, employment status, age, gender, education, household size, and marital status.

Germany

Data on German household risk preferences are taken from the German Socio-Economic Panel (GSOEP), a representative panel of the German adult population collected annually since 1984 with over 20,000 people in each wave; see Schupp and Wagner (2002) and Wagner et al. (2007) for detailed information on the GSOEP.

Risk preferences are measured in the GSOEP using the question "Are you generally a person who is fully prepared to take risks, or do you try to avoid taking risks? Please choose a number on a scale from zero (unwilling to take risks) to ten (fully prepared to take risks)". This question was asked in 2004, 2006 and 2008-2015, providing a sample of 82,789 observations. Our controls for the GSOEP include household income, employment status, age, gender, education and marital status.

The Netherlands

Data for the Netherlands are taken from the Dutch National Bank Household Survey (DHS), a representative panel survey collected annually since 1993. On average 2,000 households (4,500 people) are interviewed each year with every household member aged over 16 years responding; see Teppa et al. (2012) for more details of the DHS. To be

[15]For summary statistics of the risk preferences for all countries in our data, see Table A1.

consistent with the other data sets we restrict the sample to those 18 and over.

The DHS measures risk preferences by asking respondents to rate their agreement with six statements that relate to financial risk taking on discrete likert scales from one to seven. Examples include “I would never consider investments in shares because I find this too risky” and “If I think an investment will be profitable, I am prepared to borrow money to make this investment”. Following Warneryd (1996), responses to the six statements are combined to provide an overall measure of individual risk preferences. These data were collected between 1996-2015, providing 21,645 observations. Our control variables for the DHS include household income, age, gender, education, household size, whether the individual is retired and marital status.

United States

The United States data are taken from the American Life Panel (ALP), a nationally representative internet survey of over 6,000 American adults over the age of 18, administered by the RAND Corporation; see Pollard and Baird (2017) for more details of the ALP.

The ALP contains a number of survey modules which are administered on a subset of the survey panel. Our analyses use risk data obtained from a survey module on subjective health expectations administered in 2008, 2010, 2011 and 2013 which includes a question on risk identical to that administered in the GSOEP described above, providing a total of 1,058 observations. Our controls for the ALP include household income, employment status, age, gender, household size and marital status.

3.2 Developing countries

Kyrgyzstan

Data for Kyrgyzstan are taken from the nationally representative “Life in Kyrgyzstan Survey” conducted by the German Institute for Economic Research (DIW); see Brück et al. (2014) for details. The survey is an individual panel in which all adults living in a sample household are interviewed. The sample includes over 3,000 households and more than 8,000 people.

We use data from the four currently available waves which were conducted annually from 2010. Each wave includes a survey question on risk preferences identical to that asked in the GSOEP, described above. The four available waves provide a sample of 9,400 observations. Our controls for the Life in Kyrgyzstan Survey include household income, employment status, age, gender, education, marital status, literacy, number of living rooms in the dwelling, health status of the respondent and available water source information.

Malawi

To study Malawi we use data from The Malawi Rural Livelihood Survey (MRLS) which is a nationally representative panel survey of households conducted by the World Bank and the National Statistical Office of Malawi. Households were interviewed 4 times between November 2012 and November 2013. For further details see Beegle et al. (2017).[16] The survey includes 2,912 households.

Each wave of the survey contained a question on risk preferences asked of both the head of the household and the spouse. In particular, a non-incentivised variant of the Eckel and Grossman (2002) risk task was utilised. Respondents were asked to choose between six possible risky options. For each possible option, the payoff varied depending on the outcome of a coin toss. The payoffs for the first option were equal, while for the other options the payoff for one outcome increased while it decreased for the other, thereby increasing the riskiness of the option.

As controls we include employment status, age, gender, education, marital status, literacy, whether the household owns a mobile phone, household access to electricity, whether the household has access to enough food and the number of hours worked. As the survey does not measure income or consumption, the latter variables provide proxies for income. The four available waves provide a sample of 3,367 observations.

Thailand and Vietnam

The data for Thailand and Vietnam are taken from the "Impact of shocks on the vulnerability to poverty: Consequences for development of emerging Southeast Asian economies" project funded by the German Research Foundation. Both surveys are household panels (surveying the household head) with very similar sampling procedures and designs. Each panel comprises around 2,000 households that are representative of rural areas in Thailand and Vietnam respectively, and are intended to exclude capital cities, thereby being representative of developing regions of these countries; see Hardeweg et al. (2013) for full details of these data sets.

Risk preferences are again measured with the same instrument used in the GSOEP. We use the 2008, 2010 and 2013 waves of these surveys, providing samples of 1,751 and 1,767 observations in Thailand and Vietnam respectively. Our controls for both surveys include employment status, age, gender, education, marital status, literacy, household size, dwelling size, per capita consumption, available water source in the dwelling and health problems of the respondent.

[16]The purpose of the survey was to study Malawi's large-scale public works program (PWP).

3.3 Validity of risk preference measures

We refer to the self-reported measures of willingness to take risk described above as *risk preferences*, in line with many previous studies (Falk et al., 2018; Hanaoka et al., 2018; Schildberg-Hörisch, 2018; Dohmen et al., 2017; Cheung, 2015; Chuang and Schechter, 2015; Becker et al., 2012; Einav et al., 2012; Frey et al., 2017).

The internal and external validity of these and other similar self–reported measures of risk preferences has been extensively documented in previous studies. Risk preferences in the HILDA are highly correlated with risky asset holdings (Cardak and Wilkins, 2009), and these correlations also hold for the same type of risk question in the United States taken from the Survey of Consumer Finances (Shum and Faig, 2006). The GSOEP risk preference data has been validated using incentivized experiments by Becker et al. (2012) and Dohmen et al. (2011), who also show that the risk preference measure is strongly predictive of actual risky behaviour reported in the survey. The DHS measure of risk preferences has similarly been validated in several studies, including Warneryd (1996), Borghans and Golsteyn (2006), Kapteyn and Teppa (2011) and Salamanca et al. (2020). The risk measure for Thailand has been validated by Hardeweg et al. (2013), who find the survey measure of risk preferences is both correlated with an experimental measure of willingness to take risk and predictive of other risk taking behaviour such as the purchase of lottery tickets. While the risk preference measures from the ALP, Life in Kyrgystan survey, and the Vietnam data have not been validated in similar ways, they are themselves based on the same survey instruments used in Thailand and Germany which, as mentioned above, have been shown to have strong external validity. Similarly, while the risk preference measure in Malawi has not be explicitly validated, it is a relatively common measure of risk widely utilised in the literature and is seen to be relatively easier to understand than other experimental procedures (Charness et al., 2013).

4 Results

4.1 Developed countries

We first present the standard test-retest (OLS) estimates of equation (1) for our developed country samples, both with and without controls ($X_{i,t}$), in Table 2. Estimates of δ range from 0.47 (United States) to 0.68 (Netherlands) and are in line with existing test-retest correlation estimates in the literature that range from 0.13 to 0.68 for time horizons varying from a few days to five years (e.g., Meier and Sprenger, 2015; Schildberg-Hörisch, 2018). The key result in Table 2 is that the null hypothesis of $\delta = 1$ is rejected for each country considered. While such test-retest estimates are common in the literature, they should be treated with caution since OLS estimates of equation (4) will be biased—likely downwards—as discussed in Section 2.

We address the potential bias in these OLS estimates with our enhanced test-retest estimation approach. Results for our developed country samples are presented in three parts in Table 3. In the top panel we include estimates of the correlation between preferences (δ) with standard errors in parentheses and p-values for the test of stability ($H_0 : \delta = 1$) in brackets. Failure to reject this null implies risk preferences are stable at the intra-individual level. In the middle panel of Table 3 we present the estimated variances of idiosyncratic shocks (σ_η^2) and measurement error (σ_ε^2), and the noise to signal ratio (s). In the bottom panel we display whether the estimated model includes control variables, and show the p-values of an F-test of the joint significance of these controls ($H_0 : \beta = 0$), when included.

Focusing first on the model without controls, our preferred specification, estimates of the stability parameter (δ) range from 0.878 to 0.994 across the four developed countries considered (see the odd-numbered columns in Table 3). The inclusion of controls ($X_{i,t}$) leads to very little change – estimates of δ are marginally lower (even-numbered columns in Table 3). We are unable to reject the hypothesis that $\delta = 1$ for Germany with a p-value of 0.21 without controls and 0.34 with controls. However, for Australia, the Netherlands and the United States we reject the hypothesis that $\delta = 1$, although for Australia and the Netherlands the point estimate is very close to 1, suggesting that risk preferences change very slowly. These estimates are markedly closer to unity than the OLS estimates of δ presented in Table 2, highlighting both the bias present in estimates that do not account for the correlation between $P_{i,t-1}$ and $\upsilon_{i,t}$ in (4) and the importance of implementing the enhanced test-retest approach adopted here.

Another key element of the enhanced test-retest model is that it provides estimates of the role of idiosyncratic shocks and measurement error in the evolution of observed preferences over time through their respective variances and the noise to signal ratio. Given that the risk preference data has been normalized to have a unit variance, the variance estimates can be interpreted as changes in standard deviations of observed preferences. With the exception of the United States (0.32), estimates of the variance of idiosyncratic shocks range between 0.06 (Netherlands) and 0.08 (Australia) implying such shocks play only a small part in the variation of risk preferences, contributing at most 8% of a standard deviation. The estimate of σ_η^2 for the United States suggests that risk preferences there are more affected by idiosyncratic shocks though the lower estimate of δ (0.88) implies the impacts of those shocks are less permanent. Estimates of the variance in measurement error range from 0.18 (United States) to 0.37 (Australia) leading to high noise to signal ratios of up to 0.64. This suggests that measurement error is an important source of variation in observed risk preferences in developed economies.

Tests of the joint significance of the controls ($X_{i,t}$) reject the null hypothesis H_0 :

$\beta = 0$ for all countries. Since estimates of δ are very close to unity, an interpretation of the test of the significance of β is as a test of the role of the controls in explaining *changes* in risk preference. That is, if $\delta = 1$, equation (4) can be rearranged as $P_{i,t} - P_{i,t-1} = \beta'(X_{i,t}) + \upsilon_{i,t}$ and β informs us of the impact of $(X_{i,t})$ on the change in risk preferences rather than the level of risk preferences. Rejecting the null hypothesis $H_0 : \beta = 0$ suggests that the included controls are correlated with the change in risk preferences. However, most importantly, we find that the inclusion of the controls $(X_{i,t})$ have almost no impact on the estimates of δ, suggesting that the controls do not play a large role in changes in risk preferences.[17]

Overall, these results suggest risk preferences in three of the four developed countries exhibit either stability (Germany) or change very slowly (Australia and the Netherlands), with most of the variability due to measurement error. This implies even in developed countries, measurement error is important and potentially problematic. We further investigate results for the United States in Section 5.

4.2 Developing countries

Similar to the developed countries case, we begin our analysis of the developing countries in our sample with the standard test-retest (OLS) estimates of equation (1), presented in Table 4. Since risk preferences are collected every second or third year in Thailand and Vietnam, we report an annualized estimate of the stability parameter for these countries to ensure comparability of results. Similarly, since risk is collected around every 3 months in Malawi, we also report an annualised stability parameter. We further investigate the wave structure of the data in Section 4.3. Estimates of δ presented in Table 4 range between 0.19 (Malawi) and 0.45 (Vietnam) with the null hypothesis that $\delta = 1$ rejected for each country considered. However, these estimates are also similar to findings in the test-retest literature noted above and it is possible these low values may be due to the bias discussed in Section 2.

As for the developed countries studied above, we address the possibility of bias in these OLS estimates by applying the enhanced test-retest approach to equation (4) for Kyrgyzstan, Malawi, Thailand and Vietnam, both with and without controls. Results are presented in Table 5. Other parameter estimates and test results presented in Table 5 are analogues of those presented in Table 3.

Results in Table 5 contrast with those found for the developed countries. Estimates of δ for all four developing countries considered range between 0.30 and 0.85 (without

[17] One could argue that controls such as income may be co-determined with risk, making them potentially poor controls. We do not consider this to be an issue as we show that in all cases, adding controls has little impact on $\hat{\delta}$. Further, as a robustness check, we only include controls that are strictly predetermined (i.e., those that are time invariant such as gender or measured in earlier waves, before the measurement of risk) and show that this does not change our estimates of δ. Results are available on request.

controls). The null hypothesis of stability, $H_0 : \delta = 1$, is rejected for all countries at the 1% level.[18] The results imply risk preferences in these developing countries are not stable at the intra-individual level. Estimates of δ also change very little with the inclusion of controls leaving conclusions qualitatively unchanged. One important similarity to the developed country results is that the estimates of δ presented in Table 5 are still much higher than the OLS estimates reported in Table 4. This again highlights the bias present in the standard test-retest (OLS) estimates and the value of implementing the enhanced test-retest approach adopted here.

Turning to the impact of idiosyncratic shocks on risk preferences, we find that they contribute as much as 6 standard deviations to the variation of observed risk preferences. For all developing countries, these estimates are much larger than estimates of σ_η^2 for the developed countries studied here, with the exception of the United States. We can only speculate as to why idiosyncratic shocks seem much more important in these countries, but possible explanations include more volatile personal circumstances, as well as greater impacts of natural disasters and a lack of formal social insurance. This may be driven by a lack of risk mitigating infrastructure while social safety nets may not be as widely available as in developed countries—especially for the Thai, Malawi and Vietnamese samples which are focused on rural populations, with their agricultural economies more exposed to weather shocks. To the extent that these shocks are not captured in the data and cannot be insured for by people in these poorer countries, they may lead to larger estimates of σ_η^2.

Measurement error and the noise to signal ratio are also high in developing countries. Measurement error accounts for between $0.17 - 0.77$ standard deviations of the variation in observed risk preferences. Similar to findings in Chuang and Schechter (2015), these results quantify the large impact of measurement error on self reported risk preferences in developing countries.

Overall, these results suggest risk preferences in the developing countries studied here are unstable, with both measurement error and idiosyncratic shocks playing important roles.

4.3 Robustness

4.3.1 Economy wide shocks

We consider the possibility that the instability observed in developing countries may be a result of aggregate economy wide shocks that impact the whole population. Such shocks may have less impact or not exist at all in developed countries. To examine this further, we re-estimate our models with the inclusion of time fixed effects. This allows

[18] In Thailand the δ parameter is far from 1 but the p-value is below 0.05. This is clearly a result of a high standard error rather than an $H_0 : \delta = 1$, we interpret this p-value as such.

us to account for aggregate shocks that change risk for each specific time period. We find results are qualitatively unchanged, suggesting that such aggregate shocks do not provide an explanation for the instability observed in these developing countries.

4.3.2 The effect of restricting time periods

An important difference between the datasets of the developing and developed countries used in this study is the number of waves and lag structures. We deal with this in our estimates by standardising them. As a robustness check of the impact of these differences in datasets, we re-estimate our main model for Australia, Germany and the Netherlands with sample restrictions that mimic features of the developing country datasets.

In the odd numbered columns of Table 6 we present results of estimating the model using samples restricted to have the same lag structure as Thai and Vietnamese datasets (i.e., 3 waves of data with a 2 and 3 year gap at close to the same time period). The even-numbered columns present the results of estimating the model based on samples restricted to having three consecutive waves (2015, 2014, and 2013), which is comparable to the period during which most of the developing country data was collected. The results are very similar to those in Table 3, with estimates of δ for Australia and the Netherlands of around 0.97 and for Germany of around 0.99, while the other parameter estimates are also very similar to our original set of results. For instance, the variance of idiosyncratic shocks ranges between 0.02 (Germany) and 0.12 (Netherlands) implying a small roles in the variation of risk preferences.[19] These results suggest that differences in the lag structure and the number of waves are not the source of the estimated differences in the δ parameter between developed and developing countries.

4.3.3 Harris and Tzavalis test

While estimates of δ are close to 1 for Australia, the Netherlands and Germany, a test of the stability of risk preferences ($H_0 : \delta = 1$) for these countries using the estimates and standard errors reported in Table 3 may be problematic. As discussed in Section 2.4, if estimates of δ are super-consistent, it could be misleading (i.e., a higher chance of a type II error) to test for the time-invariance of preferences using standard t-tests. An alternative approach is to use the balanced panel unit root test and asymptotic distributions developed by Harris and Tzavalis (1999). We do this for those countries where $\delta > 0.95$, (i.e. those that exhibit the highest levels of estimated stability) as we argue these are the set of countries where super-consistency is the most relevant. Table 7 reports the Harris and Tzavalis (1999) tests for all possible balanced panels for the three relevant countries. These tests have the null hypothesis that preferences are time-invariant ($H_0 : \delta = 1$). To maximize the number of observations in each balanced

[19] Results are qualitatively unchanged by the inclusion of controls.

panel, we (i) regress each preference on the individual characteristics and a complete set of year dummies; (ii) follow Levin et al. (2002) in interpreting the residuals of that regression as a measure of preferences net of the cross-sectional influence of individual characteristics and common cross-sectional time effects; (iii) construct balanced panels of size T which include all individuals with T adjacent residuals; and (iv) calculate OLS and IV estimates of δ, and their corresponding small-sample z-statistic from Harris and Tzavalis (1999), defined as $z = (\hat{\delta} - 1)\sqrt{\frac{N}{2\{(T-1)(T-2)\}}}$.

The first two columns of Table 7 report the number of periods (T) and observations (N) in each of these balanced panels. The next three columns report the OLS estimates of the δ parameter, the z-statistic, and the p-value of the Harris-Tzavalis test. The last six columns report the IV estimates of the stability parameter, the idiosyncratic and measurement error variances, the noise-to-signal ratio, the z-statistic, and the p-value of the Harris-Tzavalis test. Focusing on the IV specification, the table shows, with few exceptions, that the IV estimates of δ are consistent with those reported in Table 3. In nearly all panels for Australia and the Netherlands, the test rejects the null of stable preferences, but similar to our baseline estimates, $\widehat{\delta}$ is close to 1 in nearly all cases. In Germany, for 3/4 of the panel models we cannot reject the null of stable preferences at $p < 0.01$. Finally, unsurprisingly, the Harris-Tzavalis test always rejects the null of time-stable preferences when based on OLS estimates.

Overall, the Harris-Tzavalis test results are similar to those of our baseline model, indicating we should reject the hypothesis that risk preferences remain time-stable in Australia and the Netherlands but not in Germany. However, the IV estimates of the time-variation of preferences suggest that even if preferences for 2/3 countries are not fully time-stable, they move very slowly from year to year and thus behave as if stable. Importantly, these results suggest that super-consistency is unlikely to have led us to type II errors when testing the stability of risk preferences based on the initial estimates presented in Table 3.

5 The role of income

Our results highlight important differences in the stability of risk preferences between developed and developing countries. Another important finding is that the stability parameter δ for the United States is relatively low compared to the other developed countries studied. In this section, we provide a possible explanation for these results.

We hypothesize that both differences may be explained by differences in incomes and social safety nets available to the vulnerable. For instance, low income individuals with relatively limited access to social safety nets may be more exposed and susceptible to shocks that may alter risk preferences, such as their ability to recover from natural disasters (Cameron and Shah, 2015) and the inability to afford treatment for illness or injury (Decker, 2016).[20]

Our data allows us to explore the stability of risk preferences for most of our low income and high income countries separately, as well as for low income and high income people within each of these countries, and we expect stark variation on both dimensions. Developing countries in our data are markedly poorer than developed countries, especially considering the rural nature of the samples in Malawi, Thailand and Vietnam. According to the World Bank, the population share living on less than \$5.5 a day in 2011 international purchasing-parity adjusted dollars is 0.7% in Australia, 0.2% in Germany, 0.5% in The Netherlands and 2.0% in the United States. These figures are much higher in the developing countries studied here with 61.3% in Kyrgyzstan, 97% in Malawi, 8.6% in Thailand and 23.6% in Vietnam. Many of these developing countries are also unequal, but even in some of the developed countries in our data we expect low wealth levels for a meaningful share of respondents; notably in the United States where the Gini Coefficient, relative poverty rates and poverty gap are by far the highest among the developed countries investigated (OECD, 2020). For example, the relative poverty rate is 18% in the United States, compared to 12% in Australia, 10% in Germany and 8% in the Netherlands.

Institutional variations between countries can exacerbate heterogeneity in the stability of risk preferences between low-income and high-income individuals, and one of the most significant institutional differences is the availability of a social safety net. Only limited social benefits are available in the developing countries we study (World Bank, 2018), implying people have little social insurance against income and wealth shocks, raising the likelihood that such shocks change risk preferences. Conversely, most of the developed countries considered here, with the possible exception of the United States, have strong social safety nets (OECD, 2020; Jusko, 2016; Faricy, 2015; Garfinkel and Smeeding, 2010). Although United States social spending is not unusually small, it has a strong reliance on a combination of employer-provided benefits and relatively weak cash transfer programs for the poor (Garfinkel and Smeeding, 2010), with a large proportion of social spending going to the relatively wealthy rather than the poorest (Faricy, 2015), especially when compared to other developed countries we consider (OECD, 2020). As such, the provision of social safety nets for the poor in the United States is relatively limited and access to social insurance less universal when compared to the other developed counties in our data. We argue that this may be one reason risk preferences in the United States are less stable than in other developed countries in our data.

[20]This is different from the effect of income itself on risk preferences, which could potentially be identified by the coefficient on income in our regressions. The interpretation of this coefficient, however, depends on whether $\hat{\delta}$ is close to zero or one; see Section 3.3 for a detailed discussion. Moreover, income could be endogenous in these models, which further complicates the interpretation of the coefficient. For all these reasons, we refrain from interpreting the estimated income coefficients or any other controls in this paper.

Bringing together all the above arguments, we can formulate four further hypotheses for our models: (i) low income people are likely to have more unstable risk preferences (lower δ) compared to high income people; (ii) within each country, the variance of idiosyncratic shocks to risk preferences (σ_η^2) should be larger for low income people; (iii) the differences between the δ of low and high income people should be larger in the United States than in other developed countries; and (iv) the δ for richer people in the United States should be comparable to the δ's in the other developed countries.

To provide suggestive evidence on these hypotheses, we first re-estimate our model separately for the lowest and highest 50% of the income distribution. This allows us to compare the δ parameter estimates for each income group within each country and to see if any differences correspond to differences in income distributions between countries. We focus on income rather than wealth because income is observed in all developed countries in our data but wealth is not. Although, for the developing countries, we use income in Kyrgyzstan and consumption in Thailand (as the survey does not measure income) but we exclude Malawi and Vietnam from this analysis.[21] As a robustness check we also estimate our model with the lowest and highest 30%.

Risk preference stability estimates for the top and bottom 50% of the income distribution are presented for developed countries in Table 8. For countries with similarly low Gini coefficients (Australia at 30.3, Germany at 27.0, and The Netherlands at 30.3 as measured in OECD (2020)), low income people generally have a lower estimated coefficient on past preferences ($\hat{\delta}$) than those with high incomes, though the differences are relatively small. However for the United States, with a much higher Gini Coefficient of 45.0, low income people have a much lower stability parameter ($\hat{\delta}$) than high income people. These differences are even larger when comparing the top and bottom 30% of the income distribution ($\hat{\delta} = 0.702$) for low income and ($\hat{\delta} = 0.956$) for high income (see Appendix Table A2 for results split by the top and bottom 30%). This suggests that those in the top 30% of the income distribution in the United States sample have an estimated stability parameter similar to those of other developed countries in our data, suggesting the relatively low overall stability estimate for the United States reported in Table 3 may be related to income inequality. Conversely, the lowest income people in the United States have a coefficient on past preferences parameter comparable to some developing countries in our sample; 0.70 for those in the bottom 30% of the distribution.

Results split by the top and bottom 50% for the available developing countries are presented in Table 9. We find results consistent with the developed countries studied. In Thailand, for example, we find $\hat{\delta} = 0.56$ for those in the bottom part of the distribution

[21] The Malwai Rurual Livelihood Survey provides no data on income, consumption or wealth, while the model for Vietnam fails to estimate as the δ is close to zero at which point the parameter variances are not computable. This could be due to a combination of lower sample sizes for this split estimation and higher measurement error. As we note in this subsection, we should interpret results from the developing countries with caution.

and $\hat{\delta} = 0.84$ for those in the upper part.[22] We do however believe the results for the developing countries should be treated very cautiously. The very low level of income even for the top part of the distribution (e.g., in Kyrgyzstan the wealthiest have incomes of only $2,609 USD per year) means that the impact of shocks and access to social security is likely to be similar between the top and the bottom part of the distribution. Further, except for Kyrgyzstan, we do not have a measure of income or wealth, and its not clear that consumption is an adequate proxy.

In nearly all countries the variance of idiosyncratic shocks (σ_η^2) also plays a bigger role for low compared to high income people. These differences are small for Australia, Germany and the Netherlands. For the United States, however, the variance of idiosyncratic shocks is much larger for low income people compared to high income people. This evidence is again consistent with the idea that idiosyncratic shocks to risk preferences are greater for low income people, especially in countries with less generous social security systems. Similarly, for Thailand, our estimates of σ_η^2 for the lowest income group is markedly larger than that of the top income group.

5.1 Discussion

The patterns of stability and the influence of idiosyncratic shocks across the income distribution highlighted in Table 8 may also offer an explanation for why the stability of risk preferences is lower in the US and developing countries. Developing countries differ from developed countries in three key ways. They have (i) lower income levels; (ii) less functional institutions; and (iii) different cultures. The United States is a particularly useful case study for the role of income as it comprises groups of people with very high and very low incomes, while institutions and culture are more uniform across the population. The positive relationship between income and risk preference stability suggested by our results may therefore help explain why we observe lower risk preference stability in developing countries than in developed countries.

The present study encounters similar challenges to the broader literature on risk preferences and their stability due to the use of self-reported risk attitude measures. While many of the measures used in this study have been shown to predict actual risky behaviors and incentivized experimental choices, it is not clear if they measure risk preference parameters or risk taking behaviors. Experimental measures used in other studies that capture parameters also generally do not integrate the participants' wealth or account for the effect of wealth on their risk taking behavior. Our study and its analysis of the role of income may shed some light on the interpretation of these widely used self-reported measures.

The greater instability of responses to these risk measures among lower income

[22]See Appendix Table A3 for results based on the top and bottom 30% of the income distribution.

individuals would be consistent with these measures representing absolute risk aversion. If what respondents have in mind when answering general risk questions is absolute risk aversion, then changes in observed preferences should manifest more among relatively low income individuals. To illustrate this, consider individuals with decreasing absolute risk aversion.[23] We may interpret instability of observed preferences in two ways.

First, a change in observed preference may be interpreted as a movement along the utility function. The concavity of utility at low wealth levels implies absolute risk aversion is more responsive to wealth changes for low wealth relative to higher wealth individuals. Therefore, regular fluctuations in wealth (e.g. from pay day, rent payments) would result in relatively larger swings in risk preferences for low wealth people, leading to more unstable preferences and to a larger idiosyncratic shock variance; see for example Akesaka et al. (Forthcoming).

Second, we may interpret a change in observed preference as a change in the risk aversion parameter. As an example, consider a constant relative risk aversion utility function for wealth w: $u(w) = \frac{w^{1-\alpha}}{1-\alpha}$.[24] Using this utility function, a change in risk preference may be interpreted as a change in the relative risk aversion coefficient (α). Any change in this coefficient is likely to be much less obvious for those who have higher wealth levels as the utility function would remain largely flat and the change in absolute risk aversion ($\frac{\alpha}{w}$) would be minimal. An equivalent change in the relative risk aversion coefficient would be much more observable for those with low wealth as the corresponding change in absolute risk aversion would be relatively larger.

Our results split by income show that, indeed, those with lower incomes (and presumably lower levels of wealth) have more unstable responses to questions regarding risk. As we control for income in our empirical analysis and quantify the variance of idiosyncratic shocks, our results point towards the instability we observe in self-reported risk measures being more than just movements along the utility function, but rather, changes in the risk preference parameter. However, this would also suggest that risk aversion parameters may be harder to identify at higher levels of wealth. Further research is required to better understand the nature of such risk preference measures. With this concern in mind, our results appear consistent with these self-reported risk measures capturing the respondent's absolute risk aversion.

[23]Decreasing absolute risk aversion is a common assumption. It is consistent with decreasing, constant, and some increasing relative risk aversion utility functions.

[24]Constant relative risk aversion is often assumed in theoretical and empirical studies (Dohmen et al. (2011); Barsky et al. (1997)) with empirical support for the assumption provided by Chiappori and Paiella (2011).

6 Conclusion

In this paper we develop the enhanced test-retest approach. The model provides an approach to test for the stability of preferences. It also provides metrics for important features of the short run dynamics of preferences that should be considered separately, including their predictable change based on observable characteristics, the role of idiosyncratic shocks on risk preferences, and the significance of measurement error, all issues of great interest in the analysis of self-reported risk preferences and their stability. For applied economists, our method is simple to implement and the paper comes with accompanying Stata code, `dynreg`, which we are happy to provide to researchers.

Importantly, the paper also provides a novel application of the model by testing the stability of risk preferences for a group of developed and developing countries. Covering eight countries, this is the largest and most comprehensive empirical analysis of the stability of individual risk preferences from which three key findings emerged.

First, risk preferences are stable or move very slowly over short time periods in developed countries (except for low–income people in the US) but are far less stable in developing countries. The implication is that theoretical and empirical models that rely on the assumption of stable risk preferences provide a good approximation of choices and behavior in developed countries, especially for high-income people, yet such models might behave less well in developing countries. A corollary is that policy analyses and calculations derived from such models are less likely to be effective in developing countries. An important implication for aid targeting developing countries is that efforts to influence the risky choices of people such as entrepreneurs or farmers may have short lived effects or appear ineffective as a consequence of the instability of risk preferences identified here. Similarly, welfare calculations, many of which implicitly assume preference stability, are likely to be misleading in developing countries and policy evaluations that rely on these calculations should be suitably adjusted. For empirical economists and policymakers, our results suggest that risk preferences should be remeasured regularly in longer term studies and policy interventions, particularly in developing countries.

Second, idiosyncratic shocks are important in explaining variation in risk preferences, and more so in developing countries and for those individuals with low income. One explanation for this is that people in less developed settings and with lower incomes may be less protected from shocks by formal institutions, physical infrastructure and social safety nets. This suggests an additional benefit of expanding social safety nets and social insurance. The secondary impact on risk preference stability will make future policy implementation and welfare evaluations simpler and more effective.

Finally, measurement error explains a large fraction of the variation in observed risk preferences in both developed and developing countries. Recent papers focus on the issue of measurement error and how it can be pivotal for inference on fundamental economic preferences and personality traits (e.g., Gillen et al., 2018). Our findings here place further emphasis on the importance of quantifying measurement error across survey instruments and contexts in order to develop a set of standard measurement instruments for risk preferences that minimize measurement error while being widely applicable. The methods and analyses presented in this paper provide a useful framework to do this.

References

Akesaka, M., P. Eibich, C. Hanaoka, and H. Shigeoka (Forthcoming): "Temporal instability of risk preference among the poor: Evidence from payday cycles," Tech. rep., American Economic Journal: Applied Economics.

Arellano, M. and S. Bond (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," The Review of Economic Studies, 58, 277–297.

Barsky, R., F. Juster, M. Kimball, and M. Shapiro (1997): "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study," The Quarterly Journal of Economics, 112, 537–579.

Beauchamp, J. P., D. Cesarini, and M. Johannesson (2017): "The psychometric and empirical properties of measures of risk preferences," Journal of Risk and uncertainty, 54, 203–237.

Becker, A., T. Deckers, T. Dohmen, A. Falk, and F. Kosse (2012): "The relationship between economic preferences and psychological personality measures," Annual Review of Economics, 4, 453–478.

Beegle, K., E. Galasso, and J. Goldberg (2017): "Direct and indirect effects of Malawi's public works program on food security," Journal of Development Economics, 128, 1–23.

Berchtold, A. (2016): "Test–retest: Agreement or reliability?" Methodological Innovations, 9, 2059799116672875.

Bland, M. and D. Altman (1986): "Statistical Methods For Assessing Agreement Between Two Methods Of Clinical Measurement," The Lancet, 327, 307–310, originally published as Volume 1, Issue 8476.

Borghans, L. and B. Golsteyn (2006): "Time discounting and the body mass index: Evidence from the Netherlands," Economics and Human Biology, 4, 39–61.

Browning, M. and P.-A. Chiappori (1998): "Efficient intra-household allocations: A general characterization and empirical tests," Econometrica, 1241–1278.

Brück, T., D. Esenaliev, A. Kroeger, A. Kudebayeva, B. Mirkasimov, and S. Steiner (2014): "Household survey data for research on well-being and behavior in Central Asia," Journal of Comparative Economics, 42, 819 – 835.

Bucciol, A. and R. Miniaci (2018): "Financial risk propensity, business cycles and perceived risk exposure," Oxford Bulletin of Economics and Statistics, 80, 160–183.

Cameron, A. C. and P. K. Trivedi (2005): Microeconometrics: Methods and Applications, Cambridge university press.

Cameron, L. and M. Shah (2015): "Risk-taking behavior in the wake of natural disasters," Journal of Human Resources, 50, 484–515.

Cardak, B. A. and R. Wilkins (2009): "The determinants of household risky asset holdings: Australian evidence on background risk and other factors," Journal of banking & Finance, 33, 850–860.

Carlsson, F., O. Johansson-Stenman, and P. K. Nam (2014): "Social preferences are stable over long periods of time," Journal of Public Economics, 117, 104 – 114.

Carroll, C. and A. Samwick (1997): "The nature of precautionary wealth," Journal of Monetary Economics, 40, 41–71.

Charness, G., U. Gneezy, and A. Imas (2013): "Experimental methods: Eliciting risk preferences," Journal of Economic Behavior and Organization, 87, 43–51.

Cheung, S. L. (2015): "Risk Preferences Are Not Time Preferences: On the Elicitation of Time Preference under Conditions of Risk: Comment," American Economic Review, 105, 2242–60.

Chiappori, P.-A. and M. Paiella (2011): "Relative risk aversion is constant: Evidence from panel data," Journal of the European Economic Association, 9, 1021–1052.

Chuang, Y. and L. Schechter (2015): "Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results," Journal of Development Economics, 117, 151–170.

Cicchetti, D. (1994): "Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instrument in Psychology," Psychological Assessment, 6, 284–290.

Cobb-Clark, D. A. and S. Schurer (2012): "The stability of big-five personality traits," Economics Letters, 115, 11 – 15.

——— (2013): "Two Economists' Musings on the Stability of Locus of Control," The Economic Journal, 123, F358–F400.

Dasgupta, U., L. Gangadharan, P. Maitra, and S. Mani (2017): "Searching for preference stability in a state dependent world," Journal of Economic Psychology, 62, 17–32.

Dean, M. and A. Sautmann (2020): "Credit Constraints and the Measurement of Time Preferences," Review of Economics and Statistics, forthcoming, 1–14.

Decker, S., . S. H. (2016): "Health shocks and risk aversion." Journal of health economics, 50, 156–170.

Dercon, S. and L. Christiaensen (2011): "Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia," Journal of development economics, 96, 159–173.

Dohmen, T., A. Falk, B. H. Golsteyn, D. Huffman, and U. Sunde (2017): "Risk attitudes across the life course," The Economic Journal, 127, F95–F116.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. Wagner (2011): "Individual risk attitudes: Measurement, determinants, and behavioral consequences," Journal of the European Economic Association, 9, 522–550.

Dohmen, T., H. Lehmann, and N. Pignatti (2016): "Time-varying individual risk attitudes over the Great Recession: A comparison of Germany and Ukraine," Journal of Comparative Economics, 44, 182–200.

Duersch, P., D. Romer, and B. Roth (2017): "Intertemporal stability of uncertainty preferences," Journal of Economic Psychology, 60, 7 – 20, stability of Preferences.

Einav, L., A. Finkelstein, I. Pascu, and M. R. Cullen (2012): "How General Are Risk Preferences? Choices under Uncertainty in Different Domains," The American Economic Review, 102, 2606–2638.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018): "Global Evidence on Economic Preferences*," The Quarterly Journal of Economics, 133, 1645–1692.

Faricy, C. G. (2015): "Welfare for the wealthy: Parties, social spending, and inequality in the United States," .

Fraley, R. C. and B. W. Roberts (2005): "Patterns of continuity: a dynamic model for conceptualizing the stability of individual differences in psychological constructs across the life course." Psychological review, 112, 60.

Frey, R., A. Pedroni, R. Mata, J. Rieskamp, and R. Hertwig (2017): "Risk preference shares the psychometric structure of major psychological traits," Science Advances, 3, e1701381.

Garfinkel, Irwin, L. R. and T. Smeeding (2010): "Wealth and welfare states: Is America a laggard or leader?" .

Gillen, B., E. Snowberg, and L. Yariv (2018): "Experimenting with Measurement Error: Techniques with Applications to the Caltech Cohort Study," Journal of Political Economy, 127, 1826–1863.

Golsteyn, B. and H. Schildberg-Hörisch (2017): "Challenges in research on preferences and personality traits: Measurement, stability, and inference," Journal of Economic Psychology, 60, 1 – 6, stability of Preferences.

Grubb, M. A., A. Tymula, S. Gilaie-Dotan, P. W. Glimcher, and I. Levy (2016): "Neuroanatomy accounts for age-related changes in risk preferences," Nature communications, 7, 13822.

Guiso, L., P. Sapienza, and L. Zingales (2018): "Time varying risk aversion," Journal of Financial Economics, 128, 403–421.

Hahn, M. H. and J. P. Haisken-DeNew (2013): "Panel W hiz and the Australian Longitudinal Data Infrastructure in Economics," Australian Economic Review, 46, 379–386.

Haisken-DeNew, J. P. and M. H. Hahn (2010): "PanelWhiz: Efficient Data Extraction of Complex Panel Data Sets-An Example Using the German SOEP," Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts-und Sozialwissenschaften, 130, 643–654.

Hanaoka, C., H. Shigeoka, and Y. Watanabe (2018): "Do risk preferences change? evidence from the great east japan earthquake," American Economic Journal: Applied Economics, 10, 298–330.

Hardardottir, H. (2017): "Long term stability of time preferences and the role of the macroeconomic situation," Journal of Economic Psychology, 60, 21 – 36, stability of Preferences.

Hardeweg, B., L. Menkhoff, and H. Waibel (2013): "Experimentally Validated Survey Evidence on Individual Risk Attitudes in Rural Thailand," Economic Development and Cultural Change, 61, 859–888.

Harris, R. D. and E. Tzavalis (1999): "Inference for unit roots in dynamic panels where the time dimension is fixed," Journal of Econometrics, 91, 201–226.

Haushofer, J. and E. Fehr (2014): "On the psychology of poverty," Science, 344, 862–867.

Jusko, K. (2016): "Saftey Net," in State of the Union - The Poverty and Inequality Report 2016, ed. by A. Chmielewski and S. Reardon, Stanford Center on Poverty and Inequality, 25 – 31.

Kapteyn, A. and F. Teppa (2011): "Subjective measures of risk aversion, fixed costs, and portfolio choice," Journal of Economic Psychology, 32, 564–580.

Karlan, D., R. Osei, I. Osei-Akoto, and C. Udry (2014): " Agricultural decisions after relaxing credit and risk constraints," The Quarterly Journal of Economics, 129, 597–652.

Kirby, K. N., R. Godoy, V. Reyes-Garćia, E. Byron, L. Apaza, W. Leonard, E. Perez, V. Vadez, and D. Wilkie (2002): "Correlates of delay-discount rates: Evidence from Tsimane'Amerindians of the Bolivian rain forest," Journal of Economic Psychology, 23, 291–316.

Levin, A., C.-F. Lin, and C.-S. J. Chu (2002): "Unit root tests in panel data: asymptotic and finite-sample properties," Journal of Econometrics, 108, 1–24.

Liebenehm, S. (2018): "Temporal Stability of Risk Attitudes and the Impact of Adverse Shocks—A Panel Data Analysis from Thailand and Vietnam," World Development, 102, 262–274.

Liu, E. M. (2013): "Time to change what to sow: Risk preferences and technology adoption decisions of cotton farmers in China," The Review of Economics and Statistics, 95, 1386–1403.

Liu, E. M. and J. Huang (2013): "Risk preferences and pesticide use by cotton farmers in China," Journal of Development Economics, 103, 202–215.

Markowitz, H. (1952): "Portfolio selection," The Journal of Finance, 7, 77–91.

Meier, A. N. (2022): "Emotions and risk attitudes," American Economic Journal: Applied Economics, 14, 527–58.

Meier, S. and C. D. Sprenger (2015): "Temporal stability of time preferences," Review of Economics and Statistics, 97, 273–286.

Merton, R. C. (1973): "An intertemporal capital asset pricing model," Econometrica: Journal of the Econometric Society, 867–887.

Mundlak, Y. (1978): "On the pooling of time series and cross section data," Econometrica, 46, 69–85.

OECD (2020): "Inequality (indicator)," .

Pollard, M. S. and M. Baird (2017): The RAND American Life Panel: Technical Description, RAND.

Roberts, B. and W. DelVecchio (2000): "The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies." Psychological bulletin, 126, 3.

Sahm, C. R. (2012): "How much does risk tolerance change?" The Quarterly Journal of Finance, 2.

Sakha, S. (2019): "Determinants of risk aversion over time: Experimental evidence from rural Thailand," Journal of Behavioral and Experimental Economics, 80, 184–198.

Salamanca, N., A. de Grip, D. Fouarge, and R. Montizaan (2020): "Locus of control and investment in risky assets," Journal of Economic Behavior & Organization, 177, 548–568.

Schildberg-Hörisch, H. (2018): "Are Risk Preferences Stable?" Journal of Economic Perspectives, 32, 135–54.

Schupp, J. and G. G. Wagner (2002): "Maintenance of and innovation in long-term panel studies: The case of the German Socio-Economic Panel (GSOEP)," AStA Advances in Statistical Analysis, 86, 163–175.

Shou, Y., M. Sellbom, and H.-F. Chen (2022): "Fundamentals of Measurement in Clinical Psychology," in Comprehensive Clinical Psychology (Second Edition), ed. by G. J. Asmundson, Oxford: Elsevier, 13–35, second edition ed.

Shum, P. and M. Faig (2006): "What explains household stock holdings?" The Journal of Banking & Finance, 30, 2579–2597.

Stigler, G. J. and G. S. Becker (1977): "De gustibus non est disputandum," The American Economic Review, 67, 76–90.

Summerfield, M., A. Bevitt, S. Freidin, M. Hahn, N. La, N. Macalalad, M. O'Shea, N. Watson, R. Wilkins, and M. Wooden (2017): "HILDA user manual–release 16," Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Teppa, F., C. Vis, et al. (2012): "The CentERpanel and the DNB household survey: Methodological aspects," Tech. rep., Netherlands Central Bank, Research Department.

Wagner, G. G., J. R. Frick, and J. Schupp (2007): "The German Socio-Economic Panel Study (SOEP): Scope, Evolution and Enhancements," SOEPpapers on Multidisciplinary Panel Data Research 1, DIW Berlin, The German Socio-Economic Panel (SOEP).

Warneryd, K. (1996): "Risk attitudes and risky behavior," Journal of Economic Psychology, 17, 749–770.

Watson, D. (2004): "Stability versus change, dependability versus error: Issues in the assessment of personality over time," Journal of Research in Personality, 38, 319–350.

Weber, M., E. U. Weber, and A. Nosić (2012): "Who takes risks when and why: Determinants of changes in investor risk taking," Review of Finance, 17, 847–883.

Wooldridge, J. M. (2010): Econometric analysis of cross section and panel data, MIT press.

World Bank (2018): The State of Social Safety Nets, Washington , D.C: World Bank.

Table 2: OLS estimates of coefficient on past preferences parameter: Developed countries

	Australia		Germany		Netherlands		United States	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coefficient on past preferences (δ)	0.571	0.520	0.586	0.558	0.681	0.660	0.515	0.469
	(0.005)	(0.006)	(0.004)	(0.004)	(0.008)	(0.008)	(0.030)	(0.032)
$H_0 : \delta = 1$	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Controls	No	Yes	No	Yes	No	Yes	No	Yes
$H_0 : \beta = 0$		[0.00]		[0.00]		[0.000]		[0.001]
R^2	0.32	0.35	0.34	0.35	0.46	0.47	0.25	0.30
Observations	67,536	67,536	82,789	82,789	21,645	21,645	1,058	1,058

Notes: Estimates based on an OLS model. Standard errors in parentheses and p-values in brackets. $H_0 : \beta = 0$ refers to the test of joint significance for the control variables.

Table 3: Estimates of risk preference stability in developed economies based on model in Equation (4)

	Australia		Germany		Netherlands		United States	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coefficient on past preferences (δ)	0.962	0.947	0.994	0.991	0.964	0.961	0.878	0.869
	(0.007)	(0.008)	(0.009)	(0.008)	(0.006)	(0.007)	(0.037)	(0.040)
$H_0 : \delta = 1$	[0.000]	[0.000]	[0.215]	[0.337]	[0.000]	[0.000]	[0.001]	[0.001]
Idiosyncratic var. (σ_η^2)	0.078	0.080	0.074	0.074	0.063	0.062	0.324	0.285
Measurement error var. (σ_ε^2)	0.374	0.374	0.357	0.357	0.290	0.291	0.175	0.227
Noise to signal ratio (s)	0.640	0.640	0.601	0.601	0.408	0.410	0.245	0.342
Controls	No	Yes	No	Yes	No	Yes	No	Yes
$H_0 : \beta = 0$		[0.000]		[0.000]		[0.000]		[0.000]
R^2	0.17	0.19	0.18	0.18	0.38	0.39	0.19	0.24
Observations	67,536	67,536	82,789	82,789	21,645	21,645	1,058	1,058

Notes: Variance estimates based on (6). Standard errors in parentheses and p-values in brackets. $H_0 : \beta = 0$ refers to the test of joint significance for the control variables.

Table 4: OLS estimates of coefficient on past preferences parameter: Developing countries

	Kyrgyzstan		Malawi		Thailand		Vietnam	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coefficient on past preferences (δ)	0.438	0.400	0.190	0.186	0.391	0.339	0.448	0.383
	(0.011)	(0.011)	(0.017)	(0.017)	(0.300)	(0.311)	(0.276)	(0.292)
$H_0 : \delta = 1$	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Controls	No	Yes	No	Yes	No	Yes	No	Yes
$H_0 : \beta = 0$		[0.00]		[0.00]		[0.00]		[0.00]
R^2	0.18	0.21	0.04	0.07	0.00	0.07	0.01	0.07
Observations	9,400	9,400	3,367	3,367	1,401	1,401	1,619	1,619

Notes: Estimates based on an OLS model. Standard errors in parentheses and p-values in brackets. $H_0 : \beta = 0$ refers to the test of joint significance for the control variables. Estimates of δ are annualized as risk preferences are observed time intervals other than 1 year: Malawi (three months), Thailand and Vietnam (three years).

Table 5: Estimates of risk preference stability in developing economies based on model in Equation (4)

	Kyrgyzstan		Malawi		Thailand		Vietnam	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coefficient on past preferences (δ)	0.853	0.867	0.296	0.301	0.718	0.548	0.489	0.477
	(0.028)	(0.036)	(0.266)	(0.269)	(0.134)	(0.325)	(0.109)	(0.206)
$H_0 : \delta = 1$	[0.000]	[0.000]	[0.000]	[0.000]	[0.035]	[0.164]	[0.000]	[0.000]
Idiosyncratic var. (σ_η^2)	0.388	0.367	6.040	5.295	0.214	0.561	1.849	1.014
Measurement error var. (σ_ε^2)	0.370	0.370	0.371	0.408	0.766	0.740	0.172	0.426
Noise to signal ratio (s)	0.651	0.650	0.634	0.745	3.370	2.928	0.296	1.303
Controls	No	Yes	No	Yes	No	Yes	No	Yes
$H_0 : \beta = 0$		[0.000]		[0.000]		[0.000]		[0.000]
R^2	0.019	0.029	0.04	0.07	0.00	0.078	0.00	0.09
Observations	9,400	9,400	3,367	3,367	1,751	1,751	1,767	1,767

Notes: Variance estimates based on (6). Standard errors in parentheses and p-values in brackets. $H_0 : \beta = 0$ refers to the test of joint significance for the control variables. Estimates of δ are annualized as risk preferences are observed time intervals other than 1 year: Malawi (three months), Thailand and Vietnam (three years).

Table 6: Restricted sample of the estimates of risk preference stability in developed economies based on model in Equation (4)

	Australia		Germany		Netherlands	
	(1)	(2)	(3)	(4)	(5)	(6)
Coefficient on past preferences (δ)	0.969	0.934	0.993	0.992	0.972	0.964
	(0.019)	(0.022)	(0.004)	(0.013)	(0.008)	(0.031)
$H_0 : \delta = 1$	[0.000]	[0.000]	[0.087]	[0.506]	[0.000]	[0.242]
Idiosyncratic var. (σ_η^2)	0.053	0.055	0.021	0.109	0.080	0.127
Measurement error var. (σ_ε^2)	0.360	0.395	0.389	0.296	0.212	0.234
Noise to signal ratio (s)	0.597	0.775	0.763	0.442	0.279	0.324
Controls	No	No	No	No	No	No
Consecutive Lags	No	Yes	No	Yes	No	Yes
R^2	0.17	0.13	0.08	0.28	0.38	0.41
Observations	10,064	11,644	22,742	13,908	2,589	1,240

Notes: In odd numbered columns we restrict the sample to have the same lag structure as Vietnam and Thailand, using 3 waves that are two and three years apart. In even numbered columns, results are based on a sample restricted to three consecutive waves (2015, 2014 and 2013). Variance estimates are based on (6). Standard errors in parentheses and p-values in brackets.

Table 7: Adaptation of the Harris and Tzavalis (1999) panel unit root test

		OLS			IV					
					Australia					
Panel size(T)	Obs	δ	z	p-val	δ	(σ_η^2)	(σ_ε^2)	s	z	p-val
3	4479	0.430	-22.204	0.001	0.992	0.058	0.510	0.961	-0.546	0.293
4	11060	0.487	-46.494	0.001	0.960	0.071	0.402	0.757	-7.263	0.001
5	18325	0.507	-73.050	0.001	0.913	0.108	0.363	0.650	-28.739	0.001
6	4800	0.577	-40.523	0.001	0.937	0.119	0.320	0.463	-13.815	0.001
7	20307	0.581	-93.392	0.001	0.949	0.060	0.309	0.522	-27.882	0.001
				Germany						
3	12060	0.538	-27.738	0.001	0.979	0.094	0.383	0.626	-2.303	0.011
4	4576	0.577	-23.483	0.001	0.995	0.091	0.345	0.539	-0.560	0.288
5	5375	0.543	-36.027	0.001	0.992	0.063	0.375	0.622	-1.518	0.065
6	50520	0.567	-119.710	0.001	1.013	0.054	0.347	0.613	9.066	0.001
				Netherlands						
3	1359	0.631	-8.240	0.001	0.924	0.120	0.270	0.415	-2.798	0.003
4	1300	0.606	-11.578	0.001	0.940	0.032	0.341	0.538	-3.728	0.001
5	1220	0.651	-13.847	0.001	0.987	0.039	0.307	0.476	-1.120	0.131
6	1290	0.652	-15.961	0.001	0.956	0.081	0.289	0.389	-5.035	0.001
7	1323	0.674	-17.807	0.001	0.931	0.131	0.249	0.337	-9.651	0.001
8	1280	0.607	-22.020	0.001	0.991	0.026	0.334	0.520	-1.486	0.069
9	1071	0.627	-20.491	0.001	0.959	0.084	0.315	0.456	-7.134	0.001
10	1150	0.619	-24.039	0.001	0.967	0.044	0.308	0.489	-6.675	0.001
11	1144	0.604	-26.385	0.001	0.958	0.016	0.348	0.507	-9.596	0.001
12	660	0.460	-26.489	0.001	0.983	0.012	0.384	0.935	-3.280	0.001
13	1287	0.623	-28.859	0.001	0.992	0.015	0.283	0.550	-2.290	0.011
14	1120	0.726	-21.129	0.001	0.982	0.027	0.254	0.313	-5.298	0.001
15	660	0.593	-21.944	0.001	0.955	0.042	0.289	0.498	-11.089	0.001
16	1056	0.727	-20.185	0.001	0.970	0.055	0.228	0.274	-9.990	0.001
17	714	0.709	-19.600	0.001	1.008	0.030	0.252	0.354	2.402	0.008
18	468	0.734	-12.706	0.001	1.007	0.047	0.252	0.291	1.855	0.032
19	247	0.531	-18.538	0.001	0.888	0.133	0.318	0.681	-21.786	0.001
20	240	0.769	-6.770	0.001	0.970	0.044	0.163	0.170	-6.146	0.001
21	189	0.429	-12.146	0.001	1.007	0.017	0.300	0.483	1.320	0.093
22	110	0.215	-10.168	0.001	2.138	0.115	0.143	0.392	172.920	0.001
23	253	0.757	-8.218	0.001	0.982	0.050	0.181	0.202	-4.382	0.001

Note: Harris and Tzavalis (1999) panel unit root tests for various balanced panels. Cross-sectional means the effect of controls (e.g., income) subtracted from each preference to decrease their effect on the test, following Levin et al. (2002). z-statistics are calculated assuming an asymptotic variance of $2[(T-1)(T-2)]^{-1}$. σ_η^2, σ_ε^2, and s calculated via non-linear least squares using information from 2 lags and netting out $X_{i,t}$.

Table 8: Estimates of risk preference stability by income of the top 50% and bottom 50% in developed economies based on model in Equation (4)

	Australia		Germany		Netherlands		United States	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Bottom 50%	Top 50%	Bottom 50%	Top 50%	Bottom 50%	Top 50%	Bottom 50%	Top 50%
Coefficient on past preferences (δ)	0.948	0.964	0.979	0.999	0.974	0.970	0.839	0.893
	(0.012)	(0.009)	(0.011)	(0.009)	(0.013)	(0.008)	(0.081)	(0.040)
$H_0 : \delta = 1$	[0.00]	[0.00]	[0.057]	[0.916]	[0.048]	[0.000]	[0.047]	[0.008]
Idiosyncratic var. (σ_η^2)	0.085	0.073	0.069	0.061	0.048	0.069	0.388	0.256
Measurement error var. (σ_ε^2)	0.368	0.380	0.371	0.303	0.313	0.266	0.247	0.182
Noise to signal ratio (s)	0.948	0.964	0.670	0.550	0.511	0.346	0.368	0.277
Controls	No	No	No	No	No	No	No	No
R^2	0.10	0.19	0.14	0.22	0.29	0.44	0.09	0.21
Observations	33073	34463	21622	24467	8118	10943	332	726

Notes: In this table we split income and estimate the top and bottom 50% by income. Variance estimates based on (6). Standard errors in parentheses and p-values in brackets.

Table 9: Estimates of risk preference stability by income of the top 50% and bottom 50% in developing economies based on model in Equation (4)

	Kyrgyzstan		Thailand	
	(1)	(2)	(3)	(4)
	Bottom 50%	Top 50%	Bottom 50%	Top 50%
Coefficient on past preferences (δ)	0.826	0.890	0.561	0.840
	(0.042)	(0.038)	(0.273)	(0.176)
$H_0 : \delta = 1$	[0.00]	[0.00]	[0.107]	[0.362]
Idiosyncratic var. (σ_η^2)	0.373	0.400	0.729	0.111
Measurement error var. (σ_ε^2)	0.388	0.348	0.655	0.770
Noise to signal ratio (s)	0.755	0.556	1.961	3.441
Controls	No	No	No	No
R^2	0.00	0.03	0.01	0.01
Observations	4444	4854	869	882

Notes: In this table we split income and estimate the top and bottom 50% by income. Variance estimates based on (6). Standard errors in parentheses and p-values in brackets.

Table A1: Summary Statistics: Risk

					Percentiles			
Country	N	Mean	SD	Min	25th	50th	75th	Max
Developed Countries								
Australia	67535	1.62	0.69	1	1	2	2	4
Germany	82789	4.57	2.28	0	3	5	6	10
Netherlands	21645	5.24	1.04	1	4.5	5.33	6	7
United States	1058	4.34	2.22	0	3	5	6	10
Developing Countries								
Kyrgyzstan	9400	4.80	3.04	0	2	5	7	10
Malawi	3367	3.65	1.74	1	2	4	5	6
Thailand	1751	4.71	2.84	0	3	5	6	10
Vietnam	1767	5.89	2.6	0	5	6	8	10

Notes: This table reports summary statistics for the non standardised measure of risk for the estimating sample.

Table A2: Estimates of risk preference stability by income of the top 30% and bottom 30% in developed economies based on model in (4)

	Australia		Germany		Netherlands			United States
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Bottom 30%	Top 30%	Bottom 30%	Top 30%	Bottom 30%	Top 30%	Bottom 30%	Top 30%
Coefficient on past preferences (δ)	0.944	0.960	0.973	1.003	0.969	0.970	0.702	0.956
	(0.015)	(0.011)	(0.015)	(0.010)	(0.019)	(0.011)	(0.165)	(0.060)
$H_0 : \delta = 1$	[0.000]	[0.000]	[0.079]	[0.778]	[0.105]	[0.005]	[0.071]	[0.459]
Idiosyncratic var. (σ^2_η)	0.062	0.072	0.080	0.046	0.066	0.074	1.169	0.211
Measurement error var. (σ^2_ε)	0.365	0.375	0.380	0.295	0.312	0.257	0.00	0.137
Noise to signal ratio (s)	0.818	0.611	0.679	0.541	0.529	0.327	0.000	0.207
Controls	No	No	No	No	No	No	No	No
R^2	0.08	0.20	0.13	0.24	0.26	0.45	0.15	0.22
Observations	20768	21352	12036	14923	4282	6673	165	313

Notes: In this table we split the sample by income and estimate the model for the top and bottom 30% of the distribution. Variance estimates based on (6). Standard errors in parentheses and p-values in brackets.

Table A3: Estimates of risk preference stability by income of the top 30% and bottom 30% in developing economies based on model in Equation (4)

	Kyrgyzstan		Thailand	
	(1)	(2)	(3)	(4)
	Bottom 30%	Top 30%	Bottom 30%	Top 30%
Coefficient on past preferences (δ)	0.873	0.869	0.731	0.878
	(0.060)	(0.046)	(0.180)	(0.219)
$H_0 : \delta = 1$	[0.66]	[0.004]	[0.136]	[0.577]
Idiosyncratic var. (σ_η^2)	0.280	0.403	0.195	0.144
Measurement error var. (σ_ε^2)	0.362	0.366	0.786	0.697
Noise to signal ratio (s)	0.721	0.580	3.425	2.109
Controls	No	No	No	No
R^2	0.00	0.05	0.01	0.01
Observations	2647	2966	521	529

Notes: In this table we split income and estimate the model for the top and bottom 30% of the distribution. Variance estimates based on (6). Standard errors in parentheses and p-values in brackets.