



PhD Thesis

PhD Candidate: Siria Angino

The Importance of Being Clustered

PhD in Economics

XXIX Cycle

LUISS Guido Carli

Supervisor: Prof. Giuseppe Ragusa

DISCLAIMER

This PhD thesis by Siria Angino is submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics. It may be freely reproduced fully or partially, with citation of the source. This is without prejudice to the rights of LUISS Guido Carli University to reproduction for research and teaching purposes, with citation of the source.

*To my students,
for that sparkle of understanding in their eyes
made every effort worth it,*

*and to Elena
for her deep friendship,
and wise stubbornness
that pushed me towards
my most precious treasures.*

Contents

Introduction	9
I A Few-Cluster-Robust Test for Weak Instruments	10
1 Introduction	10
2 Clusters, Few Clusters	13
3 Instrumental Variables, Weak Instruments	17
4 Weak Instruments and Few Clusters	23
5 Empirical Example	27
6 Conclusions	30
7 Appendix	31
8 Bibliography	43
II Sharing or gambling over losses?	45
1 Introduction	45
2 Literature Review	47
3 Experimental Design	53
4 Results	56
5 Conclusions	72

6	Appendix	74
7	Bibliography	85

Introduction

The two papers that constitute the thesis you are about to read are very different. One is a methodological paper that presents a simple yet effective way to perform inference in IV models when data are clustered in few groups; the second investigates experimentally how individuals prefer to allocate a loss, in particular when they can decide either to share it or to gamble and avoid it completely.

As you can immediately see, any attempt to link tightly these papers would be a stretch. There is however a thin thread between the two, the one that brought to the name on the title page.

If you recognize the poet's words as true, and no man is an island¹, then the idea of clustering will sound sensible to your ears. When clustering, being it by correcting standard errors in a regression, or identifying more heterogeneous groups, we acknowledge that observations are similar in some sense, connected, and some of them are more tightly so.

This feature can complicate the analysis, or it can unveil new patterns in the aggregate results. Whatever the valence of this fact, it remains a fact. Being aware of it may lead to more accurate, meaningful findings, and disregarding or even underestimating these links comes with an expensive price.

Of course, it is not just about individuals. Firms, communities, cities, each of them is unique. And yet, under the right perspective, they can be grouped and form something more homogeneous.

In the papers that follow, despite their differences, taking care of clusters is essential.

Thank you in advance for the attention you will dedicate to this work.

¹Meditation XVII, John Donne

I A Few-Cluster-Robust Test for Weak Instruments

with G. Ragusa

Abstract

In the usual IV regression models, the quality of estimation and inference dramatically depends on the relevance of the set of instruments. To test such condition, most empiricists rely on the first-stage F -statistic as suggested by Stock and Yogo (2005a). However, this method is not valid for clustered data, especially when there are few clusters.

Recently, weak instrument-robust methods for inference have been proposed, but none of them deals with the latter case. In this paper, we propose a simple method to perform inference which is robust both to the presence of weak instruments and few clusters in the case of a single endogenous regressor.

Keywords: Instrumental variable, Weak Instrument, Few Clusters, Wild Bootstrap T-Procedure

1 Introduction

Consciously or not, empirical research often features regressions with clustered data. Observations in a given cluster are correlated, either for observable factors, unobservable factors or both, which causes the failure of the “independently distributed” assumption. On the other hand, observations in different clusters are

deemed to be uncorrelated. Examples of clusters are geographical areas (regions, states) and industrial sectors.

The relatedness of the data within each cluster needs to be taken into account. If data are treated as i.i.d., the quantity of information available to the researcher is overestimated. As a consequence, the usual standard errors of the coefficients are incorrect - typically too small - and inference is invalid.

Methodologies to consistently estimate the covariance matrix of the coefficients for these cases have been developed. There is however a strong caveat: consistency is achieved only if the number of clusters tends to infinity with respect to the cluster size. In practice, this roughly translates in having more than 50 clusters².

Such condition may fail in applied research. The Global Industry Classification Standard (GICS) clusters firms in 10 sectors, the International Standard Industrial Classification (ISIC) in 12³. Entities in the EU, as citizens as firms, can be grouped at member state level, 27 after 2016 Brexit⁴. The Public Republic of China administers 33 provincial-level divisions⁵. The USA Census Bureau defines four statistical regions, with nine divisions⁶.

In all these cases, even though a cluster-robust matrix is used, it remains underestimated and the null hypothesis is over-rejected. Bootstrapping can alleviate or even solve the few-cluster problem (Cameron and Miller, 2015), but it is not readily extendable to frameworks different from OLS.

The clustered structure of the data makes testing the relevance condition in IV settings problematic. Practitioners usually rely on the $\hat{F} > 10$ Rule of Thumb by Staiger and Stock (SS, 1994), but this test is unreliable in case of few clusters

²This threshold is to be increased in case of unbalanced clusters - i.e., differently sized.

³See msci.com/gics for GICS and unstats.un.org/unsd/cr/registry/isic-4.asp for ISIC.

⁴The European NUTS (Nomenclature of territorial units for statistics) subdivides the economic territory of the EU into regions at three different levels. NUTS-1 divides Germany in 16 regions (corresponding to the Länder), Italy in 5, Spain in 7, France in 9, UK in 12. The maximum number of regions for a state in NUTS-2 is reached by Germany (39), with all other states divided in less than 30 regions each. See "Eurostat-European Commission. Regions in the European Union. Nomenclature of territorial units for statistics.", NUTS 2010/EU-27, Publications Office of the European Union, 2011.

⁵See China Statistical Yearbook 2015.

⁶See census.gov/econ/census/help/geography/regions_and_divisions.html

for two reasons. First, the first-stage estimated F -statistic is upward-biased, as a result of the underestimation of the covariance matrix of the coefficients. Second, the use of the thresholds tabled in Stock and Yogo (SY, 2005a), derived under i.i.d. and homoskedastic assumptions, is not justified, neither theoretically nor computationally.

The alternative practitioners face is the use of weak instrument-robust inference. However, the presence of few clusters still poses a threat to the quality of the results.

A survey of papers published from 2010 to 2016 in the issues of leading applied journals (the Review of Economics and Statistics, the American Economic Journal, the American Economics Review, the Journal of Political Economy) reveals that 14 papers include IV estimation together with a small number of clusters (from 8 to 57, with a median of 47.5)⁷. The first-stage F s range from 9.74 to 89.5, with a median of 17.42. Figure 1 plots the number of clusters in each paper against the first-stage $\hat{F}$ ⁸.

There were 28 papers using IV with a high number of clusters (from 70 to 49343, and a median of 313)⁹. As explained below, clustering at a low level may fail to take into account correlations among observations, resulting once again in wrong estimates for the covariance matrix of the coefficients. Clustering at a higher levels would thus be recommended in theory: in practice, up to now, no suitable tool was available to deal with few clusters in the IV framework.

In this paper, we present a simple weak instrument-robust procedure that relies on a combination of OLS and bootstrapping to produce adequate confidence intervals also in presence of few clusters. In particular, we start from the test by

⁷Papers that include the words “instrument”, “IV” or “two stage least squares/TSLS” in the abstract were first selected; hence, the use of clusters was checked. As a result of this procedure, our estimate for the number of papers including IV and few clusters is conservative.

⁸Three papers do not report such statistic or report a p-value equal to 0, and are not included in the graph.

⁹Besides them, ten papers include an unclear number of clusters, while 5 use two-way clustering.

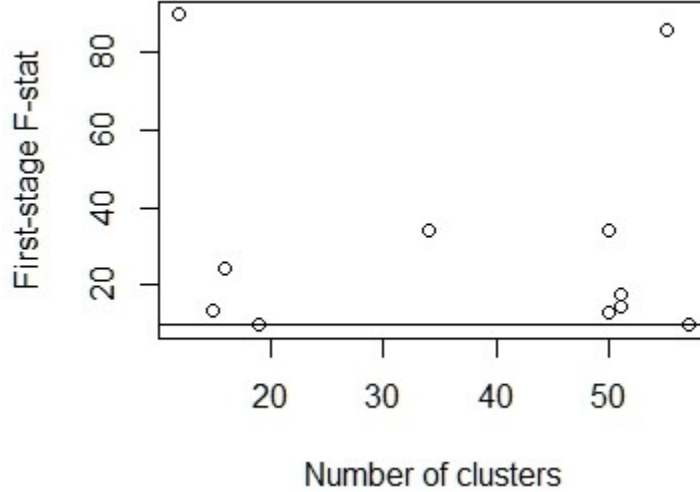


Figure 1: Number of clusters vs estimated first-stage F -statistic

Chernozhukov and Hansen (2008) and apply the Cluster Bootstrap-T procedure as in Cameron, Gelbach and Miller (2008).

The rest of the paper is organized as follows. Clusters are presented in Section 2. The linear IV regression model and the weak instrument problem are reviewed in Section 3, together with weak instrument-robust inference. Section 4 presents the joint problem, our method and simulation results. In Section 5, we reconsider two papers, “The China syndrome: Local labor market effects of import competition in the United States”, by Autor, Dorn and Hanson (2013), and “Soil Endowments, Female Labor Force Participation and the Demographic Deficit of Women in India”, by E. Carranza (2014), and apply our method. Section 6 concludes.

2 Clusters, Few Clusters

The OLS regression assumes spherical errors, which translates in two requirements: homoskedasticity and no cross-observation correlation. Clustered data are a blunt transgression to both.

With microdata, it is rather natural to identify or hypothesize the presence

of groups, or *clusters*. Within each cluster, observations are correlated among each other. Entities in different clusters, instead, are completely uncorrelated. The definition of *cluster*¹⁰ additionally requires that each entity in the population belongs to one and only one cluster (i.e. all entities are in a cluster, with no crossovers).

Examples of clusters are geographical areas, such as states, regions or cities, occupations, industrial sectors. Panel data are an example of cluster data, where each entity is a cluster: observations for the same entity in different times are correlated, while observations for different entities are deemed to be uncorrelated.

If our data are drawn from “populations with a grouped structure” (Moulton, 1986), observations are correlated either for observable or unobservable characteristics. Consider the following model:

$$y_{ig} = \beta x_{ig} + v_{ig} \quad (1)$$

$$= \beta x_{ig} + \alpha_g + u_{ig} \quad (2)$$

$$g = 1, 2, \dots, G \quad i = 1, 2, \dots, N$$

In this model, G is the total number of clusters and N is the size of each of them, which is constant. The total sample size is thus $S = NG$.

As in Moulton, the error term v_{ig} is additive and can be divided in two parts: α_g is the unobserved cluster effect, u_{ig} is the idiosyncratic error. The unobserved cluster effect is the same for all the observations in a given cluster.

The regressor x_{ig} can either be totally individual-specific, cluster-fixed¹¹ or a mixture of the two.

If the clustered nature of the data is not taken into account, the uncertainty

¹⁰In this paper, we do not consider spatial correlation. In particular, we assume equi-correlated observations within each cluster.

¹¹Some variables may be inserted in the regression at aggregate (cluster) level; for example, employment growth at state level can help explaining individual wages.

embedded in the regression is underestimated, as the set of available information is considered wider than the actual one. The $\hat{\beta}$ stays unbiased and consistent under the usual assumptions, but its covariance matrix does not. In particular, the standard errors for $\hat{\beta}$ are underestimated, which translates in non-conservative inference results.

Luckily, cluster-robust estimators for the covariance matrix are available, and popular statistical packages include them¹². Unluckily, these robust estimators are consistent only if the number of clusters goes to infinity with respect to the cluster size.

With a limited number of clusters (there is general consensus in defining “few” less than 50 clusters) the covariance matrix is again underestimated and Wald tests generally over-reject.

Cameron and Miller (2015) identify three methods to perform inference in these cases:

- Bootstrapping at cluster level with asymptotic refinements;
- Correcting the cluster-robust covariance matrix (for example, multiplying standard errors by $\sqrt{\frac{G}{G-1}}$);
- Considering more adequate critical levels, for example those of a t -student distribution with $G - 1$ degrees of freedom.

We focus on the first methodology. In particular, we employ the Wild Cluster Bootstrap T-Procedure (WCBT). Bootstrapping provides asymptotic refinement for asymptotically pivotal statistics, that is, statistics whose asymptotic distribution does not depend on any unknown parameter, as it is for Wald tests.

We illustrate the WCBT procedure in a simple example. Suppose the interest lies in testing the null hypothesis $H_0 : \beta = \beta_0$ in equation 1 at α significance level.

¹²In Stata, *ivreg2* allows for cluster-robust estimator of the covariance matrix in both stages, while *ivreg* only clusters in the second stage.

As this is a scalar, we apply bootstrap to a t-test. All independent variables except for x are partialled out.

1. Estimate the model and compute the t-statistic $t_b(\beta) = \frac{\hat{\beta} - \beta_0}{se(\hat{\beta})}$.
2. Form the residuals imposing the null hypothesis: $\tilde{v}_{ig} = y_{ig} - \beta_0 x_{ig}$.
3. Apply the Wild Bootstrap T-Procedure B times:
 - (a) For each $g = 1, 2, \dots, G$, build $\tilde{v}_{ig}^* = \tilde{v}_{ig}$ or $\tilde{v}_{ig}^* = -\tilde{v}_{ig}$ with equal probability, hence $\tilde{y}_{ig}^* = \beta_0 x_{ig} + \tilde{v}_{ig}^*$.
 - (b) Regress $\tilde{y}_{ig}^*$ on x_{ig} and store $t_b^*(\beta)$ for $\beta = \beta_0$.
4. Compute the $1 - \alpha$ empirical quantile of $t_b^*(\beta)$, that we call $q_{1-\alpha} t_b^*(\beta)$.
5. Check whether $|t_b(\beta)| < q_{1-\alpha} t_b^*(\beta)$.

This method works fine for $G \geq 10$. For fewer clusters, Webb (2013) suggests to weight the residuals $\tilde{v}_{ig}$ with a six-point distribution $\{-\sqrt{1.5}, -\sqrt{1}, -\sqrt{0.5}, \sqrt{0.5}, \sqrt{1}, \sqrt{1.5}\}$, rather than the two-points Rademacher distribution $\{-1, 1\}$.

The requirement of a large number of clusters (with respect to the cluster size) may seem similar to the large sample one. Actually, the solutions to these issues - if any - are sensibly different.

The crucial point is that the true number of clusters (the “cluster population”) is fixed by nature and datasets generally sample from the whole cluster population; i.e., they contain observations from each cluster, so that we cannot add any other¹³.

¹³Assume that we have data for individuals in each US states. Individuals within each state are correlated, while there is no correlation between observations in different states. In other words, the correct clustering is at the state level. In order to increase the number of clusters, we should have/create an additional state.

Adding observations within each cluster does neither solve nor alleviate the problem. Actually, a larger cluster size aggravates it, creating a trade-off with the sample size, that we still want as large as possible.

So, while it is at least conceivable to collect more entity observations, including observations from new clusters once all of them have been covered is impossible.

Could one increase the number of clusters choosing a “lower level” to cluster at? For example, if one decides to cluster at US county-level, instead of the state-level, the number of clusters would be 3144 instead of 50.

However, one must recognize that there is a correct level to cluster at. The risk of clustering at a low level is to leave unaccounted for some of the correlation between observations. This, in turn, brings us back to an incorrect estimation of the uncertainty about our model¹⁴.

In practice, it is common to cluster at progressively broader levels and stop when the changes in the covariance matrix are relatively small.

As a final remark, if the regressor of interest is i.i.d. (randomly assigned within cluster) and consequently the within-cluster correlation is 0, clustering is unnecessary, even if the unobserved cluster effect is present (i.e., the within-cluster (or *intracluster*) correlation of v_{ig} is not 0).

3 Instrumental Variables, Weak Instruments

The violation of the OLS assumption of exogeneity is probably the most common and serious challenge in econometric analysis. The presence of omitted variables, measurement errors and/or simultaneous causality leads in fact to inconsistent estimators.

¹⁴Consider two individuals, one from Josephine County, the other from Curry County, both in Oregon. These counties are adjacent. Clustering at county-level, we do not account for the existence of possible unobservable and observable characteristics these two individuals share, such as culture and economic system. This is better done by clustering at state-level.

In IV models, a set of instruments is used to disentangle an exogenous variation in the problematic regressor of interest, x , and estimate its causal effect on the dependent variable y . The endogeneity problem is thus solved using the IV model and the TSLS estimation, provided that two conditions are satisfied.

The exogeneity condition requires the set of instruments to be uncorrelated with the regression error term or, said differently, to affect the dependent variable only through the endogenous regressor.

The second condition, relevance, demands the set of instruments and the endogenous regressor to be correlated instead, once all other exogenous variables have been taken into account.

The presence of both conditions ensures consistency of the IV estimator. However, a third requirement guarantees its good properties: the set of instruments and the endogenous regressor must be “strongly” correlated.

Weak correlation produces a bias in the same direction as the OLS one and a generally non-normal distribution, also in case of large samples, see Bound, Jaeger and Baker (1995). This, in turn, complicates inference procedures: point estimates, hypothesis testing and confidence intervals are unreliable.

To check whether the available set of instruments is “good enough” to obtain desirable properties for β^{IV} , the most common test in applied econometrics dates back to Staiger and Stock (1994) and finds its final form in Stock and Yogo (SY, 2005a). It relies on the first-stage F -statistic, since this is tightly correlated to the strength of the instruments, which is measured by the concentration parameter μ^2 .

For an IV model with i.i.d. observations and M instruments:

$$y_i = \beta x_i + v_i \tag{3}$$

$$x_i = z_i \gamma + \xi_i \tag{4}$$

the concentration parameter is defined as:

$$\mu^2 = \frac{\gamma' Z' Z \gamma}{\sigma_\xi}$$

where Z is the $(NG) \times M = S \times M$ instrument matrix. Consider now the F -statistic for the hypothesis that the whole coefficient vector is zero, $\gamma = 0$.

$$F = \frac{\hat{\gamma}' Var(\hat{\gamma})^{-1} \hat{\gamma}}{M}$$

Under homoskedasticity, MF has a non-central χ^2 distribution with M degrees of freedom and non-centrality parameter μ^2 . If the sample size S is large, the estimated F -statistic, $\hat{F}$, and the true one, F , are close. It follows that

$$\hat{F} \simeq \frac{\mu^2}{M}$$

Higher values of F are thus associated with stronger instruments. In SY, the estimated F -statistic is compared with a suitable threshold, which depends on the number of instruments and the quality of the β^{tols} properties the researcher desires. If $\hat{F}$ is bigger than the threshold, the set of instruments is deemed “strong enough”.

As the readers probably acknowledge, the SY test is often reduced in practice to the well-known rule of thumb where the threshold is fixed at 10.

SY also offered two working definitions of “weakness” for the set of instruments. One is framed in terms of the relative bias of the IV estimator with respect to the OLS one, while the other concerns the size of a Wald test based on the IV estimator.

The bias definition compares the bias of the IV estimator with respect to the OLS one, both taken in absolute terms.

$$B_S = \frac{|E[\beta^{\hat{IV}}] - \beta|}{|E[\hat{\beta}] - \beta|}$$

Call B the asymptotic relative bias: $B = \lim_{s \rightarrow \infty} B_S$.

Consider now the worst case, which means that the highest possible degree of endogeneity, measured by ρ , is set. The maximal asymptotic relative bias is $B^{max} =$

$\max_{\rho} B$. A weak instrument set consists of instruments that potentially leads to a maximal asymptotic relative bias¹⁵ greater than some value b ¹⁶.

A set of instruments is strong under the size definition if the size of the Wald test is close to its nominal value α for all possible configurations of the IV regression model. In our case, with a single endogenous regressor, a simple t-test is sufficient. Define the rejection rate R_S :

$$R_S = Pr_{\beta_0}[|t^{IV}| > z_{\alpha/2}]$$

where $z_{\alpha/2}$ is the α – *level* critical value of the standard normal distribution. Call R the asymptotic rejection rate, $R = \lim_{s \rightarrow \infty} R_S$. The rejection rate depends on the degree of endogeneity, ρ , and SY consider the worst-case scenario, $R^{max} = \max_{\rho} R$. A weak set of instruments consists of instruments that potentially lead to a maximal rejection rate greater than some value r for a test with nominal value α ¹⁷.

3.1 $\hat{F} > 10$ Rule of Thumb

In the light of what has been presented in the previous paragraph, linking the first-stage F -statistic and the properties of the $\beta^{\hat{IV}}$ distribution easily follows. There exists some threshold for F such that, above it, the relative bias of $\beta^{\hat{IV}}$ is smaller than b or such that the actual size of the an α -sized test does not exceed r , with a given confidence level (SY uses 95%).

Once the researcher fixes b or r , depending on the characteristics of the $\beta^{\hat{IV}}$

¹⁵The maximal IV relative bias is also the maximal absolute bias in standardized units (see SY for further details).

¹⁶Formally, for $\mathcal{Z} = \{\gamma, \sigma_v^2, \sigma_{\xi}^2\}$, a set of instruments is weak under the relative bias definition if:

$$\mathcal{W}_{bias} = \{\mathcal{Z} : B^{max} \geq b\}$$

¹⁷Formally, for $\mathcal{Z} = \{\gamma, \sigma_v^2, \sigma_{\xi}^2\}$, a set of instruments is weak under the size definition if:

$$\mathcal{W}_{bias} = \{\mathcal{Z} : R^{max} \geq r\}$$

distribution one is interested in and the degree of desired conservatism, the first-stage F -statistic is calculated and compared with the thresholds tabled in SY. If $\hat{F}$ is greater than this threshold, the set of instruments is considered strong.

Under the bias definition, for $b = 10\%$, the threshold is always smaller than 11.50, regardless of the number of instruments. Since the IV tradition had already taken 10 as a good threshold, and this is close to 11.50, the rule of thumb used in practice is often limited to checking whether $\hat{F} > 10$. For a smaller bias, $b = 5\%$, the threshold mainly doubles.

The size definition is more demanding in terms of F -statistic. For a maximal size of 10%, given a nominal value of the test $\alpha = 5\%$, the threshold increases non-linearly in the number of instrument - 16.38 with $M = 1$, 19.93 for $M = 2$, 38.54 for $M = 10$, 62.30 for $M = 20$.

3.2 Robust methods

The SY method relies heavily on i.i.d. and homoskedasticity. To the best of our knowledge, there is no theoretical or analytical support to any direct or indirect extension neither to the heteroskedastic case, nor to the clustered case: the relationship between the concentration parameter and the first-stage F -statistic is no longer exact. Still, researchers frequently report heteroskedastic-robust F -statistics when using TSLS and compare them with the SY homoskedastic thresholds.

Limited information maximum likelihood (LIML) and Fuller estimators are more robust to weak instruments than TSLS, but not fully.

Actually, there cannot exist any estimator which is consistent regardless of the strength of the instruments, as its distribution depends on μ^2 . Under weak instruments, its value cannot be consistently estimated. In the limiting case when $\gamma = 0$, which means that the strength of the instrument is null, β^{IV} is not identified, hence not consistent, and no valid t-test based on such estimator can be conceived. The intuition is straightforward: if the set of instruments is uncorrelated with the endogenous regressor, any value of β found using such instruments is consistent

with the data (see Dufour, 1997).

However, fully robust inference procedures have recently gathered more attention. These tests are designed in a way that they maintain the correct size regardless of the strength of the set of instruments.

The strand of literature we follow is based on statistics whose distribution does not depend on the concentration parameter. In particular, we focus on the approach by Chernozhukov and Hansen (CH, 2008).

For the model described in 3 and 4, we can write:

$$y_i = \beta x_i + z_i \theta + v_i \quad (5)$$

Under $H_0 : \beta = \beta_0$, a way to express this is

$$y_i - \beta_0 x_i = z_i \theta + v_i \quad (6)$$

and, if z influences y only through x , and it is uncorrelated with the error term (exogeneity condition), then a simple Wald test on the significance of θ tests H_0 , too.

Such test can be made robust to heteroskedasticity and clustered data by using the appropriate covariance matrix estimator. However, this is not yet robust to the presence of few clusters.

As one may notice, the correlation between x and z does not enter the test. Said otherwise, the distribution of the Wald statistic for θ does not depend on the concentration parameter.

By exploiting the fact that testing and confidence sets are two sides of the same coin, the repetition of the test just described for a grid of values for β_0 allows to build valid $(1 - \alpha)$ confidence intervals. Such confidence intervals contain all those values of β_0 such that H_0 cannot be rejected at significance level α .

In case there are more instruments than endogenous variables, this simple procedure may use information inefficiently. Nonetheless, as noticed by CH in their

survey of the IV literature, even though a little dated¹⁸, most papers use as many instruments as endogenous regressors, so that the simple procedure described above is adequate.

The main disadvantage of CH is that a rejection of H_0 can arise either because $\beta \neq \beta_0$ or because the instrument exogeneity condition fails. This is a weakness that CH shares with the Anderson-Rubin (AR) test, which is based on a similar intuition but uses LIML instead (Stock, Wright and Yogo, 2002).

More efficient procedures are available. For a single endogenous regressor, Moreira (2003) shows that the AR test is the uniformly most powerful unbiased test, but its power deteriorates quickly the number of instruments. Another strand of literature focuses on performing inference conditionally on the sufficient statistics for μ^2 . The Conditional Likelihood Ratio, for example, has better power properties than other weak instrument-robust tests.

However, the strength of the CH method is its reliance on a simple sequence of Wald tests in the OLS framework and, as we show in Section 4.1, it can be easily improved to take into account the presence of few clusters.

4 Weak Instruments and Few Clusters

Our main interest is inference in IV models with few clusters. With a single endogenous variable, the model can be written as:

$$y_{ig} = \beta x_{ig} + v_{ig} = \beta x_{ig} + \alpha_g + u_{ig} \quad (7)$$

$$x_{ig} = z_{ig}\gamma + \xi_{ig} = z_{ig}\gamma + \eta_g + \epsilon_{ig} \quad (8)$$

Using IV, assessing the relevance of the set of instruments is problematic. The $\hat{F} > 10$ Rule of Thumb as in SY is invalid for two reasons: the first-stage statistic is

¹⁸They consider issues of the American Economic Review from March 1999 to March 2004, of the Journal of Political Economy from February 1999 to June 2004 and of the Quarterly Journal of Economics from February 1999 to February 2004.

upward biased, and reliance on thresholds derived under i.i.d. and homoskedastic assumptions is not justified, neither theoretically nor computationally.

The alternative is to perform inference with weak instrument-robust methods which, withal, are not robust to the presence of few clusters. We solve this problem with bootstrap and present in Section 4.1 a simple procedure that relies on a sequence of OLS regressions.

4.1 FC-WI-robust inference

In this section, we present our procedure to perform inference on β in a way which is robust to both few clusters and weak instruments (FC-WI).

We assume we have a single endogenous variable, so that the parameter β is a scalar, and z_{ig} is a vector containing M instruments. Additional exogenous regressors are partialled out.

Under $H_0 : \beta = \beta_0$, we rewrite 7 and 8 as

$$y_{ig} - \beta_0 x_{ig} = z_{ig}\theta + v_{ig} \quad (9)$$

As previously explained, the exogeneity of the instruments (exclusion restriction) implies $\theta = 0$. Testing the significance of θ equals testing $\beta = \beta_0$. However, this must be corrected for the presence of few clusters.

The procedure to construct FC-WI-robust confidence intervals is the following:

1. Select a set, $\mathcal{B}$, of potential values for β .
2. For each $b \in \mathcal{B}$, build $\tilde{y}_{ig} = y_{ig} - bx_{ig}$, estimate the model $\tilde{y}_{ig} = z_{ig}\theta + v_{ig}$ using a cluster-robust covariance matrix and store $F_b(\theta)$, the F -statistic for $\theta = 0$.
3. Apply the Wild Bootstrap T-Procedure B times:

- (a) For each $g = 1, 2, \dots, G$, build $\tilde{y}_{ig}^* = \tilde{y}_{ig}$ or $\tilde{y}_{ig}^* = -\tilde{y}_{ig}$ ¹⁹ with equal probability (in case $G < 10$ use Webb's six-points distribution for the weights).
 - (b) Regress $\tilde{y}_{ig}^*$ on z_{ig} and store $F_b^*(\theta)$ for $\theta = 0$.
4. Compute the $1-\alpha$ empirical quantile of $F_b^*(\theta)$, that we call $q_{1-\alpha}F_b^*(\theta)$. Check whether $F_b(\theta) < q_{1-\alpha}F_b^*(\theta)$. The confidence region will include all those b s for which this condition holds.

It must be noticed that such confidence set is not necessarily convex, not symmetric (Finlay and Magnusson, 2009).

A word of caution: researchers must pay attention in case the number of instruments is too close to the number of clusters. In particular, the model is not identified for $M \geq G$ and, for M slightly less than G , standard errors explode as in the cases of OLS imperfect multicollinearity.

4.2 Simulation

4.2.1 Data

In our Montecarlo simulations, we generate data with a specific concentration parameter using its link with the first-stage F -statistic, even though there is no longer a perfect correspondence between the two.

The set of M orthogonal and normally distributed instruments is characterized by an intraclass correlation coefficient (ICC) ι_z . The ICC accounts for the relatedness of clustered data by comparing the variance within clusters with the variance between clusters. Mathematically, it is the ratio between the within-cluster variance divided by the total one.

¹⁹As we must impose H_0 , and under H_0 $\theta = 0$, the new y equals the residuals.

If $\iota_z = 1$, the instrument is fully fixed within each cluster. On the contrary, for $\iota_z = 0$, each observation of the instrument is distributed independently from the others. Notice that the ICC cannot take negative values.

First-stage errors ξ_{ig} and second-stage errors v_{ig} (see Equations 7 and 8) are generated from standard normal distributions, with ICCs ι_ξ and ι_v . We also impose a correlation $\rho_{\alpha\eta}$ between the cluster-fixed parts of the errors, α_g and η_g , and a correlation $\rho_{u\epsilon}$ between the idiosyncratic parts, u_{ig} and ϵ_{ig} .

We simulate data both under homoskedasticity and heteroskedasticity. In the latter case, the variance of ϵ_{ig} is made proportional to $\sum_m z_{ig}^2$.

Additional details are provided in Appendix 7.1.

4.2.2 Results

In the simulation, we compare our FC-WI-robust method to the Chernozhukov-Hansen method without any few-cluster correction and to the standard TSLS procedure with a cluster-robust covariance matrix, both in the homoskedastic and in the heteroskedastic case. Results are displayed in Appendix 7.3.

The number of Montecarlo iterations is fixed at 5000. For each iteration, the Wild Bootstrap T-Procedure is repeated $B = 500$ times. When the number of clusters is set lower than 10, the six-point distribution is used instead of the Rademacher one.

The test behaves well regardless of the strength of the set of instruments, measured by the F -statistic. The addition of the Wild Bootstrap T-Procedure definitely improves performances, especially with very few clusters and with multiple instruments.

The Tables also include the percentage of estimated F -statistics greater than the threshold of 10. This percentage stays high even with relatively low ICCs, both for the instrument and the error term, and a small F -statistic. With multiple instruments, the probability to overestimate the actual F -statistic, and thus the strength of the set of instruments, is close to 1.

5 Empirical Example

In this section, we apply the FC-WI-robust method to “The China syndrome: Local labor market effects of import competition in the United States”, by Autor, Dorn and Hanson, published in the American Economic Review, 2013, and “Soil Endowments, Female Labor Force Participation and the Demographic Deficit of Women in India”, by E. Carranza, published in the American Economic Journal: Applied Economics, 2014. We start from the latter.

Carranza seeks to explain the unbalanced and heterogeneous sex ratio across Indian districts by differences in women’s relative employment in agriculture (in India, 72% of the population is rural). To get rid of potential endogeneity issues, she exploits the depth of the land tillage, proxied with soil textures, to instrument the demand for female workers, thus their economic value to the household. Deep tillage is possible in loamy soils, but not in clayey soils. In the latter case, operations such as fertilizing, weeding and transplanting are necessary, and these are typically performed by women.

The sample contains 584 districts located in 35 states. Those with complete information and actually employed in the IV regression are 365, distributed in 16 states. We rework Carranza’s results clustering at the state level. The average state contains 22 districts, with a standard deviation of 13. Six instruments are used (not accounting for included instruments, also present in the main equation).

We focus on instrumenting the share of female agricultural workers to the total of rural workers²⁰. The ICCs for the set of instruments are reported in Table 1. The average coefficient is 0.39, which is rather high. This suggests that a failure to cluster is going to have meaningful effects on inference.

As in Table 2, the first-stage F -statistic using an heteroskedastic-robust estimator is 23.58, which decreases to 14.83 when we use a cluster-robust estimator.

²⁰Considering laborers instead of workers, the first-stage F -statistic on Z is equal to 12.24 using an heteroskedastic-robust estimator but drops to 3.99 using a cluster-robust one. Since we would like to show that the threshold of 10 is not enough to guarantee good results in the IV regression, we prefer to consider *FLFP* for workers, which results in a higher first-stage F -statistic.

The $\hat{F} > 10$ Rule of Thumb would thus suggest to proceed with the second stage.

Carranza finds that increasing the relative employment of women as workers by ten percentage points should increase the presence of women (on 1000 men) by 28 units. As the average of number of girls per 1000 boys is 924.27, a increase of 9% would suffice to erase the deficit, defined as a sex ratio outcome below 950 per 1000 boys²¹. The effects is statistically significant at 1% level.

Table 1: Intraclass Correlation Coefficients for the instruments

Instrument	ICC	Instrument	ICC
% of loamy soils	0.42	% of scheduled tribes	0.38
% of clayey soils	0.40	% of scheduled castes	0.52
District area	0.44	% of rural population	0.17

Table 2: TSLS results

Statistics	Ht.	Cl.
First-stage F -stat	23.58	14.83
β	281.7	
t-stat on β	7.06	3.84
se on β	39.88	73.43

In Table 3, we compare the 95% confidence intervals obtained using different methods. For the FC-WI-robust one, we run 5000 bootstrap repetitions on a set of 801 potential values for β (from -100 to 700) with one-unit intervals.

According to TSLS and CH procedures, $FLFP$ has a significant effect on the child sex ratio. Our method, however, presents a wider confidence interval. Using the FC-WI-robust method, the female relative employment does not have a statistically significant impact on the child sex ratio.

²¹The average share of female workers over the total should move from 29% to 38%.

Table 3: 95% Confidence Intervals for β

Method	95% CI
TSLS- Ht.	(204; 360)
TSLS - Cl.	(138; 426)
CH	(238;390)
FC-WI	(-30;552)

The paper by Autor, Dorn and Hanson seeks to use changes in Chinese imports by other high-income countries to instrument those by US and analyze the effect of the latter on the change in the manufacturing employment share of the working-age population in each commuting zone.

The sample size is 1444, with 48 US states as clusters. The average state contains about 30 observations, with a standard deviation of 21. A first difference model with weights is used; this is not problematic for our procedure, but one has to be careful in using weights when partialling out other regressors, too.

Only one instrument is used. Its intraclass correlation, differently from the previously analyzed paper, is only 0.045, and the number of clusters is not that small: this suggests that results will be mildly affected by a change in the inference procedure.

We analyze the authors' preferred specification²².

Table 4: TSLS results

Statistics	Ht.	Cl.
First-stage F -stat	533.32	47.64
β	-0.596	
t-stat on β	-11.05	-6.04
se on β	0.054	0.099

²²It is contained in Column 6, Table 3

Table 5: 95% Confidence Intervals for β

Method	95% CI
TSLS- Ht.	(-0.70;-0.49)
TSLS - Cl.	(-0.79;-0.40)
CH	(-0.86;-0.55)
FC-WI	(-0.85;-0.50)

As we expected, there is little difference among the different methods: the 95% confidence intervals are rather similar, with those evaluated using the FC-WI-robust and the Chernozhukov-Hansen procedures moderately more distant from 0 than the TSLS approach with clusters.

The effect of China imports by US on employment stays negative and significant, as found by Autor, Dorn and Hanson.

6 Conclusions

With clustered data, and especially few clusters, the usual methods to assess the quality of a set of instruments in IV are not reliable. In particular, the empiricists' favorite test on the F -statistic by Stock and Yogo fails to detect weak instruments because a) such statistic is overestimated and b) the thresholds reported in SY are found under restrictive assumptions such as i.i.d. and homoskedastic data. It follows that the size of an α test on β in the second stage is far from being the correct one (the relative bias, too, is going to be higher than the expected).

Some procedures have been proposed to circumvent the problem posed by weak instruments. Chernozhukov-Hansen (2008) procedure is particularly easy to apply, since it relies on a series of OLS regressions and it is easy to understand.

We have improved this methodology to account for the presence of few clusters by using the Wild Bootstrap T-Procedure. This addition effectively ameliorates the precision of inference.

7 Appendix

7.1 Simulating data

For commodity, we copy here Equations 7 and 8:

$$\begin{aligned} y_{ig} &= \beta x_{ig} + \alpha_g + u_{ig} = \beta x_{ig} + v_{ig} \\ x_{ig} &= z_{ig}\gamma + \eta_g + \epsilon_{ig} = z_{ig}\gamma + \xi_{ig} \end{aligned}$$

In our simulations, the set of M normally distributed instruments, each one orthogonal to the others, has zero mean and fixed variance $\sigma_z^2=1$. We consider here a single instrument z_{ig} , but the generalization to a set easily follows.

The instrument z_{ig} is the sum of two parts, one of whom is cluster-fixed:

$$z_{ig} = \tilde{z}_{ig} + \tilde{z}_g$$

Exploiting the ICC definition, we generate each part of the instrument from the following distributions:

$$\begin{aligned} \tilde{z}_{ig} &\sim N(0, (1 - \iota_z)\sigma_z^2) \\ \tilde{z}_g &\sim N(0, \iota_z\sigma_z^2) \end{aligned}$$

so that:

$$ICC(z) = \frac{\sigma_{\tilde{z}_g}^2}{\sigma_{\tilde{z}_g}^2 + \sigma_{\tilde{z}_{ig}}^2} = \frac{\iota_z\sigma_z^2}{\iota_z\sigma_z^2 + (1-\iota_z)\sigma_z^2} = \iota_z$$

Generating the errors requires the additional difficulty of making them correlated across the two stages, which allows us to build an endogenous x_{ig} . Take $w_{1,ig}$, $w_{2,ig}$, $w_{1,g}$ and $w_{2,g}$ standard normal vectors, the last two being cluster-fixed.

$$\begin{aligned}
\epsilon_{ig} &= \begin{cases} \sqrt{(1-\iota_\xi)\sigma_\xi^2}w_{1,ig} & \text{if homosk.} \\ \sqrt{(1-\iota_\xi)\sigma_\xi^2}\sqrt{\sum_m z_{ig}^2}w_{1,ig} & \text{if heterosk.} \end{cases} \\
\eta_g &= \sqrt{\iota_\xi\sigma_\xi^2}w_{1,g} \\
u_{ig} &= \begin{cases} \sqrt{(1-\iota_v)\sigma_v^2}(\rho_{u\epsilon}w_{1,ig} + \sqrt{1-\rho_{u\epsilon}^2}w_{2,ig}) & \text{if homosk.} \\ \sqrt{(1-\iota_v)\sigma_v^2}(\sqrt{\sum_m z_{ig}^2}\rho_{u\epsilon}w_{1,ig} + \sqrt{1-\rho_{u\epsilon}^2}w_{2,ig}) & \text{if heterosk.} \end{cases} \\
\alpha_g &= \sqrt{\iota_v\sigma_v^2}(\rho_{\alpha\eta}w_{1,g} + \sqrt{1-\rho_{\alpha\eta}^2}w_{2,g})
\end{aligned}$$

The resulting correlation between the main equation and the reduced form errors is:

$$Corr(v_{ig}, \xi_{ig}) = \rho_{\alpha\eta}\sqrt{\iota_v\iota_\xi} + \rho_{u\epsilon}\sqrt{(1-\iota_v)(1-\iota_\xi)}$$

We impose a specific first-stage F -statistic setting γ as

$$\gamma = ki_m$$

exploiting:

$$F = \frac{\gamma' \text{Var}(\hat{\gamma})^{-1} \gamma}{M}$$

Under homoskedasticity:

$$\kappa = \sqrt{\frac{\sigma_\xi^2(1 + (N-1)\iota_\xi\iota_z)F}{NG\sigma_z^2}}$$

Under heteroskedasticity:

$$\kappa = \sqrt{\frac{\sigma_\xi^2 F}{NG\sigma_z^2} \left((1-\iota_\xi) \frac{3+M-1}{M} + \iota_\xi + (N-1)\iota_\xi\iota_z \right)}$$

7.2 More on γ

7.2.1 Under homoskedasticity

We call V the variance of this object:

$$\begin{aligned}\sqrt{NG}(\hat{\gamma} - \gamma) &= \left(\frac{1}{NG} \sum_{g=1}^G \sum_{i=1}^N z'_{ig} z_{ig} \right)^{-1} \frac{1}{\sqrt{NG}} \sum_{g=1}^G \sum_{i=1}^N z'_{ig} (\epsilon_{ig} + \eta_g) \\ V &= \left[\frac{1}{NG} \sum_{g=1}^G \sum_{i=1}^N E(z'_{ig} z_{ig}) \right]^{-1} \left[\frac{1}{NG} \sum_{g=1}^G E(z'_g \Omega z_g) \right] \\ &\quad \left[\frac{1}{NG} \sum_{g=1}^G \sum_{i=1}^N E(z'_{ig} z_{ig}) \right]^{-1}\end{aligned}$$

Then, we split calculations in two parts :

$$\frac{1}{NG} E[z'_{ig} z_{ig}] = \sigma_z^2 I_m$$

$$\begin{aligned}\frac{1}{NG} E[z'_{ig} (\eta_g + \epsilon_{ig}) (\eta_g + \epsilon_{ig})' z_{ig}] &= \frac{1}{NG} E(\Sigma_g^G \Sigma_i^N z_{ig}^2 (\eta_g + \epsilon_{ig})^2 + \Sigma_g^G (\Sigma_i^N \Sigma_{j, j \neq i}^N z_{ig} z_{jg} \eta_g^2) I_M) \\ &= \frac{1}{NG} (NG \sigma_z^2 \sigma_\xi^2 + G(N^2 - N) \iota_z \sigma_z^2 \iota_\xi \sigma_\xi^2) I_M \\ &= \frac{1}{NG} NG \sigma_z^2 \sigma_\xi^2 (1 + (N - 1) \iota_z \iota_\xi) I_M \\ &= \sigma_\xi^2 \sigma_z^2 (1 + (N - 1) \iota_z \iota_\xi) I_M\end{aligned}$$

$$\begin{aligned}Var(\hat{\gamma}) &= \frac{V}{NG} = \frac{1}{NG} \frac{1}{(\sigma_z^2)^2} \sigma_\xi^2 \sigma_z^2 (1 + (N - 1) \iota_z \iota_\xi) \\ &= \frac{\sigma_\xi^2 (1 + (N - 1) \iota_\xi \iota_z)}{NG \sigma_z^2} I_M\end{aligned}$$

$$F = \frac{\gamma' Var(\hat{\gamma})^{-1} \gamma}{M} = \frac{k^2 i_M' i_M}{M} \left(\frac{NG \sigma_z^2}{\sigma_\xi^2 (1 + (N - 1) \iota_\xi \iota_z)} I_M \right)^{-1}$$

$$= \frac{k^2 M}{M} \left(\frac{NG\sigma_z^2}{\sigma_\xi^2(1+(N-1)\iota_\xi\iota_z)} \right)^{-1}$$

As γ can be rewritten as ki_m , for a given F -statistic:

$$\kappa = \sqrt{\frac{\sigma_\xi^2(1+(N-1)\iota_\xi\iota_z)F}{NG\sigma_z^2}}$$

7.2.2 Under heteroskedasticity

First, for simplicity, we call:

$$v_{ig} = \frac{(1-\iota_\xi)\sigma_\xi^2 \sum_m z_{ig}^2}{M\sigma_z^2} + \iota_\xi\sigma_\xi^2$$

$$v_g = \iota_\xi\sigma_\xi^2$$

We call V the variance of this object:

$$\begin{aligned} & \sqrt{NG}(\hat{\gamma} - \gamma) \\ &= \left(\frac{1}{NG} \sum_{g=1}^G \sum_{i=1}^N z'_{ig} z_{ig} \right)^{-1} \frac{1}{\sqrt{NG}} \sum_{g=1}^G \sum_{i=1}^N z'_{ig} (\epsilon_{ig} + \eta_g) \end{aligned}$$

$$V = \Theta \left[\frac{1}{NG} \sum_{g=1}^G \sum_{i=1}^N E(z'_{ig} \Omega z_{ig}) \right] \Theta$$

$$\text{where } \Theta = \left[\frac{E(Z'Z)}{NG} \right]^{-1}.$$

The external parts become:

$$\Theta = \left[\frac{NG\sigma_z^2}{NG} I_M \right]^{-1} = \frac{1}{\sigma_z^2} I_M$$

While the inner part becomes:

$$\frac{1}{NG} E(Z' \Omega Z) =$$

$$\frac{1}{NG}E \begin{bmatrix} \sum_i \sum_g z_{1,ig}^2 v_{ig} + \sum_g \sum_i \sum_j z_{1,ig} z_{1,jg} v_g & 0 & 0 \\ 0 & \dots & 0 \\ \dots & 0 & 0 \\ 0 & 0 & \sum_i \sum_g z_{M,ig}^2 v_{ig} + \sum_g \sum_i \sum_j z_{M,ig} z_{M,jg} v_g \end{bmatrix}$$

All the elements on the diagonal turn out to be the same:

$$\begin{aligned} & E[\sum_i \sum_g z_{m,ig}^2 v_{ig} + \sum_g \sum_i \sum_j z_{m,ig} z_{m,jg} v_g] \\ &= E[\sum_i \sum_g z_{m,ig}^2 (\frac{(1-\iota_\xi)\sigma_\xi^2 \sum_{\bar{m}}^M z_{\bar{m},ig}^2}{M\sigma_z^2} + \iota_\xi \sigma_\xi^2) \\ & \quad + \sum_g \sum_i \sum_j z_{m,ig} z_{m,jg} \iota_\xi \sigma_\xi^2] \\ &= \frac{(1-\iota_\xi)\sigma_\xi^2}{M\sigma_z^2} E[\sum_i \sum_g z_{m,ig}^2 \sum_{\bar{m}} z_{\bar{m},ig}^2] + \iota_\xi \sigma_\xi^2 E[\sum_i \sum_g z_{m,ig}^2] \\ & \quad + \iota_\xi \sigma_\xi^2 E[\sum_g \sum_i \sum_j z_{m,ig} z_{m,jg}] \\ &= \frac{(1-\iota_\xi)\sigma_\xi^2}{M\sigma_z^2} E[\sum_i \sum_g \sum_{\bar{m}} z_{m,ig}^2 z_{\bar{m},ig}^2] + NG\iota_\xi \sigma_\xi^2 \sigma_z^2 + (N^2 - N)G\iota_\xi \sigma_\xi^2 \iota_z \sigma_z^2 \\ &= \frac{(1-\iota_\xi)\sigma_\xi^2}{M\sigma_z^2} E[\sum_i \sum_g (z_{m,ig}^4 + z_{m,ig}^2 z_{\bar{m},ig}^2 \dots)] + NG\iota_\xi \sigma_\xi^2 \sigma_z^2 \\ & \quad + (N-1)NG\iota_\xi \sigma_\xi^2 \iota_z \sigma_z^2 \\ &= \frac{(1-\iota_\xi)\sigma_\xi^2}{M\sigma_z^2} NG(3\sigma_z^4 + (M-1)\sigma_z^4) + NG\iota_\xi \sigma_\xi^2 \sigma_z^2 + NG(N-1)\iota_\xi \sigma_\xi^2 \iota_z \sigma_z^2 \\ &= NG\sigma_\xi^2 \sigma_z^2 [\frac{(1-\iota_\xi)}{M}(3\sigma_z^2 + (M-1)\sigma_z^2) + \iota_\xi + (N-1)\iota_\xi \iota_z] \\ &= NG\sigma_\xi^2 \sigma_z^2 [\frac{(1-\iota_\xi)}{M}(3+M-1) + \iota_\xi + (N-1)\iota_\xi \iota_z] \end{aligned}$$

Putting the pieces together:

$$\begin{aligned} Var(\hat{\gamma}) &= \frac{V}{NG} \\ &= \frac{1}{(NG)^2 \sigma_z^4} NG\sigma_\xi^2 \sigma_z^2 [\frac{(1-\iota_\xi)}{M}(3+M-1) + \iota_\xi + (N-1)\iota_\xi \iota_z] I_M \\ &= \frac{\sigma_\xi^2}{NG\sigma_z^2} [\frac{(1-\iota_\xi)}{M}(3+M-1) + \iota_\xi + (N-1)\iota_\xi \iota_z] I_M \end{aligned}$$

We obtain:

$$\begin{aligned} F &= \frac{\gamma' Var(\hat{\gamma})^{-1} \gamma}{M} \\ &= \frac{k^2 i_M' i_M}{M} (\frac{\sigma_\xi^2}{NG\sigma_z^2} [\frac{(1-\iota_\xi)}{M}(3+M-1) + \iota_\xi + (N-1)\iota_\xi \iota_z])^{-1} \\ &= \frac{k^2 M \sigma_\xi^2}{MNG\sigma_z^2} [\frac{(1-\iota_\xi)}{M}(3+M-1) + \iota_\xi + (N-1)\iota_\xi \iota_z] \\ &= \frac{k^2 \sigma_\xi^2}{NG\sigma_z^2} [\frac{(1-\iota_\xi)}{M}(3+M-1) + \iota_\xi + (N-1)\iota_\xi \iota_z] \end{aligned}$$

As γ can be rewritten as ki_m , for a given F -statistic:

$$\kappa = \sqrt{\frac{\sigma_\xi^2 F}{NG\sigma_z^2} \left((1 - \iota_\xi)^{\frac{3+M-1}{M}} + \iota_\xi + (N-1)\iota_\xi \iota_z \right)}$$

7.3 Simulations

We report below the results of the Montecarlo simulations described in Section 4.2.

Table 6: 5,000 Simulations from DGP with Group Level Homoskedastic Random Errors, 1 instrument.

Parameters						F-stat		TSLS t-test		CRVE CH		Wildbootstrap CH		
ρ	F	G	N	ι_z	$\iota_u = \iota_\epsilon$	$\hat{F}$	$\hat{F} > 10$	$\hat{F} > F$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
0.9	5	5	30	0.9	0.9	34.026	0.471	0.644	0.209	0.253	0.236	0.294	0.038	0.103
0.9	5	20	30	0.9	0.9	8.605	0.304	0.569	0.121	0.154	0.098	0.157	0.047	0.099
0.9	5	35	30	0.9	0.9	7.271	0.248	0.540	0.106	0.137	0.075	0.136	0.047	0.099
0.9	5	50	30	0.9	0.9	6.767	0.223	0.519	0.099	0.128	0.066	0.119	0.051	0.020
0.9	10	5	30	0.9	0.9	59.043	0.665	0.665	0.200	0.244	0.224	0.285	0.042	0.108
0.9	10	20	30	0.9	0.9	15.960	0.573	0.573	0.105	0.140	0.103	0.162	0.052	0.106
0.9	10	35	30	0.9	0.9	13.771	0.546	0.546	0.091	0.126	0.077	0.125	0.050	0.097
0.9	10	50	30	0.9	0.9	12.780	0.540	0.540	0.089	0.119	0.073	0.128	0.052	0.105
0.9	20	5	30	0.9	0.9	112.897	0.847	0.692	0.201	0.256	0.218	0.284	0.044	0.106
0.9	20	20	30	0.9	0.9	30.239	0.867	0.595	0.093	0.137	0.089	0.143	0.051	0.093
0.9	20	35	30	0.9	0.9	25.939	0.869	0.559	0.075	0.118	0.078	0.136	0.054	0.103
0.9	20	50	30	0.9	0.9	24.349	0.885	0.541	0.077	0.109	0.069	0.122	0.054	0.102
0.9	50	5	30	0.9	0.9	266.739	0.981	0.707	0.224	0.282	0.233	0.293	0.042	0.111
0.9	50	20	30	0.9	0.9	73.975	0.996	0.598	0.092	0.140	0.100	0.152	0.055	0.105
0.9	50	35	30	0.9	0.9	62.680	0.999	0.576	0.075	0.123	0.072	0.129	0.050	0.102
0.9	50	50	30	0.9	0.9	59.299	0.999	0.556	0.067	0.117	0.069	0.120	0.051	0.103
0.9	75	5	30	0.9	0.9	434.396	0.992	0.694	0.224	0.285	0.230	0.292	0.044	0.101
0.9	75	20	30	0.9	0.9	109.904	1.000	0.605	0.094	0.147	0.098	0.154	0.051	0.105
0.9	75	35	30	0.9	0.9	94.600	1.000	0.597	0.074	0.126	0.072	0.129	0.046	0.099
0.9	75	50	30	0.9	0.9	87.742	1.000	0.561	0.071	0.119	0.069	0.128	0.052	0.102
0.9	5	5	30	0.3	0.3	17.586	0.442	0.649	0.153	0.193	0.148	0.217	0.054	0.108
0.9	5	20	30	0.3	0.3	7.916	0.280	0.553	0.104	0.137	0.076	0.133	0.052	0.101
0.9	5	35	30	0.3	0.3	7.040	0.238	0.529	0.100	0.127	0.064	0.120	0.051	0.101
0.9	5	50	30	0.3	0.3	6.664	0.219	0.521	0.100	0.131	0.063	0.116	0.055	0.102
0.9	10	5	30	0.3	0.3	32.360	0.673	0.673	0.140	0.184	0.142	0.207	0.056	0.112
0.9	10	20	30	0.3	0.3	14.683	0.572	0.572	0.096	0.128	0.078	0.129	0.055	0.103
0.9	10	35	30	0.3	0.3	13.013	0.545	0.545	0.087	0.121	0.067	0.118	0.054	0.101
0.9	10	50	30	0.3	0.3	12.481	0.537	0.537	0.089	0.121	0.065	0.120	0.056	0.107
0.9	20	5	30	0.3	0.3	66.074	0.895	0.703	0.132	0.185	0.141	0.200	0.052	0.107
0.9	20	20	30	0.3	0.3	28.102	0.889	0.607	0.078	0.118	0.072	0.133	0.051	0.105
0.9	20	35	30	0.3	0.3	24.828	0.893	0.559	0.071	0.106	0.063	0.119	0.052	0.104
0.9	20	50	30	0.3	0.3	23.643	0.896	0.550	0.072	0.100	0.059	0.110	0.051	0.100
0.9	50	5	30	0.3	0.3	160.420	0.993	0.718	0.146	0.202	0.149	0.209	0.059	0.111
0.9	50	20	30	0.3	0.3	67.031	0.998	0.609	0.076	0.122	0.077	0.135	0.052	0.104
0.9	50	35	30	0.3	0.3	60.232	0.998	0.585	0.061	0.108	0.070	0.122	0.056	0.105
0.9	50	50	30	0.3	0.3	57.154	0.999	0.563	0.061	0.103	0.059	0.116	0.050	0.104
0.9	75	5	30	0.3	0.3	250.378	0.999	0.713	0.145	0.209	0.147	0.214	0.052	0.110
0.9	75	20	30	0.3	0.3	101.355	1.000	0.619	0.076	0.123	0.081	0.137	0.055	0.108
0.9	75	35	30	0.3	0.3	89.773	1.000	0.589	0.062	0.112	0.065	0.120	0.050	0.102
0.9	75	50	30	0.3	0.3	86.764	1.000	0.581	0.061	0.107	0.058	0.111	0.051	0.097

Table 7: 5,000 Simulations from DGP with Group Level Homoskedastic Random Errors, 4 instruments.

Parameters						F-stat		TSLS t-test		CRVE CH		Wildbootstrap CH		
ρ	F	G	N	ι_z	$\iota_u = \iota_\epsilon$	$\hat{F}$	$\hat{F} > 10$	$\hat{F} > F$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
0.9	5	5	30	0.9	0.9	5.569e + 05	0.954	0.984	0.289	0.350	0.907	0.926	0.046	0.092
0.9	5	20	30	0.9	0.9	1.354e + 01	0.511	0.837	0.197	0.262	0.305	0.382	0.054	0.104
0.9	5	35	30	0.9	0.9	9.307e + 00	0.338	0.781	0.176	0.233	0.198	0.275	0.055	0.106
0.9	5	50	30	0.9	0.9	8.461e + 00	0.292	0.764	0.187	0.247	0.142	0.216	0.051	0.101
0.9	10	5	30	0.9	0.9	5.330e + 05	0.985	0.985	0.262	0.325	0.917	0.935	0.044	0.090
0.9	10	20	30	0.9	0.9	2.457e + 01	0.837	0.837	0.141	0.198	0.305	0.386	0.055	0.103
0.9	10	35	30	0.9	0.9	1.730e + 01	0.767	0.767	0.138	0.195	0.200	0.277	0.056	0.108
0.9	10	50	30	0.9	0.9	1.502e + 01	0.738	0.738	0.129	0.180	0.146	0.221	0.052	0.104
0.9	20	5	30	0.9	0.9	2.729e + 06	0.997	0.985	0.236	0.294	0.914	0.929	0.052	0.097
0.9	20	20	30	0.9	0.9	4.707e + 01	0.980	0.842	0.126	0.177	0.316	0.401	0.050	0.099
0.9	20	35	30	0.9	0.9	3.301e + 01	0.984	0.775	0.108	0.162	0.193	0.267	0.047	0.100
0.9	20	50	30	0.9	0.9	2.889e + 01	0.985	0.736	0.100	0.154	0.136	0.208	0.048	0.094
0.9	50	5	30	0.9	0.9	5.945e + 06	1.000	0.987	0.221	0.284	0.909	0.932	0.045	0.093
0.9	50	20	30	0.9	0.9	1.200e + 02	1.000	0.850	0.115	0.170	0.322	0.398	0.055	0.101
0.9	50	35	30	0.9	0.9	8.019e + 01	1.000	0.780	0.094	0.154	0.191	0.275	0.052	0.104
0.9	50	50	30	0.9	0.9	7.038e + 01	1.000	0.746	0.087	0.147	0.150	0.222	0.052	0.107
0.9	75	5	30	0.9	0.9	3.855e + 06	1.000	0.985	0.220	0.281	0.906	0.926	0.048	0.096
0.9	75	20	30	0.9	0.9	1.704e + 02	1.000	0.855	0.109	0.165	0.308	0.390	0.052	0.101
0.9	75	35	30	0.9	0.9	1.195e + 02	1.000	0.778	0.081	0.130	0.179	0.262	0.051	0.096
0.9	75	50	30	0.9	0.9	1.038e + 02	1.000	0.742	0.078	0.137	0.152	0.221	0.056	0.104
0.9	5	5	30	0.3	0.3	7.847e + 04	0.934	0.982	0.230	0.296	0.807	0.842	0.049	0.099
0.9	5	20	30	0.3	0.3	1.091e + 01	0.446	0.833	0.189	0.241	0.201	0.283	0.053	0.102
0.9	5	35	30	0.3	0.3	8.426e + 00	0.288	0.773	0.173	0.231	0.131	0.204	0.049	0.108
0.9	5	50	30	0.3	0.3	7.791e + 00	0.238	0.745	0.179	0.237	0.111	0.182	0.054	0.108
0.9	10	5	30	0.3	0.3	8.066e + 05	0.985	0.985	0.191	0.254	0.793	0.830	0.044	0.089
0.9	10	20	30	0.3	0.3	2.030e + 01	0.833	0.833	0.129	0.188	0.197	0.283	0.050	0.101
0.9	10	35	30	0.3	0.3	1.565e + 01	0.773	0.773	0.118	0.173	0.126	0.202	0.047	0.101
0.9	10	50	30	0.3	0.3	1.410e + 01	0.742	0.742	0.118	0.173	0.095	0.159	0.047	0.097
0.9	20	5	30	0.3	0.3	4.529e + 08	0.998	0.985	0.167	0.232	0.796	0.829	0.044	0.090
0.9	20	20	30	0.3	0.3	3.898e + 01	0.990	0.841	0.104	0.163	0.194	0.275	0.053	0.102
0.9	20	35	30	0.3	0.3	3.015e + 01	0.992	0.788	0.099	0.154	0.126	0.200	0.050	0.102
0.9	20	50	30	0.3	0.3	2.682e + 01	0.990	0.732	0.087	0.140	0.103	0.167	0.051	0.099
0.9	50	5	30	0.3	0.3	7.455e + 05	1.000	0.987	0.150	0.217	0.786	0.824	0.053	0.103
0.9	50	20	30	0.3	0.3	9.658e + 01	1.000	0.842	0.080	0.137	0.191	0.272	0.053	0.099
0.9	50	35	30	0.3	0.3	7.342e + 01	1.000	0.788	0.076	0.129	0.126	0.199	0.051	0.100
0.9	50	50	30	0.3	0.3	6.604e + 01	1.000	0.746	0.071	0.129	0.104	0.167	0.055	0.103
0.9	75	5	30	0.3	0.3	2.408e + 08	1.000	0.985	0.159	0.225	0.795	0.827	0.052	0.110
0.9	75	20	30	0.3	0.3	1.419e + 02	1.000	0.849	0.081	0.131	0.194	0.278	0.052	0.106
0.9	75	35	30	0.3	0.3	1.091e + 02	1.000	0.782	0.073	0.125	0.122	0.199	0.045	0.096
0.9	75	50	30	0.3	0.3	9.870e + 01	1.000	0.745	0.070	0.121	0.099	0.167	0.050	0.095

Table 8: 5,000 Simulations from DGP with Group Level Homoskedastic Random Errors, 10 instruments.

Parameters															F-stat		TSLs t-test		CRVE CH		Wildbootstrap CH	
ρ	F	G	N	ι_z	$\iota_u = \iota_\epsilon$	$\tilde{F}$	$\hat{F} > 10$	$\hat{F} > F$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$								
0.9	5	20	30	0.9	0.9	44.477	0.938	0.994	0.301	0.380	0.880	0.910	0.051	0.102								
0.9	5	35	30	0.9	0.9	16.297	0.729	0.972	0.305	0.390	0.608	0.685	0.050	0.102								
0.9	5	50	30	0.9	0.9	11.713	0.549	0.951	0.287	0.376	0.432	0.532	0.058	0.107								
0.9	10	20	30	0.9	0.9	80.686	0.995	0.995	0.199	0.278	0.884	0.915	0.049	0.098								
0.9	10	35	30	0.9	0.9	29.904	0.970	0.970	0.192	0.258	0.614	0.684	0.045	0.097								
0.9	10	50	30	0.9	0.9	21.653	0.941	0.941	0.197	0.272	0.445	0.544	0.052	0.095								
0.9	20	20	30	0.9	0.9	157.782	1.000	0.995	0.144	0.215	0.878	0.908	0.059	0.108								
0.9	20	35	30	0.9	0.9	57.086	0.999	0.968	0.137	0.193	0.607	0.680	0.051	0.103								
0.9	20	50	30	0.9	0.9	41.372	1.000	0.938	0.123	0.187	0.435	0.527	0.054	0.104								
0.9	50	20	30	0.9	0.9	382.566	1.000	0.994	0.113	0.179	0.883	0.915	0.050	0.104								
0.9	50	35	30	0.9	0.9	138.688	1.000	0.966	0.101	0.162	0.622	0.700	0.051	0.105								
0.9	50	50	30	0.9	0.9	100.323	1.000	0.941	0.093	0.156	0.453	0.553	0.056	0.100								
0.9	75	20	30	0.9	0.9	563.743	1.000	0.994	0.113	0.172	0.881	0.912	0.055	0.109								
0.9	75	35	30	0.9	0.9	206.286	1.000	0.968	0.089	0.153	0.601	0.685	0.052	0.099								
0.9	75	50	30	0.9	0.9	148.097	1.000	0.932	0.084	0.145	0.432	0.528	0.049	0.092								
0.9	5	20	30	0.3	0.3	23.968	0.884	0.993	0.273	0.364	0.664	0.742	0.054	0.099								
0.9	5	35	30	0.3	0.3	12.363	0.613	0.961	0.274	0.368	0.376	0.488	0.049	0.096								
0.9	5	50	30	0.3	0.3	10.076	0.445	0.940	0.275	0.363	0.275	0.372	0.051	0.102								
0.9	10	20	30	0.3	0.3	44.608	0.991	0.991	0.186	0.259	0.685	0.756	0.049	0.096								
0.9	10	35	30	0.3	0.3	23.086	0.968	0.968	0.172	0.248	0.393	0.488	0.053	0.098								
0.9	10	50	30	0.3	0.3	18.580	0.941	0.941	0.176	0.254	0.267	0.365	0.048	0.099								
0.9	20	20	30	0.3	0.3	87.841	1.000	0.994	0.131	0.194	0.677	0.749	0.057	0.102								
0.9	20	35	30	0.3	0.3	44.728	1.000	0.967	0.124	0.189	0.375	0.480	0.056	0.097								
0.9	20	50	30	0.3	0.3	35.734	1.000	0.936	0.123	0.184	0.258	0.357	0.057	0.103								
0.9	50	20	30	0.3	0.3	213.582	1.000	0.992	0.104	0.159	0.680	0.747	0.054	0.101								
0.9	50	35	30	0.3	0.3	108.209	1.000	0.966	0.084	0.142	0.372	0.478	0.045	0.092								
0.9	50	50	30	0.3	0.3	86.643	1.000	0.936	0.083	0.146	0.263	0.359	0.058	0.104								
0.9	75	20	30	0.3	0.3	313.851	1.000	0.993	0.090	0.150	0.687	0.760	0.053	0.105								
0.9	75	35	30	0.3	0.3	160.845	1.000	0.966	0.082	0.135	0.391	0.491	0.055	0.102								
0.9	75	50	30	0.3	0.3	128.781	1.000	0.941	0.079	0.139	0.271	0.372	0.049	0.100								

Table 9: 5,000 Simulations from DGP with Group Level Heteroskedastic Random Errors, 1 instrument.

Parameters										F-stat		TSLs t-test		CRVE CH		Wildbootstrap CH	
ρ	F	G	N	ι_z	$\iota_u = \iota_\epsilon$	$\hat{F}$	$\hat{F} > 10$	$\hat{F} > F$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	
0.9	5	5	30	0.9	0.9	31.451	0.466	0.645	0.194	0.234	0.224	0.285	0.043	0.105	0.043	0.105	
0.9	5	20	30	0.9	0.9	8.469	0.289	0.549	0.117	0.150	0.095	0.149	0.050	0.097	0.050	0.097	
0.9	5	35	30	0.9	0.9	7.391	0.255	0.527	0.109	0.147	0.075	0.129	0.050	0.099	0.050	0.099	
0.9	5	50	30	0.9	0.9	7.005	0.235	0.539	0.105	0.137	0.069	0.125	0.052	0.105	0.052	0.105	
0.9	10	5	30	0.9	0.9	70.601	0.676	0.676	0.203	0.257	0.224	0.286	0.044	0.109	0.044	0.109	
0.9	10	20	30	0.9	0.9	15.709	0.567	0.567	0.102	0.139	0.094	0.147	0.0526	0.097	0.0526	0.097	
0.9	10	35	30	0.9	0.9	13.535	0.544	0.544	0.096	0.126	0.080	0.141	0.052	0.107	0.052	0.107	
0.9	10	50	30	0.9	0.9	12.872	0.537	0.537	0.088	0.117	0.069	0.123	0.052	0.101	0.052	0.101	
0.9	20	5	30	0.9	0.9	287.863	0.852	0.684	0.207	0.263	0.231	0.290	0.043	0.112	0.043	0.112	
0.9	20	20	30	0.9	0.9	30.636	0.872	0.599	0.093	0.132	0.096	0.147	0.051	0.099	0.051	0.099	
0.9	20	35	30	0.9	0.9	25.751	0.875	0.565	0.080	0.116	0.078	0.130	0.053	0.099	0.053	0.099	
0.9	20	50	30	0.9	0.9	24.362	0.884	0.548	0.069	0.105	0.067	0.119	0.046	0.098	0.046	0.098	
0.9	50	5	30	0.9	0.9	282.443	0.982	0.707	0.224	0.289	0.233	0.294	0.038	0.108	0.038	0.108	
0.9	50	20	30	0.9	0.9	72.447	0.995	0.589	0.095	0.148	0.094	0.155	0.052	0.098	0.052	0.098	
0.9	50	35	30	0.9	0.9	62.627	0.998	0.581	0.083	0.131	0.083	0.141	0.058	0.112	0.058	0.112	
0.9	50	50	30	0.9	0.9	59.287	0.998	0.560	0.060	0.108	0.069	0.123	0.047	0.101	0.047	0.101	
0.9	75	5	30	0.9	0.9	443.000	0.990	0.695	0.203	0.270	0.210	0.274	0.043	0.102	0.043	0.102	
0.9	75	20	30	0.9	0.9	108.502	1.000	0.613	0.090	0.141	0.093	0.153	0.051	0.097	0.051	0.097	
0.9	75	35	30	0.9	0.9	92.761	1.000	0.570	0.075	0.130	0.076	0.133	0.049	0.102	0.049	0.102	
0.9	75	50	30	0.9	0.9	89.086	1.000	0.579	0.069	0.120	0.071	0.124	0.052	0.106	0.052	0.106	
0.9	5	5	30	0.3	0.3	19.126	0.440	0.653	0.158	0.200	0.156	0.213	0.060	0.114	0.060	0.114	
0.9	5	20	30	0.3	0.3	7.813	0.272	0.556	0.106	0.142	0.077	0.130	0.058	0.100	0.058	0.100	
0.9	5	35	30	0.3	0.3	6.956	0.236	0.529	0.107	0.140	0.067	0.120	0.054	0.097	0.054	0.097	
0.9	5	50	30	0.3	0.3	6.694	0.222	0.526	0.096	0.126	0.059	0.106	0.051	0.096	0.051	0.096	
0.9	10	5	30	0.3	0.3	36.648	0.674	0.674	0.137	0.181	0.144	0.209	0.055	0.105	0.055	0.105	
0.9	10	20	30	0.3	0.3	14.501	0.578	0.578	0.091	0.125	0.080	0.138	0.056	0.107	0.056	0.107	
0.9	10	35	30	0.3	0.3	12.904	0.551	0.551	0.076	0.104	0.061	0.113	0.049	0.096	0.049	0.096	
0.9	10	50	30	0.3	0.3	12.349	0.540	0.540	0.088	0.120	0.059	0.115	0.051	0.101	0.051	0.101	
0.9	20	5	30	0.3	0.3	59.369	0.892	0.707	0.143	0.196	0.144	0.216	0.055	0.116	0.055	0.116	
0.9	20	20	30	0.3	0.3	27.744	0.890	0.593	0.077	0.116	0.081	0.133	0.056	0.101	0.056	0.101	
0.9	20	35	30	0.3	0.3	24.836	0.892	0.566	0.077	0.114	0.069	0.123	0.057	0.103	0.057	0.103	
0.9	20	50	30	0.3	0.3	23.636	0.895	0.540	0.075	0.110	0.065	0.118	0.054	0.104	0.054	0.104	
0.9	50	5	30	0.3	0.3	156.093	0.994	0.711	0.143	0.201	0.149	0.208	0.057	0.114	0.057	0.114	
0.9	50	20	30	0.3	0.3	67.240	0.999	0.606	0.074	0.115	0.078	0.126	0.054	0.101	0.054	0.101	
0.9	50	35	30	0.3	0.3	60.214	0.998	0.581	0.064	0.107	0.067	0.121	0.053	0.103	0.053	0.103	
0.9	50	50	30	0.3	0.3	56.894	1.000	0.558	0.057	0.099	0.058	0.114	0.049	0.102	0.049	0.102	
0.9	75	5	30	0.3	0.3	240.846	0.999	0.725	0.153	0.210	0.152	0.216	0.056	0.115	0.056	0.115	
0.9	75	20	30	0.3	0.3	100.277	1.000	0.622	0.080	0.131	0.080	0.138	0.056	0.107	0.056	0.107	
0.9	75	35	30	0.3	0.3	89.348	1.000	0.595	0.066	0.113	0.067	0.115	0.053	0.100	0.053	0.100	
0.9	75	50	30	0.3	0.3	85.076	1.000	0.565	0.059	0.104	0.063	0.114	0.053	0.103	0.053	0.103	

Table 10: 5,000 Simulations from DGP with Group Level Heteroskedastic Random Errors, 4 instruments.

Parameters										F-stat		TSLS t-test		CRVE CH		Wildbootstrap CH	
ρ	F	G	N	ι_z	$\iota_u = \iota_\epsilon$	$\hat{F}$	$\hat{F} > 10$	$\hat{F} > F$		$\alpha = 0.05$	$\alpha = 0.10$		$\alpha = 0.05$	$\alpha = 0.10$		$\alpha = 0.05$	$\alpha = 0.10$
0.9	5	5	30	0.9	0.9	6.358e + 05	0.952	0.984		0.281	0.348		0.913	0.929		0.052	0.101
0.9	5	20	30	0.9	0.9	1.363e + 01	0.519	0.835		0.202	0.259		0.312	0.390		0.051	0.101
0.9	5	35	30	0.9	0.9	9.340e + 00	0.347	0.780		0.195	0.257		0.205	0.274		0.059	0.108
0.9	5	50	30	0.9	0.9	8.075e + 00	0.261	0.734		0.179	0.235		0.146	0.211		0.048	0.104
0.9	10	5	30	0.9	0.9	7.375e + 05	0.981	0.981		0.267	0.330		0.909	0.928		0.047	0.091
0.9	10	20	30	0.9	0.9	2.408e + 01	0.829	0.829		0.139	0.195		0.317	0.399		0.053	0.103
0.9	10	35	30	0.9	0.9	1.699e + 01	0.765	0.765		0.132	0.187		0.198	0.274		0.053	0.107
0.9	10	50	30	0.9	0.9	1.491e + 01	0.727	0.727		0.128	0.186		0.143	0.222		0.054	0.097
0.9	20	5	30	0.9	0.9	1.041e + 06	0.997	0.984		0.241	0.301		0.908	0.925		0.050	0.101
0.9	20	20	30	0.9	0.9	4.733e + 01	0.979	0.826		0.134	0.190		0.317	0.402		0.049	0.099
0.9	20	35	30	0.9	0.9	3.254e + 01	0.984	0.764		0.107	0.158		0.183	0.256		0.054	0.100
0.9	20	50	30	0.9	0.9	2.833e + 01	0.981	0.723		0.098	0.150		0.142	0.214		0.054	0.102
0.9	50	5	30	0.9	0.9	1.332e + 07	1.000	0.982		0.236	0.295		0.910	0.931		0.049	0.099
0.9	50	20	30	0.9	0.9	1.129e + 02	1.000	0.835		0.104	0.153		0.318	0.392		0.050	0.102
0.9	50	35	30	0.9	0.9	7.956e + 01	1.000	0.765		0.088	0.140		0.201	0.277		0.059	0.106
0.9	50	50	30	0.9	0.9	6.837e + 01	1.000	0.723		0.084	0.143		0.143	0.220		0.054	0.097
0.9	75	5	30	0.9	0.9	2.865e + 08	1.000	0.984		0.216	0.284		0.900	0.917		0.046	0.099
0.9	75	20	30	0.9	0.9	1.699e + 02	1.000	0.840		0.100	0.159		0.315	0.395		0.055	0.105
0.9	75	35	30	0.9	0.9	1.180e + 02	1.000	0.774		0.085	0.136		0.201	0.274		0.057	0.112
0.9	75	50	30	0.9	0.9	1.028e + 02	1.000	0.728		0.083	0.142		0.147	0.219		0.053	0.100
0.9	5	5	30	0.3	0.3	1.514e + 05	0.815	0.924		0.313	0.378		0.776	0.814		0.041	0.088
0.9	5	20	30	0.3	0.3	6.155e + 00	0.144	0.519		0.276	0.335		0.178	0.256		0.047	0.099
0.9	5	35	30	0.3	0.3	4.971e + 00	0.062	0.405		0.266	0.328		0.123	0.187		0.056	0.104
0.9	5	50	30	0.3	0.3	4.646e + 00	0.042	0.368		0.277	0.343		0.099	0.158		0.053	0.098
0.9	10	5	30	0.3	0.3	3.024e + 06	0.923	0.923		0.231	0.297		0.785	0.818		0.051	0.105
0.9	10	20	30	0.3	0.3	1.071e + 01	0.444	0.444		0.193	0.256		0.194	0.279		0.058	0.111
0.9	10	35	30	0.3	0.3	8.664e + 00	0.308	0.308		0.178	0.235		0.120	0.195		0.049	0.101
0.9	10	50	30	0.3	0.3	7.968e + 00	0.244	0.244		0.174	0.238		0.098	0.164		0.056	0.102
0.9	20	5	30	0.3	0.3	2.445e + 08	0.981	0.927		0.188	0.250		0.784	0.823		0.051	0.101
0.9	20	20	30	0.3	0.3	1.998e + 01	0.861	0.398		0.135	0.191		0.201	0.281		0.054	0.106
0.9	20	35	30	0.3	0.3	1.595e + 01	0.808	0.224		0.116	0.171		0.121	0.187		0.053	0.098
0.9	20	50	30	0.3	0.3	1.474e + 01	0.788	0.165		0.118	0.168		0.096	0.167		0.050	0.100
0.9	50	5	30	0.3	0.3	2.242e + 09	0.999	0.921		0.153	0.216		0.778	0.817		0.051	0.096
0.9	50	20	30	0.3	0.3	4.759e + 01	1.000	0.361		0.098	0.152		0.197	0.275		0.047	0.101
0.9	50	35	30	0.3	0.3	3.808e + 01	0.999	0.178		0.090	0.144		0.117	0.190		0.053	0.101
0.9	50	50	30	0.3	0.3	3.499e + 01	0.999	0.105		0.080	0.132		0.094	0.159		0.046	0.096
0.9	75	5	30	0.3	0.3	1.231e + 07	1.000	0.918		0.154	0.221		0.775	0.815		0.048	0.098
0.9	75	20	30	0.3	0.3	7.024e + 01	1.000	0.340		0.086	0.136		0.178	0.262		0.052	0.097
0.9	75	35	30	0.3	0.3	5.719e + 01	1.000	0.179		0.084	0.139		0.128	0.199		0.060	0.109
0.9	75	50	30	0.3	0.3	5.259e + 01	1.000	0.096		0.073	0.127		0.092	0.162		0.048	0.096

Table 11: 5,000 Simulations from DGP with Group Level Heteroskedastic Random Errors, 10 instruments.

Parameters										F-stat			TSLS t-test			CRVE CH			Wildbootstrap CH		
ρ	F	G	N	ι_z	$\iota_u = \iota_\epsilon$	$\hat{F}$	$\hat{F} > 10$	$\hat{F} > F$		$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.10$
0.9	5	20	30	0.9	0.9	41.542	0.918	0.992		0.310	0.392	0.392	0.874	0.907	0.907	0.051	0.099	0.099	0.051	0.099	0.099
0.9	5	35	30	0.9	0.9	15.818	0.713	0.973		0.307	0.396	0.396	0.611	0.687	0.687	0.049	0.103	0.103	0.049	0.103	0.103
0.9	5	50	30	0.9	0.9	11.188	0.516	0.934		0.290	0.376	0.376	0.425	0.517	0.517	0.046	0.093	0.093	0.046	0.093	0.093
0.9	10	20	30	0.9	0.9	74.053	0.989	0.989		0.206	0.282	0.282	0.878	0.910	0.910	0.049	0.097	0.097	0.049	0.097	0.097
0.9	10	35	30	0.9	0.9	28.877	0.964	0.964		0.201	0.271	0.271	0.607	0.687	0.687	0.052	0.098	0.098	0.052	0.098	0.098
0.9	10	50	30	0.9	0.9	20.627	0.932	0.932		0.191	0.269	0.269	0.443	0.532	0.532	0.048	0.098	0.098	0.048	0.098	0.098
0.9	20	20	30	0.9	0.9	148.831	1.000	0.991		0.152	0.221	0.221	0.874	0.907	0.907	0.059	0.104	0.104	0.059	0.104	0.104
0.9	20	35	30	0.9	0.9	54.351	0.999	0.959		0.134	0.206	0.206	0.598	0.674	0.674	0.052	0.100	0.100	0.052	0.100	0.100
0.9	20	50	30	0.9	0.9	39.695	0.999	0.928		0.130	0.197	0.197	0.438	0.534	0.534	0.045	0.093	0.093	0.045	0.093	0.093
0.9	50	20	30	0.9	0.9	358.794	1.000	0.993		0.117	0.185	0.185	0.880	0.909	0.909	0.057	0.108	0.108	0.057	0.108	0.108
0.9	50	35	30	0.9	0.9	132.271	1.000	0.966		0.100	0.154	0.154	0.614	0.690	0.690	0.054	0.108	0.108	0.054	0.108	0.108
0.9	50	50	30	0.9	0.9	95.954	1.000	0.929		0.093	0.152	0.152	0.438	0.534	0.534	0.050	0.098	0.098	0.050	0.098	0.098
0.9	75	20	30	0.9	0.9	545.401	1.000	0.992		0.110	0.179	0.179	0.874	0.909	0.909	0.053	0.101	0.101	0.053	0.101	0.101
0.9	75	35	30	0.9	0.9	197.018	1.000	0.958		0.099	0.155	0.155	0.615	0.692	0.692	0.049	0.099	0.099	0.049	0.099	0.099
0.9	75	50	30	0.9	0.9	142.577	1.000	0.917		0.087	0.140	0.140	0.443	0.530	0.530	0.043	0.094	0.094	0.043	0.094	0.094
0.9	5	20	30	0.3	0.3	7.723	0.227	0.660		0.655	0.731	0.731	0.610	0.689	0.689	0.049	0.095	0.095	0.049	0.095	0.095
0.9	5	35	30	0.3	0.3	4.550	0.030	0.340		0.659	0.732	0.732	0.346	0.433	0.433	0.053	0.103	0.103	0.053	0.103	0.103
0.9	5	50	30	0.3	0.3	3.793	0.007	0.203		0.650	0.722	0.722	0.243	0.338	0.338	0.057	0.109	0.109	0.057	0.109	0.109
0.9	10	20	30	0.3	0.3	12.472	0.520	0.520		0.424	0.512	0.512	0.613	0.686	0.686	0.044	0.099	0.099	0.044	0.099	0.099
0.9	10	35	30	0.3	0.3	7.290	0.175	0.175		0.428	0.515	0.515	0.340	0.430	0.430	0.050	0.104	0.104	0.050	0.104	0.104
0.9	10	50	30	0.3	0.3	6.188	0.077	0.077		0.417	0.513	0.513	0.238	0.330	0.330	0.056	0.106	0.106	0.056	0.106	0.106
0.9	20	20	30	0.3	0.3	22.596	0.895	0.451		0.258	0.341	0.341	0.622	0.695	0.695	0.051	0.103	0.103	0.051	0.103	0.103
0.9	20	35	30	0.3	0.3	13.035	0.682	0.098		0.264	0.347	0.347	0.339	0.435	0.435	0.057	0.106	0.106	0.057	0.106	0.106
0.9	20	50	30	0.3	0.3	11.017	0.546	0.032		0.263	0.344	0.344	0.232	0.325	0.325	0.056	0.107	0.107	0.056	0.107	0.107
0.9	50	20	30	0.3	0.3	51.326	1.000	0.384		0.149	0.218	0.218	0.619	0.700	0.700	0.051	0.101	0.101	0.051	0.101	0.101
0.9	50	35	30	0.3	0.3	30.141	0.998	0.058		0.147	0.214	0.214	0.345	0.432	0.432	0.054	0.101	0.101	0.054	0.101	0.101
0.9	50	50	30	0.3	0.3	25.214	0.999	0.007		0.144	0.218	0.218	0.242	0.337	0.337	0.055	0.104	0.104	0.055	0.104	0.104
0.9	75	20	30	0.3	0.3	78.151	1.000	0.399		0.132	0.196	0.196	0.634	0.704	0.704	0.052	0.103	0.103	0.052	0.103	0.103
0.9	75	35	30	0.3	0.3	44.096	1.000	0.053		0.118	0.186	0.186	0.343	0.432	0.432	0.056	0.108	0.108	0.056	0.108	0.108
0.9	75	50	30	0.3	0.3	37.291	1.000	0.008		0.117	0.185	0.185	0.230	0.327	0.327	0.048	0.100	0.100	0.048	0.100	0.100

8 Bibliography

- [1] David H Autor, David Dorn, and Gordon H Hanson. The china syndrome: Local labor market effects of import competition in the united states. *The American Economic Review*, 103(6):2121–2168, 2013.
- [2] John Bound, David A Jaeger, and Regina M Baker. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American statistical association*, 90(430):443–450, 1995.
- [3] A Colin Cameron, Jonah B Gelbach, and Douglas L Miller. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427, 2008.
- [4] A Colin Cameron and Douglas L Miller. A practitioner s guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372, 2015.
- [5] A Colin Cameron and Pravin K Trivedi. *Microeconometrics: methods and applications*. Cambridge university press, 2005.
- [6] Eliana Carranza. Soil endowments, female labor force participation, and the demographic deficit of women in india. *American Economic Journal: Applied Economics*, 6(4):197–225, 2014.
- [7] Victor Chernozhukov and Christian Hansen. The reduced form: A simple approach to inference with weak instruments. *Economics Letters*, 100(1):68–71, 2008.
- [8] Jean-Marie Dufour. Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica: Journal of the Econometric Society*, pages 1365–1387, 1997.

- [9] Keith Finlay, Leandro M Magnusson, et al. Implementing weak-instrument robust tests for a general class of instrumental-variables models. *Stata Journal*, 9(3):398, 2009.
- [10] Marcelo J Moreira. A conditional likelihood ratio test for structural models. *Econometrica*, 71(4):1027–1048, 2003.
- [11] Brent R Moulton. Random group effects and the precision of regression estimates. *Journal of econometrics*, 32(3):385–397, 1986.
- [12] Douglas O Staiger and James H Stock. Instrumental variables regression with weak instruments, 1994.
- [13] James H Stock, Jonathan H Wright, and Motohiro Yogo. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20(4), 2002.
- [14] James H Stock and Motohiro Yogo. Testing for weak instruments in linear iv regression. *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, 2005.
- [15] Matthew D Webb et al. Reworking wild bootstrap based inference for clustered errors. *Queen's Economics Department Working Paper No*, 1315:18, 2013.
- [16] Jeffrey M Wooldridge. Cluster-sample methods in applied econometrics. *American Economic Review*, pages 133–138, 2003.

II Sharing or gambling over losses?

Abstract

This paper investigates experimentally whether individuals prefer to share an exogenous loss in a deterministic way or to gamble over it. In particular, in some scenarios subjects face an equal allocation of the loss, in others a social lottery with an equal chance to suffer it entirely. The loss is implemented after the endowment is earned. We find that the loss domain does not affect subjects' behavior directly, but it decreases their probability to play in a competitive way.

Keywords: Loss, risk attitude, social preferences.

1 Introduction

In every society, be it small or large, resources are to be allocated. Different disciplines have wondered about the most efficient way to do so, or the fairest one. A strand of literature in behavioral economics, still alive and kicking, has focused on the cornerstone of the allocation process: preferences. How do individuals prefer to distribute resources among agents?

The word “resource” has a positive connotation, but alas! life is not a bed of roses. Expenditures, damages, liabilities - these, too, are “resources” to be shared. Therefore, research cannot avoid to venture in the obscure domain of losses, even though convincingly implementing losses in a laboratory is not an easy task.

In this paper, we analyze how individuals desire and thus decide²³ to allocate an exogenous loss between themselves and another person. The Dictator Game

²³In revealed preferences we trust.

and its modified version, the Probabilistic Dictator Game, serve the purpose well, and they are played in the experiment. However, individuals often do not have the power to fully establish the allocation, be it in terms of outcomes or chances. Because of this, the main focus is on binary choices.

In particular, we consider situations in which one of the alternatives (or both) can be considered “fair”, in the sense that either the loss or the chances to suffer it are equally split. Choices in isolation are also observed to control for risk attitudes.

Our main interest lies in investigating experimentally whether individuals prefer to settle for a deterministic allocation or to let a gamble decide who bears the whole loss, depending on the situation. Moreover, we hope to shed a light on the criteria individuals follow when picking their preferred alternative. With respect to isolation, do people change their choices if these affect someone else’s payoff? If so, is this due to equality concerns or does something else enter the stage? Finally, do people playing with losses react to social context as those dealing with gains?

It is important to underline that the focus here is not on which alternative is fairer from a philosophical point of view, or more efficient. The interest lies solely in establishing what the individuals prefer.

As a matter of fact, the social lottery creates maximum inequality. It is a zero-sum game, hence, if one does not end up losing the whole endowment, the other does. However, it gives a chance not to suffer any loss, to the satisfaction of loss averse individuals. Moreover, when chances are equally split, this alternative can be perceived as fairer than an unequal deterministic allocation (like Machina’s mom, as a third party decision maker, would suggest²⁴).

In the deterministic alternative, both individuals bear a part of the loss to eliminate risk and ensure that each one still has something at the end. In cases of equal allocation, because of inequality aversion, this alternative may seem preferred without any doubt: nevertheless, under losses, individuals tend to like risk. In cases of unequal outcomes, envy or inequality aversion may push individuals toward a

²⁴Machina (1989) uses the parable of a ‘mom’ who must give a non-divisible treat to one of her two children, and prefers a coin toss to giving the treat directly to one or the other.

more promising or equitable gamble.

Returning to the problem of implementing losses in a laboratory, doing so after the subjects are directly provided with the money is likely to generate undesired distortions in their behavior, as they may perceive this money as “manna from heaven” and not as their own. In order to avoid this, our subjects perform a useful task, yet within everyone’s mean, to earn the initial amount of money and reinforce the endowment effect.

As you may have realized from this introduction, we are going to jackknife into a Bermuda Triangle made of risk attitudes, social preferences and losses.

We find no evidence of greater generosity under losses when subjects can freely decide the final allocation. Facing binary choices instead, three types of subjects can be identified: the majority is driven by competitiveness, a significant part compromises for the equal allocation, but does not act against outcome inequality choosing the fair gamble instead, and a small percentage acts as a selfish decision maker.

We also find that the domain does not affect behaviors within each group, but changes the probability to belong to one of them: under losses, less individuals behave in a competitive way.

Section 2 contains a review of the relevant literature. In Section 3 the experimental design is illustrated. Analysis and results can be found in Section 4. Section 5 wraps up and concludes.

2 Literature Review

We know much about individual choices in the domain of gains. Unfortunately (or maybe not), human behavior seems impervious to fully predictable patterns and well-established results in the domain of gains cannot be directly shifted - or capsized - in the domain of losses.

The problem of researching within this domain is that having participants experiencing losses in the laboratory - or deceiving them in believing so - is unethical

and unpractical for future recruitment. Individuals willing to risk a loss of money are also likely to induce a strong selection bias. On the other hand, the use of hypothetical losses makes the experiment vulnerable to easy criticism.

A third way is the "loss after endowment" procedure, in which the experimenter initially provides the subjects with an amount of money that covers all possible losses that may arise during the experiment. The risk in this case is that individuals may fail to feel this money as their own and consider it as "manna from heaven" or "house money" instead. This is already a problem in experiments dealing with gains: in measuring risk attitudes, for example, participants are likely to result more risk loving than they are in their everyday life (Thaler and Johnson, 1990).

Some researches investigate this issue. Findings are mixed. Etchart-Vincent and L'Haridon (2011) claim that there is no significant difference in risk aversion under hypothetical losses, losses after endowment and real losses²⁵, while answers with positive stakes are not the same under the three procedures. Other studies remain skeptical (as Weber and Zuchel, 2005).

An original alternative is to put to work the old adage "time is money" and use wasted time as a substitute for losses, as in Berger et al. (2012).

The method implemented in this paper, "loss after earned endowment", consists in giving an initial endowment to the subjects conditional on the completion of some task, as suggested by Etchart-Vincent and l'Haridon (2011). In this way, the individuals feel entitled to the money and perceive losses as a real ones.

This practice is already in use in some Dictator Games with gains (see Cherry et al., 2002, or Oxoby and Spraggon, 2008). Predictably, the use of earned endowments affects results: individuals tend to be less generous in social games and more cautious when measuring risk attitude.

Having clarified how to implement losses in the experiment, we examine individuals' behavior in this domain.

²⁵Scenarios where "real losses" are implemented are mixed with scenarios with gains.

2.1 Social Preferences and Deterministic Losses

According to Prospect Theory, the disutility associated with losing a certain amount of money is greater than the utility of gaining that same amount. Individuals are then more motivated to avoid losses than to obtain equivalent gains. Hence, one could expect individuals to behave more selfishly when allocating losses.

Baquero et al. (2013) run two Dictator Games with gains and losses after endowment²⁶ and find that the dictators' behavior is similar. Individuals are slightly more generous sharing losses (by 3.5%), a "compassion effect" which is small but significant.

In the same paper, they run two Ultimatum Games. Now, if individuals dislike more the idea of losing everything than the idea of gaining nothing, offers which are equivalently unequal from a numerical point of view should be accepted more often if they concern losses.

Expecting this, proposers should be less generous and favor themselves more under losses than under gains. Moreover, risk seeking in the loss domain suggests that proposers prefer to take a gamble and risk their unfair offer to be rejected, whereas an equal split would seem a safe but less profitable offer.

Baquero et al. find opposite results. Equivalently unequal offers are rejected more often under losses: the maximum loss subjects are willing to bear is 6.18% less than the corresponding gain offer. Proposers, anticipating this, are fairer in the first place (by 6.07%). Buchan et al. (2005) confirm these findings.

One caveat is that both studies use a strategy method and some relevant emotional responses that only appear facing actual offers may be lacking. Moreover, in Buchan et al., each participant plays both as the proposer and the responder. Guth et al. (1982) show that individuals tend to be more generous when they play both roles.

However, Zhou and Wu (2011) present fake offers to the responders and find stronger responses to losses. Punishing a greedy dictator possibly helps responders

²⁶Both players are given an initial (unearned) amount, and a lottery decides which condition (gain or loss) is implemented.

to counteract negative emotions associated with being treated unfairly. Because of loss aversion, both this feeling of unfairness and the punishment that can be put in place are more severe, *ceteris paribus*. For example, when rejecting an offer (-3,-7), the responder imposes on the proposer an extra loss of 7; when he rejects (7,3), he makes the proposer's gain decreases by 7. For loss aversion, the loss of 7 is felt more strongly than the missed gain of 7.

On the other "side" of the Ultimatum Game, Leliveld et al. (2008) test the assumption that proposers act more generously because of fear of rejection and strategic self-interest. They find evidence against it and propose instead the do-no-harm principle (Baron, 1995) as an explanation for this equality hunger in situations involving losses.

The do-no-harm principle states that individuals are reluctant to harm another person to benefit themselves. As inflicting a loss might be considered more harmful than withholding a gain, when playing with losses fairness becomes more important - and self-interest less decisive.

Leliveld et al. also consider the distance between behaviors and norms. They modify the relationship between the proposer's payoff on the responder's decision. Under gains, when the responder's decision is less important for the proposer's payoff, individuals' behavior departs significantly from the social norm towards self-interest. Under losses, however, norms and behaviors always match and the final outcome stays fair.

Using wasted time as a substitute for losses, Berger et al. find a similar pattern. In this version of the Ultimatum Game, the proposer has 60 minutes of waiting time²⁷ to allocate. In case of acceptance, each one waits for the agreed waiting time to receive the show-up fee. In case of rejection, both have to wait for 60 minutes. The results confirm that fair divisions are typical, with a mean of 30 minutes "offered"²⁸.

²⁷"Waiting" means that participants have to sit in an isolated place, hidden by separate visual covers, and they are not allowed to take any personal belonging with them: they can do nothing except waiting.

²⁸The most extreme offer is 40 minutes and it is the only rejected offer.

Even though one may be reluctant to compare losses in terms of money and in terms of time, all the evidence points in the same direction²⁹.

Guo et al. (2013) show that, at the neural level, the fairness-related neural network is activated when playing with losses, but not with gains. Subjects physically experience more unfairness and a stronger desire to sanction the violation of the social relevant norm in the loss domain.

What emerges from these studies is an increased demand for fairness in presence of losses. The decision maker's loss aversion is counterbalanced and overcome by the feeling that allocating losses unfairly is *really* unfair - more than with gains. People dislike losses, but they dislike even more imposing them on others in an unfair way.

At the very basis of this unfairness aversion for losses there may be an evolutionary explanation. As in Zhou and Wu:

In such situations, the need to abide to social norms seems to be more urgent in that the violation of norms may threaten the survival of species.

2.2 Social Preferences and Risky Gains

Preferences concerning the allocation of gains have surely received more attention than their counterpart in the loss domain. However, the subset of studies that analyze their interaction with risk attitudes is not numerous.

In the Probabilistic Dictator Game by Krawczyk and Le Lec (2008), subjects are asked to share chances to win a resource - an amount of money - with another participant. In the treatment of our interest, once chances are shared, a random draw decides which subject gets the whole resource. In this case, subjects are found

²⁹Kern and Chugh (2009) and De Dreu, Lualhati and McCusker (1994) find that, when framing a situation in terms of losses, individuals behave more unethically and are more outcome-oriented, respectively. However, subjects only face hypothetical choices in these experiments.

more selfish than when they share deterministically (they give 13.3% of the total amount vs 8.3% of chances).

A related study is Kocher, Krawczyk and Le Lec (KKL, 2013), where an amount of money of 50 euros has to be allocated between two subjects. The decision maker can either opt for sharing the resource in a deterministic way ($x\%$ of the resource to him, $(100 - x)\%$ to the second player) or for allocating it with a social lottery ($x\%$ chances to win the entire resource to him, $(100 - x)\%$ to the other). The x is fixed in each scenario.

Notice that if risk averse subjects, which should be the majority under gains, do not have any concern for others, they choose the safe option since it has the same expected value as the lottery. If they care for others', too, and are inequality averse, they would stick to the deterministic option, because gambling always generates ex-post maximum inequality.

A majority of subjects actually follow this prediction: no matter the value of x , the safe option is usually preferred. Then, there is a substantial minority whose behavior is affected by their relative position. In a disadvantaged position, where $x < 25$, they risk more than in isolation (this is confirmed by Bolton and Ockenfelds, 2010). When equality of payoffs can be implemented by the deterministic outcome, they still prefer to gamble. Finally, in advantaged position, they take less risk.

For this group of individuals, results point towards the existence of a social reference point, the other's payoff, as suggested by Linde and Sonnemans (2012), which in turn triggers responses predicted by Prospect Theory. When decision makers are in a disadvantaged position, they frame their situation as a loss and are willing to risk more, and vice versa.

However, notice that these individuals tend to choose the gamble even when the safe option implies equality of outcomes. A possible explanation is that gloating, the utility from being ahead of others, is stronger, in absolute value, than the disutility of being behind (Bault et al, 2008), a sort of reflection of loss aversion in the social context. If this is true, individuals prefer to gamble and try to be ahead.

Another related explanation is that these individuals value a lot the mere idea

of winning the game by being ahead, independently of the absolute difference in the final payoffs.

Summing up, most participants always choose the safe option, while another substantial part chooses the lottery when they are in an unfavorable position, in an attempt to evade it.

3 Experimental Design

The experiment is composed of two treatments, “Loss” and “Gain”, the latter being ancillary to the first. Each subject is assigned to only one treatment, hence a between subject design allows to compare decisions under gains and losses. The experiment was held in three sessions, one in July and two in September 2016, in CESARE laboratory (Rome). Instructions were given in Italian to ease comprehension. The z-Tree software was used (Fischbacher 2007).

The Loss treatment is explained first. In order to implement losses during the experiment, subjects initially complete a task and get paid a flat rate of 20 euros for it. During the task, they are aware that other stages will follow, but they do not know what these will envisage.

The task consists in filling up a dataset with answers to a survey provided by other students on paper sheets long before the experiment. Subjects are asked to insert answers (letters from A to E) in a form appearing on the screen of their desktops. Only the id number of the responder and the answers are provided, preventing any influence on their choices in the following stages.

The task was designed to enhance the endowment effect as much as possible, so that losses in later stages are actually perceived as money out of one’s pocket. Its usefulness was stressed to the subjects, as they are informed that the dataset would be used in other independent studies. At the same time, it was within everyone’s mean.

We did not control the quality of the insertion during the experiment because this would invalidate the perceived usefulness of the task. However, an ex-post

examination of the dataset created by the subjects reveals that only one subject “randomized” the insertion and was consequently excluded from the analysis³⁰.

Due to the laboratory policy, it is not possible to materially hand in the money to the subjects at the end of the task and then ask them to stay until the end of the experiment. However, they are again informed that they have earned 20 euros before proceeding.

At this point, subjects are given new instructions with the payment system (illustrated below). The experiment can now be ideally divided in three parts.

The first one is the core of the experiment. Subjects are matched anonymously in pairs and presented with a series of 13 binary choices in random order. For each of them, they are asked whether they prefer to allocate a loss of 20 euros sharing it deterministically or gambling to avoid it entirely.

The deterministic allocation is represented as $(-l, -20 + l)$, where the decision maker loses l and the other subject covers the remaining loss. The social lottery is expressed as $(q, (-20, 0); 1 - q, (0, -20))$, where q corresponds to the scenario in which the decision maker loses everything and the other loses nothing $(-20, 0)$, and vice versa for $1 - q$.

There are two scenarios types, in which subjects decide between:

- Eqd - an equal deterministic allocation and the lottery with different chances to win;
- Eqp - a deterministic allocation and a lottery with equal chances to win;

The scenario in which the choice is between an equal deterministic allocation and a lottery with equal chances to win is common to both scenario types.

The safe outcome ranges from -4 to -16, while probabilities range from 0.2 to 0.8. Extremely low or high values of the probability, and hence the corresponding

³⁰We used the fact some answers sheets were the same across sessions. Over a total of 500 sheets inserted (with repetitions) 288 could be double-checked and only 4 of them contained mistakes.

deterministic allocations, are not used in order to avoid probability over/under weighting as suggested by Prospect Theory.

Probabilities sum up to one so that the expected value of the loss from an ex-ante point of view (the one a social decision maker could have) is equal to the sure loss in the deterministic allocation. The total amount of wealth stays constant across scenarios and alternatives, so that social efficiency measured in these terms is not at stake. Also cooperation and strategic interactions among individuals are completely absent.

In the second stage, subjects play a standard Dictator Game over a loss 20 euros and a Probabilistic Dictator Game as in Krawczyk and Le Lec (2008), in which they split chances to completely avoid the loss. They are also asked which game they have preferred between the latter two.

In the third stage, subjects play the same 13 scenarios as before, but in isolation.

For payment, one scenario among the social binary choices, the dictator games and the isolation binary choices is randomly selected by the computer for all subjects. If the scenario is played in isolation, each subject's choice is directly implemented; otherwise, one subject is selected with equal probability in each pair and his/her decision is implemented. If the decision maker's choice includes a lottery, the computer randomizes one last time and outputs the winner.

The final payoff is made of the 20 euros earned with the initial task, plus or minus whatever is earned in the following stage of the experiment, plus a show-up fee of 5 euros.

After the final payment is privately revealed and before they receive it, subjects answer a questionnaire on demographics, political/religious beliefs and the clarity of the experiment itself. Those who, on a scale from 0 to 4, responded 0 (*Very little*) or 1 (*Little*) to the question "How clear where the instructions of the experiment?" were excluded.

In the Gain treatment, subjects do not complete any initial task. They start immediately with the social binary choices. Everything else is the same as in the Loss treatment, both in terms of choices (framed in terms of gains, of course) and the selection of the scenario for payment purposes. The final payment in this case is equal to what subjects earned during the experiment plus the show-up fee of 5 euros.

3.1 Sample description

A total of 110 subjects were recruited. The final dataset includes 100 subjects, quite evenly distributed between the two treatments. As stated before, we excluded one subject who randomized the insertion of answers in the initial task and 9 who stated that instructions were unclear (4 in the Loss treatment). Subjects were currently enrolled as or had been students at LUISS university.

Table 12: Data description

	Loss	Gain	p-value
N	49	51	
Female (%)	40.82	43.14	0.816
Age (mean)	22.20	23.27	0.022
Clarity (mean)	3.25	3.31	0.644
No of past experiments (mean)	4.41	5.61	0.081

4 Results

In the following analysis, for easiness of comparison between the two treatments, we “translate” loss outcomes in terms of the money the subjects would still have after the loss is implemented: $(-l, -20 + l)$ is now represented as $(d, 20 - d)$, where

$d = 20 - l$. For example, the allocation $(-4, -16)$ becomes $(16, 4)$.

Similarly, probabilities are restated in terms of chances to remain with 20 euro, hence the social lottery under losses is now identified with $(p, 1 - p)$ where $p = 1 - q$:

$$(q, (-20, 0); 1 - q, (0, -20)) \rightarrow (p, (20, 0); 1 - p, (0, 20))$$

For example, the lottery in terms of q $(0.2, 0.8)$, with a 20% probability to lose everything for the decision maker, becomes $(0.8, 0.2)$ in terms of p , where 80% is the chance to remain with 20 euros.

Nothing changes in the gain scenarios. Under this notation, *ceteris paribus*, a selfish agent prefers higher d and p regardless of the treatment. Notice that the subjects' payoff structure is the same between treatment for the same scenario.

The analysis first deals with subjects' allocation preferences when they have maximum freedom of movement, namely the two Dictator Games. After that, binary choices between the deterministic allocation and the social lottery are considered.

4.1 Dictator Games

Figure 2 represents the distributions for the two games. In order to ease comparison between them, for the Probabilistic Dictator Game the expected value of the amount the subjects keep for themselves is reported instead of the probabilities. The green lines are the average amount kept by the decision makers, while the red ones are the median values. Some statistics are provided in Table 13.

Table 13: Dictator Games statistics

	Losses			Gains		
	Mean	Median	% Complete Selfish	Mean	Median	% Complete Selfish
DDG	17.6	19.0	49.0	16.8	19.0	41.2
PDG	18.0	19.8	44.9	17.6	19.8	43.1

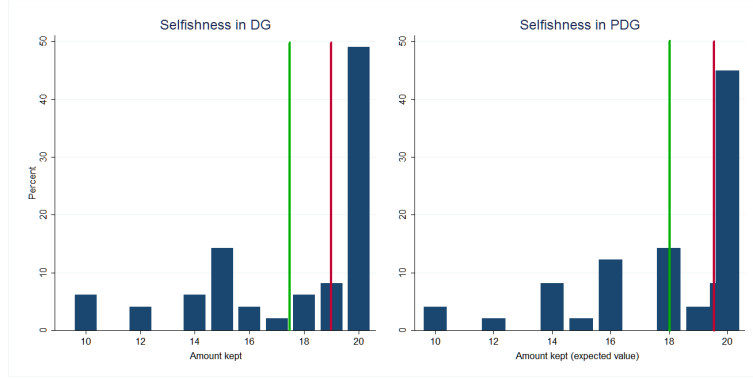


Figure 2: Deterministic and Probabilistic DG, Losses

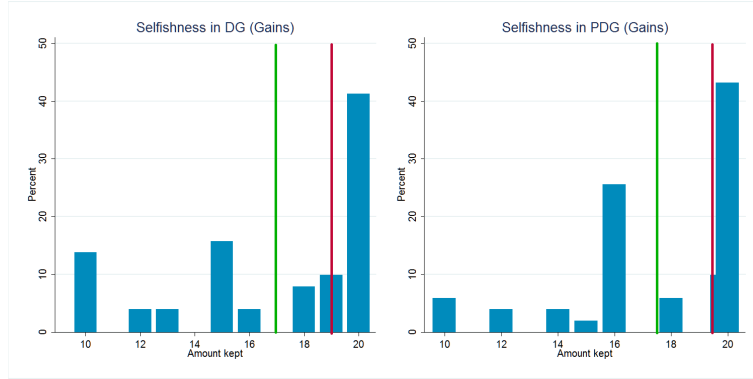


Figure 3: Deterministic and Probabilistic DG, Gains

In the Deterministic Dictator Game (DDG) with losses, half of the subjects allocate the entire loss to the other individual. Only 6% (3 subjects) share the loss equally.

In the DDG with gains, a smaller percentage of subjects, yet non significantly so, take the whole amount for themselves, and 3% share it equally. No differences are found between means, medians and entire distributions with respect to losses.

In the Probabilistic Dictator Game (PDG) with losses, 45% of subjects make sure to win the lottery. Interestingly, 8% keep 99% of chances, probably in an attempt to mitigate one own conscience by giving the other some probability, however tiny. This type of subjects, “type 99”, and the completely selfish ones together

amount for 53% of the sample. Two individuals (4%) allocate chances equally.

In the PDG with gains, figures are very close to the case of losses: 43% of subjects keep all the chances to win the lottery, 9% are “type 99”, and about 6% split chances in half. Means, medians and distributions are again similar .

On average, in the DDG subjects keeps 12.25% of the loss, while in the PDG the percentage of the loss in expected terms lowers to 8.90%. However, the paired t-test reveals that the difference is not statistically significant. Distributions, too, are similar according to the Wilcoxon signed-rank test. Individuals are as altruistic (or as selfish) sharing deterministically as they are sharing probabilistically.

On the contrary, under gains, the percentage given by the decision maker to the other subject in DDG (15.78%) and in PDG (11.96%) are statistically different (p-value=0.036): as found by Kocher and Le Lec, the average donation is more generous when subjects share deterministically. Distributions are mildly different (p-value=0.100).

Hence, allocation choices on sure gains are not distinguishable from those on sure losses, and the same holds true for the distribution of chances, similar between the two domains. However, while individuals allocate sure gains more generously than chances in the same domain, no differences emerge between the allocations under losses.

The correlation between the allocations in the two games under losses is 0.72 (0.73 under gains). They are plotted one against the other in Figure 4: the 45-degree line is in green. The STATA option ‘jitter’ is used to add spherical noise to the plotted data and give an idea of the actual density of each point.

Allocation choices concentrate in the upper-right quadrant of the picture, the most selfish one. Around 39% of subjects are completely selfish in both games, 47% if also “type 99” subjects are considered; only 2 subjects (4%) share equally both losses and chances to lose.

Similarly, under gains, 37% of the subjects take all for themselves, while 3 subjects (6%) split equally in both circumstances.

Interestingly, when asked to state one’s preference for the DDG or the PDG, the

loss sample perfectly splits in two, while the percentage of subjects preferring the first is much higher under gains: 74.51%. This difference (p-value=0.015) may be a signal that individuals feel more “comfortable” with risk under losses. However, the preferred game is not a relevant predictor for altruism, neither in gains nor in losses.

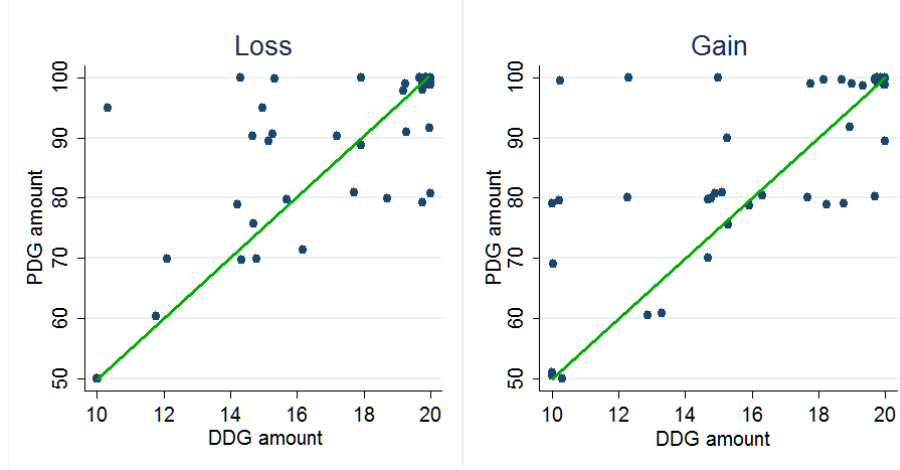


Figure 4: Deterministic vs Probabilistic DG, Losses and Gains

4.2 Sharing or gambling?

In this Section, we analyze subjects’ choices when they are asked to opt either for the deterministic allocation or for the social lottery. A summary for each scenario, divided by domain and context, can be found in Appendix 6.1.

The Eqd scenario under losses is considered first. Here, the deterministic allocation is the equal split while the probability to completely avoid the loss varies.

The majority of individuals prefer the deterministic allocation when the chances to win the social lottery are below 50%, as the red bars in Figure 5(a) show.

We run a random effect logit model for the probability to select the lottery over the deterministic alternative in each scenario. This allows to take into account the clustered nature of the data, as choices made by one subject are very likely to be correlated one another. We report average marginal effects in Table 14, while logit

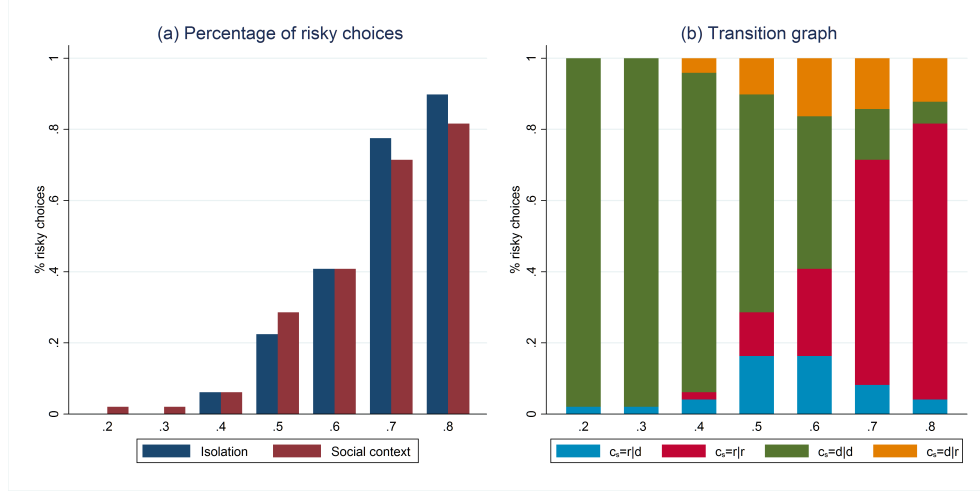


Figure 5: Egd scenarios, Losses

models coefficients are in Appendix 6.2, Table 18.

The percentage of risky choices increases with the probability associated to them, as one would expect from selfish individuals. Indeed, risk levels in social context are similar to those in isolation.

At this point, one needs to investigate whether the presence of other subjects does not influence choices at all, i.e. individuals are unresponsive to the social context, or whether the overall picture masks some interesting features of the data, such as heterogeneous behaviors. The transition graph in Figure 5(b) hints at the second explanation.

This graph describes the percentage of choices in the social context conditional on choices in isolation for each scenario. It is immediately apparent that the overall unresponsiveness to the social context is the result of conflicting forces: for medium/high probabilities to win, around 25-30% of choices change from isolation in each scenario. Similar percentages switch in opposite directions.

Moving to gains, the majority of subjects prefer the deterministic, equal allocation unless the probability to win the social lottery is very high ($p > 60\%$). The

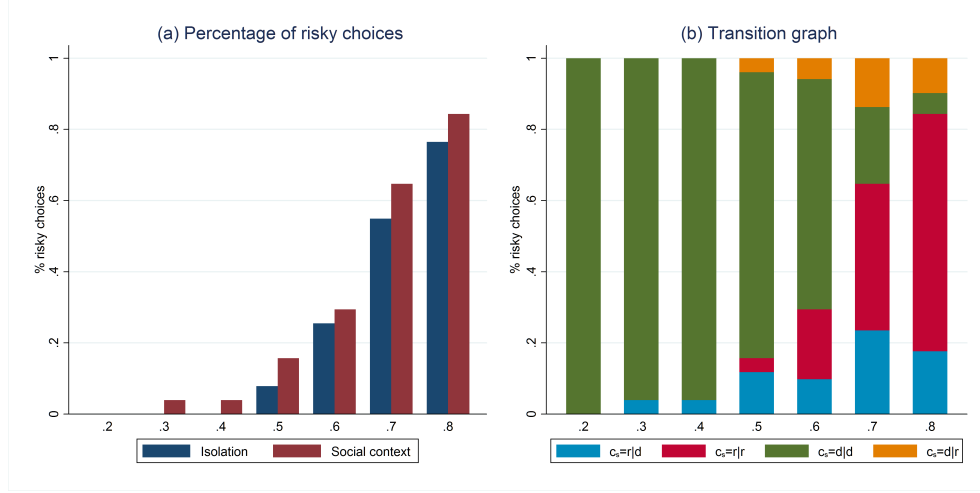


Figure 6: Egd scenarios, Gains

trend of risky choices again increases in p .

Results in Table 14 show that more risk - around 6 percentage points - is taken in social context. Individuals tend to compete more to obtain the whole prize. The magnitude of this effect is far from being negligible: the average percentage of lottery choices is 23.5% in isolation, hence the social context induces individuals to take 25% more risk.

As under losses, the transition graph reveals heterogeneous behaviors. In this domain, however, shifting to the deterministic option is a milder phenomenon, while the percentages of switches towards the social lottery are similar to losses. This results in the increase in social risk taking highlighted by the econometrics results.

The aggregate regression partially summarizes our findings so far: the interaction of social context and losses is negative and significant at 10% level, and completely undoes the positive effect of being in social context.

An important piece of information to draw from the aggregate regression is that, in isolation, the percentage of risky choices increases by 10% in the loss domain. In order to check whether the reflection effect as predicted by Prospect Theory

Table 14: Equal outcomes, average marginal effects

	Loss	Gain	Aggregate
E(lottery)	0.0694*** (0.000)	0.0689*** (0.000)	0.0688*** (0.000)
social	-1.62e-18 (1.000)	0.0606** (0.004)	0.0617** (0.006)
loss			0.107** (0.004)
social x loss			-0.0617 (0.052)
Observations	686	714	1400

p-values in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

applies, we focus on the number of risky choices in Eqd isolation scenarios, $reqd_i$. This provides a good proxy for risk attitudes, since subjects behave rather regularly in isolation, switching from the safe option to the risky one only once.

Since we do not allow to state indifference between the two alternatives, we use two definitions of risk loving individual: a strict one, that includes individuals with $reqd > 4$, and a slack one, with $reqd \geq 4$.

The two-sample Wilcoxon rank-sum test finds that distributions of risk taking levels differ between domains (p-value=0.012). The average numbers of risky choices, 2.37 under losses and 1.65 under gains, are statistically different, too (p-value=0.0063). However, the average subject under losses is not risk loving, no matter the definition (p-value>0.99).

We do not confirm the reflection effect described by Prospect Theory. Nonetheless, we do find a higher percentage of risk loving individuals under losses (p-value=0.017 for the slack definition, 0.074 for the strict one). Our findings agree with Holt and Laury (2005): risk aversion is less common in the loss domain.

On the other side, in social context the percentage of risky choices does not differ between domains. We return on this point later on with more details.

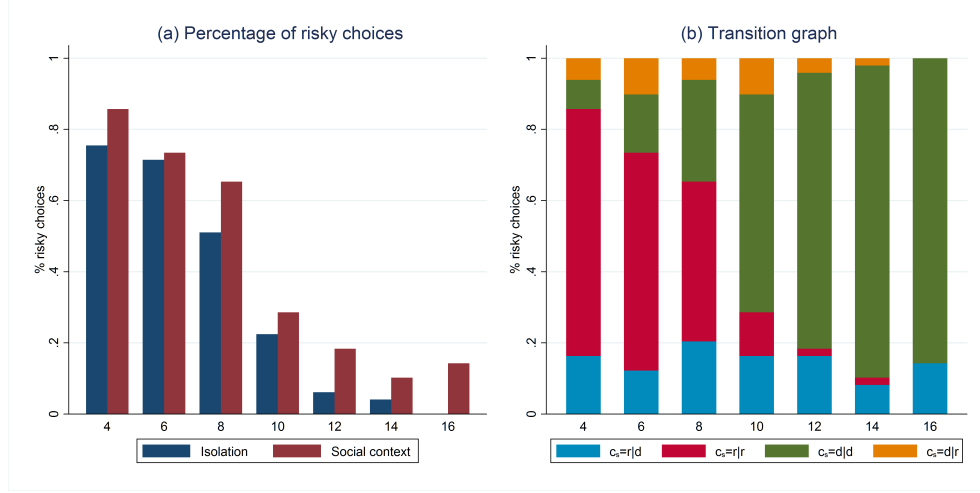


Figure 7: Eqp scenarios, Losses

In Eqp scenarios, where the loss can be allocated by means of a fair gamble against a deterministic allocation, the first is preferred by a large majority of subjects as long as the deterministic amount is below 10 euros. The trend is decreasing in d , as it is in isolation.

Table 15 shows that the social context induces subjects to opt for the lottery more often than in isolation. In particular, this effect amounts to 10 percentage points. Considering that the average percentage of risky choices in Eqp scenarios is 33%, the reaction to the social context is remarkably strong and could be related to the perceived fairness of the social lottery as well as to competitive behavior.

The transition graph in Figure 7(b) reveals a stable percentage of choices switching to the social lottery (an average of 15%). The magnitude of these shifts prevails on the few movements in the opposite direction (average of 5.5%), which decrease alongside the value of the deterministic option.

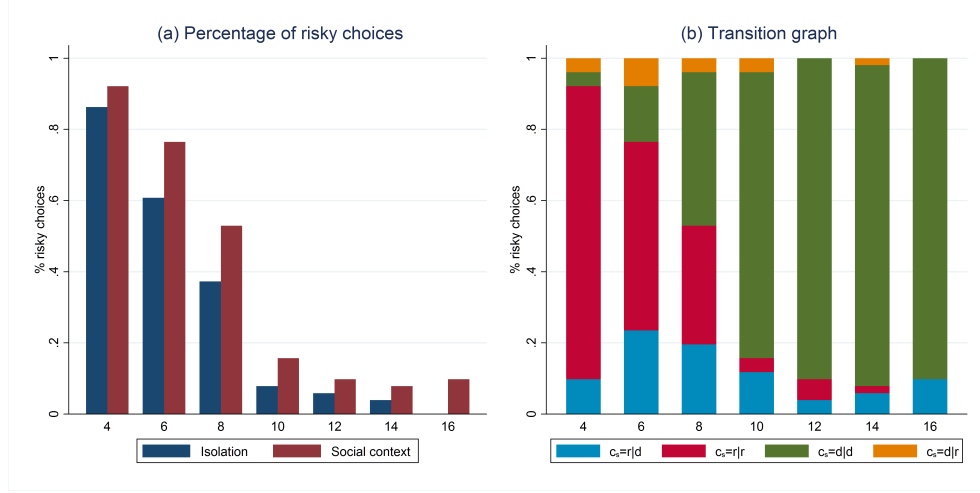


Figure 8: Eqp scenarios, Gains

The situation is rather similar under gains. A large majority of subjects prefer the deterministic allocation when it guarantees an equal split or a more favorable share, and a similar pattern emerges in isolation, too.

The effect of switching to the social context is very stable between domains (10 percentage points again). The 28.9% of choices in isolation are lotteries, thus the effect is again large.

The transition graph is similar to the one under losses: an average of 3% switch to the deterministic allocation, while 12% do the opposite.

The aggregate regression reveals that, in isolation, risk taking levels are the same under gains and losses. The same result emerges if we use the number of risky choices in Eqp scenarios, $reqp_i$. The evidence from both scenario types suggests that losses affect risk attitudes only through one's perception - or weighting - of probabilities, and that $p = 50\%$ is a fixed point.

As in Eqd scenarios, the domain does not affect risk attitudes in social context.

Summing up, percentages of risky choices in social context are similar between gains and losses. Their increment with respect to isolation is similar in importance

Table 15: Equal chances, average marginal effects

	Loss	Gain	Aggregate
det outcome	-0.0675*** (0.000)	-0.0699*** (0.000)	-0.0687*** (0.000)
social	0.102*** (0.000)	0.0958*** (0.000)	0.0975*** (0.000)
loss			0.0396 (0.389)
social x loss			0.00321 (0.927)
Observations	686	714	1400

p-values in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

and magnitude in both domains when one of the alternatives in the fair gamble. In scenarios entailing the equal allocation, instead, the domain matters: under losses, the social lottery is chosen as much as the gamble in isolation, while under gains the percentage of risky choices increases in social context.

However, for Eqd scenarios especially, overall results may mask some important features of the data that it is worth investigating in more detail.

4.2.1 Heterogeneity

Transition graphs suggest that heterogeneous behaviors are present, and their analysis could shed a new light on aggregate results presented so far. We run a k-means clusters analysis on the subjects' choices for both scenario types, in isolation as well as in social context. We identify three clusters, whose composition is reported in Table 16. Figures 9 and 10 shows the percentages of risky choices in both scenario types. Transition graphs by group are in Appendix 6.3.

Group 1 chooses to gamble less often than in isolation, but only when the equal split can be implemented. In Eqp scenarios instead, their behavior is similar between contexts. Group 2 includes competitive subjects, that consistently take more risk in social contexts under both scenario types. Group 3 gathers few subjects

Table 16: Percentage of subjects in each group by domain

Group	Loss	Gain	Aggregate
1	38.8 (21)	23.5 (15)	31.0 (36)
2	49.0 (22)	64.7 (30)	57.0 (52)
3	12.2 (6)	11.8 (6)	12.0 (12)*

who behave similarly in isolation and in social context.

Group 3 is the smallest group: there are 12 subjects, evenly distributed between gains and losses. The unresponsiveness of this group is linked to selfishness, as it is sustained later on. However, a separate story must be told for two subjects, both females, who always prefer the safe option. They also behave rather altruistically in the dictator games, as they keep for themselves 14 and 15 euros in the deterministic one and 14 and 16 in the probabilistic one, below the means and medians of both games. These two subjects can thus be classified as altruistic and risk averse. Because of the small sample size of Group 3, the inclusion of these outliers alters significantly the statistics, hence tests on this group neglects them.

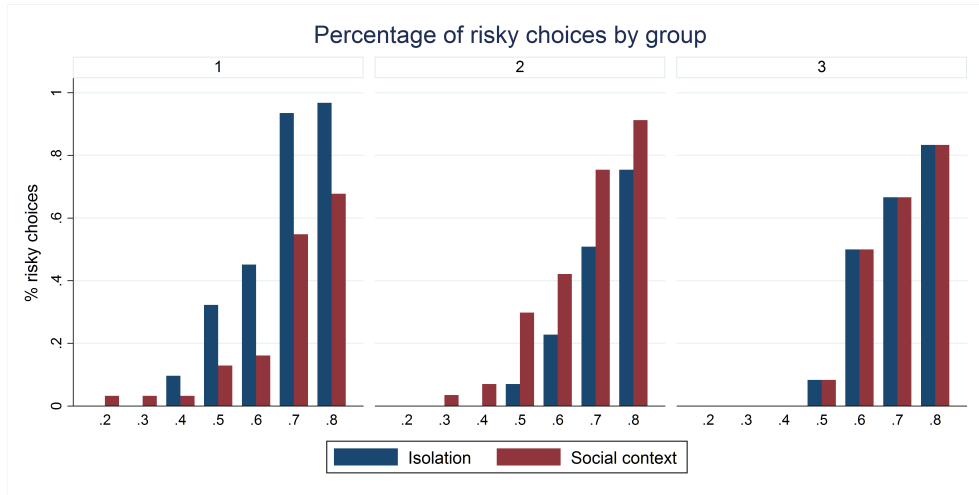


Figure 9: Eqd scenarios

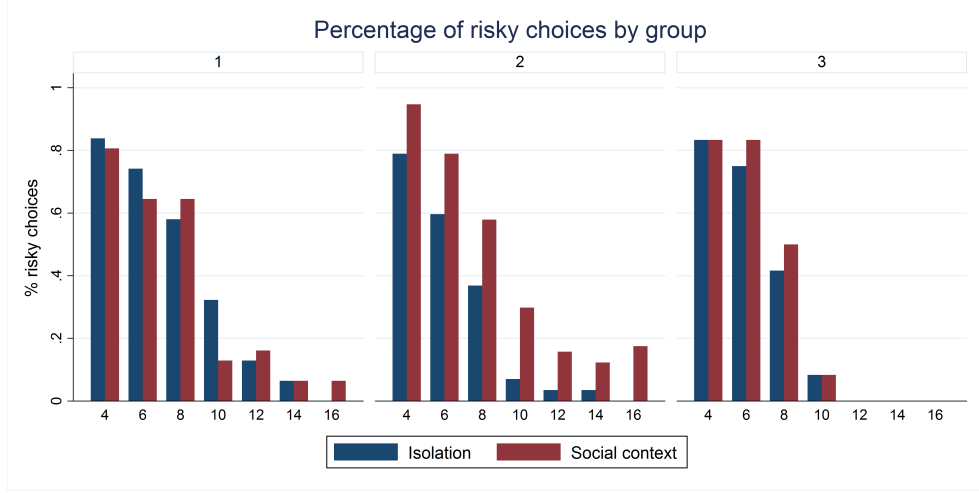


Figure 10: Eqp scenarios

The remaining subjects exhibit a behavior which is consistent with the one of a rational, selfish decision maker. To check for this, we consider their choices in the dictator games. These subjects are more selfish than the members of the other groups in both games (p-value < 0.001 for DDG, 0.005 for PDG) and they are more often completely so (0.009 for DDG, 0.050 for PDG).

Group 1 and 2 do not differ in altruistic behaviors as measured by the social preferences controls, similarly to what happens in KKL. Also other covariates, such as measures of wealth, number of siblings, parents' education, religious/political beliefs are unable to assign individuals to one of these two groups. We examine the effect of gender below.

It is rather plausible to assume that Group 2 is motivated by a competitive behavior, which is not linked to altruistic concerns and risk attitudes (even though, in Eqp scenarios, increased risk taking may be reinforced by the fact that the social lottery gives fair chances). Such competitiveness has rational features: consider Figure 11(b), where the solid line represents the percentage of switches to the social lottery for each Eqp scenarios, calculated over the number of choices that in isolation were deterministic. This percentage increases as the social lottery gets

more convenient.

Similarly, in Eqp scenarios, as one's share in the deterministic allocation increases, the probability to switch to the social lottery decreases, as shown in Figure 12(b).

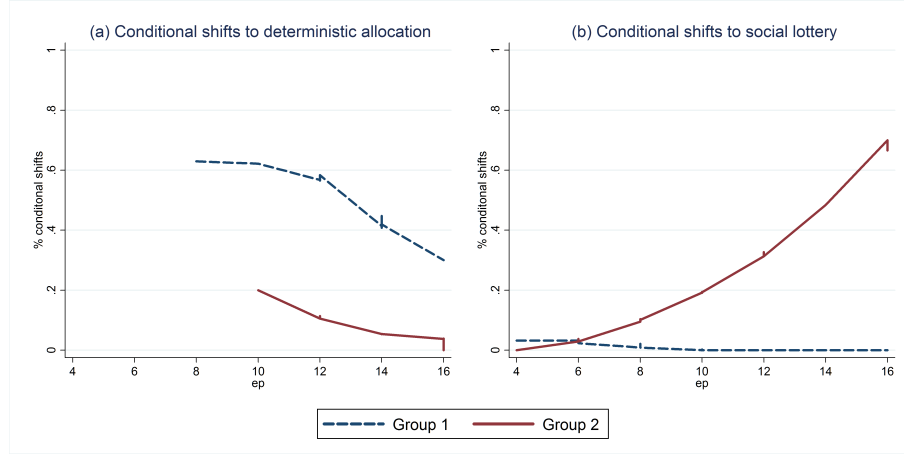


Figure 11: Eqp scenarios, by group

Percentages of conditional switches for low probabilities are missing since no subject chooses the lottery in isolation in those scenarios.

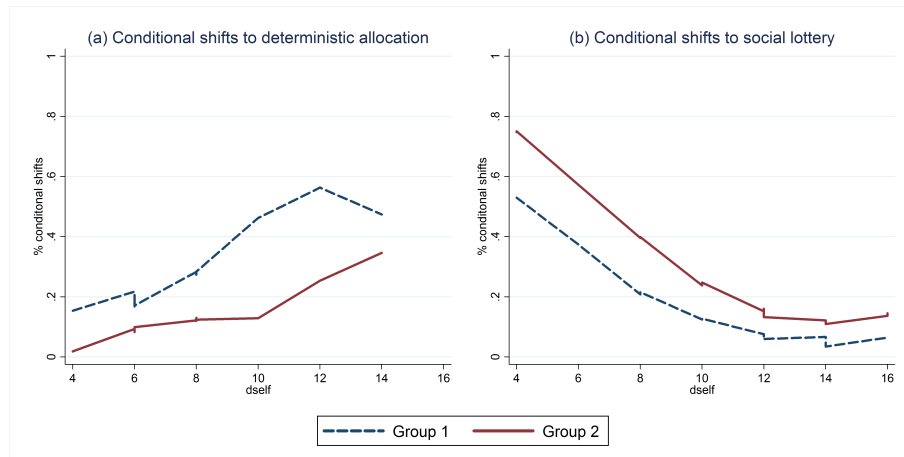


Figure 12: Eqp scenarios, by group

Percentages of conditional switches for high deterministic outcomes are missing since no subject chooses the lottery in isolation in those scenarios.

Factors behind Group 1 behavior are more ambiguous. The stronger preference for the deterministic allocation in Eqd could be driven by fear of losing the social lottery - a shame in gambling and failing against another individual -, by a preference for the fair outcome over the gamble, or by aversion for chances inequality

The first explanation fails to account for the overall unresponsiveness in Eqp scenarios, even though some heterogeneity here remains despite clustering, as the transition graphs in Figure 14 reveals³¹. An average of 9.2% of choices switch toward the deterministic alternative; fear of competition may justify them.

The second explanation is linked to a preference for equality that does not materialize when the amount (positive or negative) is freely allocated, but seem to emerge when the choice is constrained to a binary set. Rather than by a pure desire for the equal allocation, subjects may be attracted by the focal point it represents in terms of social norms. This preference for the equal allocation, in fact, seems not to imply a deterministic inequality aversion, as the results in the Eqp scenarios indicate.

The percentage of switches to the equal allocation in Eqd scenarios conditional on the the number of risky choices in isolation, shown in Figure 11, decreases as the social lottery becomes more convenient. This excludes the hypothesis of inequality aversion linked to chances to win the social lottery.

In conclusion, combining the first two explanations, the equal allocation can be seen as a socially acceptable compromise that Group 1 often prefers to the social lottery, reversing what risk attitudes would prescribe.

In Eqd scenarios, we observe more gambling in social context under gains, but not under losses. At this point, one may wonder whether this is due to the composition of the sample - to the fact that more subjects behave competitively under gains - or to more aggressive risk taking by competitive subjects in the same domain (and/or to a milder risk reduction by the others).

As Group 3 does not react to the social context and its size in percentage points

³¹The small sample size does not allow to perform an additional, sensible clustering analysis for this group.

is the same under both domains, we restrict our attention to the remaining groups.

Table 16 suggests that Group 1 and Group 2 are nearly equally sized under losses, while under gains the second is nearly twice the first: the p-value on the significance of the difference is 0.001. In addition, we test that Group 1 is significantly smaller under gains than under losses, and we cannot reject such hypothesis at 5% level (p-value=0.044). This confirms the presence of a composition effect.

In order to check whether this is topped by a change in behaviors, once we condition for the groups size, we introduce a new type of variable that measure how “aggressive” one’s behavior is. We define rdd_i as the ratio of switches to the social lottery over the total of safe options selected in isolation for each individual in Eqd scenarios. This describes how often each individuals switches to the risky option, considering how many times there is the chance to do so. This variable is not significantly different between domains, both considering each group separately and in aggregation.

We also control for the possibility of a more aggressive risk reduction in social context under losses. The variable is rdd_i the ratio of switches to the deterministic option over the total of lotteries selected in isolation for each individual in Eqd scenarios. Again, no difference between domains was found.

Table 17:

Group	Losses				Gains				Aggregate			
	rdd	rdl	rpdl	rpl	rdd	rdl	rpdl	rpl	rdd	rdl	rpdl	rpl
1	43.7	0.9	23.1	16.8	54.5	2.8	29.2	5.0	47.7	1.6	25.4	12.2
2	6.0	19.3	9.8	32.7	6.4	18.4	5.5	24.7	6.2	18.8	7.1	28.2
3	0	0	0	0	0	0	0	5.6	0	0	0	3.3

In Table 17 we include also $rpdl_i$ and rpl_i , the ratio of switches to the deterministic option and to the social lottery respectively over the total of lotteries/safe options selected in isolation in Eqp scenarios.

We conclude that the observed difference between the effect of social context in gains and losses in Eqd scenarios is solely due to the change in relative percentage of competitive subjects between domains. Said differently, the domain influences the probability to behave in a competitive manner or not; however, conditionally

on this, subjects take similar choices under gains and losses.

It is interesting to notice that, under losses, the percentage of females in Group 1 is significantly higher than in the other groups in aggregate (p-value=0.007) and than Group 2 in particular (p-value=0.003), while under gains this percentage drops to the common level³². Losses thus seem to influence especially women's probability to behave in a competitive manner, leveraging their altruistic preferences³³ or their weaker taste for competition (Niederle and Vesterlund, 2007), or both.

5 Conclusions

Loss aversion suggests that the allocation of a negative resource may be a more delicate task than its counterpart in the domain of gains. This paper shed a light on how individuals prefer to allocate losses between themselves and another person, in particular when the decision entails a deterministic split and a social lottery. Crucially, losses are implemented after the endowment is earned completing an effort task that the subjects may perceive as actually useful.

We find that, on average, people prefer to share the loss than to gamble over it. The more convenient the social lottery is with respect to the deterministic allocation, the more individuals are tempted to aim at the whole pie.

Once controlling for choices in isolation, clustering is used to unmask heterogeneous behaviors: facing the equal split, the majority reacts to the social context in a competitive way, a smaller group opts more strongly for the deterministic allocation, probably seen as a good compromise of social norms and caution to avoid losing vis à vis another subjects, and very few subjects mimic choices in isolation. In aggregate, this results in a level of risky choices which is similar to isolation.

³²Aggregating both domains, the percentage of women in Group 1 is significantly higher than in other groups (p-value=0.004).

³³Women are usually found more generous than men, and our results confirm this: in the dictator games they keep less than men and are less often completely selfish (p-values on the significance of the difference in the average donation equal to 0.011 in DDG, 0.049 in PDG; in the percentage of complete selfish subjects <0.001 in DDG, 0.007 in PDG.)

When the social lottery gives fair chances to win the whole prize, the positions mainly reduce to two: unresponsiveness or more competition. In this case, overall, more risk is taken when playing with, or better against, someone else.

Apart from the case of selfish decision makers, social preferences, as elicited by the dictator games, do not help in predicting one's membership to one of the other two groups.

The domain does not impact generosity as measured by the deterministic and the probabilistic dictator games. Similarly, the overall percentage of risky choices in social context, either facing the fair split or the fair gamble, stays constant between gains and losses. However, playing in the loss domain decreases the probability to behave in a competitive manner.

Whether the do-no-harm principle applies more strongly to the domain of losses remains an open question. The evidence presented in this papers suggests that it crucially depends on the context, and that concerns for others may have indirect, unexpected ways to show up.

6 Appendix

6.1 Percentages of risky choices by scenario, context and domain

Type	Scenario		Loss % risky choices		Gain % risky choices	
	d	q	I	S	I	S
Both	(10,10)	(0.5,05)	22.4	28.6	9.8	15.7
Eqd	(10,10)	(0.2,08)	0.0	2.0	0	0
Eqd	(10,10)	(0.3,0.7)	0	2.0	0	3.9
Eqd	(10,10)	(0.4,0.6)	6.1	6.1	0	3.9
Eqd	(10,10)	(0.6,0.4)	40.8	40.8	25.5	29.4
Eqd	(10,10)	(0.7,0.3)	77.6	71.4	54.9	64.7
Eqd	(10,10)	(0.8,0.2)	89.8	81.6	76.5	84.3
Eqd	Total		33.8	33.2	23.5	28.9
Eqp	(4,16)	(0.5,05)	75.5	85.7	86.3	92.2
Eqp	(6,14)	(0.5,05)	71.4	73.5	60.8	76.5
Eqp	(8,12)	(0.5,05)	51.0	65.3	37.3	52.9
Eqp	(12,8)	(0.5,05)	6.1	18.4	5.9	9.8
Eqp	(14,6)	(0.5,05)	4.1	10.2	3.9	7.8
Eqp	(16,4)	(0.5,05)	0	14.3	0	9.8
Eqp	Total		33.0	42.3	28.9	37.8

6.2 Additional regressions

Table 18: Equal outcomes

	Loss	Gain	Aggregate
E(lottery)	0.760*** (0.000)	0.935*** (0.000)	0.831*** (0.000)
social	-1.78e-17 (1.000)	0.823** (0.005)	0.745** (0.007)
loss			1.291** (0.005)
social x loss			-0.745 (0.054)
Constant	-9.475*** (0.000)	-13.05*** (0.000)	-11.64*** (0.000)
lnsig2u			
Constant	0.978** (0.003)	1.422*** (0.000)	1.178*** (0.000)
Observations	686	714	1400

p-values in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: Equal chances

	Loss	Gain	Aggregate
det outcome	-0.596*** (0.000)	-0.851*** (0.000)	-0.696*** (0.000)
social	0.903*** (0.000)	1.167*** (0.000)	0.989*** (0.000)
loss			0.401 (0.390)
social x loss			0.0325 (0.927)
Constant	4.318*** (0.000)	5.835*** (0.000)	4.704*** (0.000)
lnsig2u Constant	1.169*** (0.000)	1.590*** (0.000)	1.326*** (0.000)
Observations	686	714	1400

p-values in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 20: Equal outcomes

	Group 1	Average marginal effects	Group 2	Average marginal effects
E(lottery)	0.923*** (0.000)	0.0688*** (0.000)	0.972*** (0.000)	0.0695*** (0.000)
social	-1.959** (0.002)	-0.146** (0.001)	1.730*** (0.000)	0.124*** (0.000)
loss	1.810* (0.050)	0.135* (0.041)	0.790 (0.219)	0.0564 (0.216)
social x loss	-0.475 (0.532)	-0.0354 (0.532)	-0.0454 (0.928)	-0.00325 (0.928)
Constant	-11.78*** (0.000)		-13.94*** (0.000)	
lnsig2u Constant	1.386*** (0.001)		1.531*** (0.000)	
Observations	434	434	966	966

p-values in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 21: Equal chances

	Group 1	Average marginal effects	Group 2	Average marginal effects
det outcome	-0.684*** (0.000)	-0.0689*** (0.000)	-0.742*** (0.000)	-0.0686*** (0.000)
social	-0.401 (0.439)	-0.0404 (0.438)	1.480*** (0.000)	0.137*** (0.000)
loss	0.838 (0.338)	0.0844 (0.333)	-0.0814 (0.891)	-0.00753 (0.891)
social x loss	0.242 (0.711)	0.0244 (0.711)	0.428 (0.341)	0.0396 (0.339)
Constant	5.076*** (0.000)		4.882*** (0.000)	
Insig2u Constant	1.382*** (0.000)		1.409*** (0.000)	
Observations	434	434	966	966

p-values in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6.3 Additional Graphs

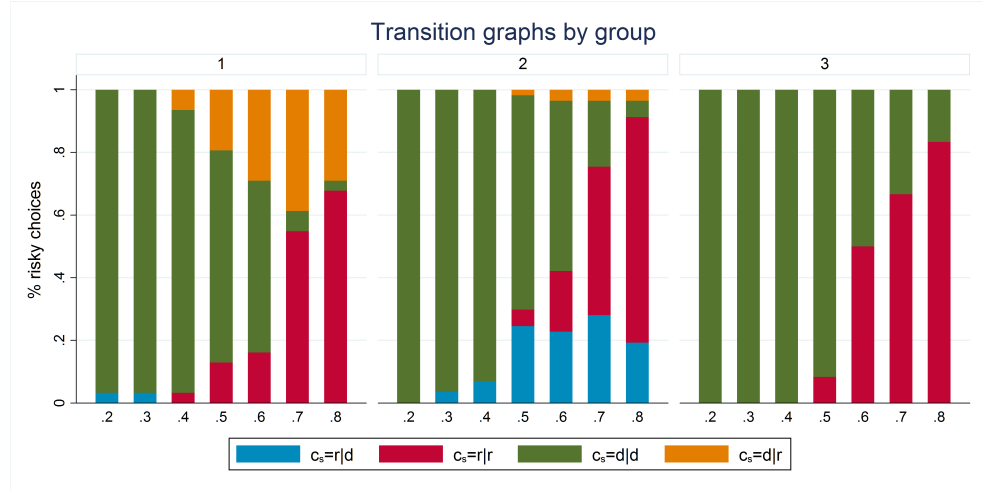


Figure 13: Egd scenarios

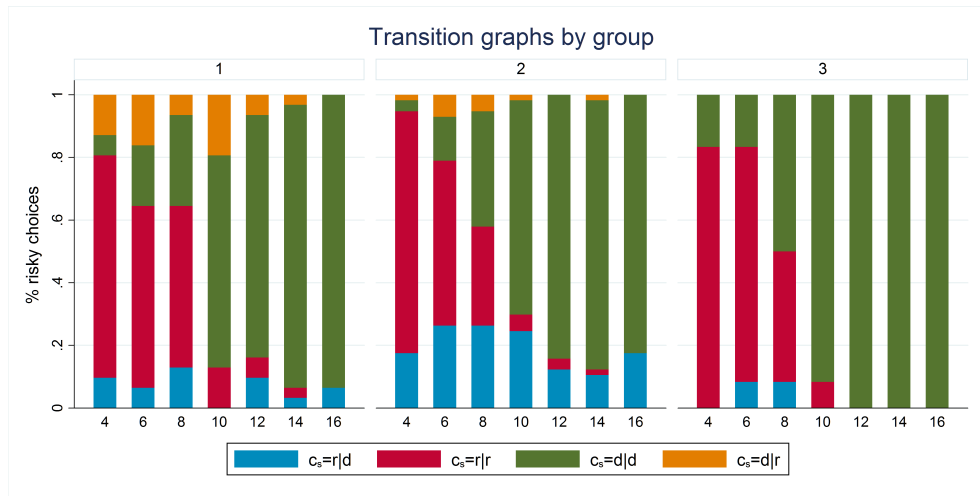


Figure 14: Eqp scenarios

6.4 Instructions

INSTRUCTIONS (Losses)

Welcome.

You may have noticed a small stack of papers on your workstation. These are answers to a survey on different themes, handed out to a sample of students. These answers are going to build the dataset for an economic analysis, independent from today's experiment.

In the first part of this experiment, you are asked to report the answers on each paper on the forms that will appear on the display once you click "OK". We kindly ask you to pay close attention in inserting the answers, since the accuracy of the following economic analysis depends on them.

Each question (identified on the papers as "Question") can only have one answer. In case the answer is ambiguous or wrong (multiple crosses, deletions), insert in the system the answer which you deem more correct, trying to interpret the will of the respondent. In these cases, please write down ID, Set and Question on the white paper available on your work station.

The reward for the completion of the task is 20 euros.

When you have inserted all the answers, you will move to the following stage of the experiment, for which you will be given and read new instructions.

Please, do not communicate with other participants for the whole duration of the experiment. If you do not abide by the rules, you will be excluded from the experiment without any compensation.

If you have any question about the instructions, raise your hand at the end of the reading and a responsible will answer in private manner.

Good luck with your work!

INSTRUCTIONS II (Losses)

Welcome to the second stage of the experiment.

In the stage you are about to participate to, different scenarios will be presented. In each of them, you will be asked to choose the alternative you prefer.

Scenarios are all the same for all participants and, as you will see, are of two types. In first type scenarios, the decisions you will take concern you and another participant, who will change randomly in each scenario and whose identity will never be revealed. In second type scenarios, the decision to take only concerns you.

At the end of the experiment, the computer will randomly select one scenario for all the participant for payment purposes.

If the selected scenario is of the first type, the computer will choose with equal probability one between you and the other participant you were associated to in that particular scenario and the decision of the selected person is the one that is implemented.

If the selected scenario is of the second type, your decision in that scenario is immediately implemented.

At the end you will receive the earning in the selected scenario plus the show-up fee of 5 euros and the earning in the first stage.

Before collecting the payment, you have to compile a short survey that will appear on the desktop display. After you have answered all the questions, please remain seated until the experimenter calls you to hand you your payment. At that moment, remember to bring with you the receipt, the pen, the answer sheets you found on your work station and the paper where you wrote down eventual ambiguities in the insertion stage.

Once again, if you have any question about the instructions, raise your hand at the end of the reading and a responsible will answer in private manner.

INSTRUCTIONS (Gains)

Welcome.

In the stage you are about to participate to, different scenarios will be presented. In each of them, you will be asked to choose the alternative you prefer.

Scenarios are all the same for all participants and, as you will see, are of two types. In first type scenarios, the decisions you will take concern you and another participant, who will change randomly in each scenario and whose identity will never be revealed. In second type scenarios, the decision to take only concerns you.

At the end of the experiment, the computer will randomly select one scenario for all the participant for payment purposes.

If the selected scenario is of the first type, the computer will choose with equal probability one between you and the other participant you were associated to in that particular scenario and the decision of the selected person is the one that is implemented.

If the selected scenario is of the second type, your decision in that scenario is immediately implemented.

Your final earning for the experiment is equal to the earning in the selected scenario plus the show-up fee of 5 euros.

Before collecting the payment, you have to compile a short survey that will appear on the desktop display. After you have answered all the questions, please remain seated until the experimenter calls you to hand you your payment. At that moment, remember to bring with you the receipt and the pen you found on your work station.

Please, do not communicate with other participants for the whole duration of the experiment. If you do not abide by the rules, you will be excluded from the experiment without any compensation.

If you have any question about the instructions, raise your hand at the end of the reading and a responsible will answer in private manner.

ID **070**
Set 1

QUESTION 1	☒	☐ B	☐ C	☐ D	
QUESTION 2	☐ A	☒	☐ C	☐ D	
QUESTION 3	☐ A	☐ B	☒	☐ D	☐ E
QUESTION 4	☒	☐ B	☐ C	☐ D	☐ E
QUESTION 5	☐ A	☒	☐ C		
QUESTION 6	☒	☐ B	☐ C	☐ D	
QUESTION 7	☐ A	☐ B	☐ C	☒	
QUESTION 8	☐ A	☒	☐ C	☐ D	
QUESTION 9	☐ A	☐ B	☒	☐ D	
QUESTION 10	☒	☐ B	☐ C	☐ D	☐ E
QUESTION 11	☐ A	☐ B	☒		
QUESTION 12	☒	☐ B	☐ C	☐ D	☐ E

Figure 15: Answer sheet fac-simile

QUESTION TYPE - SOCIAL CONTEXT

In the insertion stage you earned 20 euros.

In this scenario, a loss of 20 euros hits you and the other person you are paired with.

Which of the following options do you prefer to share the loss of 20 euros?

Option A:

I lose (10) euros.

The other person loses the remaining
10 euros.

Option B:

I lose the 20 euros with probability
(50)%.

I give to the other person a chance to
lose the 20 euros equal to (50)%.

QUESTION TYPE - ISOLATION

In the insertion stage you earned 20 euros.

In this scenario, which of the following options do you prefer?

Which of the following options do you prefer to share the loss of 20 euros?

Option A:

I lose (10) euros.

Option B:

I lose the 20 euros with probability
(50)%, I lose 0 with probability (50)%.

DETERMINISTIC DICTATOR GAME

In the insertion stage you earn 20 euros. In this scenario, a loss of 20 euros hits you and the other person you are paired with. how do you want to share the loss?
(The sum of the two losses must be equal to 20)

I lose _ _ _

The other person loses _ _ _

PROBABILISTIC DICTATOR GAMES

In the insertion stage you earn 20 euros. In this scenario, a loss of 20 euros hits you and the other person you are paired with. how do you want to share the probability to suffer the whole loss? (The sum of the two probabilities must be equal to 100)

I lose 20 euros with probability _ _ _

The other person loses 20 euros with
probability _ _ _

PREFERENCE

Which of the two latter scenarios have you preferred?

Dividing the loss

Dividing the probabilities to suffer the
loss.

7 Bibliography

- [1] Armenak Antinyan. Loss and other-regarding preferences: Evidence from dictator game. *Department of Management, Università Ca'Foscari Venezia Working Paper*, 3, 2014.
- [2] Guillermo Baquero, Willem Smit, and Luc Wathieu. The generosity effect: Fairness in sharing gains and losses. Technical report, ESMT Working Paper, 2013.
- [3] Jonathan Baron. Blind justice: Fairness to groups and the do-no-harm principle. *Journal of Behavioral Decision Making*, 8(2):71–83, 1995.
- [4] Nadege Bault, Giorgio Coricelli, and Aldo Rustichini. Interdependent utilities: how social ranking affects choice behavior. *PloS one*, 3(10):e3477, 2008.
- [5] Roger Berger, Heiko Rauhut, Sandra Prade, and Dirk Helbing. Bargaining over waiting time in ultimatum game experiments. *Social science research*, 41(2):372–379, 2012.
- [6] Gary E Bolton and Axel Ockenfels. Betrayal aversion: Evidence from brazil, china, oman, switzerland, turkey and the united states: Comment. *The American Economic Review*, pages 628–633, 2010.
- [7] Nancy Buchan, Rachel Croson, Eric Johnson, and George Wu. Gain and loss ultimatums. *Advances in Applied Microeconomics*, 13:1–23, 2005.
- [8] Jeffrey V Butler. Inequality and relative ability beliefs. Technical report, Einaudi Institute for Economic and Finance (EIEF), 2013.
- [9] Gary Charness and Matthew Rabin. Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869, 2002.
- [10] Todd L Cherry, Peter Frykblom, and Jason F Shogren. Hardnose the dictator. *American Economic Review*, pages 1218–1221, 2002.

- [11] H David, David Dorn, and Gordon H Hanson. The china syndrome: Local labor market effects of import competition in the united states. *The American Economic Review*, 103(6):2121–2168, 2013.
- [12] Carsten KW De Dreu, Joselito C Lualhati, and Christopher McCusker. Effects of gain-loss frames on satisfaction with self–other outcome-differences. *European Journal of Social Psychology*, 24(4):497–510, 1994.
- [13] Nathalie Etchart-Vincent and Olivier l Haridon. Monetary incentives in the loss domain and behavior toward risk: An experimental comparison of three reward schemes including real losses. *Journal of Risk and Uncertainty*, 42(1):61–83, 2011.
- [14] Peter H Farquhar. State of the art-utility assessment methods. *Management science*, 30(11):1283–1300, 1984.
- [15] Nicola Frignani and Giovanni Ponti. Risk versus social preferences under the veil of ignorance. *Economics Letters*, 116(2):143–146, 2012.
- [16] Alexia Gaudeul. Social preferences under uncertainty. Technical report, Jena Economic Research Papers, 2013.
- [17] Xiuyan Guo, Li Zheng, Lei Zhu, Jianqi Li, Qianfeng Wang, Zoltan Dienes, and Zhiliang Yang. Increased neural responses to unfairness in a loss context. *Neuroimage*, 77:246–253, 2013.
- [18] Charles A Holt and Susan K Laury. Risk aversion and incentive effects. *American economic review*, 92(5):1644–1655, 2002.
- [19] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, pages 263–291, 1979.
- [20] Mary C Kern and Dolly Chugh. Bounded ethicality the perils of loss framing. *Psychological Science*, 20(3):378–384, 2009.

- [21] Martin Kocher, Michal Krawczyk, and Fabrice Le Lec. Sharing or gambling? on risk attitudes in social contexts. 2013.
- [22] Michal Krawczyk and Fabrice Le Lec. Social decisions under risk. evidence from the probabilistic dictator game. *University of Amsterdam, working paper*, 2008.
- [23] Susan Laury and Charles A Holt. Further reflections on prospect theory. *Andrew Young School of Policy Studies Research Paper Series*, (06-11), 2005.
- [24] Marijke C Leliveld, Ilja van Beest, Eric van Dijk, and Ann E Tenbrunsel. Understanding the influence of outcome valence in bargaining: A study on fairness accessibility, norms, and behavior. *Journal of Experimental Social Psychology*, 45(3):505–514, 2009.
- [25] Jona Linde and Joep Sonnemans. Social comparison and risky choices. *Journal of Risk and Uncertainty*, 44(1):45–72, 2012.
- [26] Robert J Oxoby and John Spraggon. Mine and yours: Property rights in dictator games. *Journal of Economic Behavior & Organization*, 65(3):703–713, 2008.
- [27] Richard H Thaler and Eric J Johnson. Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management science*, 36(6):643–660, 1990.
- [28] Martin Weber and Heiko Zuchel. How do prior outcomes affect risk attitude? comparing escalation of commitment and the house-money effect. *Decision Analysis*, 2(1):30–43, 2005.
- [29] Xiaolin Zhou and Yan Wu. Sharing losses and sharing gains: increased demand for fairness under adversity. *Journal of Experimental Social Psychology*, 47(3):582–588, 2011.

ACKNOWLEDGMENTS

Once again, I must thank Professor Giuseppe Ragusa for sharing his ideas and intuitions, and for mentoring and trusting me.

Thanks to Professors Giovanni Ponti and Daniela di Cagno, who made me feel welcome in the behavioral economics world.

Thanks to mum and dad, for their unconditional love and for having showed me how wonderful knowledge can be.

Thanks to Michele, for being a perfect partner in crime.

Thanks to my whole family, than can make me feel their love even through the ether.

Thanks to Annalaura, for being my sunshine: no achievement could rival finding you.

Thanks to Francesca, for being at my side, my best friend, be it fair or stormy weather.

Thanks to Jacopo, for sharing good times, bad movies and the weirdest adventures.

Thanks to Martina, who always find a way to be there for me, no matter how far we are.

Thanks to Giuseppe, Raluca, Francesco and Marzio, who made me *ad honorem* XXX cycle Phd student, and to all people at the 5th floor, for the invaluable coffee breaks, shared smiles and mutual support.

Thanks to all those who, in one way or another, have been part of this adventure of mine.

Once again, I cannot but acknowledge how lucky I am. All the errors are and remain my own.