Genes, History and Economics

Genes, History and Economics

Björn Wallace

Dissertation for the Degree of Doctor of Philosophy, Ph.D.
Stockholm School of Economics, 2011

KEYWORDS: Behavioral economics; Experimental economics; Behavioral genetics; Ultimatum game; Risk; Dictator game; Behavioral anomalies; Economic history; Slave trades; Longevity.

Genes, History and Economics

ISBN 978-91-7258-853-0

PRINTED BY:
Intellecta Infolog, Göteborg 2011

DISTRIBUTED BY:
The Research Secretariat
Stockholm School of Economics
Box 6501, SE-113 83 Stockholm, Sweden
www.hhs.se

Preface

This report is a result of a research project carried out at the Economics department at the Stockholm School of Economics (SSE). This volume is submitted as a doctor's thesis at SSE. The author has been entirely free to conduct and present his research in his own ways as an expression of his own ideas. SSE is grateful for the financial support which has made it possible to fulfill the project.

Filip Wijkström
Associate Professor
SSE Director of Research

Till min familj

Acknowledgements

By the standards of economics I have an unusually high number of coauthors. As a consequence I am indebted to many. Two thirds of this dissertation stems from a research project that I, my supervisor Magnus Johannesson, and David Cesarini have been working on for the last five years. For me, Magnus has been the perfect supervisor, and David is as good a coauthor as anyone could wish for. Without their remarkable energy and productivity, this dissertation would not have been what it now is. Particularly important contributions were also made by Margherita Bottero, Chris Dawes, Erik Lindqvist and Örjan Sandewall. In addition to being a coauthor Margherita provided me with moral support and displayed unwonted patience.

I would also like to take the opportunity to thank the faculty at the Department of Economics, and in particular David Domeij, Tore Ellingsen, Richard Friberg, Juanna Joensen, Jörgen Weibull and Mark Voorneveld. I am also grateful to all my colleagues and friends among the graduate students at SSE and SDPE, especially Anna Dreber Almenberg, Sara Formai, Tobias Laun, Erik Mohlin, André Romahn, Linus Siming and Robert Östling. Anna is not only a friend, future coauthor, and poker impresario, for 3 years she was my landlady. Carin Blanksvärd, Ritva Kiviharju and Lilian Öberg provided excellent administrative assistance throughout the years.

Outside the rather narrow world of the Department, Paul Lichtenstein, Patrik Magnusson, Arvid Malm, Tino Sanandaji and Patrik Ivert made important contributions to this dissertation. I am also very grateful to all those who participated in the experiments and surveys that form the empirical foundation for much of this dissertation.

Finally, I would like to thank my friends and family. In particular, my mother, grandmother, brother and father who always helped and supported me.

Contents

1. Introduction

This dissertation consists of six chapters that span a very diverse set of topics. Yet, it has two unifying themes, economics and biology, that tie it together. The first four chapters present the principal findings from a project that was initiated jointly with David Cesarini and Magnus Johannesson, and that applies the twin method from behavioral genetics to economics. The last two chapters instead use a simple regression framework and evidence from biological anthropology to investigate recent claims regarding the effects of child bearing and past slave trades.

2. Genes and economics

There is a small, but rapidly growing, literature studying the genetic and environmental origins of economic behavior and outcomes (Bowles et al., 2005; Beauchamp et al., 2011). Until recently, this literature focused exclusively on outcomes, and in particular income. In chapters 1-4 we instead focus on economic behavior and decision-making.

Previous behavioral genetic work outside the domains of economics has changed the way that we think about a number of behavioral traits. In this literature it is typically found that i) variation is heritable ii) genetic factors are more important than family environment iii) a large fraction of variation cannot be explained by neither genes nor family environment (Turkheimer, 2000; Plomin et al., 2009). However, compared to many other disciplines, and psychology in particular, economics is lagging behind. In fact, as recently as 2009 the leading text book in behavioral genetics described economics as "still essentially untouched by genetic research" (Plomin et al., 2009, p. 353). Hopefully, the chapters in this dissertation can help to improve on this somewhat unsatisfactory state of the art.

Chapters 1 and 2 study economic decision-making in the laboratory using the twin method. More specifically, we study the ultimatum and dictator games alongside risky gambles, using same-sex twin pairs as our subject pool. Given a few additional assumptions, the fact that identical twins have, in expectation, a twice as high coefficient of genetic relatedness as fraternal twins implies that we can study the genetic and environmental contributions to variation in behavior by studying twin correlations in observed choices. Chapters 3 and 4 apply the same method to actual portfolio choices associated with a far-reaching pension reform, as well as to a set of standard behavioral anomalies. Taken together, these four chapters provide strong evidence in favor of the hypothesis that genes influence economic decision-making. Thus, economic behavior does not appear to be much different from other types of behavior.

3. Economics and history

The last two chapters of the dissertation turn to the past, rather than genes, in an effort to evaluate recent findings regarding two important welfare outcomes. In chapter 5 we investigate Nunn's (2008) claim that past slave trades had a negative impact on current economic performance in Africa. By extending the sample period back in time we demonstrate that this relationship was not significant in 1960. In addition, by applying Nunn's method to an episode of large scale slave raiding in Italy,

we demonstrate that there exists a similar negative relationship across Italian regions, although it becomes insignificant when geographical controls are included. Intriguingly, going back to 1960, the coefficient on slave raids for Italy also has a similar time trend to that for Africa. Taking these facts, and our reading of the historical and anthropological literature, which is much different from that of Nunn, into account we do not find much support for the hypothesis that the African slave trades had a negative impact on current economic performance.

Finally, chapter 6 investigates the large and negative relationship between giving birth to a son, rather than a daughter, and maternal longevity that was documented in a Sami hunter-gatherer population from Finland (Helle et al., 2002). Using a substantially larger sample of pre-industrial Swedish Sami we find no evidence in favor of such a relationship.

4. Brasklapp

Five of the chapters in this dissertation (Ch. 1-4 & 6) are slightly altered versions of previously published papers (Wallace et al., 2007; Cesarini et al., 2009 a, b; 2010; 2011). Unfortunately, the fact that earlier versions of the chapters were prepared as separate articles for five different journals means that they can at times appear both repetitive, and in terms of notation and formatting, somewhat inconsistent. I apologize to the reader for these inconveniences.

References

Beauchamp, J. P., D. Cesarini, M. Johannesson, M. van der Loos, P. Koellinger, P. J. F. Groenen, J. H. Fowler, N. Rosenquist, A. R. Thurik, and N. A. Christakis, "Genetics and Economics," *J. Econom. Perspect.*, forthcoming (2011).

Bowles, S., H. Gintis, and M. Osborne Groves (Eds), *Unequal Chances: Family Background and Economic Success* (Princeton, NJ: Princeton University Press, 2005).

Cesarini, D., C. T. Dawes, M. Johannesson, P. Lichtenstein, and B. Wallace, "Genetic variation in preferences for giving and risk-taking," *Quarterly Journal of Economics* 124 (2009), 809-842.

Cesarini, D., E. Lindqvist, and B. Wallace, "Is there an adverse effect of sons on maternal longevity?," *Proceedings of Royal Society B* 276 (2009), 2081-2084.

Cesarini, D., M. Johannesson, P. Lichtenstein, Ö. Sandewall, and B. Wallace, "Genetic Variation in Financial Decision Making," *Journal of Finance* 65 (2010), 1725–1754.

Helle, S., V. Lummaa, and J. Jokela, "Sons reduced maternal longevity in preindustrial humans," *Science* 296 (2002), 1085.

Nunn, N., "The Long-Term Effects of Africa's Slave Trades," *Quarterly Journal of Economics* 123 (2008), 139-176.

Cesarini, D., M. Johannesson, P. K. E. Magnusson, and B. Wallace, "The behavioral genetics of behavioral anomalies," *Management Science*, forthcoming (2011).

Plomin, R., J. C. DeFries, and G. E. McClearn, and P. McGuffin, *Behavioral Genetics*, 5th ed. (New York, NY: Worth Publishers, 2009).

Wallace, B., D. Cesarini, P. Lichtenstein, and M. Johannesson, "Heritability of ultimatum game responder behavior," *Proceedings of the National Academy of Sciences* 104 (2007), 15631-15634.

PAPER 1

Heritability of ultimatum game responder behavior

Björn Wallace David Cesarini Paul Lichtenstein Magnus Johannesson

Abstract. Experimental evidence suggests that many people are willing to deviate from materially maximizing strategies in order to punish unfair behavior. Even though little is known about the origins of such fairness preferences, it has been suggested that they have deep evolutionary roots and that they are crucial for maintaining and understanding cooperation among non-kin. Here we report the results from an ultimatum game, played for real monetary stakes, using twins recruited from the population-based Swedish Twin Registry as our subject pool. Employing standard structural equation modeling techniques, we estimate that >40% of the variation in subjects' rejection behavior is explained by additive genetic effects. Our estimates also suggest a very modest role for common environment as a source of phenotypic variation. Based on these findings, we argue that any attempt to explain observed ultimatum bargaining game behavior that ignores this genetic influence is incomplete.

1. Introduction

It is frequently pointed out that humans exhibit unusually high rates of cooperation among non-kin (Nowak, 2006), and it has been suggested that one important factor for enhancing cooperation is that humans appear willing to forego material payoffs to punish unfair behavior (Fehr, 2002; Fehr & Fischbacher, 2003; de Quervain et al., 2004). Such fairness preferences have been widely studied by using experimental games, and in particular the ultimatum game (Güth et al., 1982; Fehr & Schmidt, 1999; Camerer, 2003).

In the ultimatum game, two subjects are assigned the role of either proposer or receiver, and then they bargain over a sum of money (the "cake"). The proposer makes an offer on how to divide the cake. If the receiver accepts the proposer's offer, the

A slightly different version of this paper was published as: Wallace, B., D. Cesarini, P. Lichtenstein, and M. Johannesson, "Heritability of ultimatum game responder behavior," *Proceedings of the National Academy of Sciences* 104 (2007), 15631-15634. We thank Terry Burnham, Ernst Fehr, Herbert Gintis, Erik Lindqvist, Moses Ndulu, Martin Nowak, Paul Schrimpf, Lucius Vinicius, and Robert Östling for helpful comments. This work was supported by the Jan Wallander and Tom Hedelius Foundation and the Swedish Research Council. The Swedish Twin Registry is also supported by grants from the Swedish Research Council, the Ministry for Higher Education, and AstraZeneca.

players are paid accordingly, whereas if the offer is rejected, both players receive a zero payoff. In a one-shot game, rational and money-maximizing responders should accept any positive offer because the alternative is a zero payoff. Two stylized facts about responder behavior emerge from the ultimatum game literature: first, that unfair offers are often rejected and second, that the acceptance threshold varies substantially between individuals (Fehr & Schmidt, 1999; Camerer, 2003). The average responder behavior has been shown to be relatively stable across Western cultures (Roth et al., 1991), whereas more variation has been observed among non-Western small-scale societies (Henrich et al., 2004).

Although there is a voluminous literature discussing the cultural and evolutionary origins of observed fairness preferences, the relative social and genetic contributions have hitherto been left unexplored. In this work, we use the classical twin design to estimate the heritability of the propensity to reject unfair offers in the ultimatum bargaining game. In doing so, we not only provide the first decomposition of the social and genetic contributions to ultimatum game rejection behavior but also to behavior in experimental games in general. The virtue of the twin design is that by comparing monozygotic (MZ) twins, who share the same set of genes, and dizygotic (DZ) twins, whose genes are imperfectly correlated, we can estimate the proportion of variance in phenotype due to genetic, shared, and nonshared environmental effects (Neale & Cardon, 1992).

2. Results

In Figure 1 we report the distribution of acceptance thresholds for MZ and DZ twins. As can be seen, the distributions are very similar, and we cannot reject the null hypothesis of identical distributions (χ^2 test; χ^2 value = 4.292, df = 5, n = 653, p = 0.491).[1] The average acceptance threshold is 32.68 SEK, implying that on average a responder demands ≈33% of the cake to accept the proposer's offer in the ultimatum game.[2] Figure 2 illustrates the correlation in acceptance thresholds within MZ and DZ twin pairs. The Spearman rank correlation is 0.39 (95% confidence interval, 0.26–0.49) for MZ twin pairs and -0.04 (95% confidence interval, -0.25 to 0.18) for DZ twin pairs. The difference in the correlation coefficients is highly significant (t test; t = 3.212, p

[1] There is too little variation in the proposal stage to estimate model parameters with reasonable precision, and hence we focus on responder behavior. The equality of distributions of acceptance thresholds by zygosity was tested with a design-based independence test, which takes the correlation between twins into account. The χ^2 statistic is adjusted by using the second-order correction of Rao and Scott (1984).

[2] To estimate the mean acceptance threshold, the acceptance threshold is set at the middle value of the acceptance threshold intervals.

< 0.01, two-sided) and thus provides strong evidence in favor of a genetic effect on ultimatum game responder behavior.

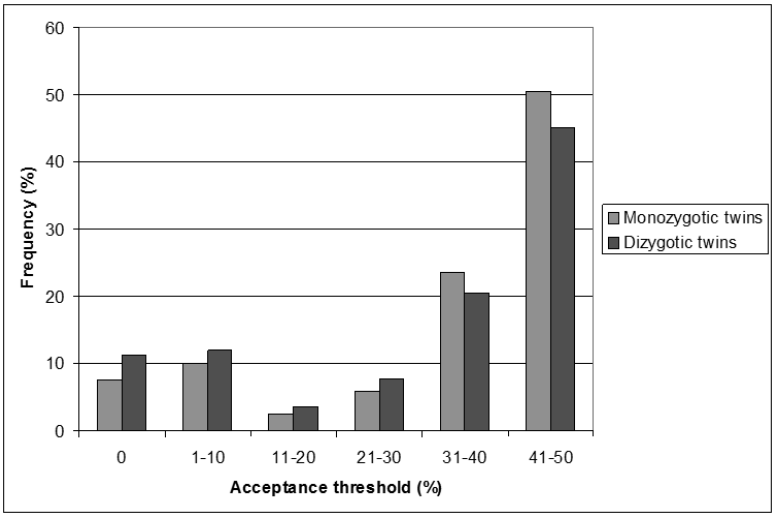


Figure 1. Distribution of acceptance thresholds in the ultimatum game for MZ and DZ twins.

We also use a standard threshold model to decompose phenotypic variation into additive genetic effects, common environmental effects, and nonshared environmental effects. The model assumes that there is an underlying, continuous and normal distribution of fairness preferences, with arbitrary cutoffs (thresholds) observed by the researcher. In Table 1 we present maximum likelihood estimates of the model parameters as well as its nested submodels.[3] Additive genetic effects are estimated to account for 42% of the variation in the best-fitting model, and this estimate is significantly different from zero ($p < 0.01$). The point estimate of the effect of common environment is zero, and its confidence interval has an upper boundary at 21%, suggesting that common environmental influences are at most a moderately important source of

[3] The maximum likelihood estimation is implemented in Mx, a numerical optimizer for behavior genetics (Neale et al., 2002). We estimate a threshold model based on the response categories for the acceptance threshold in the experiment. The programming code for the threshold model we estimate is adapted from the digital scripts library at the University of Amsterdam (Posthuma & Boomsma, 2005). The algorithm that estimates confidence intervals for the parameters is explained in some detail by Neale and Miller (1997).

variation. The remaining variation is accounted for by nonshared environment. The estimate of nonshared environment also includes any measurement error.

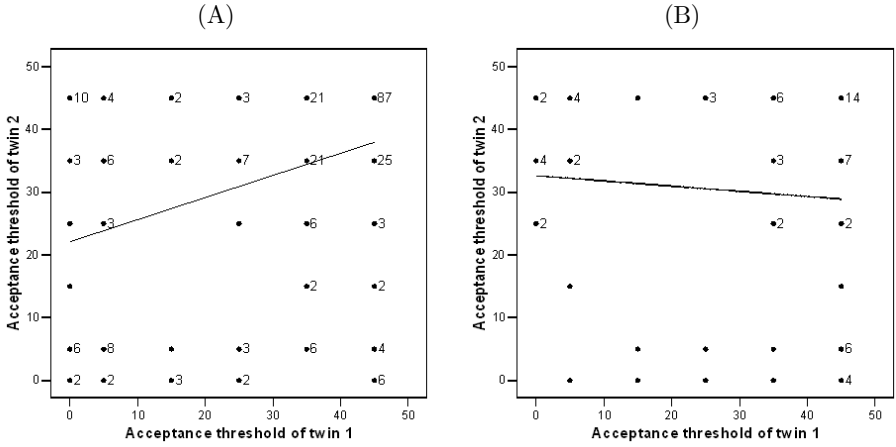


Figure 2. Ultimatum game acceptance thresholds for MZ (A) & DZ (B) pairs.

Table 1. Maximum likelihood estimates of the structural equation ACE model and its submodels (95% confidence intervals in parentheses)[4]

Model	χ^2	p	df	MSE	A	C	E
ACE	79.44	0.08	63	0.02	0.42 (0.17 − 0.54)	0.00 (0.00 − 0.21)	0.58 (0.46 − 0.72)
AE	79.44	0.09	64	0.02	0.42 (0.28 − 0.54)		0.58 (0.46 − 0.72)
CE	86.96	0.01	64	0.04		0.32 (0.19 − 0.4)	0.68 (0.56 − 0.81)
E	109.30	0.00	65	0.04			1

[4] Note: The genetic contribution (A) is highly significant in the full ACE model with a point estimate of 42% of the variation. The common environmental contribution (C) has a point estimate of zero and is not significant. The unique environment (E) is always the largest component. The AE submodel is the best-fitting model; comparing this model with the ACE model, we cannot reject the null hypothesis of a zero effect of common environment (χ^2 test; χ^2 value = 0, df = 1, p = 1). The ACE model significantly outperforms the CE submodel (χ^2 test; χ^2 value = 7.52, df = 1, p < 0.01) and the E submodel (χ^2 test; χ^2 value = 29.88, df = 2, p < 0.001).

Finally, we consider the results stratified by sex because pooling may not always be appropriate. Our data set is fairly unbalanced with respect to sex, approximately three-quarters of subjects being women. There was no evidence that pooling by sex is inappropriate because we could not reject the null hypothesis that the distribution of the acceptance threshold is the same for men and women (χ^2 test; χ^2 value = 2.653, df = 5, n = 653, p = 0.778). When we estimate separate models for men and women, the estimates of additive genetic effects were practically identical: 0.412 for women and 0.435 for men.

3. Discussion

This work has demonstrated that genetic influences are important determinants of rejection behavior in the ultimatum game. In our best-fitting model, additive genetic effects account for 42% of the observed variation in responder behavior, and our point estimate for common environment is zero. However, note that the estimate for the genetic effect should be considered to be a lower boundary because it presumes perfect reliability in the measurement of responder behavior. If there is noise in eliciting acceptance thresholds, the estimate of additive genetic effects will be downward-biased.

Our finding of substantial genetic effects on ultimatum game rejection behavior is consistent with previous research in behavioral genetics and neuroscience. For instance, survey based studies repeatedly find sizeable genetic effects on a wide range of economically relevant social attitude variables such as personality and political preferences (Bouchard et al., 1990; Bouchard & McGue, 2003; Rushton, 2004; Alford et al., 2005). Furthermore, recent studies have shown that the responder stage in the ultimatum game is associated with increased activation in the dorsolateral prefrontal cortex as well as responders' circulating testosterone levels (Sanfey et al., 2003; Knoch et al., 2006; Burnham, 2007). The dorsolateral prefrontal cortex is a brain region whose structure is under pronounced genetic influence (Gray & Thompson, 2004; Toga & Thompson, 2005), as are individual testosterone levels (Harris et al., 1998).

The etiology of fairness preferences has been intensely debated. Some authors have argued that the willingness to engage in costly punishment reflects fundamental and universal fairness preferences that evolved through a process of gene–culture coevolution in early modern Homo sapiens, preferences that are considered crucial for maintaining cooperation among non-kin (Fehr & Fischbacher, 2003; Boyd et al., 2003). Others have argued that indirect reciprocity models based on an evolved psychology for reputation management provide a more parsimonious explanation (Gale et al., 1995; Nowak, 2006; Nowak & Sigmund, 1998; Nowak et al., 2000; Milinski et al., 2002;

Nowak & Sigmund, 2005; Burnham & Johnson, 2005). Although our results are consistent with an evolutionary origin of fairness preferences, it is important to remember that heritability measures the genetically determined variation around some average behavior. Hence, it does not provide us with any direct evidence with regard to the evolutionary dynamics that brought it about.

However, the fact that there seems to be a substantial genetically determined heterogeneity in ultimatum game rejection behavior is an interesting finding in and of itself. First of all, it suggests that economic theory and policy need to address the potential importance of genetic influences on economic preferences, behavior, and outcomes (Zizzo, 2003). It also suggests that the current debate on the evolutionary origins of ultimatum game rejection behavior, and the preferences it proxies, should perhaps be brought into the broader context of the evolution of variation in personality. If we observe a mean phenotypic expression in behavior with substantial genetic variation, then any evolutionary model of its origin must not only account for average phenotype but also for phenotypic variation.

In humans as well as in other species there is substantial genetically determined variation for a great number of personality traits (Dall et al., 2004; Nettle, 2006; Penke et al., 2007). Unfortunately, this variation is not well understood because the observed patterns seldom correspond to the simplest evolutionary genetic models of fixation. Many authors have therefore resorted to the default explanation that observed variation is nonadaptive (Dall et al., 2004). This is, however, not necessarily the case, and a number of recent papers have examined these questions in more detail. For instance, Dall et al. emphasize the combination of frequency-dependent selection and state-dependent behavioral specialization as a possible source of adaptive variation, whereas Penke et al. (2007) argue that balancing selection by environmental heterogeneity explains individual variation in personality.[5]

Finally, on a more general note, our findings suggest that it is time to take seriously the proposition that experimental behavior may be substantially heritable, which may explain why, despite ample experimental evidence, the origins of individual variation remain elusive, and most attempts to find theoretically appealing and empirically stable correlates to experimentally derived preferences have had only mixed success.

[5] One interpretation, suggested to us by a referee, is that the high heritability casts doubt on accounts of ultimatum rejections relying on notions of evolutionary disequilibrium. In light of our findings, it seems plausible that the trait has been under stabilizing selection in the post-Pleistocene era.

References

Alford, J. R., C. L. Funk, and J. R. Hibbing, "Are Political Orientations Genetically Transmitted?," *American Political Science Review* 99 (2005), 153-167.

Bouchard, T. J. Jr, D. T. Lykken, M. McGue, N. L. Segal, and A. Tellegen, "Sources of human psychological differences: the Minnesota Study of Twins Reared Apart," *Science* 250 (1990), 223-228.

Bouchard, T. J. Jr, and M. McGue, "Genetic and environmental influences on human psychological differences," *Journal of Neurobiology* 54 (2002), 4–45.

Boyd, R., H. Gintis, S. Bowles, and P. J. Richerson, "The evolution of altruistic punishment," *Proceedings of the National Academy of Sciences* 100 (2003), 3531-3535

Burnham, T. C., "High-testosterone men reject low ultimatum game offers," *Proceedings of Royal Society B* 274 (2007), 2327-2330.

Burnham, T. C., and D. D. P. Johnson, "The Biological and Evolutionary Logic of Human Cooperation," *Analyse & Kritik* 27 (2005), 113–135.

Camerer, C. F., *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton, NJ: Princeton University Press, 2003).

Dall S. R. X., A. I. Houston, and J. M. McNamara, "The Behavioral Ecology of Personality: Consistent Individual Differences from an Adaptive Perspective," *Ecology Letters* 7 (2004), 734–739.

de Quervain, D. J.-F., U. Fischbacher, V. Treyer, M. Schellhammer, U. Schnyder, A. Buck, and E. Fehr, "The Neural Basis of Altruistic Punishment," *Science* 305 (2004), 1254-1258.

Fehr, E., and K. M. Schmidt, "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics* 114 (1999), 817-868.

Fehr, E., and S. Gächter, "Altruistic punishment in humans," *Nature* 415 (2002), 137-140.

Fehr, E., and U. Fischbacher, "The nature of human altruism," *Nature* 425 (2003), 785–791.

Gale, J., K. G. Binmore, and L. Samuelson, "Learning to be imperfect: The ultimatum game," *Games and Economic Behavior* 8 (1995), 56-90.

Gintis, H., S. Bowles, R. Boyd, and E. Fehr, "Explaining altruistic behavior in humans," *Evolution and Human Behavior* 24 (2003), 153-172.

Gray, J. R., and P. M. Thompson, "Neurobiology of intelligence: science and ethics," *Nature Reviews Neuroscience* 5 (2004), 471-482.

Güth, W., R. Schmittberger, and B. Schwarze, "An experimental analysis of ultimatum bargaining," *Journal of Economic Behavior & Organization* 3 (1982), 367-388.

Harris, J. A., P. A. Vernon, and D. I. Boomsma, "The Heritability of Testosterone: A Study of Dutch Adolescent Twins and Their Parents," *Behavior Genetics* 28 (1998), 165–171.

Henrich J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, and H. Gintis (Eds.), *Foundations of Human Sociality: Economic Implications and Ethnographic Evidence from Fifteen Small-Scale Societies* (Oxford, UK: Oxford University Press, 2004).

Knoch, D., A. Pascual-Leone, K. Meyer, V. Treyer, and E. Fehr, "Diminishing Reciprocal Fairness by Disrupting the Right Prefrontal Cortex," *Science* 314 (2006), 829-832.

Lichtenstein, P., P. F. Sullivan, S. Cnattingius, M. Gatz, S. Johansson, E. Carlström, C. Björk, M. Svartengren, A. Wolk, L. Klareskog, U. de Faire, M. Schalling, J. Palmgren, and N. L. Pedersen, "The Swedish Twin Registry in the Third Millennium: An Update," *Twin Research and Human Genetics* 9 (2006), 875-882.

Milinski, M., D. Semmann, and H.-J. Krambeck, "Reputation helps solve the 'tragedy of the commons'," *Nature* 415 (2002), 424-426.

Neale, M. C., and L. R. Cardon, *Methodology for Genetic Studies of Twins and Families* (Dordrecht, NL: Kluwer Academic Publishers, 1992).

Neale, M. C., and M. B. Miller, "The Use of Likelihood-Based Confidence Intervals in Genetic Models," *Behavior Genetics* 27 (1997), 113–120.

Neale, M. C., S. M. Boker, G. Xie, and H. H. Maes, *Mx: Statistical Modeling* (Richmond, VA: VCU Department of Psychiatry, 2002).

Nettle, D. "The behavioural ecology of personality: consistent individual differences from an adaptive perspective," *American Psychologist* 61 (2006), 622-631.

Nowak, M. A., "Five Rules for the Evolution of Cooperation," *Science* 314 (2006), 1560–1563.

Nowak, M. A., and K. Sigmund, "Evolution of indirect reciprocity by image scoring," *Nature* 393 (1998), 573-577.

Nowak, M. A., and K. Sigmund, "Evolution of indirect reciprocity," *Nature* 437 (2005), 1291-1298.

Nowak, M. A., K. M. Page, and K. Sigmund, "Fairness Versus Reason in the Ultimatum Game," *Science* 289 (2000), 1773-1775.

Penke, L., J. J. A. Denissen, and G. F. Miller, "The evolutionary genetics of personality," *European Journal of Personality* 21 (2007), 549-587.

Posthuma, D., and D. I. Boomsma, "Mx scripts library: structural equation modeling scripts for twin and family data," *Behavior Genetics* 35 (2005), 499-505.

Rao, J. N. K., and Alastair J. Scott, "On Chi-Squared Tests for Multiway Contingency Tables with Cell Proportions Estimated from Survey Data," *Annals of Statistics* 12 (1984), 46-60.

Roth, A. E., V. Prasnikar, M. Okuno-Fujiwara, and S. Zamir, "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study," *American Economic Review* 81 (1991), 1068-1095.

Rushton, J. P., "Genetic and environmental contributions to pro-social attitudes: a twin study of social responsibility," *Proceedings of Royal Society B* 271 (2004), 2583-2585.

Sanfey, A. G., J. K. Rilling, J. A. Aronson, L. E. Nystrom, J. D. Cohen, "The Neural Basis of Economic Decision-Making in the Ultimatum Game," *Science* 300 (2003), 1755-1758.

Toga, A. W., and P. M. Thompson, "Genetics of Brain Structure and Intelligence," *Annual Reviews of Neuroscience* 28 (2005), 1-23.

Zizzo, D. J., "Empirical evidence on interdependent preferences: nature or nurture?," *Cambridge Journal of Economics* 27 (2003), 867-880.

Appendix: Materials and methods

This work was undertaken in collaboration with the Swedish Twin Registry at Karolinska Institutet. The registry is the largest twin registry in the world and has been described in detail elsewhere (Lichtenstein et al., 2006). Same-sex twins born 1960–1985 were solicited by e-mail and recruited in all major Swedish cities through the summer and fall of 2006. A condition for participation was that both twins in a pair be able to attend the same experimental session. In total, 658 individuals (71 DZ and 258 MZ pairs of twins) participated. Zygosity was assigned by questionnaire items that have been shown to have a reliability of up to 98% (Lichtenstein et al., 2006).

In the first stage of the experiment, all subjects played the role of proposer and were asked to divide 100 SEK (approximately $15) between themselves and a randomly selected anonymous counterpart not partaking in the same experimental session. In the second and final stage, all subjects played the role of responder and were once again matched with a randomly selected anonymous counterpart, different from that in stage 1, not partaking in the same session. We used the strategy method to extract acceptance thresholds (Camerer, 2003). Each subject determined whether he or she would accept or reject every possible proposal in multiples of 10% before learning the actual proposal. This method allows for the recovery of the entire strategy of each participant. We then recorded the lowest offer that the responder indicated a willingness to accept in the range of offers between 0% and 50%. In this region, all subjects, except for the very few idiosyncratic responses referred to below, exhibit simple, monotonic behavior. The mean acceptance threshold is therefore uniquely defined.[6]

Of the 658 participants, two failed to respond to the ultimatum game question and had to be dropped from the analysis. In addition, eight subjects provided inconsistent responses.[7] Of these inconsistent responses, five were clarified by an e-mail follow-up question (two subjects did not respond to the e-mail, and one subject wanted to keep the inconsistent answer). Thus, our dataset in the analysis consists of 653 individual

[6] Previous experimental research has shown that it is fairly common to observe some subjects rejecting both low (<50%) and high (>50%) offers (Gintis et al., 2003). In the parlance of experimental economics, these subjects are "hyperfair." In our data, slightly fewer than one-third of the subjects exhibit such behavior.

[7] For example, a response in which an individual accepts an offer of 10% but then rejects an offer of 20% is considered inconsistent because an acceptance threshold is not uniquely defined on the interval from 0 to 50%.

observations (511 MZ twins and 142 DZ twins), and 324 complete twin pairs (253 MZ pairs and 71 DZ pairs).[8]

[8] Furthermore, our results are robust to the manner in which we deal with these inconsistent responses. Dropping the five inconsistent responses clarified by e-mail has no discernible effect on the heritability estimates or their significance levels.

PAPER 2

Genetic variation in preferences for giving and risk taking

David Cesarini Christopher T. Dawes Magnus Johannesson
Paul Lichtenstein Björn Wallace

Abstract. In this paper, we use the classical twin design to provide estimates of genetic and environmental influences on experimentally elicited preferences for risk and giving. Using standard methods from behavior genetics, we find strong prima facie evidence that these preferences are broadly heritable and our estimates suggest that genetic differences explain approximately twenty percent of individual variation. The results thus shed light on an important source of individual variation in preferences, a source that has hitherto been largely neglected in the economics literature.

1. Introduction

Writing in 1875, the prolific Francis Galton concluded the first scientific inquiry into the behavior of twins by remarking that "There is no escape from the conclusion that nature prevails enormously over nurture" (Galton 1875, p. 576). In fact, Galton was so taken with his results that he continued, "My only fear is that my evidence seems to prove too much and may be discredited on that account, as it seems contrary to all experience that nurture should go for so little." Although his methodology would

A slightly different version of this paper was published as: Cesarini, D., C. T. Dawes, M. Johannesson, P. Lichtenstein, and B. Wallace, "Genetic variation in preferences for giving and risk taking," *Quarterly Journal of Economics* 124 (2009), 809-842. The paper has benefited from discussions with Coren Apicella, Samuel Bowles, Terry Burnham, Bryan Caplan, Tore Ellingsen, James Fowler, Jon Gruber, Garett Jones, Moses Ndulu, Matthew Notowidigdo, Niels Rosenquist, Paul Schrimpf, Steven Pinker, Tino Sanandaji, Örjan Sandewall, Vernon Smith and Robert Östling. Thanks to Larry Katz and six anonymous referees for very helpful comments. We thank the Jan Wallander and Tom Hedelius Foundation, the Swedish Research Council, and the Swedish Council for Working Life and Social Research for financial support. The Swedish Twin Registry is supported by grants from the Swedish Research Council, The Ministry for Higher Education and Astra Zeneca. Rozita Broumandi and Camilla Björk at the Swedish Twin Registry responded to a number of queries, for which we are grateful. Patrik Ivert, Niklas Kaunitz and Benjamin Katzeff provided excellent research assistance. We are grateful to a number of colleagues, especially Håkan Jerker Holm and Fredrik Carlsson, for help with subject recruitment outside Stockholm. The paper was completed while David Cesarini was visiting the Research Institute of Industrial Economics in Stockholm; he gratefully acknowledges their hospitality.

be considered dubious, if not flawed, by modern standards, Galton's work laid the conceptual basis for behavior genetics (Bouchard & Propping, 1993; Plomin et al., 2001a), the study of genetic and environmental influences on variation in human behavior. Today there is ample of evidence for the importance of genetic influences ("nature") on variation in human behavioral traits. However, the debate about the rather nebulous concepts "nature" and "nurture" still rages.

In economics, there is a small but growing research field using behavior genetic techniques. The seminal paper is due to Taubman (1976), who employed the twin design to estimate the heritability of earnings for U.S. males. Later papers in this procession, based on either twins or adoptees, include Behrman and Taubman (1989), Sacerdote (2002, 2007), Plug and Vijverberg (2003), Björklund, Lindahl and Plug (2006), and Björklund, Jäntti and Solon (2007). In short, these studies find that both "nature" and "nurture" are important determinants of life outcomes and they uniformly corroborate the importance of genetic influences on educational attainment and earnings.[1]

Some recent work in economics also focuses on the issue of intergenerational transmission of preferences. Cipriani, Giuliani and Jeanne (2007) report mother–son correlations for contributions in a standard public goods game and find no significant associations, interpreting this as evidence that peer effects influence contributions. Dohmen et al. (2006), on the other hand, use survey evidence on attitudinal questions and find modest intergenerational correlations in self-reported trust and risk attitudes. Naturally, these papers suffer from the limitation that it is impossible to separately identify genetic (parents passing on genes for a certain trait to their biological children) and cultural transmission.

In this paper, we move beyond the computation of intergenerational correlations and offer a direct test of the hypothesis that economic preferences are under genetic influence. We elicit preferences experimentally with a subject pool of twins recruited from the population-based Swedish Twin Registry. The virtue of this approach is that by comparing monozygotic (MZ) twins, who share the same set of genes, to dizygotic (DZ) twins, whose genes are imperfectly correlated, we can estimate the proportion of variance in experimental behavior due to genetic and to shared and unique environmental effects. The measures of economic preferences that we use are based on *de facto* observed experimental behavior under controlled circumstances with financial incentives attached to performance. For risk taking, we also present some

[1] For an extensive collection of essays on the intergenerational transmission of economic opportunity, see the volume edited by Bowles, Gintis and Osborne Groves (2005).

supplementary survey-based evidence derived from hypothetical questions that have been behaviorally validated (Dohmen et al., 2005, 2006).

This paper is the first to use the twin methodology to study (i) experimentally elicited risk preferences and (ii) giving behavior in a dictator game. Outside economics, two papers have used the twin methodology to shed light on individual variation in the ultimatum game (Wallace et al., 2007) and the trust game (Cesarini et al., 2008). Two other previous papers used twins as a subject pool (Loh & Elliott, 1998; Segal & Hershberger, 1999) but the experiments therein were designed to test whether cooperation varied by genetic relatedness, as predicted by inclusive fitness theory (Hamilton, 1964). Therefore, twins played against their cotwins, and consequently it is not possible to estimate heritability from these studies.

We find strong evidence that preferences for risk taking and giving are broadly heritable. Our point estimates from the best-fitting models suggest that approximately twenty percent of individual variation can be explained by genetic differences. Furthermore, our results suggest only a modest role for common environment as a source of variation. We argue that the significance of these results extends well beyond documenting an important, but hitherto largely ignored, source of preference heterogeneity. For example, although it is widely accepted that parent–offspring correlations in isolation cannot be used to discriminate between theories of genetic and cultural transmission, much economic research is carried out under the presumption that genetic transmission is small enough so that it can be safely ignored. Such an assumption is not consistent with our findings.

Importantly, the estimates we report are in line with the behavior genetics literature, where survey based studies have documented substantial genetic influences on variation in economically relevant abilities, preferences, and behaviors such as intelligence (Bouchard et al., 1990), personality (Jang, Livesley & Vernon, 1996), addiction (True et al., 1997), prosociality (Rushton et al., 1986; Rushton, 2004), sensation seeking (Stoel, De Geus & Boomsma, 2006), religiosity (Bouchard et al., 1999; Kirk et al., 1999; Koenig et al., 2005), political preferences (Alford, Funk & Hibbing 2005), and political participation (Fowler, Dawes & Baker, 2008). The remainder of this paper is structured as follows: in sections 2 and 3, we describe the method and the experiments used in detail; in section 4, we report the results; and in section 5, we discuss our findings. Section 6 concludes.

2. Data collection

2.1. Subject recruitment. The study was undertaken in collaboration with the Swedish Twin Registry at Karolinska Institutet.[2] The registry, which is the largest twin registry in the world, has been described in detail elsewhere (Lichtenstein et al., 2006). All of our invitees were same-sex twin pairs that had previously participated in the Webbased survey STAGE, an acronym for "the Study of Twin Adults: Genes and Environment." This survey was administered between November 2005 and March 2006 to all twins born in Sweden between 1959 and 1985, and it attained a response rate of 61%. Its primary purpose was to study environmental and genetic influences on a number of diseases (Lichtenstein et al., 2006), but it also contains self-reported data on marital, employment, and fertility status, as well as information on the frequency of twin contact. To allow further examination of the effects of our methods of recruitment on the representativeness of our sample, we also merged the STAGE cohort with a specially requested data set of socioeconomic and demographic variables compiled by Statistics Sweden.

In a first recruitment effort, during the summer and fall of 2006, a total of 658 twins (71 DZ and 258 MZ pairs) participated in the Swedish cities of Göteborg, Helsingborg, Kristianstad, Linköping, Lund, Malmö, Norrköping, Örebro, Stockholm, Uppsala, and Västerås. Due to the relatively small sample of DZ twins, a second round of data collection took place in February 2008. Both MZ and DZ twins were invited to participate, but DZ twins were pursued somewhat more vigorously, with personalized invitations and reminders being sent out to those who did not respond. This recruitment effort was successful in augmenting the sample size of DZ twins, and the complete data set comprises 920 twins: 141 DZ pairs and 319 MZ pairs. A vast majority of subjects, approximately 80%, are female. For the second data collection round, twins were recruited in the cities of Borlänge, Göteborg, Helsingborg, Jönköping, Lund, Malmö, Örebro, Stockholm, Umeå, Uppsala, Västerås, and Växjö. In all of the experimental sessions a condition for participation was that both twins in a pair be able to attend the same session. Moreover, invitations were extended only to twins who were both domiciled in the same city, or its surrounding areas. Zygosity was resolved by questionnaire items that have been shown to have a reliability of somewhere between 95% and 98% (Lichtenstein et al., 2006).

2.2. Experimental procedures. When subjects arrived at an experimental session they were seated apart and given general instructions orally. They were asked not to talk to one another during the experiment and to alert the experimenter if they had

[2] The study and subject recruitment were approved by the Ethics Committee for Medical Research in Stockholm.

any questions (questions were rare and were answered in private). Subjects were also told about the strong norm against deception in experimental economics. After having filled out a form with information for the administration of payments, subjects were given instructions for the first experiment (the modified dictator game; see below). There were no time constraints, so when all participants finished making their decisions, the next set of instructions was handed out. Subjects participated in a total of five different experiments. The experiment phase was followed by a short questionnaire with survey questions, a personality test, and a test of cognitive ability. On average, experimental sessions lasted a little more than an hour and average earnings were 325 SEK (exchange rate; $1 is about 6 SEK).

2.3. Giving. We used a modified dictator game to measure preferences for giving ("altruism").[3] In a standard dictator game (Forsythe et al., 1994), a subject decides how to split a sum of money between herself and another person (see Camerer [2003] for an overview of dictator game results). A variant of this approach first used by Eckel and Grossman (1996) is that the subject decides how to allocate a sum of money between herself and a charity. As donations to charity may be related to empathy and altruism more strongly than donations in the standard dictator game, we opted for this approach. Fong (2007) has demonstrated that empathy is a more important motivation for dictator game giving when recipients are perceived to be in great need (in their case welfare recipients). In the present study subjects decided how to allocate 100 SEK (about $16) between themselves and a charity called "Stadsmissionen." Stadsmissionen's work is predominantly focused on helping the homeless in Sweden. All subjects responded to the dictator game question and are included in the analysis below (319 MZ pairs and 141 DZ pairs).

2.4. Risk taking. To measure risk aversion, subjects were presented with six choices, each between a certain payoff and a 50/50 gamble for 100 SEK (about $16). The certain payoffs were set to 20, 30, 40, 50, 60, or 80 SEK. After subjects had made their six choices, one of these was randomly chosen for payoff by rolling a die. The gamble was resolved with a coin toss in front of the participants. The measure of risk aversion determines seven intervals for the certainty equivalent of the gamble. A similar question has been used by Holt and Laury (2002). Nineteen subjects provided inconsistent responses (2% of the total sample), and these were dropped (leaving 307

[3] Independently, Bardsley (2008) and List (2007) have shown that augmenting the choice set of the dictator to allow him or her to take money from the partner dramatically reduces generosity. This suggests that people's behavior in the standard dictator game is sensitive to cues about social norms in experimental settings. Regardless of one's favored interpretation of giving in dictator games, we will provide evidence suggesting that such giving is heritable.

MZ pairs and 135 DZ pairs for the analysis).[4] We refer to this measure as risk aversion and it is our primary measure of risk preferences.

We supplement this first measure of risk preferences with two hypothetical questions designed to measure risk attitudes. The first question, which we denote risk investment, asks the subjects to assume that they have won 1 million SEK on a lottery and that they are then given the opportunity to invest some of this money in a risky asset with an equal probability of doubling the investment or losing half the investment. Subjects can then choose between six different levels of investments: 0, 200,000, 400,000, 600,000, 800,000, or 1 million SEK. This question is similar to the question with real monetary payoffs, but involves much larger (although hypothetical) stakes. The second question, risk assessment, measures general risk attitudes on a 0–10 scale, where 0 is complete unwillingness to take risks and 10 is complete willingness to take risks. This scale question measures general risk attitudes rather than monetary risk attitudes. Dohmen et al. (2005) showed that all of these three measures of risk attitudes are significantly related to each other, and established the behavioral validity of the two hypothetical questions with respect to real risk taking.

3. Twin methodology

Comparing the behavior of identical and nonidentical twins is a form of quasi-controlled experiment. MZ and DZ twins differ in their genetic relatedness. If a trait is heritable, then it must be the case that the correlation in MZ twins is higher than the correlation in DZ twins. We start by examining the MZ and DZ correlations. Such an examination serves two purposes. A number of authors (Loehlin, 1965; Goldberger, 1977, 1979), have noted that moving from a crude comparison of correlations to a full-fledged variance decomposition requires making some strong independence and functional form assumptions. A first purpose is therefore to examine whether a significant difference in correlations exists. This serves as a diagnostic of whether the traits in question are under genetic influence. Second, as explained below, the workhorse models in behavior genetics do imply certain restrictions on the MZ and DZ correlations. Correlations that fall significantly outside the space of permissible correlations are therefore an indication of model misspecification and the raw correlations can be used to test for such misspecification. To explain why, it is necessary to introduce some basic concepts from behavior genetics (see Chapter 3 in Neale and Maes [2004]). By phenotype, we simply mean the observed outcome variable. The location of a gene on

[4] An inconsistent response is one in which the certainty equivalent is not uniquely defined; that is, an individual who chooses 20 SEK rather than the gamble in the first question and then chooses the gamble rather than 30 SEK in the second question. Such behavior is a strong indication that the subject either has misunderstood the question, or has failed to take it seriously.

a chromosome is known as a locus. Alleles are the alternative forms of a gene that may occupy the same locus on a chromosome. Finally, the genotype of an individual is the alleles he or she has at a locus. Suppose that the phenotype of twin $j \in \{1, 2\}$ in family i can be written as the sum of four independent influences,

$$\chi_{ij} = C_{ij} + E_{ij} + A_{ij} + D_{ij}, \tag{3.1}$$

where C_{ij} is the common environmental factor, E_{ij} is the individually experienced unique environment factor, A_{ij} is an additive genetic factor, and D_{ij} is a dominance factor. Common environmental influences are defined as those influences shared by both twins, for example the home environment, so that $C_{i1} = C_{i2}$. Unique environmental influences, by contrast, are defined as environmental experiences idiosyncratic to each twin.

Behavior geneticists distinguish between additive genetic effects and dominance effects. For an intuitive illustration of the difference, consider the simple case where there are two possible alleles, a_1 and a_2, so that each individual, getting one allele from each parent, has genotype (a_1, a_1), (a_1, a_2), or (a_2, a_2). Dominance is then present whenever the effect of having genotype (a_1, a_2) is not equal to the mean effect of genotypes (a_1, a_1) and (a_2, a_2). In other words, dominance can be thought of as an interaction effect.

Because the influences are assumed to be independent, the model predicts that the covariance in MZ twins is equal to

$$Cov_{MZ} = \sigma_A^2 + \sigma_D^2 + \sigma_C^2 \tag{3.2}$$

because identical twins share the same genes and were reared together. The phenotypic covariance between DZ twins is derived in Mather and Jinks (1977) as

$$Cov_{DZ} = \frac{1}{2}\sigma_A^2 + \sigma_D^2 + \sigma_C^2 \tag{3.3}$$

The coefficients of genetic relatedness for DZ twins in equation (3.3) thus imply that DZ twins share half the additive genetic effects and a quarter of the dominance effects.

Notice that parameters of this model cannot be identified with twin data alone, because we have one equation less than the number of parameters to be estimated. This ambiguity is typically resolved in twin research by assuming that all gene action is additive, so that $\sigma_D^2 = 0$. Behavior geneticists distinguish between broad heritability, defined as $(\sigma_A^2 + \sigma_D^2)/(\sigma_A^2 + \sigma_D^2 + \sigma_C^2 + \sigma_E^2)$, and narrow heritability, defined simply as $\sigma_A^2/(\sigma_A^2 + \sigma_D^2 + \sigma_C^2 + \sigma_E^2)$. The identifying restriction that σ_D^2 equals zero can be tested by examining whether ρ_{DZ} is at least half of ρ_{MZ}, and the greatest difference in

correlation allowed by the model arises when $\sigma_C^2 = 0$ and $\sigma_A^2 = 0$, in which case ρ_{MZ} is four times greater than ρ_{DZ}.

In our empirical analysis, we start by comparing the correlations of MZ and DZ twins using the bootstrap. Letting N_{MZ} be the number of complete MZ pairs, we draw N_{MZ} pairs with replacement 1,000 times and calculate both parametric and nonparametric correlation each time. We proceed analogously for DZ twins and then create a 1,000×1 vector where the DZ correlation is subtracted from the MZ correlation for each draw. This gives a distribution for the difference in correlations between the two samples. The p-value for the test of the hypothesis that the two correlations are equal is then the number of negative entries in the vector divided by 1,000. The use of a one-sided test is theoretically justified in our case because the notion that the DZ correlation could be greater than the MZ correlation is not a particularly interesting alternative hypothesis. We also use the same bootstrap technique to test the hypothesis that the DZ correlation is at least half as large as the MZ correlation. The result of the latter exercise will inform our choice of identifying restrictions.

For our two main outcome variables, we estimate mixed-effects Bayesian ACE models.[5] We report results treating outcome variables as continuous as well as ordinal. Using the same notation as before, the model is written as

$$y_{ij}^* = \chi_{ij} \tag{3.4}$$

where χ_{ij} is the sum of genetic, shared environment, and unshared environment random effects. For MZ twins the latent variable is the sum of three random effects,

$$\chi_{ij}^{MZ} = A_i + C_i + E_{ij}, \tag{3.5}$$

where A_i is the family genetic factor, C_i is the family-shared environment factor, and E_{ij} is the individually experienced unshared environment factor. For DZ twins the latent variable is a function of four random effects variables,

$$\chi_{ij}^{DZ} = A_{1i} + A_{2ij} + C_i + E_{ij}, \tag{3.6}$$

where A_{1i} is the family genetic factor shared by both twins, A_{2ij} is the individually inherited genetic factor that is unique to each twin, and C_i and E_{ij} are the same as for MZ twins. In the continuous models, we take the outcome variables in the experiment to be y_{ij}^*. In the ordered models, the outcome variables are instead modeled under

[5] Researchers have increasingly used Bayesian methods, implemented using Markov chain Monte Carlo (MCMC) algorithms, to estimate the variance components in *ACE* models. The likelihood functions in genetic models often present computational challenges for maximum likelihood approaches because they contain high-dimension integrals that cannot be evaluated in closed form and thus must be evaluated numerically. For a detailed discussion of Bayesian *ACE* models, we refer to van den Berg, Beem and Boomsma (2006).

the assumption that y_{ij}^* is not directly observed. Instead, the observed variable y_{ij} is assumed to be one of $k+1$ ordered categories separated by k thresholds that are estimated as part of the model. The three risk measures naturally fall into categories, and hence these categories are used in the analysis. A visual inspection of Figure 1 shows that the distribution of dictator game responses is roughly trimodal, with peaks at the three focal points: donating the entire endowment, donating half the endowment, or keeping the entire endowment. Approximately 80% of responses are in one of those three categories. Consequently we construct an ordinal variable where individuals who donate between 0 and 33 are coded as 0, individuals who donate between 33 and 66 are coded as 1, and individuals who donate more than 66 are coded as 2. We use the variances of the random effects to generate estimates of heritability, common environment, and unique environment. Because the underlying components are not constrained, the estimated proportions can range anywhere from 0 (the component has no effect on variance) to 1 (the component is solely responsible for all observed variance).

Replicating the methods used in this literature, we assume that our unobserved random effects are normally distributed and independent:

$$A \sim N\left(0, \sigma_A^2\right), \tag{3.7}$$

$$A_1 \sim N\left(0, \sigma_A^2/2\right), \tag{3.8}$$

$$A_2 \sim N\left(0, \sigma_A^2/2\right), \tag{3.9}$$

$$C \sim N\left(0, \sigma_C^2\right), \tag{3.10}$$

$$E \sim N\left(0, \sigma_E^2\right). \tag{3.11}$$

The variance of A_1, the family genetic effect for DZ twins, is fixed to be half the variance of A, the family genetic effect for MZ twins, reflecting the fact that MZ twins on average share twice as many genes as DZ twins. Moreover, DZ twins are also influenced by individually specific genes A_2 that are drawn from the same distribution as the shared genes, because on average half their genes are shared and half are not. These assumptions about the genetic variance help to distinguish shared genes from the shared environment variable C, which is assumed to have the same variance for both MZ and DZ twin families, and the residual unique environment variable, E, from which a unique draw is made for each individual. The contribution of a variance component is simply estimated as $\sigma_i^2/(\sigma_E^2 + \sigma_A^2 + \sigma_C^2)$, where $i \in \{A, C, E\}$.[6]

[6] If we tried to estimate all three components of variance simultaneously in the ordered model, it would not be identified, so we fix the variance of the unshared environment σ_E^2 to be one.

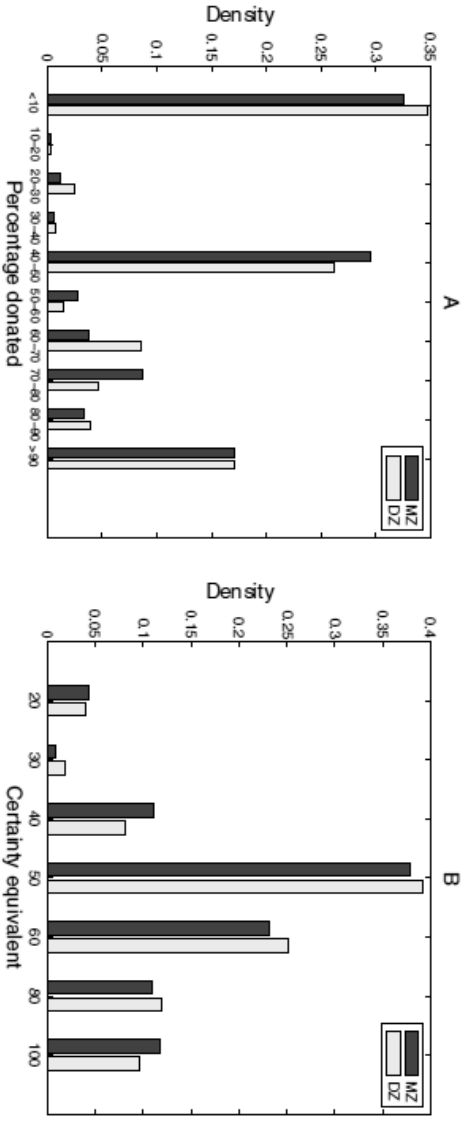


Figure 1. Panel A: The distribution of giving (percent donated), by zygosity. Panel B: The distribution of risk aversion (certainty equivalent), by zygosity

We estimate three types of models in addition to the ACE model. An AE model accounts for only heritability and common environment, a CE model accounts for only common and unique environment, and an E model accounts for only unique environment. Procedurally, the difference between the ACE model and these submodels is that one or more variances are restricted to equal zero. Estimating submodels allows testing whether the parameter restriction results in a significant deterioration in fit. For example, in the AE model, the random effect for the common environment is not estimated. To compare the fit of ACE, AE, CE, and E models we used the deviance information criterion (DIC), a Bayesian method for model comparison analogous to the Akaike information criterion (AIC) in maximum likelihood estimation. Models with smaller DIC are considered to have better out-of-sample predictive power (Gelman et al., 2004). The DIC is defined as the sum of deviance ($Dbar$), a measure of model fit, and the effective number of parameters (pD), which captures model complexity.[7]

In our Markov chain Monte Carlo procedure we use vague, or flat, prior distributions to ensure that they do not drive our results. For the thresholds, τ_i, we use a mean-zero normal distribution with variance 1,000,000, and for the precision parameters associated with σ_A^2, σ_E^2, and σ_C^2, we use a Pareto distribution with shape parameter equal to 1 and scale parameter equal to 0.001, which is the equivalent of putting a uniform (0, 1,000) prior on the variances. A Pareto distribution has proven to work well for variance components in genetic models (Burton et al., 1999; Scurrah, Palmer & Burton, 2000). In addition, we use convergence diagnostics to make sure that the stationary posterior distribution has been reached. To ensure that the models converged to their target posterior distribution, we began sampling from the joint posterior distribution after convergence was established using the Brooks and Gelman (1998) statistic (values of less than 1.1 on all parameters indicate convergence). For all of the models the "burn-in" period was 100,000 iterations and the chains were thinned by 100.

4. Results

In Table 1 we report some background statistics. On average, subjects donated 54% of their endowment in the dictator game to the charity and the average certainty equivalent in the risky gamble was 52.[8] Results from the first hypothetical question

[7] Letting θ be the parameter vector, y the data, p the likelihood function, and $f(y)$ a standardizing term which is a function of the data alone, the deviance is defined as $D(\theta) = -2\ln(p(y|\theta)) + 2\ln f(y)$. Then $Dbar$ is defined as $Dbar = E_\theta(D(\theta))$ and pD is defined as $pD = Dbar - D(\overline{\theta})$, where $\overline{\theta}$ is the expectation of θ. The deviance information criterion can then be calculated as DIC $= pD + Dbar$ For further details, see Spiegelhalter et al. (2002).

[8] To facilitate interpretation, in Table 1 we define the certainty equivalent as the midpoint between the lowest sure amount that the subject is willing to accept and the category immediately below. For

reveal that subjects invest on average 31% of their endowment. Finally, on a scale from 0 to 10, subjects report an average willingness to take risks of just above 5. Tests of equality for all four variables fail to reject the null hypothesis that the MZ and DZ means are equal at the 5% level. To give an impression of individual variation in responses, in Figure 1 we plot histograms of the distributions for risk aversion and giving, separately, for DZ and MZ twins. A visual inspection reveals that there is ample variation in responses and fails to lend much support to the hypothesis that the frequency distributions vary by zygosity.

Table 1. Experimental behavior[9]

		MZ	DZ	p-value
Giving	Mean	53.60	54.43	.77
	S.D.	37.27	37.94	
	n	638	282	
Risk aversion	Mean	52.38	51.88	.71
	S.D.	18.53	17.80	
	n	625	276	
Risk investment	Mean	30.25	33.19	.08
	S.D.	21.22	21.28	
	n	638	279	
Risk assessment	Mean	4.98	5.25	.07
	S.D.	1.98	1.96	
	n	636	279	

In Table 2, we report parametric and nonparametric correlations for MZ and DZ twins. Pearson correlations do not differ appreciably from Spearman correlations. These correlations convey a lot of information, and because a purely environmental model cannot account for any differences between MZ and DZ correlations, they serve as a preliminary diagnostic of whether the preferences in question are in part under genetic influence. For giving, the Spearman correlation is 0.319 for MZ twins and 0.106 for DZ twins, consistent with a genetic effect. Similarly, for risk aversion, the Spearman correlation is 0.222 for MZ twins and 0.025 for DZ twins, whereas for risk investment, the corresponding figures are 0.264 and 0.096. However, for risk assessment, the separation is larger, with an MZ correlation of 0.367 and a DZ correlation of -0.034. As the sample size is smaller for DZ twins, these correlations are estimated with less precision, yielding wider confidence intervals. Yet, when the equality of the correlations is tested using the bootstrap, the one-sided p-value is less than 2% for giving, risk aversion, and

example, a subject who chooses the gambles at 20, 30, and 40 and then prefers 50 SEK with certainty is assigned a certainty equivalent of 45.

[9] The p-value is for the test of the hypothesis that the mean of the MZ and DZ distributions are the same. Standard errors are adjusted to take nonindependence into account (Liang & Zeger, 1986).

risk assessment. Though the MZ correlation is also higher than the DZ correlation for risk investment, the hypothetical investment question, the difference is not significant at the 5% level (p = 0.07). The robust separation of MZ and DZ correlations is illustrated in Figure 2, where we plot the response of twin 1 against the response of twin 2 separately for MZ and DZ twins. Hence, the evidence is very compelling that genes do contribute to phenotypic variation in both giving and risk aversion.

Table 2. Experimental behavior[10]

		MZ	DZ	p-value $\rho_{MZ} - \rho_{DZ}$
Giving	Spearman	0.319^{***}	0.106	0.015
		$(0.211 - 0.426)$	$(-0.067 - 0.292)$	
	Pearson	0.317^{***}	0.099	0.013
		$(0.208 - 0.424)$	$(-0.075 - 0.279)$	
	n	319	141	
Risk aversion	Spearman	0.222^{***}	0.025	0.020
		$(0.118 - 0.341)$	$(-0.150 - 0.189)$	
	Pearson	0.222^{***}	0.024	0.024
		$(0.099 - 0.342)$	$(-0.135 - 0.179)$	
	n	307	135	
Risk investment	Spearman	0.264^{***}	0.096	0.066
		$(0.149 - 0.364)$	$(-0.077 - 0.277)$	
	Pearson	0.304^{***}	0.110	0.057
		$(0.177 - 0.408)$	$(-0.079 - 0.315)$	
	n	319	139	
Risk assessment	Spearman	0.367^{***}	-0.034	0.001
		$(0.266 - 0.468)$	$(-0.217 - 0.148)$	
	Pearson	0.384^{***}	-0.043	0.001
		$(0.280 - 0.481)$	$(-0.237 - 0.139)$	
	n	317	139	

We also used the same bootstrapping method to test the null hypothesis that the DZ correlation is at least half the MZ correlation, as implied by the ACE specification. For neither risk aversion (p = .16), risk investment (p = .36), nor giving (p = .30) can we reject the null hypothesis. On the other hand, we can reject the null hypothesis for risk assessment (p = .02), suggesting that the estimation of an ACE model is inappropriate. Notice that even though we cannot reject the hypothesis at conventional

[10] ***, **, * significantly different from zero at 1%, 5%, and 10% levels. All results are bootstrapped. p-values are one-sided. 95% confidence intervals within parentheses.

levels of significance in three out of four cases, it is still striking that the estimated DZ correlations are always less than half the MZ correlations.

In what follows, we restrict our attention to the results from our experiments with monetary incentives; results for the supplemental risk measures are reported in Tables A3–A5 in the Appendix. Because we cannot reject the null hypothesis that the DZ correlation is at least half the MZ correlation for our two main experimental measures, we do not depart from the convention of estimating ACE models. In Tables 3 and 4 we present the estimates of the variance components of the ACE model and its nested submodels. Parameter estimates are similar, regardless of whether the outcome variable is treated as continuous or ordinal. The estimate of genetic influences on giving is 0.22 (0.28) in the most general version of the continuous (ordered) model. Corresponding estimates for risk aversion are 0.14 and 0.16, whereas the contribution of the common environment is closer to zero, both in our modified dictator game and for risk aversion.

It is interesting to contrast these results to those that have previously been reported for other outcome variables of interest to economists. For example, Björklund, Jäntti and Solon (2005) estimated heritability of earnings in Sweden using multiple sibling types and obtained heritability estimates for income in the range 10% to 30%, whereas Taubman's original estimates based on a sample of white U.S. war veterans were slightly higher (Taubman, 1976). The estimates for trust and trustworthiness reported in previous papers, though imprecise, are also in the neighborhood of 20% in both U.S. and Swedish data (Cesarini et al., 2008). Generally, the estimated heritabilities for our experimentally elicited preferences are a little lower than the reported broad heritabilities for personality, which tend to be around 50% (Plomin et al., 2001a), and lower still than the estimates of the heritability of IQ (Neisser et al., 1996). In making the comparison to psychological variables it is, however, important to bear in mind that the reliability of the measurement instruments used by psychometricians in IQ and personality research may be different from the reliability of behavior in economic experiments.

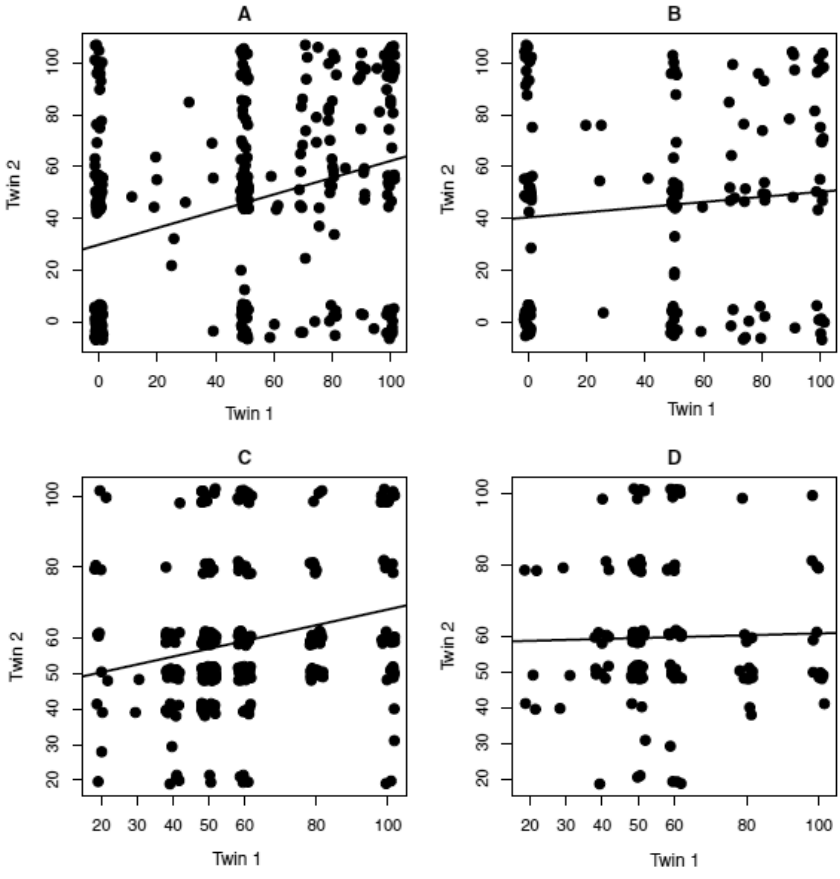


Figure 2. Scatterplots (jittered for expositional clarity) Panel A: percent donated, dictator game, MZ. Panel B: ditto, DZ. Panel C: risk aversion, certainty equivalent, MZ. Panel D: ditto, DZ.

Table 3. Results of the ACE model and its nested submodels, giving[11]

		ACE	AE	CE	E
Continuous	A	0.22	0.31	–	–
		$(0.05, 0.36)$	$(0.21, 0.40)$	–	–
	C	0.09	–	0.25	–
		$(0.01, 0.23)$	–	$(0.16, 0.33)$	–
	E	0.70	0.69	0.75	1.00
		$(0.60, 0.79)$	$(0.60, 0.79)$	$(0.67, 0.84)$	$(1.00 - 1.00)$
Dbar		4,719	4,706	4,783	5,043
pD		227.3	234.9	184.8	2.0
DIC		4,946	4,941	4,968	5,045
Ordered	A	0.28	0.39	–	–
		$(0.06, 0.46)$	$(0.27, 0.51)$	–	–
	C	0.11	–	0.32	–
		$(0.01, 0.30)$	–	$(0.21, 0.43)$	–
	E	0.61	0.61	0.68	1.00
		$(0.50, 0.73)$	$(0.49, 0.74)$	$(0.57, 0.79)$	$(1.00 - 1.00)$
Dbar		1,693	1,688	1,761	2,023
pD		236.0	238.7	189.8	2.0
DIC		1,929	1,927	1,951	2,025

In light of these results, it is not surprising to find that both for giving and for risk aversion, the diagnostics of model fit repeatedly point to the AE model as the most appropriate. Setting C to equal zero is potentially a drastic step, but it is consistent with the fairly low DZ correlations that we observe. When the AE submodel is estimated, the estimates of A for giving are 0.31 (0.39) in the continuous (ordered) models. The corresponding figure for risk aversion is 0.21 (0.25). We also report the results from CE and E models. CE models always have fit diagnostics worse than the AE and ACE models. Not surprisingly, the E model fits the data very poorly.

[11] A is the genetic contribution; C is the common environment contribution; E is the unique environment contribution. *Dbar*: Deviance. *pD*: Effective number of parameters. DIC: Bayesian deviance information criterion. 95% credible intervals within parentheses.

Table 4. Results of the ACE model and its nested submodels, risk aversion[12]

		ACE	AE	CE	E
Continuous	A	0.14	0.21	–	–
		(0.02, 0.27)	(0.11, 0.31)	–	–
	C	0.07	–	0.17	–
		(0.00, 0.18)	–	(0.08, 0.26)	–
	E	0.80	0.79	0.83	1.00
		(0.69, 0.89)	(0.70, 0.89)	(0.74, 0.93)	(1.00 – 1.00)
Dbar		7,713	7,707	7,752	7,914
pD		160.8	163.9	130.6	2.0
DIC		7,873	7,871	7,883	7,916
Ordered	A	0.16	0.25	–	–
		(0.01, 0.30)	(0.14, 0.36)	–	–
	C	0.09	–	0.20	–
		(0.01, 0.22)	–	(0.10, 0.30)	–
	E	0.75	0.75	0.80	1.00
		(0.65, 0.86)	(0.64, 0.86)	(0.70, 0.90)	(1.00 – 1.00)
Dbar		2,760	2,752	2,804	2,985
pD		181.4	186.3	149.1	5.9
DIC		2,941	2,938	2,953	2,991

4.1. Equal environment assumption. Critics of the classical twin design cite a number of alleged failures of the equal environment assumption, including that MZ twins are more likely to interact, and that parents, on average, give MZ twins more similar treatment (Pam et al., 1996). Indeed, Björklund, Jäntti and Solon (2005) have shown, using a data set with nine different sibling types, that estimates of the variance components in income do change substantially when the equal environment assumption is relaxed. In the context of research on personality and IQ, the evidence is, however, fairly convincing that any bias that arises from the equal environment assumption is not of first order. Most importantly, for measures of personality and cognitive ability, studies of MZ and DZ twins reared apart tend to produce estimates of heritability similar to those using twins reared together (Bouchard, 1998). Because studies of twins reared apart do not rely on the equal environments assumption, it is unlikely that the assumption is a major source of bias. Second, although it is true that MZ twins report a

[12] A is the genetic contribution; C is the common environment contribution; E is the unique environment contribution. *Dbar*: Deviance. *pD*: Effective number of parameters. DIC: Bayesian deviance information criterion. 95% credible intervals within parentheses.

higher frequency of contact with one another than DZ twins, twin similarity has been shown to cause greater contact rather than vice versa (Posner et al., 1996). Other studies have failed to find a significant relationship between similarity and contact. For example, one large study found that frequency of contact is not correlated with similarity in social attitudes (Martin et al. 1986). Third, the claim that the greater similarity of MZ twins is due to more uniform parental influences rests on fairly weak empirical ground. Measures of the degree of similarity in parental treatment turn out to not be correlated with similarity in IQ or other personality measures (Bouchard et al., 1990). Also, in the relatively rare cases where parents miscategorize their twins as MZ instead of DZ (or the converse), differences in cognitive ability and personality persist (Bouchard & McGue, 2003). Finally, we note that our estimated Cs are very low, and it would appear that the Bayesian estimator, if anything, overstates the importance of shared environment compared to other standard estimators.[13]

4.2. Measurement error. In the simplest case, where the studied preference is observed with mean zero random error, we can think of the unique environment component as being composed of two terms, $E_{ij} = E^*_{ij} + \epsilon_{ij}$, here ϵ_{ij} is a mean zero variable with variance σ^2_ϵ and is i.i.d. across time. Under these assumptions, it is easy to show that the estimates of A and C need to be scaled up by a factor of $1/(1-\sigma^2_\epsilon)$. For example, under the conservative assumption of a retest correlation of 0.8, this would imply a σ^2_ϵ of 0.2, and therefore the estimates of A and C would need to be scaled up by 25%, that is, to somewhere between 0.18 and 0.41 for A in our ACE models. There is surprisingly little evidence on test–retest stability in economic experiments. One recent paper (Brosig, Riechmann & Weimann, 2007) examined the temporal stability of individual behavior in modified dictator and prisoner's dilemma games and found that individual behavior is unstable across time in a given game. However, the authors used a concept of stability that is not easily mapped to an estimate of σ^2_ϵ. Other papers have estimated error rates from identical responses to items, typically finding reversal rates on the order of 10%–20% (Harless & Camerer, 1994; Hey & Orme, 1994).

4.3. Representativeness. Compared to most experimental work, our sample is an improvement in terms of representativeness because we draw our subjects from a population-based registry and not a pool of college students. Yet it is important to establish the "selectivity" of our sample. In particular, three questions arise. First,

[13] It is clear by inspection that a method of moments estimator would produce nonsensical negative estimates of common environment. When continuous ACE models are estimated using maximum likelihood in MPLUS (Muthén & Muthén, 2006) and bootstrapping the standard errors, estimated Cs are always equal to zero, and the estimated heritabilities are 0.21 for risk aversion, 0.31 for giving, 0.29 for risk investment, and 0.35 for risk assessment. All estimates of A are significant at the 5% level.

are the MZ and DZ twins who agree to participate drawn from similar environments? Second, to what extent does our method of sampling lead to overrecruitment of subjects with certain characteristics? If any such characteristics are associated with heritability, then estimates of variance components will be biased. Third, in light of the fairly skewed ratio of MZ twins to DZ twins in our sample, are there any reasons to believe that this has affected our estimates?

A basic assumption of the ACE model is that MZ twins and DZ twins are drawn from the same environment. We have already demonstrated that in terms of experimental outcomes, the MZ and DZ distributions appear to be the same. To further investigate this hypothesis, we conducted a battery of tests for equality of background variables including gender, years of education, employment status, health, income, and marital status. With the exception of age, we did not find any significant differences between the MZ and DZ samples. The results are reported in Table 5.

Table 5. MZ-DZ comparisons for background variables[14]

	MZ		DZ			
	Mean	S.D.	Mean	S.D.	p-value	Data Source
Female	0.77	0.42	0.82	0.39	0.24	Multiple
Age	34.30	7.35	35.95	7.81	0.03	Multiple
Education	13.70	2.22	13.63	2.18	0.69	Stat. Swe.
Income	201,973	152,674	217,548	119,997	0.19	Stat. Swe.
Employed full time	0.54	0.50	0.60	0.49	0.23	STAGE
Unemployed	0.03	0.18	0.04	0.19	0.80	STAGE
Self-employed	0.04	0.20	0.07	0.25	0.32	STAGE
On sickleave	0.04	0.19	0.02	0.12	0.10	STAGE
Gov. employee	0.40	0.49	0.45	0.50	0.26	STAGE
Cognitive ability	0.03	0.99	−0.06	1.02	0.30	Exp.
Emotional stability	−0.04	1.00	0.10	0.99	0.09	Exp.
Agreeableness	0.02	0.98	−0.04	1.04	0.55	Exp.
Extraversion	−0.04	0.98	0.08	1.04	0.16	Exp.
Conscientiousness	−0.02	1.01	0.04	0.98	0.55	Exp.
Health	1.87	0.81	1.88	0.79	0.86	STAGE
Marital status	0.25	0.43	0.29	0.46	0.26	Stat. Swe.
No. of children	0.70	0.99	0.76	0.99	0.55	Stat. Swe.

[14] Education refers to years of education. Income is the sum of wage income, taxable transfers, and income from own company for the year 2005 (in SEK). Employment information was gathered when the subject responded to the STAGE questionnaire. Psychological measures were adjusted to have mean 0 and standard deviation 1 for the whole sample. Health is self-reported on a scale from 1 to 5. Marital status is a dummy variable taking the value 1 if the subject is married. Number of children is number of children under 18 living in the respondent's household in the year 2005. The p-value is for the test of the hypothesis that the mean of the MZ and DZ distributions are the same. We utilized

Second, it is possible that the twins who participated are not representative of the population as a whole. Like most twin studies (Lykken, McGue & Tellegen, 1986), our method of recruitment led to an oversampling of women and of MZ twins. Comparing our participants to the STAGE cohort as a whole on a number of background variables, we find few economically interesting differences. These results are also reported in the Appendix.

A comparison to the entire STAGE cohort is only an imperfect measure of representativeness, however, because STAGE respondents are also a self-selected group. We have therefore merged our experimental data with information on educational attainment, marital status, and income from Statistics Sweden and can thus further examine how our sample compares to the population mean for the cohort born 1959 to 1985. The population marriage rate is, for women 36%, and for men 29%. This is slightly higher than what we observe in our experimental sample. For income,the population averages are close to those of our participants. On average men earn 247,000 SEK, whereas our male subjects earn 244,000 SEK. For women the corresponding figures are 181,000 and 197,000. Finally, we find that the average years of education in the cohort as a whole are 12.09 for men and 12.49 for women, which is slightly more than one year less than the average for our experimental sample.

The upshot of this discussion is that our method of sampling leads to mild overrecruitment of subjects who are younger than average, are less likely to be married, and have fewer children on average. There is also modest overrecruitment of subjects with better than average educational attainment. Is this above-average educational attainment of our subjects a source for concern? For instance, it has been suggested that the heritability of intelligence might be moderated by social stratum (Turkheimer et al. 2003), at least in children, and a similar argument might apply to the effect of educational attainment on our outcome variables. To investigate this, we modify the continuous version of our baseline model to allow for interaction between A and years of education.[15] The fit of the new model is slightly better for risk aversion and slightly worse for the other three variables, suggesting that the interaction between A and education should not be included. For risk aversion, heritability increased somewhat, to 0.21 (95% CI 0.02, 0.39), compared to the baseline model.[16]

adjusted Wald tests for equality taking into account nonindependence within twin families (Liang & Zeger, 1986).

[15] This model is $\chi_{ij}^{MZ} = A_i + \beta \times A_i \times \text{Education}_{ij} + C_i + E_{ij}$ for MZ twins and $\chi_{ij}^{DZ} = A_{1i} + A_{2ij} + \beta \times (A_{1i} + A_{2ij}) \times \text{Education}_{ij} + C_i + E_{ij}$ for DZ twins.

[16] The DIC for the risk aversion, risk investment, risk assessment, and dictator game interaction models are 7,813, 3,881, 3,698, and 4,919, respectively. New baseline models were run to account for the fact that the interaction models were based on fewer observations due to missing values for the years of education variable. The baseline DICs are 7,824, 3,872, 3,695, and 4,915.

Finally, there is a third, more subtle way in which recruitment bias may be affecting our estimates. A plausible explanation for the overrecruitment of MZ twins is that because MZ twins are in more frequent contact with each other, it is easier for them to coordinate on a date and time. The concern here is that coordination costs, or willingness to participate more generally, might be associated with behavioral similarity. If so, this will inflate correlations, leading to an upward bias in the estimates of A and C. If this form of selection is more severe for MZ or DZ twins, it will also bias the estimates of the relative importance of common environmental and genetic influences. A reasonable proxy variable for costs of coordination is the frequency of contact between twins. Self-reported data on frequency of contact are available in STAGE.[17] When we compare twins who took part in our study with those who did not, there is a practically and statistically significant difference in the anticipated direction. MZ twins who participated in the study report a frequency of contact of 260 interactions per year, whereas those who did not participate report 234 interactions per year. The corresponding figures for DZ twins are 199 and 155. These differences are highly significant. In other words, frequency of contact is a robust predictor of participation. The crucial question, however, is whether frequency of contact predicts behavioral similarity. To test this, we regress the absolute value of the within-pair difference in giving and the three measures of risk on the average self-reported frequency of contact. Controlling for zygosity, the coefficient on frequency of contact is never significant. In other words, a reasonable proxy variable for "costs of coordination" does not seem to be related to behavioral similarity.

A second robustness test is to take variables that are available for the STAGE cohort in its entirety and ask whether there are any systematic differences between subjects who participated in our experiments and those who did not, in terms of correlations. If correlations in health, income, years of education, and the numerous other variables we investigate are consistently higher in the experimental sample, this would then suggest that these are a self-selected group with greater concordance in general. The results from this exercise are reported in Table A2 of the Appendix of this paper. There is no tendency for the patterns of correlations to differ between the two groups.

[17] We construct the frequency of contact variable as follows. Subjects who report at least one interaction (by e-mail, telephone, or letter) per day are assigned a value of 365. Subjects who report less than one interaction per day are simply assigned a value equal to the number of interactions per year. Interestingly, frequency of contact also provides a falsification test of the basic twin model. Because this variable is the same for both twins in a pair, it cannot possibly be heritable. A higher MZ correlation than DZ correlation would then suggest that measurement errors are more correlated in MZ twins. Fortunately, this turns out not to be the case. In our experimental sample, the MZ correlation is 0.76 and the DZ correlation is 0.71. In STAGE as a whole, the correlations are 0.77 and 0.75.

4.4. Genetic nonadditivity. The models we use—like most behavior genetic models—assume that genes influence a trait in an additive manner. That is, the genetic effect is simply the sum of all individual effects. This is by far the most common way to achieve identification. It has long been known that the twin model suffers from parameter indeterminacy when, for example, dominance effects are present because the number of parameters to be estimated exceeds the number of independently informative equations (Keller & Coventry, 2005). The fact that our DZ correlations are less than half of the MZ correlations could be the result of sampling variation. But it could also be an indication that there is some nonadditive genetic variation present. For one of our risk measures, risk assessment, we are in fact able to reject the hypothesis that the DZ correlation is at least half the MZ correlation. In Table A5 of the Appendix to this paper, we report the results of an ADE model and show that this model fits the data better, as judged by the DIC criterion.

A more rigorous way to test for nonadditivity would be to extend the data set to include also sibling, parent–child, or even cousin data. Though our data do not contain such information, Coventry and Keller (2005) recently completed a major review of all published parameter estimates using the extended family design compared to classical twin design estimates derived from the same data. The authors report that the estimates of broad heritability in twin studies are fairly accurate. However, the classical twin design overestimates the importance of additive genetic variation and underestimates the importance of nonadditive genetic variation. Evidence from studies of adoptees points in the same direction. In a recent metastudy by Loehlin (2005), the author reports average correlations of 0.13 for personality and 0.26 for attitudes in families with children reared by their biological parents. However, the correlations for personality and attitudes are 0.04 and 0.07, respectively, between adopted children and their nonbiological parents, but 0.13 and 0.20 between adopted children and their biological parents (Loehlin, 2005). Because only additive genetic variance is transmissible across generations (Fisher, 1930), doubling the parent–child correlation produces an upper bound on the estimate of narrow heritability. The fact that this upper bound is lower than estimates derived from twin studies reinforces the point that there is probably nonadditive variation in personality and attitudes. The low DZ correlations we observe suggest that a similar situation obtains for economic preferences.

We thus concur with the conclusion in Coventry and Keller (2005), namely that the estimates from the classical twin design should not be interpreted literally, but are nevertheless very useful because they produce reasonably accurate estimates of broad heritability, and hence of genes as a source of phenotypic variation.

5. Discussion

In this paper, we have used standard behavior genetic techniques to decompose variation in preferences for giving and risk taking into environmental and genetic components. We document a significant genetic effect on risk taking and giving, with genes explaining approximately twenty percent of phenotypic variation in the best-fitting models. The estimated effect of common environment, by contrast, is smaller. Though these results are clearly in line with the behavior genetic literature (Turkheimer, 2000), the implications of these findings in the context of modern economics merit further comment.

In particular, it is important to exercise great care in interpreting the estimates of variance components. Contrary to what is sometimes supposed, they are estimates of the proportion of variance explained and thus do not shed any direct light on the determinants of average phenotype. This distinction is important. For instance, if genetic transmission in a studied population is uniform, then a trait that is primarily acquired through genes might actually show low, or zero, heritability. The same argument is true for common environment. A low estimated C could simply mean that there is little variation in how parents culturally transmit preferences or values to their children. This caveat is especially important to bear in mind when interpreting heritability estimates from a study population such as ours, where it seems plausible to assume that environmental variation between families is modest.

Like any other descriptive statistic, a heritability estimate is specific to the population for which it is estimated, and, though our findings are probably informative about heritability in other modern Western societies, we caution against further extrapolation. Variation in our study population is in all likelihood small relative to cross-country differences or historical environmental differences that could potentially generate greater variation in risk preferences and giving. Perhaps the most striking and intuitive illustration of this point comes from the study of income, which is moderately heritable in Sweden as well as in the United States (Björklund, Jäntti & Solon, 2005; Taubman 1976). In recent centuries incomes have increased manifold, and even today an individual's country of origin is by far the most important determinant of that individual's income (Sala-i-Martin, 2006). In other words, a heritability statistic says little about the malleability of a trait with respect to environmental interventions (Goldberger, 1979).

Caution should also be exercised in interpreting our estimate of unique environment (E), because it is not possible to separately identify unique environment and measurement error without knowledge of test–retest correlations (Plomin & Daniels, 1987; Plomin et al., 2001b). This is because if there is noise in the elicitation of preferences,

such noise will be subsumed under the estimate of unique environmental effects.[18] Further, a number of important sources of unique environmental effects, such as accidents, are nonsystematic in nature. The observation that the human genome could not possibly specify every synaptic connection in the brain and that random events could lead to different developmental outcomes, even in genetically identical individuals, falls into this category (Molenaar, Boomsma & Dolan, 1993; Jensen, 1997).

Economists have traditionally expressed agnosticism about the causal mechanisms behind individual differences in preferences. Although choosing to overlook genetic explanations is often well motivated on the grounds of parsimony, especially in studies taking a historical or geographical perspective, our findings, combined with the preexisting behavior genetics literature, uncover a unique and potentially important source of preference heterogeneity. Despite ample experimental evidence, the origins of individual behavioral variation in economic games have thus far remained elusive, and many attempts to find theoretically appealing and empirically stable correlates to preferences elicited experimentally have yielded contradictory results (Camerer, 2003). If preferences are indeed under moderate genetic influence, any attempt to understand heterogeneity in preferences without taking this into account will be incomplete.

Recently, much interest has been directed toward finding biological or neurological correlates of experimental behavior. Of course, this does not necessarily imply either causality or a genetically mediated association. However, the fact that many of the biological variables with known associations to individual differences in strategies or preferences are strongly heritable does lend some support, if only circumstantial, to our findings. For instance, financial risk taking has been claimed to vary over the menstrual cycle in women (Bröder & Hohmann, 2003; Chen, Katuscak & Ozdenoren, 2005) and correlates both with facial masculinity and with circulating testosterone levels in men (Apicella et al., 2008). A number of imaging studies have also explored the neural correlates of both giving and financial risk taking. One study found activation in the striatum both on receiving money and on donating to charity (Moll et al., 2006). Another study found similar activation patterns and demonstrated enhanced activation when the charitable donation was voluntary (Harbaugh, Mayr & Burghart, 2007). In the context of financial risk taking, Kuhnen and Knutson (2005) demonstrated that risk-seeking is associated with activation in the nucleus accumbens, whereas risk aversion is associated with activation in the insula. In general, brain structure is under strong genetic influence, though there are substantial regional differences in heritability

[18] This result also has implications for the genome-wide association studies that are currently under way, examining genetic variation across the human genome and behavior in experimental games. Noise in the elicitation of, for instance, social preferences is likely to frustrate these efforts. Multiple measurement would be one way of dealing with the problem.

(Thompson et al., 2001; Toga & Thompson, 2005). The same is true for hormone levels (Harris, Vernon & Boomsma, 1998; Bartels et al., 2003).

6. Conclusions

In this paper, we have presented an empirical investigation into the relative contributions of individual differences in genes and environment to observed variation in economic preferences for risk and giving. Notwithstanding the fact that all twin siblings are of the same age and were raised together in the same family, genetically identical MZ twins still exhibit much greater similarity in their preferences for risk and giving than do DZ twins. Although our results do not allow us to be as assertive as Sir Francis Galton, they do suggest that humans are endowed with genetic variation in their proclivity to donate money to charity and to take risks. By now there are a plethora of studies exploring the sources of individual variation in economic experiments and games, yet up until recently considerations of genetic influences have remained relatively absent. Here we have argued that this failure to consider genes obscures an important source of preference heterogeneity. Ultimately, we hope that a better understanding of the underlying individual genetic heterogeneity[19] in economic preferences and the adaptive pressures under which these preferences evolved will lead to a more comprehensive economic science that can bridge some of the unexplained gaps between empirical data and economic theory (Cosmides & Tooby, 1994; Burnham, 1997).

Finally, our findings suggest a number of directions for future research. In recent years we have witnessed rapid advancement in the field of molecular genetics, including the initial tentative steps toward uncovering the complex genetic architecture underlying variation in individual personality and preferences. In fact, we are aware of one paper that has already uncovered a polymorphism on the AVPR1a gene that is associated with generosity in the dictator game (Knafo et al., 2008). Two recent papers also report that carriers of the 7R allele on the Dopamine Receptor D4 gene (DRD4) take greater financial risks in laboratory experiments (Dreber et al., 2009; Kuhnen & Chiao, 2009). The identification of specific genes, or more likely combinations of genes, associated with particular traits holds promise for economic research. Most importantly, as noted by Benjamin et al. (2007), it will allow the study of interactions between genotypes and policies to better predict the consequences of policy for individuals. A second direction for future research is to look beyond the laboratory and instead consider field proxies for the underlying preferences. There are well-known issues associated with the generalizability of laboratory findings (Levitt & List, 2007), and documenting similar

[19] Genetic variation can be maintained in equilibrium for a number of reasons. For a discussion of this difficult subject in the context of personality differences, see two recent papers by Dall, Houston and McNamara (2004) and Penke, Denissen and Miller (2007).

genetic influences in the field therefore ought to be a priority. A third, and perhaps most natural, direction is to try to disentangle additive and nonadditive genetic variation. We anticipate that studies employing the extended family design will shed more light on this issue. The fairly low DZ correlations we observe provide some tentative, but far from conclusive, evidence for nonadditivity.

References

Alford, J. R., C. L. Funk, and J. R. Hibbing, "Are Political Orientations Genetically Transmitted?," *American Political Science Review* 99 (2005), 153-167.

Apicella, C. L., A. Dreber Almenberg, B. Campbell, P. Gray, M. Hoffman, and A. C. Little, "Testosterone and Financial Risk-Taking," *Evolution and Human Behavior* 29 (2008), 384–390.

Bardsley, N., "Dictator Game Giving: Altruism or Artefact?," *Experimental Economics* 11 (2008), 122–133.

Bartels, M., S. M. Van den Berg, F. Sluyter, D. I. Boomsmaa, and E. J. C. de Geus, "Heritability of Cortisol Levels: Review and Simultaneous Analysis of Twin Studies," *Psychoneuroendocrinology* 28 (2003), 121–137.

Behrman, J. R., and P. Taubman, "Is Schooling Mostly in the Genes? Nature–Nurture Decomposition Using Data on Relatives," *Journal of Political Economy* 97 (1989), 1425–1446.

Benjamin, D. J., C. F. Chabris, E. L. Glaeser, V. Gudnason, T. B. Harris, D. Laibson, L. Launer, and S. Purcell, "Genoeconomics," in M. Weinstein, J. W. Vaupel, and K. W. Wachter (Eds), *Biosocial Surveys* (Washington, DC: National Academies Press, 2007).

Björklund, A., M. Jäntti, and G. Solon, "Influences of Nature and Nurture on Earnings Variation: A Report on a Study of Various Sibling Types in Sweden," in S. Bowles, H. Gintis, and M. Osborne Groves (Eds), *Unequal Chances: Family Background and Economic Success* (Princeton, NJ: Princeton University Press, 2005).

Björklund, A., M. Jäntti, and G. Solon, "Nature and Nurture in the Intergenerational Transmission of Socioeconomic Status: Evidence from Swedish Children and Their Biological and Rearing Parents," *Advances in Economic Analysis and Policy* 7 (2007), 1753–1753.

Björklund, A., M. Lindahl, and E. Plug, "The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data," *Quarterly Journal of Economics* 121 (2006), 999–1028.

Bouchard, T. J. Jr., "Genetic and Environmental Influences on Adult Intelligence and Special Mental Abilities," *Human Biology* 70 (1998), 257–279.

Bouchard, T. J. Jr., D. T. Lykken, M. McGue, N. L. Segal, and A. Tellegen, "Sources of Human Psychological Differences: The Minnesota Study of Twins Reared Apart," *Science* 250 (1990), 223–228.

Bouchard, T. J. Jr., and M. McGue, "Genetic and Environmental Influences on Human Psychological Differences," *Journal of Neurobiology* 54 (2003), 4–45.

Bouchard, T. J. Jr., M. McGue, D. T. Lykken, and A. Tellegen, "Intrinsic and Extrinsic Religiousness: Genetic and Environmental Influences and Personality Correlates," *Twin Research* 2 (1999), 88–98.

Bouchard T. J. Jr., and P. Propping (Eds), *Twins as a Tool of Behavioral Genetics: Report of the Dahlme Workshop on What Are the Mechanisms Mediating the Genetic and Environmental Determinants of Behavior?* (Chichester, UK: John Wiley, 1993).

Bowles, S., H. Gintis, and M. Osborne Groves (Eds), *Unequal Chances: Family Background and Economic Success* (Princeton, NJ: Princeton University Press, 2005).

Bröder, A., and N. Hohmann, "Variations in Risk Taking Behavior over the Menstrual Cycle: An Improved Replication," *Evolution and Human Behavior* 24 (2003), 391–398.

Brooks, S. P., and A. Gelman, "General Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational and Graphical Statistics* 7 (1998), 434–455.

Brosig, J., T. Riechmann, and J. Weimann, "Selfish in the End? An Investigation of Consistency and Stability of Individual Behavior," MPRA Paper 2035, University Library of Munich (2007).

Burnham, T. C., *Essays on Genetic Evolution and Economics*, Ph.D. thesis, Harvard University (1997).

Burton, P. R., K. J. Tiller, L. C. Gurrin, W. O. C. M. Cookson, Q. A. W. Musk, and L. J. Palmer, "Genetic Variance Components Analysis for Binary Phenotypes Using Generalized Linear Mixed Models (GLMMS) and Gibbs Sampling," *Genetic Epidemiology* 17 (1999), 118–140.

Camerer, C. F., *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton, NJ: Princeton University Press, 2003).

Cesarini, D., C. T. Dawes, J. H. Fowler, M. Johannesson, P. Lichtenstein, and B. Wallace, "Heritability of Cooperative Behavior in the Trust Game," *Proceedings of the National Academy of Sciences* 104 (2008), 15631–15634.

Chen,Y., P. Katuscak, and E. Ozdenoren, "Why Can't a Woman Bid More Like a Man?," Mimeo, University of Michigan (2005).

Cipriani, M., P. Giuliani, and O. Jeanne, "Like Mother Like Son? Experimental Evidence on the Transmission of Values from Parents to Children," IZA Discussion Paper No. 2768 (2007).

Cosmides, L., and J. Tooby, "Better Than Rational: Evolutionary Psychology and the Invisible Hand," *American Economic Review* 84 (1994), 327–332.

Coventry, W. L., and M. C. Keller, "Estimating the Extent of Parameter Bias in the Classical Twin Design: A Comparison of Parameter Estimates from Extended Twin–Family and Classical Twin Designs," *Twin Research and Human Genetics* 8 (2005), 214–223.

Dall, S. R. X., A. I. Houston, and J. M. McNamara, "The Behavioral Ecology of Personality: Consistent Individual Differences from an Adaptive Perspective," *Ecology Letters* 7 (2004), 734–739.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde, "The Intergenerational Transmission of Risk and Trust Attitudes," IZA Discussion Paper No. 2380 (2006).

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner, "Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey," IZA Discussion Paper No. 1730 (2005).

Dreber, A., C. L. Apicella, D. T. A. Eisenberg, J. R. Garcia, R. Zamore, J. K. Lum, and B. C. Campbell, "The 7R Polymorphism in the Dopamine Receptor D4 Gene (DRD4) Is Associated with Financial Risk-Taking in Men," *Evolution and Human Behavior* 30 (2009), 85–92.

Eckel, C. C., and P. J. Grossman, "Altruism in Anonymous Dictator Games," *Games and Economic Behavior* 16 (1996), 181–191.

Fisher, Ronald A., *The Genetical Theory of Natural Selection* (Oxford, UK: Oxford University Press, 1930).

Fong, C. M., "Evidence from an Experiment on Charity to Welfare Recipients: Reciprocity, Altruism and the Empathic Responsiveness Hypothesis," *Economic Journal* 117 (2007), 1008–1024.

Forsythe, Robert, Joel L. Horowitz, N. E. Savin, and Martin Sefton, "Fairness in Simple Bargaining Experiments," *Games and Economic Behavior* 6 (1994), 347–369.

Fowler, J. H., C. T. Dawes, and L. Baker, "Genetic Variation in Political Participation," *American Political Science Review* 2 (2008), 233–248.

Galton, Francis, "The History of Twins, as a Criterion of the Relative Powers of Nature and Nurture," *Fraser's Magazine* 12 (1875), 566–576.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis* (New York, NY: Chapman & Hall/CRC, 2004).

Goldberger, A. S., "Twin Methods: A Skeptical View," in P. Taubman (Ed.), *Kinometrics: Determinants of Socioeconomic Success within and between Families* (Amsterdam: North-Holland, 1977).

Goldberger, A. S., "Heritability," *Economica* 46 (1979), 327–347.

Hamilton, W. D., "The Genetical Evolution of Social Behaviour I and II," *Journal of Theoretical Biology* 7 (1964), 1–52.

Harbaugh, W. T., U. Mayr, and D. R. Burghart, "Neural Responses to Taxation and Voluntary Giving Reveal Motives for Charitable Donations," *Science* 316 (2007), 1622–1625.

Harless, D. W., and C. F. Camerer, "The Predictive Utility of Generalized Expected Utility Theories," *Econometrica* 62 (1994), 1251–1289.

Harris, J. A., P. A. Vernon, and D. I. Boomsma, "The Heritability of Testosterone: A Study of Dutch Adolescent Twins and Their Parents," *Behavior Genetics* 28 (1998), 165–171.

Hey, J. D., and C. Orme, "Investigating Generalizations of Expected Utility Theory Using Experimental Data," *Econometrica* 62 (1994), 1291–1326.

Holt, C. A., and S. K. Laury, "Risk Aversion and Incentive Effects," *American Economic Review* 92 (2002), 1644–1655.

Jang, K. L., W. J. Livesley, and P. A. Vernon, "Heritability of the Big Five Personality Dimensions and Their Facets: A Twin Study," *Journal of Personality* 64 (1996), 577–592.

Jensen, A. R., "The Puzzle of Nongenetic Variance," in R. J. Sternberg, and E. L. Grigorenko (Eds), *Heredity, Intelligence, and Environment* (Cambridge, UK: Cambridge University Press, 1997).

Keller, M. C., and W. L. Coventry, "Quantifying and Addressing Parameter Indeterminacy in the Classical Twin Design," *Twin Research and Human Genetics* 8 (2005), 201–213.

Kirk, K. M., H. H. Maes, M. C. Neal, A. C. Heath, N. G. Martin, and L. J. Eaves, "Frequency of Church Attendance in Australia and the United States: Models of Family Resemblance," *Twin Research* 2 (1999), 99–107.

Knafo, A., S. Israel, A. Darvasi, R. Bachner-Melman, F. Uzefovsky, L. Cohen, E. Feldman, E. Lerer, E. Laiba, Y. Raz, L. Nemanov, I. Gritsenko, C. Dina, G. Agam, B. Dean, G. Bornstein, and R. P. Ebstein, "Individual Differences in Allocation of Funds in the Dictator Game Associated with Length of the Arginine Vasopressin 1a Receptor (AVPR1a) RS3 Promoter Region and Correlation between RS3 Length and Hippocampal mRNA," *Genes, Brain and Behavior* 7 (2008), 266–275.

Koenig, L. B., M. McGue, R. F. Krueger, and T. J. Bouchard Jr., "Genetic and Environmental Influences on Religiousness: Findings for Retrospective and Current Religiousness Ratings," *Journal of Personality* 73 (2005), 471–488.

Kuhnen, C. M., and J. Y. Chiao, "Genetic Determinants of Financial Risk Taking," *PLoS ONE* 4 (2009).

Kuhnen, C. M., and B. Knutson, "The Neural Basis of Financial Risk Taking," *Neuron* 47 (2005), 763–770.

Levitt, S. D., and J. A. List, "What do Laboratory Experiments Measuring Social Preferences Tell Us about the Real World," *Journal of Economic Perspectives* 21 (2007), 153–174.

Liang, K.-Y., and S. L. Zeger, "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika* 73 (1986), 13–22.

Lichtenstein, P., P. F. Sullivan, S. Cnattingius, M. Gatz, S. Johansson, E. Carlström, C. Björk, M. Svartengren, A. Volk, L. Klareskog, U. de Faire, M. Schalling, J. Palmgren, and N. L. Pedersen, "The Swedish Twin Registry in the Third Millennium: An Update," *Twin Research and Human Genetics* 9 (2006), 875–882.

List, J. A. "On the Interpretation of Giving in Dictator Games," *Journal of Political Economy* 115 (2007), 482–494.

Loehlin, J. C., "Some Methodological Problems in Cattel's Multiple Abstract Variance Analysis," *Psychological Review* 72 (1965), 156–161.

Loehlin, J. C., "Resemblance in Personality and Attitudes between Parents and Their Children: Genetic and Environmental Contributions," in S. Bowles, H. Gintis, and M. Osborne Groves (Eds), *Unequal Chances: Family Background and Economic Success* (Princeton, NJ: Princeton University Press, 2005).

Loh, C. Y., and J. M. Elliot, "Cooperation and Competition as a Function of Zygosity in 7- to 9-Year-Old Twins," *Evolution and Human Behavior* 19 (1998), 397–411.

Lykken, D. T., M. K. McGue, and A. Tellegen, "Recruitment Bias in Twin Research: The Rule of Two-Thirds Reconsidered," Behavior Genetics 17 (1986), 343–362.

Martin, N. G., L. J. Eaves, A. C. Heath, R. Jardine, L. M. Feingold, and H. J. Eysenck, "Transmission of Social Attitudes," *Proceedings of the National Academy of Sciences* 83 (1986), 4364–4368.

Mather, K., and J. L. Jinks, *An Introduction to Biometrical Genetics* (London: Chapman and Hall, 1977).

Molenaar, P. C. M., D. I. Boomsma, and C. V. Dolan, "A Third Source of Developmental Differences," *Behavior Genetics* 23 (1993), 519–524.

Moll, J., F. Krueger, R. Zahn, M. Pardini, R. de Oliveira-Souza, and J. Grafman, "Human Fronto-Mesolimbic Networks Guide Decisions about Charitable Donation," *Proceedings of the National Academy of Sciences* 103 (2006), 15623–15628.

Muthén, L. K., and B. O. Muthén, *Mplus. Statistical Analysis with Latent Variables. User's Guide*, Version 4.1 (Los Angeles, CA: 2006).

Neale, M. C., and H. H. M. Maes, *Methodology for Genetic Studies of Twins and Families* (Dordrecht, NL: Kluwer Academic, 2004).

Neisser, U., G. Boodoo, T. J. Bouchard Jr., A.W. Boykin, N. Brody, S. J. Ceci, D. F. Halpern, J. C. Loehlin, R. Perloff, R. J. Sternberg, and S. Urbina, "Intelligence: Knowns and Unknowns," *American Psychologist* 51 (1996), 77–101.

Pam, A., S. S. Kemker, C. A. Ross, and R. Golden, "The Equal Environments Assumption in MZ–DZ Twin Comparisons: An Untenable Premise of Psychiatric Genetics?" *Acta Geneteticae Medicae Gemellologiae (Roma)* 45 (1996), 349–360.

Penke, L., J. J. A. Denissen, and G. F. Miller, "The Evolutionary Genetics of Personality," *European Journal of Personality* 21 (2007), 549–587.

Plomin, R., K. Asbury, P. G. Dip, and J. Dunn, "Why Are Children in the Same Family So Different? Unshared Environment a Decade Later," *Canadian Journal of Psychiatry* 46 (2001b), 225–233.

Plomin, R. D., and D. Daniels, "Why Are Children in the Same Family So Different from Each Other?," *Behavioral and Brain Sciences* 10 (1987), 1–16.

Plomin, R. D., J. C. DeFries, G. E. McClearn, and P. McGuffin, *Behavioral Genetics*, 4th ed. (New York, NY: Freeman, 2001a).

Plug, E., and W. Vijverberg, "Schooling, Family Background, and Adoption: Is It Nature or Is It Nurture?," *Journal of Political Economy* 111 (2003), 611–641.

Posner, S. F., L. Baker, A. Heath, and N. G. Martin, "Social Contact and Attitude Similarity in Australian Twins," *Behavior Genetics* 26 (1996), 123–133.

Rushton, J. P., "Genetic and Environmental Contributions to Pro-social Attitudes: A Twin Study of Social Responsibility," *Proceedings of the Royal Society B* 271 (2004), 2583–2585.

Rushton, J. P., D. W. Fulker, M. C. Neale, D. K. B. Nias, and H. J. Eysenck, "Altruism and Aggression: The Heritability of Individual Differences," *Journal of Personality and Social Psychology* 50 (1986), 1192–1198.

Sacerdote, B., "The Nature and Nurture of Economic Outcomes," *American Economic Review* 92 (2002), 344–348.

Sacerdote, B., "How Large Are the Effects from Changes in Family Environment? A Study of Korean American Adoptees," *Quarterly Journal of Economics* 122 (2007), 119–157.

Sala-i-Martin, X., "The World Distribution of Income: Falling Poverty and Covergence, Period," *Quarterly Journal of Economics* 121 (2006), 351–397.

Scurrah, K. J., L. J. Palmer, and P. R. Burton, "Variance Components Analysis for Pedigree-Based Censored Survival Data Using Generalized Linear Mixed Models (GLMMs) and Gibbs Sampling in BUGS," *Genetic Epidemiology* 19 (2000), 127–148.

Segal, N. L., and S. L. Hershberger, "Cooperation and Competition between Twins: Findings from a Prisoner's Dilemma Game," *Evolution and Human Behavior* 20 (1999), 29–51.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde, "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 64 (2002), 583–639.

Stoel, R. D., E. J. C. De Geus, and D. I. Boomsma, "Genetic Analysis of Sensation Seeking with an Extended Twin Design," *Behavior Genetics* 36 (2006), 229–237.

Taubman, P., "The Determinants of Earnings: Genetics, Family, and Other Environments: A Study of White Male Twins," *American Economic Review* 66 (1976), 858–870.

Thompson, P., T. D. Cannon, K. L. Narr, T. van Erp, V.-P. Poutanen, M. Huttunen, J. Lönnqvist, C.-G. Standertskjöld-Nordenstam, J. Kaprio, M. Khaledy, R. Dail, C. I. Zoumalan, and A. W. Toga, "Genetic Influences on Brain Structure," *Nature Neuroscience* 4 (2001), 1–6.

Toga, A. W., and P. M. Thompson, "Genetics of Brain Structure and Intelligence," *Annual Review of Neuroscience* 28 (2005), 1–23.

True, W. R., A. C. Heath, J. F. Scherrer, B. Waterman, J. Goldberg, N. L., S. A. Eisen, M. J. Lyons, and M. T. Tsuang, "Genetic and Environmental Contributions to Smoking," *Addiction* 92 (1997), 1277–1288.

Turkheimer, E., "Three Laws of Behavior Genetics and What They Mean," *Current Directions in Psychological Science* 9 (2000), 160–164.

Turkheimer, E., A. Haley, B. D'Onofrio, M. Waldron, and I. I. Gottesman, "Socioeconomic Status Modifies Heritability of IQ in Young Children," *Psychological Science* 14 (2003), 623–628.

Van den Berg, S. M., L. Beem, and D. I. Boomsma, "Fitting Genetic Models Using Markov Chain Monte Carlo Algorithms with Bugs," *Twin Research and Human Genetics* 9 (2006), 334–342.

Wallace, B., D. Cesarini, P. Lichtenstein, and M. Johannesson, "Heritability of Ultimatum Game Responder Behavior," *Proceedings of the National Academy of Sciences* 104 (2007), 15631–15634.

Appendix

In this appendix, we provide some details on the Bayesian estimation procedure, additional information on recruitment bias, and the results for our two additional measures of risk preferences (referred to in the text as risk investment and risk assessment).

6.1. Details on estimation.

6.1.1. *Ordered models.* In the ordered models, the outcome variables are modelled under the assumption that y_{ij}^* is not directly observed. Instead, the observed variable y_{ij} is assumed to be one of $k+1$ ordered categories (0 to k):

$$\begin{aligned} y_{ij} &= 0 \;\; if \;\; y_{ij}^* \leq \tau_1, \\ y_{ij} &= 1 \;\; if \;\; \tau_1 < y_{ij}^* \leq \tau_2, \\ &\ldots \\ y_{ij} &= k \;\; if \;\; y_{ij}^* > \tau_k, \end{aligned}$$

where τ_i is an unknown threshold parameter that is estimated as part of the model. For MZ twins, the probability of observing an outcome is given by:

$$\begin{aligned} P(y_{ij} &= 0|A_i, C_i, X_i) = \Phi\left(\tau_1 - (A_i + C_i)\right), \\ P(y_{ij} &= 1|A_i, C_i, X_i) = \Phi\left(\tau_2 - (A_i + C_i)\right) - \Phi\left(\tau_1 - (A_i + C_i)\right), \\ &\ldots \\ P(y_{ij} &= k|A_i, C_i, X_i) = 1 - \Phi\left(\tau_k - (A_i + C_i)\right), \\ 0 &< \tau_1 < \ldots < \tau_k. \end{aligned}$$

where Φ is the cumulative standard normal distribution. For DZ twins, the probability is:

$$\begin{aligned} P(y_{ij} &= 0|A_{1i}, A_{2ij}, C_i, X_i) = \Phi\left(\tau_1 - (A_{1i} + A_{2ij} + C_i)\right), \\ P(y_{ij} &= 1|A_{1i}, A_{2ij}, C_i, X_i) = \Phi\left(\tau_2 - (A_{1i} + A_{2ij} + C_i)\right) - \Phi\left(\tau_1 - (A_{1i} + A_{2ij} + C_i)\right), \\ &\ldots \\ P(y_{ij} &= k|A_{1i}, A_{2ij}, C_i, X_i) = 1 - \Phi\left(\tau_k - (A_{1i} + A_{2ij} + C_i)\right), \\ 0 &< \tau_1 < \ldots < \tau_k. \end{aligned}$$

6.1.2. *ADE model.* In the ADE model, we assume that,

$$\chi_{ij}^{MZ} = A_i + D_i + E_{ij},$$

where A_i is the family genetic factor, D_i is the dominance deviation and E_{ij} is the individually-experienced unshared environment factor. For DZ twins the latent variable is a function of four random effects variables:

$$\chi_{ij}^{DZ} = A_{1i} + A_{2ij} + D_{1i} + D_{2ij} + E_{ij},$$

In order to model a correlation of 0.25 in the DZ twins for the nonadditive (dominance) genetic effects we split up the dominance component, σ_D^2, into two independent parts, and assume that,

$$\begin{aligned} D &\sim N\left(0, \sigma_D^2\right), \\ D_1 &\sim N\left(0, \frac{1}{4}\sigma_D^2\right), \\ D_2 &\sim N\left(0, \frac{3}{4}\sigma_D^2\right). \end{aligned}$$

For the precision parameter associated with σ_D^2, we use a Pareto distribution with shape parameter equal to 1 and scale parameter equal to 0.001.

6.2. Representativeness. In Table A1, we compare our participants to the STA GE cohort as a whole on a number of background variables. The STAGE cohort is very large, so it is important to distinguish statistical significance from practical significance. For health, income and employment status, we find no significant differences. We do however find that our subjects are somewhat younger than the average STAGE respondent. The difference is approximately 3.5 years for men and 1.5 years for women. We also find that participants in the experiment are less likely to be unemployed. In our experimental sample, the unemployment rate is two percentage points lower than in STAGE for women, and four percentage points lower for men. Further, marriage rates are somewhat lower, a phenomenon which is no doubt related to their lower average age. In particular, 22% of our participating men are married, as compared to 29% in STAGE. The corresponding figures for women are 28% and 33%. Participants in the experiments also, on average, have 0.25 fewer children under 18 living in their household.

While the 61% response rate in STAGE is not alarmingly low, it merits further investigation, because STAGE respondents themselves may not be fully representative of the general population.

Table A1. Comparison of experimental sample and STAGE cohort[20]

	Men			Women		
	Sample	STAGE.	p-value	Sample	STAGE.	p-value
Age	33.03	36.66	< 0.01	35.29	36.57	< 0.01
Education	13.69	12.50	< 0.01	13.67	12.78	< 0.01
Income	243,524	269,764	0.11	196,591	195,289	0.84
Employed full time	0.72	0.78	0.24	0.52	0.54	0.46
Unemployed	0.02	0.06	0.02	0.04	0.06	< 0.01
Self-employed	0.09	0.14	0.07	0.04	0.05	0.23
On sickleave	0.02	0.02	0.93	0.03	0.04	0.43
Gov. employee	0.28	0.22	0.10	0.44	0.50	0.02
Health	1.74	1.85	0.11	1.92	1.96	0.25
Marital status	0.22	0.29	0.03	0.28	0.33	0.02
No. of children	0.60	0.81	0.02	0.75	1.01	< 0.01

In private correspondence with the Swedish Twin Registry, we have learnt that there are no significant differences between participants and non-participants with respect to age or birthweight. As is common in twin studies, women are overrepresented (Lykken, McGue & Tellegen, 1980) also in STAGE, with a larger fraction of non-participants being male (58% versus 44%).

Table A2. Correlations in experimental sample and STAGE cohort[21]

	Exp. sample		STAGE	
	MZ	DZ	MZ	DZ
Education	0.68	0.43	0.69	0.45
Income	0.69	0.58	0.59	0.45
Employed full time	0.25	0.21	0.28	0.29
Self-employed	0.30	0.30	0.33	0.25
On sickleave	-0.04	-0.02	0.18	0.06
Gov. employee	0.35	0.21	0.29	0.25
Health	0.46	-0.04	0.33	0.17
Marital status	0.33	0.33	0.39	0.27
No. of children	0.51	0.44	0.53	0.38

[20] Education refers to years of education. Income is the sum of wage income, taxable transfers and income from own company for the year 2005 (in SEK). Employment information was gathered when the subject responded to the STAGE questionnaire. Health is self-reported on a scale from 1 to 5. Marital status is a dummy variable taking the value 1 if the subject is married. Number of children is number of children under 18 living in the respondent's household in the year 2005. We utilized adjusted Wald tests for equality taking into account non-independence within twin families (Liang & Zeger, 1986).

[21] Education refers to years of education. Income is the sum of wage income, taxable transfers and income from own company for the year 2005 (in SEK). Employment information was gathered when the subject responded to the STAGE questionnaire. Health is self-reported on a scale from 1 to 5.

Non-participants are also more likely to be diagnosed with a psychological disorder (4.4% versus 7.7%) or to have at least one parent born outside Sweden (16.1% versus 12.8%). Participants on the other hand are more likely to have studied after high-school (41% versus 27%).

In Table A2, we report MZ and DZ correlations on a large number of background variables for the STAGE cohort as a whole and for our experimental sample. In general, there is no tendency for the patterns of correlations to differ between the samples.

6.2.1. *Data definitions.* The data from Statistics Sweden is for the year 2005 and includes income excluding capital income (förvärvsinkomst), marital status and years of education. Unlike the STAGE data, the data from Statistics Sweden is not self-reported but registry based.

Researchers interested in the variables in STAGE are advised to contact the Swedish Registry, which maintains a web-based (but password protected) database with variable definitions.

6.3. Additional results for risk.

Table A3. Results of the ACE model and its nested submodels, risk investment[22]

		ACE	AE	CE	E
Continuous	A	0.19	0.29	–	–
		(0.01, 0.34)	(0.20, 0.39)	–	–
	C	0.10	–	0.24	–
		(0.00, 0.26)	–	(0.15, 0.33)	–
	E	0.71	0.71	0.76	1.00
		(0.62, 0.81)	(0.62, 0.80)	(0.67, 0.85)	(1.00 – 1.00)
Dbar		3,683	3,670	3,734	3,988
pD		216.4	224.5	180.9	2.0
DIC		3,900	3,894	3,915	3,990
Ordered	A	0.22	0.32	–	–
		(0.02, 0.38)	(0.21, 0.42)	–	–
	C	0.10	–	0.26	–
		(0.01, 0 .27)	–	(0.17, 0.35)	–
	E	0.68	0.68	0.74	1.00
		(0.59, 0.79)	(0.58, 0.80)	(0.65, 0.83)	(1.00 – 1.00)
Dbar		2,375	2,367	2,431	2,677
pD		221.7	226.8	182.1	4.99
DIC		2,597	2,593	2,614	2,682

Marital status is a dummy variable taking the value 1 if the subject is married. Number of children is number of children under 18 living in the respondent's household in the year 2005.

[22] A is the genetic contribution; C is the common environment contribution; E is the unique environment contribution. *Dbar*: Deviance. *pD*: Effective number of parameters. DIC: Bayesian deviance information criterion. 95% credible intervals within parentheses.

In Tables A3 and A4 we report ACE model results for risk investment and risk assessment. Since the correlations we observe for risk assessment are significantly outside the permissible space of correlations, we also estimate an ADE model for risk assessment, see Table A5. The DIC model selection criterion suggests that the ADE model better fits the data.

Table A4. Results of the ACE model and its nested submodels, risk assessment[23]

		ACE	AE	CE	E
Continuous	A	0.29	0.35	–	–
		(0.14, 0.41)	(0.25, 0.44)	–	–
	C	0.05	–	0.25	–
		(0.00, 0.17)	–	(0.17, 0.34)	–
	E	0.65	0.65	0.75	1.00
		(0.56, 0.75)	(0.56, 0.75)	(0.66, 0.84)	(1.00 – 1.00)
Dbar		3,466	3,455	3,578	3,844
pD		253.5	257.9	187.1	2.00
DIC		3,719	3,713	3,765	3,846
Ordered	A	0.33	0.38	–	–
		(0.19, 0.45)	(0.28, 0.48)	–	–
	C	0.05	–	0.28	–
		(0.00, 0.17)	–	(0.19, 0.36)	–
	E	0.62	0.62	0.72	1.00
		(0.53, 0.72)	(0.53, 0.72)	(0.64, 0.81)	(1.00 – 1.00)
Dbar		3,474	3,471	3,604	3,877
pD		279.7	279.4	204.7	9.86
DIC		3,753	3,751	3,809	3,897

[23] A is the genetic contribution; C is the common environment contribution; E is the unique environment contribution. *Dbar*: Deviance. *pD*: Effective number of parameters. DIC: Bayesian deviance information criterion. 95% credible intervals within parentheses.

Table A5. Results of the ADE model and its nested submodels, risk assessment[24]

		ADE
Continuous	A	0.05
		$(0.00 - 0.14)$
	D	0.33
		$(0.19 - 0.44)$
	E	0.63
		$(0.54 - 0.73)$
$Dbar$		$3,424$
pD		275.8
DIC		$3,700$
Ordered	A	0.04
		$(0.22 - 0.48)$
	D	0.37
		$(0.22 - 0.48)$
	E	0.59
		$(0.00 - 0.15)$
$Dbar$		$3,432$
pD		301.1
DIC		$3,733$

[24] A is the genetic contribution; D is the dominance deviation; E is the unique environment contribution. $Dbar$: Deviance. pD: Effective number of parameters. DIC: Bayesian deviance information criterion. 95% credible intervals within parentheses.

PAPER 3

Genetic variation in financial decision-making

David Cesarini Magnus Johannesson Paul Lichtenstein
Örjan Sandewall Björn Wallace

Abstract. Individuals differ in how they construct their investment portfolios, yet empirical models of portfolio risk typically account only for a small portion of the cross-sectional variance. This paper asks whether genetic variation can explain some of these individual differences. Following a major pension reform Swedish adults had to form a portfolio from a large menu of funds. We match data on these investment decisions with the Swedish Twin Registry and find that approximately 25% of individual variation in portfolio risk is due to genetic variation. We also find that these results extend to several other aspects of financial decision-making.

1. Introduction

It is well known that the composition of investment portfolios varies substantially across individuals, yet the determinants of these individual differences are not fully understood (Guiso, Haliassos & Jappelli, 2002; Curcuru et al., 2009). Explaining investor heterogeneity is relevant to a number of prominent but unsettled debates in finance. Importantly, a better understanding of the determinants of cross-sectional variance leads to facts that theories of portfolio allocation have to be consistent with. For these reasons there is a voluminous literature that tries to understand heterogeneity in portfolio composition, especially as it pertains to individuals' willingness to bear financial risks.[1]

A slightly different version of this paper was published as: Cesarini, D., M. Johannesson, P. Lichtenstein, Ö. Sandewall, and B. Wallace, "Genetic Variation in Financial Decision-Making," *Journal of Finance* 65 (2009), 1725-1754. We are grateful to two anonymous reviewers and the editors for detailed comments on an earlier draft of this manuscript. We also thank Malcolm Baker, Jonathan Beauchamp, Dan Benjamin, Terry Burnham, Magnus Dahlquist, Chris Dawes, Peter Diamond, Esther Duflo, Ernst Fehr, James Fowler, Glenn Harrison, Magnus Henrekson, Sandy Jencks, David Laibson, Matthew Notowidigdo, Steve Pischke, James Poterba, Drazen Prelec, Andrei Shleifer, and Per Strömberg for helpful discussions. IFN provided a hospitable environment in which a first draft of this paper was written. The usual caveat applies. An Internet Appendix containing sample code for the behavior genetic models is available at http://www.afajof.org/supplements.asp.

[1] See Cohn et al. (1975), Friend and Blume (1975), Morin and Suarez (1983), Poterba (1993), Bertaut (1994), Pålsson (1996), Heaton and Lucas (1997, 2000), Guiso, Haliassos and Bertaut (2002), Malmendier and Nagel (2007), Palme, Sundén and Söderlind (2007), and Curcuru et al. (2009).

This paper asks whether genetic variation can help explain heterogeneity in portfolio risk. In particular, it builds on a recent string of papers (Wallace et al., 2007; Cesarini et al., 2008a, 2009) wherein laboratory experiments designed to elicit economic preferences, including risk, are run on a sample of twins. The comparison of the behavior of monozygotic (MZ, also known as identical) twins to that of dizygotic (DZ, also known as fraternal) twins is a form of quasi-experiment. MZ twins reared together share both their family environment and their genes. In contrast, DZ twins reared together share their family environment, but their degree of genetic relatedness is no greater than that of ordinary siblings. A significantly higher observed correlation for MZ twins than for DZ twins is therefore usually taken as evidence of genetic variance in the traits studied. Indeed, this is what Wallace et al. (2007) and Cesarini et al. (2008a, 2009) find for economic preferences in their sample of twins. Further, by fitting their experimental data to standard behavior genetic models, they estimate that heritability—the share of individual variation that can be explained by genetic differences—typically is somewhere between 20% and 40%.

Eliciting risk preferences experimentally has at least two disadvantages. First, there is uncertainty about the extent to which laboratory behavior generalizes to the field (Harrison & List, 2004; Levitt & List, 2007). Second, the sample sizes in the twin studies cited above, though large by the standards of behavioral economics, still do not allow for precise inference. In this paper, we use microdata from the Swedish individualized pension savings accounts introduced in 2000 to extend the previous literature from the laboratory to the field. As part of the transition to a new pension system, virtually all adult Swedes born after 1938 had to make simultaneous investment decisions with potentially far-reaching effects on their post-retirement wealth. In particular, they had to construct an investment portfolio from a menu of almost 500 funds. We take this event, sometimes referred to as the "Big Bang" of the Swedish financial sector (Palme, Sundén & Söderlind, 2007), as a field experiment from which we can infer attitudes toward financial risk. Matching individual portfolio data to the Swedish Twin Registry, we then employ standard methods from behavior genetics and estimate the heritability of portfolio risk. Unlike small-stake gambles in the laboratory or attitudinal risk questions, the investment decisions made in the pension savings accounts can have major economic consequences.[2]

Since our data set is very large, we are able to estimate parameters with much greater precision than previous studies, which have all been based on experimental data. To our knowledge, this article is the first to use behavior genetic techniques to

[2] Poterba, Venti and Wise (2000) show the substantial effects that portfolio risk can have on the accumulation of post-retirement wealth.

document the heritability of risk-taking in the financial market, as well as outside the laboratory.[3] The primary disadvantage of using portfolio allocation data to infer risk attitudes is that the riskiness of an individual's portfolio may be determined not only by risk preference parameters, but also by differences in beliefs about future returns[4] as well as heterogeneity in susceptibility to various behavioral biases (for a review, see Barberis & Thaler, 2002). We therefore eschew the label risk preferences, preferring instead to refer to our outcome variable as a measure of attitudes toward financial risk-taking.

The estimates of heritability that we obtain match the laboratory evidence in Cesarini et al. (2009) very closely, and suggest that approximately 25% of individual variation in financial risk-taking is due to genetic variation. Further, variation in childhood rearing environment does not seem to be a major cause of differences in the willingness to bear financial risk. In additional analyses, we also find that these results extend to other aspects of financial decision-making, including the tendency to "choose" the default fund, invest in "ethical" or "socially responsible" investment funds, as well as to engage in returns-chasing behavior. An immediate implication of our findings is that the considerable parent–child similarity in both self-reported attitudes toward risk (Charles & Hurst, 2003; Hryshko, Luengo-Prado & Sorensen, 2007; Dohmen et al., 2008; Kimball, Sahm & Shapiro, 2009) and the choice of what assets to hold (Chiteji & Stafford, 1999) may not arise solely because of cultural transmission from parent to child.

Besides establishing that the willingness to take financial risks is heritable, an important result in and of itself, we believe that our findings have broader implications for efforts to understand heterogeneity in portfolio allocation. For instance, the share of individual variation explained by genes is an order of magnitude larger than the R^2s typically obtained in standard empirical models of investment behavior,[5] which rarely explain more than 5% of the variation in portfolio risk despite using rich sets of covariates (Curcuru et al., 2009). In particular, previous studies of the Swedish

[3] This paper is a revision of an earlier working paper (Cesarini et al., 2008b). As this article was going to press, we learned of a recent working paper by Barnea, Cronqvist and Siegel (2009) that finds comparable results to ours, based on Swedish twin data matched to information on portfolio holdings maintained by the Swedish Tax Agency. Barnea, Cronqvist and Siegel (2009) study stock market participation and two measures of financial risk-taking behavior. Besides twin studies on portfolio risk, a related literature in economics considers economic outcome variables such as education, income and socioeconomic status. Chief references include Taubman (1976), Behrman and Taubman (1989), Plug and Vijverberg (2003), Björklund, Lindahl and Plug (2006) and Sacerdote (2007).

[4] See the discussion in Malmendier and Nagel (2007).

[5] See, for example, Cohn et al. (1975), Friend and Blume (1975), Poterba (1993), Bertaut (1994), Pålsson, (1996), Heaton and Lucas (1997, 2000), Palme, Sundén and Söderlind (2007), and Curcuru et al. (2009).

individualized pension savings accounts (Palme, Sundén and Söderlind, 2007; Säve-Söderbergh, 2008a), the very same accounts that are used in this paper, follow the results in this literature very closely. Genetic variance, by contrast, explains approximately 25% of the cross-sectional variation in portfolio risk according to our estimates, suggesting that scholars should try to better understand how genetic differences affect financial decision-making. As of today, theories of portfolio allocation do not explicitly model genetic sources of individual variation.

Taken together, the results reported here strongly suggest that people differ genetically in their willingness to bear financial risk. Moreover, the excess similarity in the portfolio risk of MZ twins does not appear to be explained by their greater similarity in income, education, cognitive ability, or wealth. One way to interpret this finding is that two individuals facing identical budget sets might still make very different financial decisions for reasons unrelated to differences in environmental circumstances.

The paper is structured as follows. In section 2, we describe the Swedish pension reform and our data set. In section 3, we describe the twin methodology. In section 4, we present our results, and relate them to previous findings. In section 5, we investigate and discuss the robustness and generalizability of our results, and in section 6 we discuss our findings and their implications for current efforts in finance to understand heterogeneity in portfolio risk. Section 7 concludes the paper.

2. Data

2.1. The Swedish pension reform. In 1994, legislation to gradually introduce a new pension system was passed by the Swedish parliament in response to demographic challenges and underfinancing of the pay-as-you-go system that had been in place since the 1960s.[6] The new system is based on a contribution rate of 18.5% on earnings, where 2.5 percentage points accrue to mandatory individual self-directed investment accounts, one of the system's key features.

As a part of the introduction of the new system, a government body—the Premium Pension Agency—was established and assigned the responsibility of handling the individual investment accounts. Almost all adult Swedes born after 1938 were invited to decide how to invest the balance on their individualized pension savings accounts, but the system only fully applied to individuals born in 1954 and thereafter.[7] The "Big Bang" occurred towards the end of 2000, when all participants in the new system had to simultaneously decide how to invest their balances. All in all, some 68% of the

[6] See Palmer (2000) for a detailed discussion of the new system.

[7] Only Swedes whose income exceeded 36,000 SEK ($1 is roughly 8 SEK) in 1995, 36,800 in 1996, 37,000 in 1997 and 37,100 in 1998 were eligible for fund selection in the year 2000.

eligible population made an active decision. Individuals who did not make an active choice had their money invested in a default fund.

Participants could construct a portfolio consisting of no more than five funds from a very large menu of options comprising almost 500 different funds.[8] Among these funds, approximately 14% were fixed income funds, investing predominantly in government bonds, 12% were so-called mixed funds, investing in both equity and bonds, 7% were life-cycle funds, and the remaining 68% were equity funds. Of the equity funds 58% were regional funds, investing in a certain geographic region, 28% were country funds, investing in a specific country, and the remaining 14% were industry funds, investing in a particular industry.

All eligible Swedes were sent a catalogue wherein the available funds were listed with information on the management fees and investment strategies of each fund. Annual historical returns were provided, when available, for the period 1995 to 1999. In addition, the cumulative return of the fund over this period was reported in a separate column. Funds were also color-coded by risk level: from red (very high risk) to green (low risk)[9] and the standard deviation of returns was reported. The circumstances under which these investment decisions were made make them uniquely suitable for inferring attitudes toward financial risk among individuals with little or no financial literacy.

2.2. Portfolio data. Our primary measure of portfolio risk, which we denote *Risk 1*, is the average risk level of the funds invested in by an individual, with the risk of each fund measured as the (annualized) standard deviation of the monthly rate of return over the previous 3 years. The weighting of a fund is equal to that fund's share in the individual's portfolio. In cases where historical returns are not available, these values are imputed by assigning the average value of risk for similar types of funds in the sample.[10] This measure is similar to that employed in Säve-Söderbergh (2008a) and Palme, Sundén and Söderlind (2007), with the one exception that we also include individuals whose money was invested in the default fund.[11] As a second measure of

[8] The official justification for this policy was that individuals should be able to select a portfolio that is suitable given their preferences. The menu of options has now been expanded from the original 461 options to over 700 different funds. For a criticism of this feature of the system, see Cronqvist and Thaler (2004).

[9] The categories were as follows: 0–7 are low-risk funds, 8–17 are average-risk funds, 18–24 are high-risk funds, and 25 and higher are very high risk funds.

[10] The classification of funds was made by the Premium Pension Agency. Examples of types are "New Markets", "IT and Communication", and "Europe Small Enterprises". Our method of imputing missing values has no interesting effects on the estimates we report in this paper.

[11] Säve-Söderbergh (2008a) excludes individuals with the default portfolio on the grounds that its investment profile was not fully known when investment decisions were made in the fall of 2000. The reason its risk profile was not known is that it was constructed to reflect the profile of an average

risk, *Risk 2*, the standard deviation of returns is adjusted to account for covariation between funds in the same portfolio.[12] As yet another robustness check, we calculate a third risk measure, *Risk 3*, as the weighted share of equity funds in an individual's portfolio.[13] Finally, we perform analyses of the *Risk 1* variable purged from variation in a set of demographic and socioeconomic variables. These variables are obtained by matching our data to administrative records.

Our measure of active participation is a binary variable that takes the value one if the individual invested at least some fraction of his wealth in a fund other than the default fund. We further define an indicator variable for ethical investment that takes the value one if an individual chose to invest some fraction of his wealth in an "ethical" investment fund. We classify a total of 18 funds as "ethical". Of these, 9 had the word ethical in the fund name and the remaining 9 funds were classified on the basis of the fund description given in the catalogue. Funds with a self-described investment strategy favoring environmentally friendly companies were classified as ethical, as were funds declaring that they did not invest in alcohol, tobacco, and the arms industry.

Finally, we construct a proxy for returns-chasing. Cronqvist and Thaler (2004) note that the single fund that attracted the largest amount of investment upon the introduction of the Swedish individualized pension savings accounts was the fund that

investor. On the other hand, it seems reasonable to assume that people had some expectation about the future level of risk in the default fund. In practice, none of the results reported in this paper are sensitive to the inclusion of these observations. This supports the notion that individuals not actively choosing a portfolio nevertheless conveyed some information about their risk preferences.

[12] As we do not have time series for the returns of all our funds, we proceed as follows. First, for the 36 months 1997 to 1999, we average the monthly returns across funds within each type of fund in the data, hence giving us one time-series of monthly returns for each type of fund. We then calculate the correlations between the monthly returns of different types of funds during the period in question. These correlations serve as proxies for the correlations between the returns of individual funds when calculating portfolio risk based on the variances in return for each fund. As with our primary risk measure, *Risk 1*, the contribution of each fund to the portfolio risk of an individual is weighted by the amount invested in that fund by that particular individual. For individual i,

$$R_{1i} = \sum_j \alpha_{ij} r_j,$$

$$R_{2i} = \sqrt{\sum_j \sum_k \alpha_{ij}\alpha_{ik}\rho_{jk} r_j r_k},$$

where α_{ik} is the share of fund k in the portfolio of individual i, ρ_{jk} is the correlation between funds j and k (as proxied by the correlation between the types of funds to which funds j and k belong), and r_j is the risk level of fund j (measured as the annualized standard deviation of the monthly rate of return over the previous 3 years).

[13] An equity fund was defined by the Premium Pension Agency as one holding at least 75% equity investments. The polychoric correlation between the color code red (very high risk) and the definition of equity fund is 0.99. Hence, for individual i, $R_{3i} = \sum_j \alpha_{ij} q_j$ where q_j is equal to one if j is an equity fund, and zero otherwise.

had the highest reported historical performance. Though a rigorous test of returns-chasing behavior would require time series data, we constructed a crude proxy as follows. For each of the four fund categories, we first compute the mean cumulative return over the period 1995 to 1999 and the standard deviation. We then classify a fund as a "positive outlier" if it had produced a historical return that was at least one standard deviation higher than the mean cumulative return for funds in its category. For example, the average 5-year cumulative returns of equity funds was 185%, with a standard deviation of 119%. Therefore, all equity funds that reported a return in excess of 304% were classified as a "positive outlier". Eventually, we classify an individual as a "returns-chaser" if at least one of his investment funds was a "positive outlier".

2.3. Data from administrative records. The income measure used in this paper ("sammanräknad förvärvsinkomst") is defined as the sum of income earned from wage labor, own business income, pension income and unemployment compensation. Capital income is not included and the variables are not censored. To minimize the impact of transitory fluctuations, we compute the average of income earned over the 1996 to 2000 period.

Marital status is a variable that takes the value one if the individual was married in 2000. Similarly, years of education is based on educational attainment as of 2000. The data on marital status, income, and education are drawn from administrative records and should be highly reliable.[14] Unfortunately, a reliable measure of net household wealth is not available in our data set. However, until recently Sweden had a wealth tax, which applied to all households whose wealth exceeded a cutoff level of 900,000 SEK (approximately USD 125,000). For these households, data on net individual wealth holdings are available. We therefore use this variable for the year 2000 when it is available, setting the wealth of individuals in households whose wealth did not exceed the threshold equal to zero.

Finally, for most of the men in the sample, a measure of cognitive ability at age 18 is also available from conscription records. The test of cognitive ability used by the Swedish military is a standard test of general intelligence, whose psychometric properties are described in Carlstedt (2000). Recruits take four subtests (logical, verbal, spatial and technical) which, for most of the study period, were graded on a scale from 0 to 40. To construct the IQ variable, the scores are summed and then percentile-rank transformed by birth year. The percentile-ranked variable is then transformed to a standard normal distribution by taking the inverse of the standard normal distribution.

[14] However, since administrative records only contain information on legally earned and taxed income, annual income will only be an imperfect proxy for actual income.

2.4. The Swedish Twin Registry. The Swedish Twin Registry, the largest in the world, contains information on nearly all twin births in Sweden since 1886, and has been described in detail elsewhere (Lichtenstein et al., 2006). The sample used in this paper includes individuals who have participated in at least one of the Twin Registry's surveys. For these respondents, we can establish zygosity with reasonable confidence based on survey questions with proven reliability (Lichtenstein et al., 2006). In practice, roughly 90% of the twins in our data set come from one of two sources. The primary source is the web-based survey STAGE (The Study of Twin Adults: Genes and Environment). This survey was administered between November 2005 and March 2006 to all twins born in Sweden between 1959 and 1985, and it had a response rate of 60%. Data on individuals born between 1938 and 1958 come from SALT (Screening Across the Lifespan Twin study), a survey conducted by telephone in 1998. SALT had a response rate of 74% (Lichtenstein et al., 2006).[15] Though these response rates are most certainly not alarmingly low, we acknowledge that our sample may not be fully representative of the population of twins. Considering all complete same sex twin pairs born after 1938 gives a total of 7,225 female pairs, of which 3,346 are monozygotic, and 6,338 male pairs, of which 2,747 are monozygotic.

3. Method

Our primary analysis estimates the degree to which variation in our measure of portfolio risk is influenced by additive genetic factors (A), an environmental factor common to the two twins in a pair (C), and unshared environmental factors that are specific to each twin (E). Additive genetic effects are defined as the sum of the effects of individual genes influencing a trait. The assumption that genetic effects are purely additive, that is, linear, rules out possibilities such as dominant genes, where nonlinearities exist in the relationship between the amount of genetic material coding for a certain trait and the realized trait in the individual. Common environment effects are those environmental influences shared by both twins. Examples include childhood diet, schooling, parental socialization, and shared peer influences. Unshared environmental effects include influences not shared by the co-twins as well as measurement and response error.

The basic idea behind a behavior genetic decomposition is simple. MZ and DZ twins differ in their genetic relatedness but were raised under similar conditions. Therefore, evidence of greater similarity between MZ twins can be taken as evidence that the studied trait is under genetic influence.

[15] Additionally, some individuals in our sample responded to a survey sent out in 1973 (see Lichtenstein et al., 2002).

It is sometimes noted that moving from a crude comparison of correlations to a full-fledged variance decomposition requires making strong independence and functional form assumptions. Our empirical analysis therefore proceeds in two steps. First, we abstain from imposing any structural assumptions and simply compare the within-pair correlations for the risk variables in MZ and DZ twins. A measure of the statistical significance of the estimated difference between these two within-pair correlations is produced using a standard bootstrap method. Let N_{MZ} be the number of MZ pairs. We create 1,000 pseudosamples of MZ twins by randomly drawing N_{MZ} pairs with replacement 1,000 times.[16] We similarly create 1,000 pseudo-samples of DZ twins. This allows us to calculate 1,000 instances of MZ and DZ within-pair correlations, respectively, and from this to calculate 1,000 differences between MZ and DZ within-pair correlations. The observed distribution of the 1,000 realized differences is our estimated probability distribution of the difference between the MZ and DZ within-pair correlation in our original sample. The p-value for the test of the hypothesis that the two correlations are equal is then easily computed as the fraction of instances in which the difference is negative (i.e., the number of instances divided by 10).

We next proceed to a standard behavior genetic variance decomposition. The work-horse model in the behavior genetics literature, known as the ACE model, posits that additive genetic factors (A), common environmental factors (C), and specific environmental factors (E) account for all individual differences in the variable of interest. Start with the case of MZ twins. Let all variables be expressed as deviations from zero and standardize them to have unit variance. Consider a pair of MZ twins and suppose first that the outcome variable, P, can be written as the sum of two independent influences: additive genetic effects, A, and environmental influences, U. We then have that

$$P = aA + uU, \tag{3.1}$$

and, using a superscript to denote the variables for twin 2 in a pair,

$$P' = aA' + uU'. \tag{3.2}$$

Since for MZ twins $A = A'$, the covariance (which, due to our normalization, is also a correlation) between the outcome variables of the two twins is given by

$$\rho_{MZ} = a^2 + u^2 Cov(U, U')_{MZ}.$$

Now consider a DZ pair. Under the assumption that parents match randomly with respect to their values of A, so that the correlation between the additive genetic effects

[16] The term "with replacement" simply means that any pair drawn for the pseudo-sample is maintained in the pool of pairs eligible for future draws, that is, the pair is replaced by a hypothetical identical pair.

of the father and the mother is zero, it will be the case that $Cov(A, A') = 0.5$.[17] We then have that,

$$\rho_{DZ} = \frac{1}{2}a^2 + u^2 Cov(U, U')_{DZ}. \tag{3.3}$$

Finally, we impose the equal environment assumption, namely that,

$$Cov(U, U')_{MZ} = Cov(U, U')_{DZ}. \tag{3.4}$$

Under these admittedly strong assumptions, it is easy to see that heritability, the fraction of variance explained by genetic factors, is identified as $a^2 = 2(\rho_{MZ} - \rho_{DZ})$. In the standard behavior genetics framework, environmental influences are generally written as the sum of a common environmental component (C) and a non-shared environmental component (E) such that,

$$P = aA + cC + eE. \tag{3.5}$$

With this terminology, the environmental covariance component of the correlation, $u^2 Cov(U, U')$, can be written as c^2, since by definition any covariance must derive only from the common component. This allows us to write the individual variation as the sum of the three components a^2, c^2, and e^2, where a^2 is the share of variance explained by genetic differences, c^2 is the share of variance explained by common environmental influences, and e^2 is the share of variance explained by non-shared environmental influences. There are a number of ways in which the parameters of this model can be estimated. We follow standard practice and use maximum likelihood under the assumption that the outcome variables come from a bivariate normal distribution. In particular, following directly from the above derivation, we maximize the likelihood under the restriction that the variance-covariance matrix is of the form,

$$\sum = \begin{bmatrix} a^2 + c^2 + e^2 & R_i a^2 + c^2, \\ R_i a^2 + c^2 & a^2 + c^2 + e^2 \end{bmatrix}, \tag{3.6}$$

where R_i takes the value one if the observation is of an MZ pair, and 0.5 otherwise. The analyses are run in MPLUS (Muthén & Muthén, 2006), a numerical optimizer often used in behavior genetics. Throughout this paper, these models are estimated allowing the variance components to differ by gender.

Our basic empirical strategy is to estimate these variance components for the *Risk 1* and *Risk 2* measures. Additional analyses are also carried out with the risk measures purged from variation in a number of background variables, including age, income,

[17] A full derivation of the latter result can be found in any text on quantitative genetics, for instance Falconer and Mackay (1996).

education, marital status, and the proxy for wealth. For men, a measure of cognitive ability is available from conscription records and this is also used in some of the analyses.

To supplement the primary evidence on portfolio risk, we also analyze data on three additional variables that broadly capture other aspects of financial decision-making. Since participation, ethical investment, and returns-chasing are binary variables, we follow standard practice and use a threshold model. A threshold model assumes that the categories observed (for example, participation or non-participation) are merely cutoffs of some underlying distribution of the variable under study. For each twin pair, we assume that the variable has a bivariate distribution with unit variance and a correlation matrix varying as a function of zygosity, as specified in equation (3.6). We then carry out maximum likelihood estimation with respect to the variance components and the threshold.[18] The analyses of the categorical variables are conducted using the software MX (Neale et al., 2002).

4. Results

Figure 1 shows the distribution of portfolio risk for women and men, with separate bars for MZ and DZ twins. A visual inspection reveals that approximately half of the portfolios have a risk level in the range 20–25, with only small differences between the sexes in the distribution of the variable. The distributions of MZ and DZ twins also look similar, which is reassuring since the models we estimate assume that the distribution of the variable does not vary by zygosity.

[18] The maximand in the optimization problem is simply the log-likelihood of the observed data,

$$\ln L = \sum_{c=1}^{2}\sum_{i=1}^{2}\sum_{j=1}^{2} n_{ijc} \ln\left(p_{ijc}\right),$$

where n_{ijc} is the observed frequency of data in cell n_{ij} for zygosity c, and the expected proportions in each cell can be calculated by numerical integration as

$$p_{ij1} = \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} \phi(x_1, x_2, \sum_{MZ}) dx_1 dx_2,$$

$$p_{ij2} = \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} \phi(x_1, x_2, \sum_{DZ}) dx_1 dx_2,$$

where $\phi(x_1, x_2, \sum)$ is the bivariate standard normal distribution, $\sum$ is the correlation matrix whose diagonal elements are normalized to one ($a^2 + c^2 + e^2 = 1$), and t_i is the lower threshold of category i. The variable x_1 (x_2) denotes the category that twin 1 (2) belongs to. Of course, the lower threshold of category 0 is $-\infty$, and the upper threshold for category 1 is ∞. Thresholds are constrained to be the same for monozygotic and dyzygotic twins.

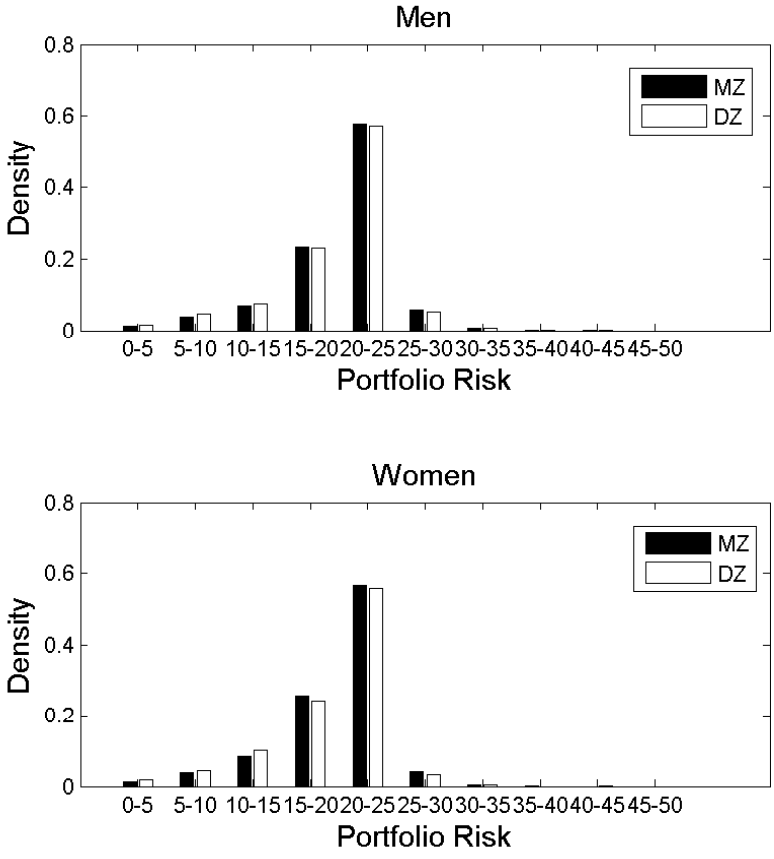


Figure 1. Portfolio risk distribution by gender and zygosity.[19]

[18] This figure shows the distribution of the portfolio risk variable separately for men and women and disaggregated by zygosity. Portfolio risk is the weighted risk of all the funds in the individual's portfolio, with the weighting of a fund equal to that fund's share in the individual's portfolio. Risk is measured as the (annualized) standard deviation of the monthly rate of return of the fund over the previous 3 years. MZ: monozygotic. DZ: dizygotic.

Table 1. Summary statistics for risk measures and supplementary variables[19]

	Women		Men		Total	
	MZ	DZ	MZ	DZ	MZ	DZ
Risk 1	19.00	18.71	19.46	19.10	19.21	18.90
S.D.	(4.24)	(4.42)	(4.40)	(4.56)	(4.32)	(4.49)
Risk 2	18.28	18.06	18.68	18.43	18.46	18.24
S.D.	(3.94)	(4.17)	(4.03)	(4.25)	(3.99)	(4.21)
Risk 3	0.77	0.77	0.81	0.80	0.79	0.78
S.D.	(0.34)	(0.34)	(0.32)	(0.33)	(0.33)	(0.34)
Active	0.71	0.68	0.71	0.67	0.71	0.68
S.D.	(0.45)	(0.46)	(0.45)	(0.47)	(0.45)	(0.47)
Ethical	0.050	0.048	0.028	0.025	0.040	0.037
S.D.	(0.218)	(0.214)	(0.164)	(0.156)	(0.197)	(0.189)
Returns chaser	0.30	0.27	0.30	0.26	0.30	0.27
S.D.	(0.46)	(0.45)	(0.46)	(0.46)	(0.46)	(0.44)
# observations	6,692	7,758	5,494	7,182	12,186	14,940

In Table 1 we report summary statistics. The average portfolio has a risk level of approximately 19. Around 70% of the individuals in the sample made an active choice and 30% exhibited returns-chasing behavior. The latter figure is quite remarkable given that only 3.5% of the funds are classified as "positive outliers". The only variable with a substantial difference between men and women is the measure of ethical investment preferences: whereas 5% of women invest in funds classified as having an ethical investment strategy, the corresponding figure for men is about 2.5%.

A first diagnostic of genetic influences can be obtained by examining the MZ and DZ correlations for the three risk measures. These sibling correlations are reported in Table 2. Again, for *Risk 1*, there are no major differences between men and women, with the MZ correlations being consistently higher than the DZ correlations. In women, the correlations are 0.27 and 0.16, and in men, they are 0.29 and 0.13. Recall that an MZ correlation captures all the determinants of portfolio risk that identical twins share; that is, genotype and shared environmental influences. Thus, the joint influence of genes and shared environment explains nearly 30% of the variation in portfolio risk. The correlations for our second and third risk measures are very similar, which demonstrates that most variation in risk is driven by differences in the share of equity in the portfolio and that adjusting for covariation in returns between the funds in a

[19] This table presents summary statistics for the main variables. MZ: monozygotic. DZ: dizygotic. Standard deviations are in parentheses. See the Appendix for other variable definitions.

portfolio does not appreciably change the results. For the supplementary variables, the MZ correlations are also consistently higher than the DZ correlations.

Table 2. With-in pair correlations[20]

		Women			Men		
		MZ	DZ	p-value	MZ	DZ	p-value
Risk 1	Pearson	0.27***	0.16***	< 0.01	0.29***	0.13***	< 0.01
	Spearman	0.28***	0.16***	< 0.01	0.30***	0.13***	< 0.01
Risk 2	Pearson	0.26***	0.15***	< 0.01	0.27***	0.12***	< 0.01
	Spearman	0.26***	0.15***	< 0.01	0.27***	0.12***	< 0.01
Risk 3	Pearson	0.26***	0.13***	< 0.01	0.24***	0.11***	< 0.01
	Spearman	0.26***	0.14***	< 0.01	0.23***	0.10***	< 0.01
Active	Polychoric	0.49***	0.29**	< 0.01	0.47***	0.22***	< 0.01
Ethical	Polychoric	0.62***	0.27***	< 0.01	0.63***	0.25***	< 0.01
Returns-chaser	Polychoric	0.42***	0.25***	< 0.01	0.37***	0.20***	< 0.01
# pairs		3,346	3,879		2,747	3,591	

It is possible that the excess MZ resemblance in portfolio risk is driven by the higher MZ correlations in background variables such as income, education, wealth, marital status, and age. In an attempt to examine this hypothesis, and to measure the share of the variation that these variables account for, Table 3 reports standard cross-sectional regressions where the dependent variable is *Risk 1*. These regressions are run separately for men and women. The first column reports the results for a simple regression of *Risk 1* on a third order age polynomial. Column 2 adds the income variable (a 5-year average for the period 1996 to 2000), the proxy for wealth, marital status, and educational attainment in years. For men, a third column with cognitive ability is also included.

[20] This table presents with-in pair correlations for *Risk 1*, *Risk 2*, *Risk 3*, and the supplementary variables separately by zygosity and sex. One sided p-values testing the equality of MZ and DZ correlations are reported. Three stars (***) denote statistical significance at the 1% level and two stars (**) denote statistical significance at the 5% level. See the Appendix for variable definitions.

Table 3. Associates of portfolio risk[21]

	Women		Men		
	Model 1	Model 2	Model 1	Model 2	Model 3
Age	2.06***	2.06***	1.74***	1.77***	1.87***
	(.20)	(.21)	(.24)	(.24)	(.87)
Age^2/10	−.46***	−.46***	−.39***	−.39***	−.41**
	(.04)	(.04)	(.05)	(.05)	(.21)
Age^3/100	.31***	.31***	.26***	.27***	.28
	(.03)	(.03)	(.03)	(.03)	(.16)
Income (in million SEK)	–	.41	–	.05	.61**
	–	(.56)	–	(.25)	(.27)
1 if married	–	−.09	–	−.20**	−.12
	–	(.07)	–	(.08)	(.10)
Wealth (in million SEK)	–	.51***	–	.06**	.02
	–	(.10)	–	(.03)	(.02)
Education (in years)	–	.07***	–	.12***	.11***
	–	(.01)	–	(.02)	(.02)
Cognitive Ability	–	–	–	–	.15***
	–	–	–	–	(.06)
R^2	0.070	0.074	0.058	0.065	0.060
No. Observations	14,450	14,383	12,676	12,601	5,895

It is clear that the only variable that explains a meaningful share of the variation in risk-taking is age, as the increment in explanatory power going from a model with just the age polynomial to a model with the five basic covariates is very modest. Other than age, education is the most robust predictor of portfolio risk. In women, the estimated coefficient is 0.07, suggesting that an additional year of education is associated with a 0.07 increase in portfolio risk, holding the other covariates constant. The corresponding coefficient for men is 0.11 and 0.12, depending on whether cognitive ability is included as a control. An increase in (our proxy for) wealth of one million SEK is associated with a 0.51 increase in portfolio risk for women. For men, the coefficient is statistically significant, but appreciably lower. Finally, the results from the specification with cognitive ability suggest that a one standard deviation increase in cognitive ability is associated with a 0.15 increase in portfolio risk.

[21] This table reports standard cross-sectional regressions where the dependent variable is *Risk 1*. These regressions are run separately for men and women. Model 1 is a simple regression on an age polynomial. Model 2 adds the income variable, a proxy for wealth, marital status, and education (in years). For men, a third column is added with cognitive ability included. Three stars (***) denote statistical significance at the 1% level and two stars (**) denote statistical significance at the 5% level.

Table 4. Results of the ACE model, 99% confidence intervals in parentheses[22]

	Risk 1	Risk 2
Women		
Genetic (a^2)	0.27^{***} $(0.11 - 0.34)$	0.27^{***} $(0.12 - 0.32)$
Common (c^2)	0.02 $(0.00 - 0.12)$	0.01 $(0.00 - 0.12)$
Unique (e^2)	0.72^{***} $(0.66 - 0.77)$	0.73 $(0.68 - 0.78)$
Men		
Genetic (a^2)	0.29^{***} $(0.19 - 0.35)$	0.27^{***} $(0.20 - 0.33)$
Common (c^2)	0.00 $(0.00 - 0.06)$	0.00 $(0.00 - 0.05)$
Unique (e^2)	0.71^{***} $(0.65 - 0.76)$	0.73^{***} $(0.67 - 0.78)$
ln(L)	$-78,410.39$	$-75,623.18$

In column 1 of Table 4 we report results from the basic ACE model without any controls. For women, heritability is estimated to be 0.27 (99% CI, 0.11 to 0.34), and for men, heritability is estimated to be 0.29 (99% CI, 0.19 to 0.35). For both sexes, most of the remaining variation comes from non-shared environment. In the second column, we show that the results are virtually unchanged when the portfolio risk is adjusted for covariation in returns between funds in a particular portfolio. Table 5 reports results for *Risk 1* residualized on the variables described in the previous section. There is no evidence that the greater resemblance of MZ twins is accounted for by their greater similarity in educational attainment, income, the wealth proxy, or marital status. The first column shows results for *Risk 1* residualized just on a third order age polynomial. Estimated heritabilities are 0.21 in women (99% CI, 0.08–0.27) and 0.23 in men (99% CI, 0.17–0.28). The estimates of c^2 are zero in both cases. The second column shows results for *Risk 1* residualized on the age polynomial, income, education, wealth, and marital status. The results are virtually identical to those reported in column 1. This is not too surprising, given that these variables explain very little of the variation in risk-taking. Finally, the last column, only available for men, adds cognitive ability to the set of controls. The estimated heritability now becomes 0.21 (99% CI, 0.11–0.28).

[22] This table shows heritability estimates for the variables *Risk 1* and *Risk 2*. A is the genetic contribution; C is the common environment contribution; E is the unique environment contribution. All models are estimated allowing the variance to differ by gender but not zygosity. Confidence intervals are constructed using the bootstrap with 1,000 draws. Three stars (***) denote statistical significance at the 1% level. See the Appendix for variable definitions.

Table 5. Results of the ACE model, 99% confidence intervals in parentheses[23]

	Model 1	Model 2	Model 3
Women			
Genetic (a^2)	0.21^{***} $(0.08 - 0.27)$	0.21^{***} $(0.08 - 0.26)$	–
Common (c^2)	0.00 $(0.00 - 0.10)$	0.00 $(0.00 - 0.08)$	–
Unique (e^2)	0.78^{***} $(0.73 - 0.84)$	0.78^{***} $(0.73 - 0.84)$	–
Men			
Genetic (a^2)	0.23^{***} $(0.17 - 0.28)$	0.23^{***} $(0.17 - 0.28)$	0.21^{***} $(0.11 - 0.28)$
Common (c^2)	0.00 $(0.00 - 0.02)$	0.00 $(0.00 - 0.01)$	0.00 $(0.00 - 0.03)$
Unique (e^2)	0.77^{***} $(0.72 - 0.82)$	0.78^{***} $(0.72 - 0.83)$	0.79^{***} $(0.72 - 0.87)$
ln(L)	$-77,661.68$	$-77,219.13$	$-15,835.15$

Finally, Table 6 shows the results of the ACE decomposition for the three supplementary proxy variables, namely active participation, ethical investment, and returns-chasing. For all three variables, there is evidence of genetic variance. Again, the differences between men and women are small, with estimates a little over 0.40 for participation, around 0.60 for ethical investment, and around 0.30 for returns-chasing.

Table 6. ACE model of supplementary variables[24]

	Active	Ethical	Returns-chaser
Women			
Genetic (a^2)	0.41^{***} $(0.23 - 0.56)$	0.60^{***} $(0.32 - 0.70)$	0.32^{***} $(0.12 - 0.48)$
Common (c^2)	0.08 $(0.00 - 0.22)$	0.00 $(0.00 - 0.25)$	0.10^{*} $(0.00 - 0.25)$
Unique (e^2)	0.51^{***} $(0.44 - 0.57)$	0.40^{***} $(0.30 - 0.51)$	0.59^{***} $(0.52 - 0.65)$
Men			
Genetic (a^2)	0.47^{***} $(0.30 - 0.53)$	0.61^{***} $(0.17 - 0.74)$	0.31^{***} $(0.09 - 0.44)$
Common (c^2)	0.00 $(0.00 - 0.13)$	0.00 $(0.00 - 0.36)$	0.05 $(0.00 - 0.22)$
Unique (e^2)	0.53^{***} $(0.47 - 0.60)$	0.39^{***} $(0.26 - 0.56)$	0.64^{***} $(0.56 - 0.71)$
ln(L)	$-16,330.82$	$-4,244.38$	$-15,928.59$

[23] This table shows heritability estimates with the *Risk 1* variable residualized on different sets of covariates. Model 1 shows results for *Risk 1* residualized on age. Model 2 shows results for *Risk 1* residualized on age, income, marital status, wealth and education. Model 3 adds cognitive ability to the list of covariates. A is the genetic contribution; C is the common environment contribution; E is the unique environment contribution. All models are estimated allowing the variance to differ by gender but not zygosity. Confidence intervals are constructed using the bootstrap with 1,000 draws. Three stars (***) denote statistical significance at the 1% level. See the Appendix for variable definitions.

[24] This table shows the heritability estimates for the supplementary variables. A is the genetic contribution; C is the common environment contribution; E is the unique environment contribution.

5. Robustness and generalizability

To establish how sensitive our main results are to departures from the underlying assumptions, we now turn to an examination of the numerous potential sources of bias, their direction, and the extent to which they may be expected to impact our findings.

5.1. Representativeness and generalizability. To ascertain how representative our sample is of the population at large, we compare it, disaggregated by zygosity and sex, to the Swedish population born between 1938 and 1978 on a number of demographic background variables. The results are reported in Table 7.[25] Respondents tend to have slightly higher incomes than the population average, but unlike other studies (Behrman, Rosenzweig & Taubman, 1994; Ashenfelter & Krueger, 1994), we do not find any economically interesting differences with respect to education. However, there is a slight tendency for participants to have higher marriage rates than the population as a whole. Finally, STAGE and SALT respondents are also somewhat older than the average for the 1938 to 1978 cohorts.

Table 7. Background variables[26]

	Women		Men		Population		
	MZ	DZ	MZ	DZ	Women	Men	Total
Income	234,363	230,560	326,272	324,824	210,000	288,000	251,000
S.D.	111,145	107,722	216,235	292,757	–	–	–
Education (yrs)	12.3	11.9	12.0	11.6	12.3	11.9	12.1
S.D.	2.6	2.7	2.8	2.9	–	–	–
Marital Status	0.52	0.55	0.55	0.56	0.52	0.48	0.50
S.D.	0.50	0.50	0.50	0.50	–	–	–
Age	48.7	51.8	50.1	52.8	46.6	46.5	46.6
S.D.	11.3	10.0	10.9	9.2	–	–	–

Obviously, it is impossible to fully establish how representative our sample is of the population as a whole. The propensity to respond to a survey is likely to be associated with a number of background characteristics that are not readily measurable, but that may nevertheless be influencing our findings, such as general motivational factors. If

All models are estimated allowing the variance to differ by gender but not zygosity. Confidence intervals are constructed using the bootstrap with 1,000 draws. Three stars (***) denote statistical significance at the 1% level, and one star (*) denotes statistical significance at the 10% level. See the Appendix for variable definitions.

[25] As is common in twin studies, women are slightly overrepresented (Lykken et al., 1990) in both STAGE and SALT, comprising 53% of our sample.

[26] This table shows summary statistics for some background variables, disaggregated by sex and zygosity. Population mean is defined as the average for individuals born 1938 to 1978. See the Appendix for variable definitions.

people with certain background characteristics are overrepresented, and if heritability is associated with these background characteristics, then the heritability estimate will be biased in the direction of this association.

In addition to asking how representative our sample of twins is, it is also important to consider whether twins as a group differ from the population as a whole with respect to unobservables. Few variables have been found to differ between twins and non-twins (Kendler et al., 1995). We can think of no convincing reason why the experience of growing up with a twin should idiosyncratically affect financial decision-making in adult life.

5.2. Equal environment assumption. Critics of the classical twin design cite a number of potential failures of the equal environment assumption, which states that shared environmental influences are not more important for monozygotic twins than for dizygotic twins. One common objection is that parents, on average, give MZ twins more similar treatment.[27] It is important to emphasize that even if MZ twins receive more similar treatment from their parents, this need not constitute a violation of the assumption; greater similarity in environment may be caused by the greater similarity in genotypes (Plomin et al., 2001). In the context of research on personality and IQ, where the equal environment assumption has been tested most rigorously, the evidence is fairly convincing that any bias that arises from this restriction is not of first order importance (Bouchard, 1998).

More importantly, for measures of personality and cognitive ability, studies of MZ and DZ twins who were reared apart tend to produce estimates of heritability similar to studies using twins reared together (Bouchard, 1998). Since studies of twins reared apart do not rely on the equal environments assumption, findings from such studies seem to validate the basic model. Also, in the relatively rare cases where parents miscategorize their twins as MZ instead of DZ (or the converse), differences in correlations of cognitive ability and personality persist (Bouchard & McGue, 2003).

5.3. Reciprocal Influences. Our baseline model assumes an absence of reciprocal influences between twins. If twins influence each other's choices positively, their degree of similarity will be inflated. Moreover, if this effect is stronger in MZ twins than in DZ twins, this will bias upward the estimate of heritability. The STAGE and SALT data sets both contain information on the frequency of contact between twins. As is commonly found in twin studies, monozygotic twins interact more than dizygotic

[27] For further criticisms of the equal environment assumption, see Joseph (2002) and the references therein.

twins. On average, MZ pairs reported 3.3 interactions per week at the time of the survey, whereas DZ pairs reported an average of 1.8 interactions per week.[28]

Running separate regressions by gender, where the dependent variable is the squared within-pair difference in portfolio risk and the independent variables are frequency of contact and zygosity, frequency of contact is a significant predictor of the within-twin-pair squared difference in portfolio risk, for both men and women. The presence of a statistically significant effect does not prove, however, that the frequency of contact is causing increased similarity. Much research has been devoted to establishing the direction of causality. Lykken et al. (1990) and Posner et al. (1996) offer some evidence suggesting that twins similar in personality tend to stay in contact with one another, and not the other way round.

One crude way to examine whether twins have communicated about their choice of funds is to ask how common it is for both twins in a pair to choose the same portfolio. Excluding pairs where both twins selected the default portfolio, of the remaining MZ twins, 8% choose the same portfolio as their co-twin. In DZ twins the corresponding figure is 3%. To further examine the sensitivity of our results to this source of bias, we conduct two robustness checks, the results of which are reported in Table 8.

Table 8. Robustness checks of the ACE model[29]

	Dropped	Matched
Women		
Genetic (a^2)	0.17^{***} $(0.07 - 0.22)$	0.17^{**} $(0.00 - 0.29)$
Common (c^2)	0.00 $(0.00 - 0.09)$	0.07 $(0.00 - 0.20)$
Unique (e^2)	0.83^{***} $(0.78 - 0.88)$	0.76^{***} $(0.70 - 0.82)$
Men		
Genetic (a^2)	0.22^{***} $(0.16 - 0.27)$	0.28^{***} $(0.11 - 0.34)$
Common (c^2)	0.00 $(0.00 - 0.03)$	0.00 $(0.00 - 0.13)$
Unique (e^2)	0.79^{***} $(0.73 - 0.84)$	0.72^{***} $(0.66 - 0.79)$
ln(L)	$-74,337.11$	$-48,555.40$

[28] We construct the frequency of contact variable as follows. Subjects who report seven or more interactions (by e-mail, telephone, or letter) per week are assigned a value of seven. All other subjects are assigned the number of interactions per week that they report. If we have data on both twins, we use the mean of the two reports.

[29] This table shows the results from the robustness checks conducted using the frequency of contact variable. In the "Dropped" column, pairs where both twins selected identical portfolios are excluded. In the "Matched" column, we stratified the data by frequency of contact into 15 groups, and for each sex and level of contact we then randomly dropped the required number of either MZ or DZ pairs to make the number of MZ and DZ pairs equal. In this restricted sample, the distribution of frequency of contact is, by construction, virtually the same in the MZ and DZ groups. All models are estimated allowing the mean and the variance to differ by gender. 99% confidence intervals in parentheses.

First, we drop all pairs in which both individuals chose the same portfolio, and rerun the analyses. Obviously, by discarding these observations, the correlations for both MZ and DZ twins will drop. Furthermore, the adjusted correlations will be downward biased if twins choosing identical portfolios are more similar than average with respect to their attitudes toward financial risk. This sample restriction produces a heritability estimate of 0.17 (99% CI, 0.07 to 0.22) in women and 0.22 (99% CI, 0.16 to 0.29) in men. Under the assumption that communication only affects choices through identical portfolios, this can serve as a lower bound for our heritability estimate in the presence of reciprocal action.

Second, we make use of our frequency of contact variable. Specifically, we stratify frequency of contact into 15 groups, and for each sex and level of contact we then randomly drop the required number of either MZ or DZ pairs to make the number of MZ and DZ pairs equal. In this restricted sample, the distribution of frequency of contact is, by construction, virtually the same in the MZ and DZ groups. Re-running the analyses on this subset of the data, the estimated heritability in women is 0.17 (99% CI, 0.00–0.29) and in men it is 0.28 (99% CI, 0.11–0.34). The finding that the heritability estimates fall only marginally is reassuring since it suggests that frequency of contact is not a major influence on our main result.[30] Our interpretation of these results is that the twins who opted for the same retirement fund would generally have chosen portfolios with similar levels of risk even without the opportunity to consult each other.

5.4. Misclassification and measurement error. We use the Swedish Twin Registry's standard algorithm to establish zygosity. The algorithm has been validated using DNA-based evidence, and studies show that misclassification is typically on the order of 2% to 5% (Lichtenstein et al., 2006). Purely random assignment error would bias heritability downward, since the difference in genetic relatedness between pairs assigned as MZ or DZ would decrease to less than one-half. However, misclassification may be non-random and related to physical similarity (notice that the questions we use to establish zygosity are solely based on assessments of physical similarity). The relevant question is then whether physical similarity is related to similarity with respect to behavior. The classical reference on this topic is Matheny, Wilson and Brown Dolan (1976), who administered two intelligence tests, two perceptual tests, one reading test, one test of speech articulation, and one personality inventory to twins and find that

Confidence intervals are constructed using the bootstrap with 1,000 draws. Three stars (***) denote statistical significance at the 1% level and two stars (**) denote statistical significance at the 5% level. See the Appendix for variable definitions.

[30] A significant drop in estimated heritability is, however, a necessary but not sufficient condition for frequency of contact to be the cause of greater similarity.

"correlations revealed no systematic relation between the similarity of appearance and the similarity of behaviors for either the identical twin pairs or the same-sex fraternal twin pairs" (p. 1573).[31] We conclude that the bias that arises due to misclassification is likely to be small and to lead to an understatement of heritability.

As in the case of misclassification, measurement error tends to bias a^2 and c^2 downwards since any such error will be subsumed under the estimate of e^2. In the simplest case, where the variable is observed with a mean zero random error with variance σ_ϵ^2 it is easy to show that the estimates of a^2 and c^2 need to be scaled up by a factor of $\frac{1}{1-\sigma_\epsilon^2}$. However, whereas measurement error is easy to conceptualize in psychometric research as the test-retest reliability of some instrument designed to measure a personality trait, it is less clear how to interpret such error in the present case, where it would presumably involve the choice of actual portfolio risk being related to factors other than willingness to take financial risks. While this is certainly likely to be the case, it is far from obvious how the reliability of actual observed risk-taking in the field could convincingly be tested.

5.5. Assortative mating. Finally, we note that the model assumes the absence of assortative mating. Dohmen et al. (2008) and Kimball, Sahm and Shapiro (2009), however, report significant positive spousal correlations for a survey-based measure of risk-taking, which is consistent with assortative mating. Positive assortative mating implies a correlation between the As of spouses, which would bias our estimates of heritability downward.

6. Discussion

In this paper, we demonstrate that for our measures of portfolio risk, MZ twins exhibit significantly greater resemblance than DZ twins. Previous work has shown that risk preferences elicited in the laboratory, using either survey questions or gambles over small stakes, are heritable (Cesarini et al., 2009). The contribution of this paper is to document similar results in a field setting uniquely suited for inferring individual attitudes toward risk-taking, a setting where standard objections about external validity do not apply (Harrison & List, 2004; Levitt & List, 2007). Moreover, our sample size is some 30 times greater than that of Cesarini et al. (2009), so estimation error is smaller. An immediate implication of our result is that the considerable parent-child similarity in both self-reported attitudes toward risk (Charles & Hurst, 2003; Hryshko, Luengo-Prado & Sorensen, 2007; Dohmen et al., 2008; Kimball, Sahm & Shapiro, 2009), and

[31] Hettema, Neale and Kendler (1995) report no significant associations between physical similarity and phenotypic resemblance in four out of the five psychological disorders they consider (the one exception is bulimia).

choice of which assets to hold (Chiteji & Stafford, 1999) is due to genetic transmission and not merely cultural transmission. For instance, the parent-child correlation in risk attitudes found in Dohmen et al.'s (2008) representative German sample imply upper bounds on heritability of approximately 0.35, and the point estimates of heritability in Cesarini et al. (2009) range from 0.14 to 0.35.[32] This consistency of results across different methodologies is reassuring because it suggests that the findings are not driven by confounding factors particular to our study. Such factors include the fact that our sample may not be fully representative (unlike the sample in Dohmen et al., 2008), or the fact that we cannot rule out the possibility that the twins have communicated about their choice of portfolio (unlike the experimental evidence in Cesarini et al. (2009) where twins always participated in the same experimental session without any opportunity to communicate with each other).

Our results, properly interpreted, may enhance our understanding of heterogeneity in portfolio allocation (Guiso et al., 2002; Curcuru et al., 2009). Heaton and Lucas (1995) conclude a survey of this literature on asset pricing by remarking that "a notable difficulty with the models that we discuss is their inability to explain heterogeneity in asset holdings across households" (p. 27). Empirical models of portfolio risk typically also do a poor job of accounting for cross-sectional variance. For example, an early paper by Cohn et al. (1975), based on a highly selected sample of educated and wealthy clients of a brokerage firm, shows that age, marital status, and income only explain a small portion of the share of an individual's wealth invested in risky assets. More recently, Curcuru et al. (2009) used data from the Survey of Consumer Finances and report that a model with seven covariates, including financial assets, income, age, and marital status, explains about 3% of the variation in the variable stocks as a share of liquid financial assets. These findings are not unique to U.S. data, but also hold in studies of the Swedish retirement accounts used in this paper (Palme, Sundén & Söderlind, 2007; Säve-Söderbergh, 2008a). For example, Palme, Sundén and Söderlind (2007) report R^2 values of 0.03 to 0.04 when regressing portfolio risk, as defined in this paper, on a rich set of demographic and socioeconomic covariates and these standard covariates also explain a small share of portfolio risk also in our sample of twins. The one exception is age, but since age is perfectly correlated within pairs, it cannot be a source of the excess resemblance of MZ twins. By contrast, genetic variance explains approximately 25% of the cross-sectional variation in portfolio risk according to our estimates.

[32] If the coefficient of genetic relatedness is 0.5, and only genes explain parent-child resemblance, then doubling the correlation will produce an estimate of heritability. If there are other nongenetic forces that can account for the correlation, then the heritability estimated from parent offspring correlations will be upward biased.

Although the results presented here suggest that there is genetic variation in the willingness to take financial risk, the specific genetic mechanisms are not yet understood. Generally speaking, one can taxonomize these mechanisms into two broad categories. One possibility is that people with different genotypes differentially select into environments that lead them to make different financial investments from those that they would have made had they not been exposed to that environment. Insofar as genotypes cause selection into environments that influence financial decision-making, this will appear as a genetic effect in the behavior genetic framework, even though the genetic effect is in fact environmentally mediated. A second possibility is that there are in fact genes with a more proximal effect on the outcome variable of interest.

It seems very likely that some of the genetic effects reported in this paper operate through genome-wide influences on variables that have been identified as important predictors of portfolio risk in the literature. For example, in a descriptive study of allocation decisions in U.S. mandatory savings accounts, Poterba and Wise (1998) report that education and income levels predict investment choices. Palme, Sundén and Söderlind (2007) and Säve-Söderbergh (2008a) report qualitatively similar findings for the Swedish retirement accounts. Yet, in the Swedish data, residualizing portfolio risk on the standard covariates considered in the empirical literature on portfolio choice does not appreciably change the estimated heritabilities. This finding is consistent with the hypothesis that two individuals who are identical in terms of income, education, wealth, and age may still make very different portfolio investment choices. That is, faced with the same budget constraints and optimization problem, individuals with different genetic endowments may still make very different investment choices. It is useful to distinguish between two distinct, though not necessarily mutually exclusive, explanations for this, which potentially have very different implications for efforts to help people make better investment decisions. One possibility is simply that there is genetic variance in the coefficient of risk aversion, as suggested by Cesarini et al. (2009). Alternatively, there may be genetic variation in the susceptibility to behavioral biases relevant to financial decision-making (Patel, Zeckhauser & Hendricks, 1991; Benartzi, 2001; Benartzi & Thaler, 2001; Barberis & Thaler, 2002; Huberman & Jiang, 2006). Future work should focus on further disentangling preference based explanations for the heritability of portfolio risk from explanations based on behavioral anomalies and computational errors.

Our paper makes some progress in this direction. Specifically, we explore the possibility that investors use the representativeness heuristic (Kahneman & Tversky, 1972) in making their investment decisions, extrapolating from past performance to forecast future returns and thus placing their money in funds that have historically had high

returns.[33] A number of papers provide evidence that cash flow into a mutual fund is highly correlated with returns in recent years (see, for example, Sapp & Tiwari, 2004; or Siri and Tufano, 1998). Relatedly, a growing literature in finance documents that people extrapolate from their own experiences in making investment decisions (Benartzi, 2001; Choi et al., 2009). Cronqvist and Thaler (2004) note that the most popular fund in the Swedish retirement system, barring the default fund, was the fund with the highest historical return. This paper provides some evidence in favor of the view that our measure of portfolio risk may not solely capture risk preferences, by showing that a proxy for returns-chasing also appears to be heritable.

Beyond their purely descriptive value, our results also favor the use of models with heterogeneity (Aiyagari, 1993; Telmer, 1993; Heaton & Lucas, 1995; Constantinides & Duffie, 1996) and challenge the common assumption in finance and economics that people are born identical and that subsequent idiosyncratic shocks are the only important source of individual variation (e.g., Mankiw, 1986; Freeman, 1996). Most models with consumer heterogeneity maintain the assumption that individuals have the same preference parameters and instead assume that the main source of heterogeneity lies in differential shocks to individual income. But, as we have noted, it is also possible that variation in portfolio risk arises for other reasons; for example, agents may differ in their ability to solve portfolio problems, people may have different beliefs about the returns of various classes of assets, or fundamental preference parameters such as the coefficient of risk aversion may vary across individuals. The latter explanation in particular is controversial. Much of the debate about such "preference heterogeneity" can be framed in terms of the question of whether or not it is scientifically meaningful to invoke differences in preferences to explain differences in economic and financial outcomes (Stigler & Becker, 1977; Caplan, 2003). The concern here is that explanations based on unobservable differences in tastes are difficult to falsify. In Stigler and Becker's (1977) words, "no significant behavior has been illuminated by assumptions of differences in tastes" (p. 89).

Biological and genetic markers are currently being included in a number of social science surveys and economists are considering how to best leverage this information (Benjamin et al., 2007). Analyses that incorporate molecular genetic information into models of portfolio choice are thus becoming feasible. The inclusion of such variables in models of portfolio choice is one possible way to discipline theories based on otherwise unobservable differences in preference parameters and thereby resolve the circularity of theories that invoke unobserved heterogeneity as an explanation. One plausible direct channel through which genes may influence risk-taking behavior is by regulating the

[33] We are grateful to a reviewer for the suggestion to explore this question.

dopaminergic pathways in the brain, which are known to regulate the anticipation of rewards (see the discussion in Dreber et al., 2009). Indeed, two recent and independent studies find that one version (allele) of the DRD4 gene, known to be involved in the regulation of the dopaminergic system, is associated with greater financial risk-taking (Dreber et al., 2009; Kuhnen & Chiao, 2009). A number of recent studies find significant relationships between risk-taking and other biological factors such as patterns of brain activation (Kuhnen & Knutson, 2005; Cardinal, 2006; Preuschoff, Bossaerts & Quartz, 2006; Knutson et al., 2008). Though it is worth emphasizing that this evidence is merely correlational, neurostudies may offer further clues about the specific genes involved in risk-taking and shed light on the complicated pathways from genes to a particular behavior. In addition, it seems likely that many of the empirical correlations that have been discovered in finance, for example, the intriguing relationship between the propensity to gamble and anomalous "preferences" for stock (Kumar, 2009), have a shared genetic source.

In a world of imperfectly informed investors a proper understanding of the origins of preference heterogeneity is important to allow advisers and policymakers to provide sound advice to investors, and if investors do indeed have heterogenous preferences, then financial advice should be tailored taking this into account (Curcuru et al., 2009). This paper's findings suggest, for example, that variation in the exogenous family environment of an individual is not likely to be a major source of individual differences, but rather that some of the individual differences are genetic in origin. It is a common misconception, however, that since genes are fixed, they must therefore be fixed in their effects. The fallacy is most easily disposed of by a simple example, originally due to Goldberger (1979). Suppose that all variation in eyesight is due to genes. A simple environmental innovation, glasses, might remove this variation at a very low cost. The point, very simply, is that genetic variation, just like environmental variation can be persistent, or easily remediable, depending on its exact source. Moreover, an environmental intervention outside the current range of environmental variation could have large effects, regardless of the heritability of a trait and the mechanisms that explain the association between genotype and risk-taking. Therefore, the fact that a trait is heritable does not imply that it is not malleable, but it does imply that it is important to try to understand why genes are associated with the trait.

An immediate implication of the results reported here is that if stable genetic correlates of risk-taking are discovered, a host of important ethical questions about "genetic discrimination" by employers and insurance companies will be raised. Some genes have already been implicated in behavioral disorders. It seems quite possible that these genes also predict factors such as the likelihood of defaulting on a loan or

engaging in risky behaviors that exacerbate moral hazard problems in principal-agent relationships with imperfect information.[34] An analogy to medical insurance is in order, where an important and difficult policy question is how or whether insurance companies should be entitled to request genetic information on an individual's vulnerability to disease (Tabarrok, 1994). Currently, it is a violation of U.S. law for insurance companies to request genetic information in order to better predict some individual characteristic that is deemed relevant.

Finally, we note that even though portfolio risk is the primary outcome variable of interest, additional analyses of supplementary variables suggest that a broader class of behaviors related to financial decision-making are heritable. In addition to the aforementioned results on returns-chasing, similar findings obtain for active participation and for the propensity to invest "ethically." A number of studies of 401(k) saving behavior find that the default options offered to households can have a large effect on investment choices (Madrian & Shea, 2001; Choi et al., 2003) and this fact is often invoked to argue that policymakers should pay special attention to default options in the design of retirement savings plans (Samuelson & Zeckhauser, 1988; Benartzi & Thaler, 2001; Thaler & Benartzi, 2004; Benartzi & Thaler, 2007). In the context of Swedish individualized retirement accounts, Engström and Westerberg (2003) report that higher education and higher income are associated with a higher likelihood of participation. Differential fixed costs are often posited as the reason why some households do not participate in financial markets, despite the normative prescription that under weak conditions on preferences investors should invest at least some portion of their wealth in stocks (Haliassos & Bertaut, 1995). Our results suggest that these fixed costs are partly genetic in origin. The propensity to invest in "ethical" funds is also found to be heritable. Ethical investment decisions in the Swedish retirement accounts have previously been studied by Säve-Söderbergh (2008b), who identifies some of the empirical correlates of investment in "ethical" funds.

7. Conclusion

In this paper we match data on the mandatory pension investment decisions made in the fall of 2000 to the Swedish Twin Registry in an attempt to estimate genetic influence on variation in financial risk-taking. Relative to the experimental and survey evidence reported in Cesarini et al. (2009), a distinct advantage of our approach is that we examine risk-taking behavior in a field setting with large financial incentives attached to performance. We find that approximately 25% of the variation in portfolio

[34] In fact, a recent paper finds that polymorphisms on the MAOA gene are associated with credit card debt (De Neve & Fowler, 2009) but the result has yet to be replicated in an independent sample.

risk is due to genetic variation. This is in line with the previous, but small, literature that documents parent-child correlations in attitudes towards risk (Charles & Hurst, 2003; Hryshko, Luengo-Prado & Sorensen, 2007; Dohmen et al., 2008; Kimball, Sahm, & Shapiro, 2009). These results are the first to document the heritability of risk-taking in financial markets, as well as outside the laboratory, and they strongly suggest that genetic variation is an important source of individual heterogeneity.

The explanatory power of the genetic effect that we find is an order of magnitude larger than the R^2s typically reported in empirical studies of portfolio choice. This suggests that standard variables included in models of portfolio choice do not adequately capture the cross-sectional variation. The fact that the excess similarity of MZ twins is not explained by excessive similarity of income, education, and other covariates suggests that even when faced with a similar portfolio optimization problem, people demonstrate systematic differences in their allocation decisions. Therefore, models of investor heterogeneity that fail to capture this feature of the data are in a sense incomplete.

Economists disagree about whether preferences should be left as a black box (Dohmen et al., 2008). That is, is it scientifically helpful for economists to try to explore the various mechanisms underlying preferences? Our answer to this question is an unambiguous yes, and our hope is that the results reported in this paper will inspire more research on the biological sources of variation in financial risk-taking. Knowledge of such factors might produce additional levers for predicting both individual and market behavior. However, we emphasize that for this enterprise to be successful, a crucial next step of this agenda will be to try to better understand why genes are associated with portfolio risk. Efforts are currently under way to include biological markers in a number of large social surveys, several of which contain data on asset holdings. Our results suggest that these genetic markers may well ultimately help us shed light on the fundamental question of why individuals differ in their willingness to take risks.

References

Aiyagari, S. R., "Explaining financial market facts: The importance of incomplete markets and transaction costs," *Federal Reserve Bank of Minneapolis Quarterly Review* 17 (1993), 17-31.

Ashenfelter, O., and A. Krueger, "Estimates of the economic return to schooling from a new sample of twins," *American Economic Review* 84 (1994), 1157-1173.

Barberis, N., and R. H. Thaler, "A survey of behavioral finance," NBER Working Paper No. W9222 (2002).

Barnea, Amir, Henrik Cronqvist, and Stephan Siegel, "Nature or nurture: What determines investor behavior?," available at http://ssrn.com/abstract=1467088 (2009).

Behrman, J. R., and P. Taubman, "Is schooling mostly in the genes? Nature-nurture decomposition using data on relatives," *Journal of Political Economy* (1989) 97, 1425-1446.

Behrman, J. R., M. R. Rosenzweig, and P. Taubman, "Endowments and the allocation of schooling in the family and in the marriage market: The Twins Experiment," *Journal of Political Economy* 102 (1994), 1131-1174.

Benartzi, S., "Excessive extrapolation and the allocation of 401(k) accounts to company stock," *Journal of Finance* 56 (2001), 1747-1764.

Benartzi, S., and R. H. Thaler, "Naive diversification strategies in defined contribution saving plans," *American Economic Review (*2001) 91, 79-98.

Benartzi, S., and R. H. Thaler, "Heuristics and biases in retirement savings behavior," *Journal of Economic Perspectives* 21 (2007), 81-104.

Benjamin, D. J., C. F. Chabris, E. L. Glaeser, V. Gudnason, T. B. Harris, D. Laibson, L. Launer, and S. Purcell, "Genoeconomics," in M. Weinstein, J. W. Vaupel, and K. W. Wachter (Eds), *Biosocial Surveys* (Washington, DC: The National Academies Press, 2007).

Bertaut, C. C., "Who holds stock in the U.S.? An empirical investigation," Working paper, University of Maryland (1994).

Björklund, A., M. Lindahl, and E. Plug, "The origins of intergenerational associations: Lessons from Swedish adoption data," *Quarterly Journal of Economics* 121 (2006), 999-1028.

Bouchard, T.. J. Jr, "Genetic and environmental influences on adult intelligence and special mental abilities," *Human Biology* 70 (1998), 257-279.

Bouchard, T. J. Jr., and M. McGue, "Genetic and environmental influences on human psychological differences," *Journal of Neurobiology* 54 (2003), 4-45.

Caplan, B., "Stigler-Becker versus Myers-Briggs: Why preference-based explanations are scientifically meaningful and empirically important," *Journal of Economic Behavior and Organization* 50 (2003), 391-405.

Cardinal, R. N., "Neural systems implicated in delayed and probabilistic reinforcement," *Neural Networks* 19 (2006), 1277-1301.

Carlstedt, B., *Cognitive Abilities: Aspects of Structure, Process and Measurement*, Ph.D. dissertation, Gothenburg University (2000).

Cesarini, D., C. T. Dawes, J. H. Fowler, M. Johannesson, P. Lichtenstein, and B. Wallace, "Heritability of cooperative behavior in the trust game," *Proceedings of the National Academy of Sciences* 104 (2008a), 3721–3726.

Cesarini, D., M. Johannesson, P. Lichtenstein, Ö. Sandewall, and B. Wallace, "Is financial risk taking behavior genetically transmitted?," IFN Working Paper No. 765 (2008b).

Cesarini, D., C. T. Dawes, M. Johannesson, P. Lichtenstein, and B. Wallace, "Genetic variation in preferences for giving and risk-taking," *Quarterly Journal of Economics* 124 (2009), 809-842.

Charles, K. K., and E. Hurst, "The correlation of wealth across generations," *Journal of Political Economy* 111 (2003), 1155-1182.

Chiteji, N. S., and F. P. Stafford, "Portfolio choices of parents and their children as young adults: Asset accumulation by African-American families," *American Economic Review* 89 (1999), 377-380.

Choi, J. J., D. Laibson, B. Madrian, and A. Metrick, "Optimal defaults," *American Economic Review Papers and Proceedings* 92 (2003), 180-185.

Choi, J. J., D. Laibson, B. Madrian, and A. Metrick, "Reinforcement learning and savings behavior," *Journal of Finance* 64 (2009), 2515-2534.

Cohn, R. A., W. G. Lewellen, R. C. Lease, and G. G. Schlarbaum, "Individual investor risk aversion and investment portfolio composition," *Journal of Finance* 30 (1975), 605-620.

Constantinides, G. M., and D. Duffie, "Asset pricing with heterogeneous consumers," *Journal of Political Economy* 104 (1996), 219-240.

Cronquist, H., and R. H. Thaler, "Design choices in privatized social-security systems: Learning from the Swedish experience," *American Economic Review* 94 (2004), 424-428.

Curcuru, S., J. Heaton, D. Lucas, and D. Moore, "Heterogeneity and portfolio choice: Theory and evidence," in Y. Ait-Sahalia, and L. P. Hansen (Eds), *Handbook of Financial Econometrics*. (Amsterdam, NL: Elsevier, 2009).

De Neve, J.-E., and J. H. Fowler, "The MAOA Gene Predicts Credit Card Debt," available at http://ssrn.com/abstract=1457224 (2009).

Dohmen, T., A. Falk, D. Huffman, and U. Sunde, "The intergenerational transmission of risk and trust attitudes," IZA Discussion Paper No. 2380 (2008).

Dreber, A., C. L. Apicella, D. T. A. Eisenberg, J. R. Garcia, R. S. Zamore, J. K. Lum, and B. Campbell, "The 7R polymorphism in the dopamine receptor D4 gene (DRD4) is associated with financial risk taking in men," *Evolution and Human Behavior* 30 (2009), 85-92.

Engström, S., and A. Westerberg, "Which individuals make active investment decisions in the new Swedish pension system?," *Journal of Pension Economics and Finance* 2 (2003), 225-245.

Falconer, D. S., and T. F. C. Mackay, *Introduction to Quantitative Genetics* (Benjamin Cummings, 1996).

Freeman, S., "Equilibrium income inequality among identical agents," *Journal of Political Economy* 104 (1996), 1047-1064.

Friend, I., and M. E. Blume, "The demand for risky assets," *American Economic Review* 65 (1975), 900-922.

Goldberger, A. S., "Heritability," *Economica* 46 (1979), 327–347.

Guiso, L., M. Haliassos, and T. Jappelli (Eds), *Household Portfolios* (Cambridge, MA: MIT Press, 2002).

Haliassos, M., and C. C. Bertaut, "Why do so few hold stock?," *Economic Journal* 105 (1995), 1110-1129.

Harrison, G. W., and J. List, "Field experiments," *Journal of Economic Literature* 42 (2004), 1009-1105.

Heaton, J., and D. Lucas, "The importance of investor heterogeneity and financial market imperfections for the behavior of asset prices," *Carnegie-Rochester Conference Series on Public Policy* (1995) 42, 1-32.

Heaton, J., and D. Lucas, "Market frictions, savings behavior and portfolio choice," *Macroeconomic Dynamics* 1 (1997), 76-101.

Heaton, J., and D. Lucas, "Portfolio choice in the presence of background risk," *Economic Journal* 110 (2000), 1-26.

Hettema, J. M., M. C. Neale, and K. S. Kendler, "Physical similarity and the equal-environment assumption in twin studies of psychiatric disorders," *Behavior Genetics* 25 (1995), 327-335.

Hryshko, D., M. J. Luengo-Prado, and B. E. Sorensen, "Childhood determinants of risk aversion: The long shadow of compulsory education," Working paper, University of Alberta (2007).

Huberman, G., and W. Jiang, "Offering versus choice in 401(k) plans: Equity exposure and number of funds," *Journal of Finance* 61 (2006), 763-801.

Isacsson, G., "Estimating the economic return to educational levels using data on twins," *Journal of Applied Econometrics* 19 (2004), 99-119.

Joseph, J., "Twin studies in psychiatry and psychology: Science or pseudoscience?," *Psychiatric Quarterly* 73 (2002), 71-82.

Kahneman, D., and A. Tversky, "Subjective probability: A judgment of representativeness," *Cognitive Psychology* 24 (1972), 411-435.

Kendler, K. S., N. G. Martin, A. C. Heath, and L. J. Eaves, "Self-report psychiatric symptoms in twins and their nontwin relatives: Are twins different?," *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 60 (1995), 588-591.

Kimball, M. S., C. R. Sahm, and M. D. Shapiro, "Risk preferences in the PSID: individual imputations and family covariation," NBER Working Paper No. W14754 (2009).

Knutson, B., G. E. Wimmer, C. M. Kuhnen, and P. Winkielman, "Nucleus accumbens activation mediates the influence of reward cues on financial risk taking," *NeuroReport* 19 (2008), 509-513.

Kuhnen, C. M., and B. Knutson, "The neural basis of financial risk taking," *Neuron* 47 (2005), 763-770.

Kuhnen, C. M., and J. Y. Chiao, "Genetic determinants of financial risk taking," *PLoS ONE* 4 (2009), e4362.

Kumar, A., "Who gambles in the stock market?," *Journal of Finance* 64 (2009), 1889-1933.

Levitt, S. D., and J. A. List, "What do laboratory experiments measuring social preferences tell us about the real world?," *Journal of Economic Perspectives* 21 (2007), 153–174.

Lichtenstein P., U. de Faire, B. Floderus, M. Svartengren, P. Svedberg, and N. L. Pedersen, "The Swedish Twin Registry: A unique resource for clinical, epidemiological and genetic studies," *Journal of Internal Medicine* 252 (2002), 184-205.

Lichtenstein, P., P. F. Sullivan, S. Cnattingius, M. Gatz, S. Johansson, E. Carlström, C. Björk, M. Svartengren, A. Volk, L. Klareskog, U. de Faire, M. Schalling, J. Palmgren, and N. L. Pedersen, "The Swedish Twin Registry in the Third Millennium: An Update," *Twin Research and Human Genetics* 9 (2006), 875–882.

Lykken, D. T., M. McGue, T. J. Bouchard Jr, and A. Tellegen, "Does contact lead to similarity or similarity to contact?," *Behavior Genetics* 20 (1990), 547-561.

Madrian, B., and D. F. Shea, "The power of suggestion: Inertia in 401(k) participation and savings behavior," *Quarterly Journal of Economics* 116 (2001), 1149-1187.

Malmendier, U., and S. Nagel, "Depression babies: Do macroeconomic experiences affect risk-taking?," Working paper, Stanford University (2007).

Mankiw, N. G., "The equity premium and the concentration of aggregate shocks," *Journal of Financial Economics* 17 (1986), 211-219.

Matheny, A. P., R. S. Wilson, and A. Brown Dolan, "Relations between twins' similarity of appearance and behavioral similarity: Testing an assumption," *Behavior Genetics* 6 (1976), 343-351.

Morin, R. A., and A. F. Suarez, "Risk aversion revisited," *Journal of Finance* 38 (1983), 1201-1216.

Muthén, L.. K., and B. O. Muthén, *Mplus. Statistical Analysis With Latent Variables. User's Guide, Version 4.1* (Los Angeles, CA: 2006).

Neale M. C, S. M. Boker , G. Xie, and H. H. Maes, *Mx: Statistical Modeling* (Richmond, VA: VCU Department of Psychiatry, 2002).

Neale, M. C., and M. M. Miller, "The use of likelihood-based confidence intervals in genetic models," *Behavior Genetics* 27 (1997), 113-120.

Palme, M., A. Sundén, and P. Söderlind, "Investment choice in the Swedish premium pension plan," *Journal of the European Economic Association* 5 (2007), 636-646.

Palmer, E., "The Swedish pension reform model: Framework and issues," The World Bank SP Discussion Paper No. 0012 (2000).

Pålsson, A.-M., "Does the degree of relative risk aversion vary with household characteristics?," *Journal of Economic Psychology* 17 (1996), 771-787.

Patel, J., R. Zeckhauser, and D. Hendricks, "The rationality struggle: Illustrations from financial markets," *American Economic Review* 81 (1991), 232-236.

Plomin, R. D., J. C. DeFries, G. E. McClearn, and P. McGuffin, *Behavioral Genetics*, 4th ed. (New York, NY: Freeman, 2001).

Plug, E., and W. Vijverberg, "Schooling, family background, and adoption: Is it nature or is it nurture?," *Journal of Political Economy* 111 (2003), 611-641.

Posner, S. F., L. Baker, A. Heath, and N. G. Martin, "Social contact and attitude similarity in Australian twins," *Behavior Genetics* 26 (1996), 123-133.

Poterba, J., "Who owns corporate stock? A report to the New York Stock Exchange," Working paper, MIT (1993).

Poterba, J., and D. Wise, "Individual financial decisions in retirement saving plans and the provision of resources for retirement," in M. Feldstein (Ed.), *Privatizing Social Security* (Chicago, IL: University of Chicago Press, 1998).

Poterba, J., M., S. F. Venti, and D. A. Wise, "Saver behavior and 401(k) retirement wealth," *American Economic Review* 90 (2000), 297-302.

Preuschoff, K., P. Bossaerts, and S. R. Quartz, "Neural differentiation of expected reward and risk in human subcortical structures," *Neuron* 51 (2006), 381–390.

Sacerdote, B., "How large are the effects from changes in family environment? A study of Korean American adoptees," *Quarterly Journal of Economics* 122 (2007), 119-157.

Samuelson, W., and R. Zeckhauser, "Status-quo bias in decision making," *Journal of Risk and Uncertainty* 1 (1988), 7-59.

Sapp, T., and A. Tiwari, "Does stock return momentum explain the "smart money" effect?," *Journal of Finance* 59 (2004), 2605-2622.

Säve-Söderbergh, J., "Self-directed pensions: Gender, risk & portfolio choices," Working paper, Stockholm University (2008a).

Säve-Söderbergh, J., "Who is willing to let ethics guide his economic decision-making? Evidence from individual investments in ethical funds," Working paper, Stockholm University (2008b).

Siri, E. R., and P. Tufano, "Costly search and mutual fund flows," *Journal of Finance* 53 (1998), 1589-1622.

Stigler, G., and G. Becker, "De gustibus non est disputandum," *American Economic Review* 67 (1977), 76-90.

Tabarrok, A., "Genetic testing: An economic and contractarian analysis," *Journal of Health Economics* 13 (1994), 75-91.

Taubman, P., "The determinants of earnings: Genetics, family, and other environments: A study of white male twins," *American Economic Review* 66 (1976), 858-870.

Telmer, C. I., "Asset-pricing puzzles and incomplete markets," *Journal of Finance* 48 (1993), 1803-1832.

Thaler, R. H., and S. Benartzi, "Save more tomorrow: Using behavioral economics to increase employee saving," *Journal of Political Economy* 112 (2004), S164–S187.

Wallace, B., D. Cesarini, P. Lichtenstein, and M. Johannesson, "Heritability of ultimatum game responder behavior," *Proceedings of the National Academy of Sciences* 104 (2007), 15631-15634.

Appendix: Variable definitions

Active	Equal to one if individual did not invest all assets in the default fund.
Age	Age, in years, in the year 2005.
Cognitive ability	Standard normal variable constructed from the raw score on the Swedish Military's test of cognitive ability. Standardized by birth year.
Education	Years of education, imputed using population averages estimated by Isacsson (2004).
Ethical	Equal to one if individual invested some fraction of her wealth in a "socially responsible" fund.
Income	Average income between 1996 and 2000, where income (sammanräknad förvärvsinkomst) is defined as the sum of income earned from wage labor, own business income, pension income, and unemployment compensation.
Marital status	Equal to one if individual was married in 2000.
Returns-chaser	Equal to one if individual invested some fraction of wealth in a fund whose cumulative returns in the 5-year period 1995 to 1999 were at least one standard deviation higher than the mean cumulative return for funds in the same category.
Risk 1	Weighted average risk level of the selected funds, with the risk of each fund measured as the annualized standard deviation of the monthly rate of return over the previous 3 years.
Risk 2	Average risk level adjusted for covariance in the returns between portfolio funds.
Risk 3	Weighted share of high risk funds in an individual's portfolio.
Wealth	Individual net wealth, as defined by Tax Authorities, in million SEK. Variable is set to zero for individuals residing in households whose net wealth did not exceed 900,000 SEK in the year 2000.

PAPER 4

The behavioral genetics of behavioral anomalies

David Cesarini Magnus Johannesson Patrik Magnusson Björn Wallace

Abstract. A number of recent papers have examined the environmental and genetic sources of individual differences in economic and financial decision making. Here we contribute to this burgeoning literature by extending it to a number of key behavioral anomalies that are thought to be of importance for consumption, savings, and portfolio selection decisions. Using survey-based evidence from more than 11,000 Swedish twins, we demonstrate that a number of anomalies such as, for instance, the conjunction fallacy, default bias, and loss aversion are moderately heritable. In contrast, our estimates imply that variation in common environment explains only a small share of individual differences. We also report suggestive evidence in favor of a shared genetic architecture between cognitive reflection and a subset of the studied anomalies. These results offer some support for the proposition that the heritable variation in behavioral anomalies is partly mediated by genetic variance in cognitive ability. Taken together with previous findings, our results underline the importance of genetic differences as a source of heterogeneity in economic and financial decision making.

1. Introduction

Behavioral economics and behavioral genetics represent two of the most successful developments in the postwar social sciences (Camerer & Loewenstein, 2004; Plomin et al., 2009). Yet, despite their obvious common denominator, there has been very little interaction between the two. A possible explanation for this state of affairs is that the intellectual roots of the "heuristics and biases" research program pioneered by Kahneman and Tversky can be found in social psychology, which typically focuses on the identification of contextual effects. Behavioral genetics, in contrast takes, context as

A slightly different version of this paper is forthcoming in *Management Science*. This paper has benefited from the feedback of the editors and an anonymous reviewer. The authors also thank seminar participants at Carnegie Mellon, London Business School, NYU Stern, Sciences Po, Royal Holloway College (University of London), Syracuse University, and University of Cambridge for helpful comments. Jonathan Beauchamp helped them with programming and offered a number of suggestions on how to improve this paper. Financial support from the Hedelius Wallander Foundation is gratefully acknowledged. All errors remain the authors' responsibility.

given and tries to understand the underlying environmental and genetic origins of individual variation (Plomin et al., 2009). The advantages of combining these two research methodologies are substantial. Behavioral economics has been successful in documenting deviations from the assumptions and predictions of the neoclassical model, but has been unsuccessful in explaining the origins of, and variation in, these deviations (Cohen & Dickens, 2002; Camerer, 2003; Fudenberg, 2006; Gigerenzer et al., 2008). Indeed, a criticism that has at times been levelled against the behavioral economics approach is that it lacks a theory of individual differences and, significantly, that efforts to find empirically robust and theoretically plausible sources of interpersonal variation have failed. The only convincing exception to this rule of which we are aware is the relationship between cognitive ability and a number of behavioral anomalies (Stanovich & West, 1998; Stanovich, 2003; Frederick, 2005; Benjamin et al., 2006).

Partly in response to these challenges, a literature that studies the neuropsychological foundations of economic behavior has developed (Camerer et al., 2005) alongside a smaller and more recent behavioral genetics literature (Cesarini et al., 2009a). Here, we contribute to the emerging behavioral genetics literature by extending it to behavioral biases. Following in the footsteps of Ellsberg (1961), Kahneman and Tversky (1972, 1984), and Tversky and Kahneman (1981, 1983), we examine a wide array of behavioral biases that are believed to influence economic decision-making. The anomalies that we study have attracted substantial interest in both economics and finance, where their relevance for explaining real-world asset pricing (Fama, 1998; Barberis & Thaler, 2003), and consumption/savings patterns (Angeletos et al., 2006) is the topic of much debate. Similarly, the prevalence of behavioral anomalies in the laboratory has sparked an intense methodological debate in psychology and economics on whether or not they are in fact artefacts of their potentially unfamiliar and unintuitive framing (Gigerenzer, 1991; Cosmides & Tooby, 1994; Kahneman & Tversky, 1996; Thaler 1999) and under what conditions they can be switched off.

Behavioral genetics offers a number of methods that can help to resolve important questions concerning the etiology of behavioral anomalies by informing us about the relative contributions of differences in genes and environments to variation in economic decision making. The simplest and most popular of these behavioral genetic methods is the twin method, which relies on comparisons of correlations between monozygotic (MZ; also know as identical) and dizygotic (DZ; also known as fraternal) twins. The twin method is a form of quasi-experiment because MZ twins reared together share both family environment and genes, whereas the genes of DZ twins reared together are no more correlated than those of any other pair of full biological siblings. Therefore, a higher MZ resemblance for a studied trait is typically interpreted as evidence for

heritability in susceptibility to that trait. Using an extensive survey of a large cohort of Swedish twins, we demonstrate that for a wide range of behavioral anomalies the responses of MZ twins are indeed more concordant than those of DZ twins.

In addition to studying intrapair correlations, we estimate standard models and find that genetic variation typically accounts for 20%–30% of the variance in the surveyed anomalies. Variation in rearing conditions, on the other hand, appear to explain at most a modest share of individual differences. The finding that variation in susceptibility to behavioral biases is moderately heritable, and hence that there are systematic individual differences, is important in and of itself. It also points to a source of behavioral heterogeneity—namely, genes—that is only beginning to be explored in economics and finance (see Benjamin et al., 2007). Recent papers have in fact documented heritabilities for risk taking, giving, and ultimatum game rejections that are quite similar to those reported here (Wallace et al., 2007; Cesarini et al., 2009a, b; Zhong et al., 2009; Zyphur et al., 2009), a consistency of results that we find reassuring.

Having established that these behavioral anomalies are heritable, we then ask if there is evidence that some of this heritable variation is mediated by genetic variance in cognitive ability. For a subset of our sample we have data on performance on Frederick's cognitive reflection test (CRT; Frederick 2005). An individual's score on the CRT is usually interpreted as measuring system 2 decision-making abilities in dual process theory. Dual process theory distinguishes between two types of thinking: system 1 thinking, which corresponds to intuitive judgments, and system 2 thinking, which applies to the class of problems that require reasoning (Sloman, 1996). The intuitive answer to each question in the CRT is wrong, and finding the correct answer requires both reflection and the mobilization of costly cognitive resources (Frederick, 2005). Using standard multivariate behavioral genetic methods, we attempt to estimate the covariance between the genetic endowment for this measure of cognitive ability and the genetic endowments for each of the behavioral anomalies. The estimates are quite imprecise, but broadly consistent with the notion that these endowments covary, which is in agreement with the mediation hypothesis.

Overall, our results suggest that researchers in economics, finance, and the decision sciences may be well advised to study further the role of biological and genetic factors in generating individual differences. Such information, beyond its intrinsic scientific value, may also be useful for understanding heterogeneity across individuals in response to policy and the different debiasing strategies that have been proposed in the literature.

The remainder of this paper is organized as follows. In sections 2 and 3 we describe the data and the behavioral anomalies included in the survey. Section 4 outlines the

behavioral genetics methodology, and in section 5 we present the results. We discuss our findings in section 6, and section 7 concludes this paper.

2. Data

The Swedish Twin Registry (STR) is the largest twin registry in the world, and it routinely administers surveys to Swedish twins (Lichtenstein et al., 2006). Here, we use data from the most recent of these surveys, Screening Across the Life-span Twin (younger) (SALTY), which is a collaborative effort between researchers in epidemiology, medicine, and economics initiated in 2007. SALTY is the first major survey of twins that features an entire section specifically devoted to economic decision making. Beginning in early 2009, the survey was sent out to 24,914 Swedish twins born between 1943 and 1958. Final reminders were sent out in the spring of 2010 to those who did not initially respond to the survey, and the data collection was completed in the summer of 2010. The survey generated a total of 11,743 responses, equalling a response rate of 47.1%. Of the respondents, 11,418 (97.2%) gave informed consent to have their answers stored and analyzed. Zygosity was resolved either by questionnaire items with high reliability or, when available, by analysis of biosamples (Lichtenstein et al., 2006). In total, our sample is comprised of 1,150 MZ, 1,245 same-sex DZ, and 1,117 opposite-sex DZ pairs. Remaining responses were from individuals whose twin siblings did not fill in the survey.

Table 1. Sample background statistics[1]

	Respondents			Non-Respondents		
	Mean	S.D.	#	Mean	S.D.	#
Birth year	1949.9	4.57	11,418	1950.4	4.59	13,171
1 if Female	0.54	0.50	11,418	0.48	0.50	13,171
1 if MZ	0.27	0.44	11,418	0.23	0.42	13,170
1 if SS DZ	0.36	0.48	11,418	0.37	0.48	13,170
1 if OS DZ	0.36	0.48	11,418	0.39	0.49	13,170
Education	11.94	2.72	11,167	11.42	2.64	12,024
Income	294,000	186,000	11,188	275,000	213,000	12,070
1 if married	0.65	0.48	11,191	0.59	0.49	12,076

For most of the respondents we were able to obtain background statistics on education, income, and marital status from administrative records by matching individuals

[1] Income is defined as the sum of income earned from wage labor, own business, pension income, and unemployment income compensation. The education variable produced by Statistics Sweden is categorical. The categorical scores are converted into years of education using the population averages in Isacsson (2004). A survey respondent is defined here as an individual who responded to the survey and gave informed consent to have his or responses analyzed. Nonrespondents are the individuals who failed to respond to the survey. SS, same sex; OS, opposite sex.

to a data set collected previously by Cesarini et al. (2010). To ascertain how representative our sample is, we present summary statistics for the background variables in Table 1. Like in other twin studies, there is some overrepresentation of women (Lykken et al., 1987). Respondents are also better educated compared to nonrespondents, the difference being approximately 0.5 years of educational attainment. There are, however, only small differences in the other background variables, and our sample is arguably more representative of its study cohort, and the general population, than most other studies in the behavioral economics literature.[2]

3. The survey

Throughout the remainder of this paper, we use the term "behavioral anomaly" to describe behaviors that are inconsistent with the standard assumptions made in economic models, e.g., about decision making under uncertainty or the discounting of future rewards. In total, we examine seven types of behavioral anomalies using 11 variables and 17 questions, most of which were derived from seminal papers in the behavioral economics literature. Below, we describe how the measures that we analyze were constructed and, when applicable, how the questions were modified to fit the format of the survey. Throughout, we consistently code the variables so that a higher value corresponds to less susceptibility to the behavioral anomaly. The English translation of the questions used can be found in Appendix A.

3.1. Loss aversion. To study loss aversion, or small-stakes risk aversion, we use three questions that represent binary choices over participation in hypothetical gambles. These questions were constructed in the mold of, but were not identical to, those used by Tversky and Kahneman (1992). In each question respondents were asked to either accept or reject a gamble that was associated with a 50% chance of losing 1,000 SEK and a 50% chance of winning either 1,500, 2,000, or 2,500 SEK.[3] It has been known at least since Arrow (1971) that an implication of expected utility theory, when the utility function is everywhere concave in wealth, is that individuals should be approximately risk neutral over small-stakes gambles. Rabin (2000) extended the insight and demonstrated that even risk aversion over modest stakes implies implausible levels of risk aversion over larger stakes. For the purpose of analysis the individual responses to the three questions were used to define four separate categories with values ranging from 0 to 3: "always loss averse" (0), "loss averse only at 2000" (1), "loss averse only

[2] For a more exhaustive discussion on representativeness in experimental economics we refer the reader to Harrison and List (2004), and similarly, for a discussion of recruitment bias in twin research, see Lykken et al. (1987).

[3] The exchange rate is approximately seven Swedish kronor (SEK) to the U.S. dollar (November 2010).

at 1,500 and above," (2) and "never loss averse" (3). Fewer than 2% of respondents provided answers that were inconsistent, in the sense that the individual accepted the gamble at either the 1,500 or 2,000 prize, but then rejected a gamble with higher monetary reward. These inconsistent responses were omitted from the analysis.[4]

3.2. Self-control. Self-control problems, typically modelled as quasihyperbolic discounting (Laibson, 1997), have received substantial attention in recent years. Here, we use two different measures of self-control problems: one that is based on trade-offs between immediate and delayed outcomes and one that is based on the subject's selfreported actual behavior.

3.2.1. *Short-term time preference.* An indication of self-control problems is excessive discounting of future outcomes. Frederick et al. (2002) provide a comprehensive survey of the different approaches to measuring discount factors that have been proposed in the literature and note that each approach is associated with difficulties. Here, we use a simple choice task in which subjects were asked to respond to three questions that represent binary hypothetical choices between an amount of money today and a larger amount of money in the future. In each question, respondents had to choose between 5,000 SEK today or a larger amount in a week, where the larger amount was either 5,500, 6,000, or 7,000 SEK. For the purpose of analysis individual responses to the three questions were aggregated and coded into four categories. Each category is represented by an integer between 0 and 3, where 0 denotes never choosing the delayed reward, and 3 denotes always choosing the delayed reward. We eliminated inconsistent responses in the analyses, but such responses were, again, rare.

3.2.2. *Procrastination.* To supplement the evidence on discounting, we also studied self-reported actual behavior. In particular, to obtain a proxy for self-control problems in everyday life we asked subjects how often they fail to pay their bills on time. There were six alternative responses to the question ranging from "never" (5) to "several times a month" (0). Self-control problems can lead to this type of procrastination behavior, and we refer to the measure as procrastination.

3.3. Default bias. Default bias, also known as status quo bias, implies that there is a bias toward choosing the default option (Samuelson & Zeckhauser, 1988). Studies of 401(k) savings behavior have, for instance, demonstrated that the default option offered to households can have pronounced effects on investment choices (Madrian & Shea, 2001; Choi et al., 2003). We used three questions on self-reported actual behavior as indicators of default bias. Two questions asked the subjects if they had

[4] However, in analyses not reported here, we verified that including the inconsistent observations by simply letting the outcome variable be the number of rejected gambles yielded substantively identical results.

changed (i) telephone and (ii) electricity providers following the recent deregulation of these industries in Sweden. It is widely known that changing providers was associated with large potential reductions in costs, even though the products offered were close to perfect substitutes. Despite this, many consumers failed to make the transition and instead stayed with the default provider. The third and final question asked if, conditional on having previously been a member, the subject had left the Church of Sweden after the separation from the state. Each response was coded with the value 1 if the individual reported a change of provider and 0 otherwise. Similarly, individuals who reported having left the Church of Sweden were coded as 1, and individuals who reported not having left were coded as 0. We compute an index of default bias by summing the responses to the three questions, so that a higher number denotes less susceptibility to status quo bias.[5]

3.4. Illusion of control. To study illusion of control we follow Langer (1975) and investigate if valuations of gambles over random outcomes are affected by the amount of control respondents perceive that they exercise over the outcome of a lottery. Respondents were asked to make a binary choice between two hypothetical lotteries. The first lottery assigned participants a ticket with higher expected value than the ticket offered in the second lottery, in which participants instead were allowed to choose their own ticket, thus introducing a tradeoff between perceived control and expected return. If respondents preferred the second lottery, they were coded as being subject to the illusion of control. The responses of individuals who preferred the first lottery, with higher expected value, were assigned the value 1.

3.5. Ambiguity aversion. To study ambiguity aversion, we use a slightly modified version of Ellsberg's (1961) urn with 30 red balls and 60 black and yellow balls of unknown proportions. Subjects were asked to choose between three hypothetical lotteries, one paying 900 SEK if a red ball was chosen, one paying 1,000 SEK if a black ball was chosen, and one paying 1,000 SEK if a yellow ball was chosen. If respondents preferred the lottery with red as the winning color they were coded as ambiguity averse. These responses were coded as 0, and all other responses as 1.

3.6. Fungibility of money. To study fungibility we use slightly modified versions of the two theater ticket questions in Tversky and Kahneman (1981). Subjects were

[5] The question about leaving the Church of Sweden may not be a good measure of default bias insofar as it is confounded with religiosity. Because MZ twins are more similar in the strength of their religious attitudes, this may generate greater concordance in the decision to leave the Church of Sweden for reasons that have nothing to do with their susceptibility to "default bias." Excluding this question from the index generates an MZ correlation of 0.20 and a DZ correlation of 0.08, compared to 0.25 and 0.10, respectively, with the question included.

asked to make a hypothetical binary choice between going to the theater or not when they had already decided to go but either lost the ticket or the money with which they had intended to buy the ticket. The responses to the two questions were aggregated. If the subjects gave answers that were incompatible with fungibility they were coded as being subject to the bias, with a value of 0. If the answers were compatible with fungibility, the responses were coded as 1.

3.7. The representativeness heuristic. In a series of classical papers, Kahneman and Tversky (see, e.g., Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1974) documented several prevalent violations of the laws of probability in human statistical reasoning. They traced these failures to the so-called "representativeness heuristic," according to which people, ignoring other relevant factors, use similarity as a basis for judgment when making probabilistic assessments. To study the "representativeness heuristic," we administered three questions that are described below. The three questions were analyzed separately, but they were also collapsed into a total "representativeness index."

3.7.1. *Conjunction fallacy.* To study the conjunction fallacy we use a reduced version of the Linda question in Tversky and Kahneman (1983). In our version, subjects only faced two alternatives instead of the original eight. If they answered that Linda was a bank teller and a feminist, they were coded as being subject to the conjunction fallacy with a value of 0. Subjects who did not commit the conjunction fallacy were coded as 1.

3.7.2. *Base rate fallacy.* To examine the base rate fallacy, according to which subjects pay insufficient attention to base rates in making probability assessments, we use a question that conceptually draws on Tversky and Kahneman's (1983) Linda question and their 1973 paper (Kahneman & Tversky, 1973) which demonstrated insensitivity to base rates. Subjects were asked to judge whether a described man was most likely to be a nurse or a professional tennis player. If respondents answered that he was more likely to be a professional tennis player, they were coded as being subject to the base rate fallacy, because this answer fails to take into account the vastly higher base rate of male nurses than male professional tennis players. Subjects who answered that he was more likely to be a nurse were coded as 1.

3.7.3. *Insensitivity to sample size.* To study the subjects' understanding of basic laws of statistics, and in particular the law of large numbers, we use a slightly modified and reduced version of the question on sex ratios and hospitals from Kahneman and Tversky (1972). Subjects were asked to assess whether it was more likely that 60% of the children born at a hospital were boys when the hospital was small rather than large. Respondents who answered that it was more likely that the unusually high

fraction of boys would occur in the larger hospital were coded as being insensitive to sample size, with a value 0. Subjects who correctly answered that it was more likely that the fraction of boys would exceed 0.6 in the smaller hospital were coded as 1.

3.7.4. *Cognitive Reflection.* Finally, we administered Frederick's (2005) cognitive reflection test to approximately half of the sample. We used exactly the same three questions as in the original article and summed the number of correct answers for the purpose of analysis.[6]

4. Methods

4.1. The ACE framework. The basic idea behind a twin study is simple. MZ and DZ twins differ in their genetic relatedness but are reared in the same family. Therefore, any greater similarity between MZ twins in some trait is usually taken as evidence that the trait is under genetic influence. The workhorse model in the behavioral genetics literature, known as the ACE model, posits that additive genetic factors (A), common environmental factors (C), and nonshared environmental factors (E) account for all individual differences in the trait of interest.

Additive genetic effects are defined as the sum of the effects of individual genes influencing a trait. The assumption that genetic effects are purely additive, i.e., linear, rules out possibilities such as dominant genes, where nonlinearities exists in the relationship between the amount of genetic material coding for a certain trait and the actual realized trait. Common environment effects are those environmental influences shared by both twins, such as the quality of local schooling, parental education, and income. Nonshared environmental effects include influences not shared by the cotwins as well as measurement and response error.

For a formal development of the ACE model in the case of MZ and DZ twins reared together, consider first a pair of MZ twins. Let all variables, including the trait, be expressed as deviations from zero and standardize them to have unit variance. Suppose first that the outcome variable can be written as the sum of two independent influences: additive genetic effects, A, and environmental influences, U. We then have that

$$P = aA + uU,$$

and, using a superscript to denote the variables for twin 2 in a pair,

$$P' = aA' + uU'.$$

[6] Regrettably, the cognitive reflection test was removed from later waves of the survey as part of a general effort to reduce the number of questions.

Because for MZ twins $A = A'$, the covariance (which, due to our normalization, is also a correlation) between the outcome variables of the two twins is given by

$$\rho_{MZ} = a^2 + u^2 Cov(U, U')_{MZ}.$$

Now consider a DZ pair. Under the assumption that parents match randomly with respect to their values of A, so that the correlation between the additive genetic effects of the father and of the mother is zero, it will be the case that $Cov(A, A') = 0.5$. We then have that

$$\rho_{DZ} = \frac{1}{2}a^2 + u^2 Cov(U, U')_{DZ}.$$

Finally, we impose the equal environment assumption, namely, that

$$Cov(U, U')_{MZ} = Cov(U, U')_{DZ}.$$

Under these, admittedly strong, assumptions it is easy to see that heritability, the fraction of variance explained by genetic factors, is identified as $a^2 = 2(\rho_{MZ} - \rho_{DZ})$. In the standard behavioral genetics framework, environmental influences are generally written as the sum of a common environmental component (C) and a nonshared environmental component (E) such that

$$P = aA + cC + eE.$$

With this terminology, the environmental covariance component of the trait correlation, $u^2 Cov(U, U')$, can be written as c^2, because by definition any covariance must derive only from the common component. This allows us to write the individual variation as the sum of three components a^2, c^2, and e^2; a^2 is the share of variance explained by genetic differences, c^2 is the share of variance explained by common environmental influences, and e^2 the share of variance explained by nonshared environmental influences. There are a number of ways in which the parameters of this model can be estimated. In particular, following directly from the above derivation, the variance–covariance matrix is of the form

$$\sum = \begin{bmatrix} a^2 + c^2 + e^2 & R_i a^2 + c^2 \\ R_i a^2 + c^2 & a^2 + c^2 + e^2 \end{bmatrix} \tag{4.1}$$

where R_i takes the value 1 if the observation is an MZ pair, and 0.5 otherwise. The ACE framework is frequently criticized for being overly simplistic, and it is indeed based on strong assumptions about the absence of assortative mating and additive gene action (Goldberger, 1977). Additionally, the functional form and independence assumptions needed to identify the model are likely to be wrong. Yet, the basic stylized facts from behavioral genetics have proven to be quite resilient to alternative modeling assumptions. In particular, empirical work using sibling types other than twins reared together (e.g., adoptees, half and full siblings and twins reared apart) tend to produce

estimates that are quite similar to twin-based estimates. For a thoughtful discussion of the various objections that have been raised against twin studies, see Bouchard and McGue (2003).

There are, however, several interpretational issues that require attention. The first is that it is quite possible that many of the genetic effects estimated in behavioral genetics may be mediated by environments (Jencks, 1980; Dickens & Flynn, 2001). An individual's environment is often endogenous to genotype, either because genes cause selection into certain environments or because genes evoke certain behavioral reactions (Jencks, 1980). For this reason, the estimates from the behavioral genetic model are often interpreted as reduced form coefficients from a more general model in which some environments are endogenous to genotype.

In Appendix B, we sketch how, under some additional assumptions, the model can be extended to the two variable case. This model, known as the bivariate, or Cholesky, decomposition (Martin & Eaves, 1977), can be used to decompose covariation between two variables into genetic, common environmental, and nonshared environmental components. It also allows us to investigate whether the genes that are correlated with a particular behavioral anomaly are in fact also correlated with the genetic endowments for cognitive reflection.

4.2. Estimation. Because the variables that we study are ordinal, we follow the standard approach in the literature, which is to estimate a threshold model. A threshold model assumes that the categories observed (for example, being susceptible to the conjunction fallacy) are cutoffs of some underlying distribution of the studied trait. For each twin pair, the distribution of the variable is assumed to have a bivariate normal distribution with unit variance and correlation varying as a function of zygosity, as specified in equation (4.1). Maximum likelihood estimation is then carried out with respect to the variance components and the threshold, which also is estimated as a part of the model. The maximand in the optimization problem is simply the log-likelihood of the observed data,

$$\ln L = \sum_{c=1}^{2} \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ijc} \ln \left(p_{ijc} \right), \tag{4.2}$$

where n_{ijc} is the observed frequency of data in cell n_{ij} for zygosity c; k and l are, respectively, the maximum number of categories of the two variables; and the expected proportions in each cell can be calculated by numerical integration as

$$p_{ij1} = \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} \phi(x_1, x_2, \sum_{MZ}) dx_1 dx_2, \tag{4.3}$$

$$p_{ij2} = \int_{t_i}^{t_{i+1}} \int_{t_j}^{t_{j+1}} \phi(x_1, x_2, \sum_{DZ}) dx_1 dx_2, \tag{4.4}$$

where $\phi(x_1, x_2, \sum)$ is the bivariate standard normal distribution, $\sum$ is the correlation matrix, whose diagonal elements are normalized to 1 ($a^2 + c^2 + e^2 = 1$), and t_i is the lower threshold of category i. The number of thresholds will be equal to the number of categories minus one, and the thresholds are estimated as part of the model. Of course, the lower threshold of category 0 is $-\infty$, and the upper threshold for the highest category is ∞. Thresholds are constrained to be the same for monozygotic and dizygotic twins,[7] but in our estimation we allow men and women to have different thresholds.

For inference, we use likelihood ratio tests, following the suggestion of Neale and Miller (1997). In particular, confidence intervals are obtained by fixing the parameter of interest at some value different from its optimal value, while simultaneously optimizing the remaining parameters. Under some regularity conditions, the distribution of the likelihood ratio test statistic $-2\ln(L_1/L_2)$, L_1 and L_2 being respectively the maximized likelihoods of the nested and the more general models, follows a $\chi^2(1)$ distribution. The parameter is displaced until the deterioration in likelihood is significant.

However, the approach we take to estimating confidence intervals is known to be conservative when the parameter value is on the boundary of the parameter space under the null hypothesis (Dominicus et al., 2006). Dominicus et al. (2006) derive the asymptotic distribution of the test statistic for the case when estimates from an ACE model are compared to a model where the a^2 or the c^2 coefficients are constrained to equal zero. They report that as a rule of thumb, p-values derived under the assumption of an asymptotic $\chi^2(1)$ distribution need to be divided by two in this context. We prefer to err on the side of caution and therefore report confidence intervals constructed using the conventional approach. The reported p-values are hence likely to be conservative and our confidence intervals too wide.

5. Results

In Table 2 we report descriptive statistics for the 11 outcome variables and the CRT along with intrapair polychoric correlations separated by the zygosity and sex composition of the pairs. As can be seen, there is a lot of variation in susceptibility to the different behavioral anomalies. For example, an overwhelming majority of subjects exhibit the base rate and conjunction fallacies, as well as loss aversion. On the other hand, suffering from the illusion of control or treating money as nonfungible is less common. Of course, these differences may simply reflect the choice of questions used

[7] This is a testable restriction that we fail to reject for all 11 variables.

to measure the anomalies, and alternative elicitation procedures might have generated different results.

If there is heritable variation in a trait, then one should expect MZ twins to exhibit greater similarity in the trait than DZ twins. As can be seen in Table 2, this is indeed what we observe in our data. MZ correlations are consistently higher than same-sex DZ correlations, and the difference is significant at the 5% level for 7 of the 11 studied anomalies. At the 10% level, only the variables that we refer to as short time preferences and base rate fallacy fail to achieve significance. These results imply that genetic differences are important for explaining heterogeneity in the susceptibility of behavioral anomalies.

The correlations of opposite-sex DZ twins are also shown in Table 2, and these correlations tend to be smaller than the correlations of same-sex DZ twins. Even though the difference between same-sex and opposite-sex twins only approaches statistical significance in one instance (fungibility of money), the evidence should be considered in its entirety. In 10 cases out of 11, same-sex twins exhibit greater similarity than opposite-sex twins. There are several plausible interpretations of these lower opposite-sex correlations. The first is that different features of the family environment are etiologically relevant in males and females, thereby depressing their similarity. This argument is only plausible in variables for which variation in rearing conditions explains a large portion of the variation. Alternatively, a different set of genes may account for the heritable variation in males and females (Neale & Maes, 2004). A third, perhaps more speculative, possibility is that sibling imitation is stronger in same-sex sibling pairs.[8]

Table 3 provides estimates from the ACE model, using only the same-sex MZ and DZ twins, along with the likelihood-ratio-based 95% confidence intervals. The point estimates for heritability are in the range 0.16–0.42, with a median estimate of 0.24. Even using the conservative likelihood-based confidence intervals, the heritability estimates are significantly different from zero, at the 5% level, for 6 of 11 variables. As is often found in this literature, variation in family environments explain a relatively small share of phenotypic variation. In fact, the median point estimate is that the family variable explains none of the variation in susceptibility to behavioral anomalies that we observe. The remaining variance component, and by far largest in all specifications, is the so-called nonshared environment, which includes all influences on the phenotype, including noise, that are independent of genotype and rearing conditions.

[8] Of course, such interactions are absent in the standard ACE model. For a review of sibling interaction models in behavior genetics, see Neale and Maes (2004, Chap. 8).

Table 2. Sample summary statistics[9]

	Mean	S.D.	Min	Max	ρ_{MZ}	#	ρ_{DZ}	p-value of $\rho_{MZ} - \rho_{DZ}$	ρ_{DZO}
Loss aversion	0.348	0.837	0	3	0.340^{***}	10,981	0.178^{***}	0.022	0.042
Short term time preference	2.671	0.751	0	3	0.246^{***}	11,121	0.164^{***}	0.128	0.101^{**}
Procrastination	4.390	0.906	0	5	0.428^{***}	11,299	0.172^{***}	< 0.01	0.124^{***}
Default bias	1.095	0.890	0	3	0.250^{***}	10,336	0.101^{**}	< 0.01	0.140^{**}
Illusion of control	0.767	0.422	0	1	0.264^{***}	11,182	0.125^{**}	0.032	0.116^{**}
Ambiguity aversion	0.378	0.485	0	1	0.214^{***}	10,887	0.128^{**}	0.100	0.080^{*}
Fungibility of money	0.946	0.226	0	1	0.416^{***}	11,188	0.174^{*}	0.061	-0.053
Conjunction fallacy	0.163	0.369	0	1	0.252^{***}	11,060	0.048	< 0.01	-0.037
Base rate fallacy	0.120	0.325	0	1	0.289^{***}	10,987	0.175^{**}	0.142	0.151^{**}
Insensitivity to sample size	0.489	0.450	0	1	0.479^{***}	10,759	0.202^{***}	< 0.01	0.030
Representativeness	0.770	0.786	0	3	0.379^{***}	10,573	0.194^{***}	< 0.01	0.044
Cognitive Reflection (CRT)	0.813	1.131	0	3	0.590^{***}	5,323	0.302^{***}	< 0.01	0.254^{***}

Variable definitions are found in the text. Variables are coded so that a higher value corresponds to less susceptibility to the anomaly. All variables are categorical and integer valued, and a higher number corresponds to less "bias." All correlations are polychoric. ρ_{MZ}, polychoric correlations in MZ twins; ρ_{DZ}, polychoric correlation in same-sex DZ twins; ρ_{DZO}, polychoric correlation in opposite-sex twins. All results are bootstrapped. The reported p-values are one sided. Variables are coded so that a higher value corresponds to less susceptibility to the anomaly. *, **, ***Significant at the 10%, 5%, and 1% levels, respectively.

Table 3. ACE estimates for behavioral anomalies[10]

	Genetic (a^2)	Common (c^2)	Non-shared (e^2)	-2·$ln(L)$
Loss aversion	0.23^{***} $(0.13 - 0.45)$	0.11 $(0.00 - 0.33)$	0.65^{***} $(0.55 - 0.76)$	$5,949.25$
Short term time preference	0.18 $(0.00 - 0.35)$	0.07 $(0.00 - 0.26)$	0.75^{***} $(0.65 - 0.86)$	$6,259.31$
Procrastination	0.18^{**} $(0.05 - 0.24)$	0.00 $(0.00 - 0.10)$	0.82^{***} $(0.76 - 0.89)$	$10,380.28$
Default bias	0.24^{***} $(0.10 - 0.31)$	0.00 $(0.00 - 0.11)$	0.76^{***} $(0.69 - 0.82)$	$10,652.72$
Illusion of control	0.24 $(0.00 - 0.34)$	0.00 $(0.00 - 0.21)$	0.76^{***} $(0.66 - 0.86)$	$5,022.15$
Ambiguity aversion	0.16 $(0.00 - 0.29)$	0.04 $(0.00 - 0.22)$	0.80^{***} $(0.71 - 0.89)$	$6,013.37$
Fungibility of money	0.39 $(0.00 - 0.55)$	0.00 $(0.00 - 0.38)$	0.61^{***} $(0.45 - 0.78)$	$1,929.90$
Conjunction fallacy	0.22^{**} $(0.01 - 0.33)$	0.00 $(0.00 - 0.16)$	0.78^{***} $(0.67 - 0.90)$	$4,002.90$
Base rate fallacy	0.27 $(0.00 - 0.43)$	0.03 $(0.00 - 0.30)$	0.71^{***} $(0.57 - 0.86)$	$3,183.68$
Insensitivity to sample size	0.42^{***} $(0.28 - 0.50)$	0.00 $(0.00 - 0.11)$	0.58^{***} $(0.50 - 0.66)$	$6,018.48$
Representativeness	0.36^{***} $(0.21 - 0.42)$	0.00 $(0.00 - 0.11)$	0.64^{***} $(0.58 - 0.71)$	$9,381.84$

This table shows heritability estimates for the 11 variables defined in the text. A is the genetic factor; C is the common environmental factor; E is the unique environmental factor. All models are estimated allowing for different thresholds in in men and women, but the thresholds are constrained to be the same in MZ and DZtwins of the same sex. Variance components are constrained to be the same in men and women. The 95% confidence intervals shown within parentheses are constructed using likelihood ratio tests, as described in Neale and Miller (1997). Variables are coded so that a higher value corresponds to less susceptibility to the anomaly.

**, ***Significant at the 5% and 1% levels, respectively.

Finally, in Table 4 we report the correlations between the behavioral anomalies and performance on the CRT, as well as the results of the bivariate model for behavioral anomalies and CRT. To a certain extent, these correlations confirm the previously established link to cognitive ability (Stanovich & West, 1998; Stanovich, 2003; Frederick, 2005; Benjamin et al., 2006). Of the 11 anomalies, 9 are positively correlated with CRT, and 6 significantly so, reinforcing the findings of Oechssler et al. (2009). Only ambiguity aversion and procrastination are negatively correlated with CRT. However, the association between CRT and procrastination is complicated by the fact that the underlying relationship is nonmonotonic.[11] Overall, magnitudes are low to moderate, the range being 0.03 to 0.41, with a median of 0.15. Thus, higher cognitive ability as proxied by CRT is weakly associated with less susceptibility to behavioral anomalies.

Turning finally to the multivariate results, we find that the heritable variation in susceptibility to eight of the studied anomalies is positively correlated with the heritable variation in CRT scores. However, most estimates are very imprecise, and only three are statistically significant at the 5% level. For these three the correlations are, on the other hand, substantial (0.8 or above) and hence strongly suggestive of a common etiology. A partial explanation for the lack of significant correlations could be that the CRT data are only available for a subset of the SALTY survey respondents and that the phenotypic correlations between cognitive reflection and the anomalies are quite low. Our power to detect significant associations is therefore limited.

6. Discussion

In this paper we report the first set of results from a survey, unique in both its scale and scope, on the economic behavior of twins. We build on an emerging literature wherein behavioral genetic techniques are used to study the genetic and environmental sources of variation in economic decision making (Barnea et al., 2010; Cesarini et al., 2008, 2009a, b; Cronqvist & Siegel, 2011; Simonson & Sela, 2011; Zhong et al., 2009; Zyphur et al., 2009; Wallace et al., 2007). We find that well-documented behavioral anomalies such as the conjunction fallacy, loss aversion, default bias, and representativeness are moderately heritable. Overall, MZ twins consistently exhibit greater resemblance for susceptibility to behavioral anomalies than do DZ twins. Typically, genetic differences account for 20%–30% of individual variation. In sharp contrast to the genetic effects, variation in common environment accounts for only a small fraction

[11] Those who responded that they fail to pay their bills "once a month" or "several times a month" do have significantly lower CRT scores than those whose responses lie in the medium categories "once every six months" or "several times every six months." However, the respondents who "never" or "once a year" fail to pay their bills on time fare worse on the cognitive reflection task than respondents in the medium category, resulting in a negative correlation.

Table 4. Correlations between behavioral anomalies and CRT, and bivariate ACE estimates of behavioral anomalies and CRT[12]

	ρ	ρ_A	ρ_C	ρ_E	$-2 \cdot ln(L)$
Loss aversion	0.03	-0.14 $(-1.00 - 0.33)$	1.00 $(-1.00 - 1.00)$	0.02 $(-0.14 - 0.18)$	10,817.15
Short term time preference	0.17^{***}	0.31 $(-0.78 - 1.00)$	1.00 $(-1.00 - 1.00)$	0.10 $(-0.06 - 0.24)$	11,009.19
Procrastination	-0.13^{***}	-0.26 $(-0.51 - 0.08)$	-1.00 $(-1.00 - 1.00)$	0.06 $(-0.07 - 0.18)$	14,310.67
Default bias	0.08^{***}	0.10 $(-0.52 - 0.43)$	1.00 $(-1.00 - 1.00)$	-0.03 $(-0.15 - 0.08)$	15,420.42
Illusion of control	0.24^{***}	0.83^{***} $(0.31 - 1.00)$	-1.00 $(-1.00 - 1.00)$	-0.03 $(-0.18 - 0.13)$	9,750.75
Ambiguity aversion	-0.19^{***}	-0.20 $(-1.00 - 1.00)$	-1.00 $(-1.00 - 1.00)$	-0.09 $(-0.22 - 0.05)$	10,762.44
Fungibility of money	0.10	0.20 $(-1.00 - 1.00)$	-1.00 $(-1.00 - 1.00)$	0.15 $(-0.10 - 0.39)$	6,700.89
Conjunction fallacy	0.14	0.34^{*} $(-0.08 - 1.00)$	-0.67 $(-1.00 - 1.00)$	0.05 $(-0.11 - 0.21)$	8,760.65
Base rate fallacy	0.15^{**}	0.41 $(-1.00 - 1.00)$	-1.00 $(-1.00 - 1.00)$	-0.01 $(-0.20 - 0.17)$	7,951.56
Insensitivity to sample size	0.41^{***}	0.88^{***} $(0.62 - 1.00)$	-1.00 $(-1.00 - 0.68)$	0.08 $(-0.06 - 0.22)$	10,608.67
Representativeness	0.33^{***}	0.80^{***} $(0.51 - 1.00)$	-1.00 $(-1.00 - 1.00)$	0.05 $(-0.07 - 0.17)$	14,017.89

This table shows polychoric correlations and bivariate heritability estimates (Martin & Eaves, 1977) between CRT scores and the 11 variables defined in the text. All models are estimated allowing for different thresholds in men and women, but the thresholds are constrained to be the same in MZ and DZ twins of the same sex. The first column shows the polychoric correlation between the behavioral anomaly and CRT. The remaining columns show the estimated correlations between the three variance components of CRT and the corresponding variance components for each behavioral anomaly. For example, the estimated correlation between the genetic endowment for cognitive reflection and each behavioral anomaly is ρ_A. The 95% confidence intervals shown within parentheses are constructed using likelihood ratio tests, as described by Neale and Miller (1997). Variables are coded so that a higher value corresponds to less susceptibility to the anomaly.
*, **, ***Significant at the 10%, 5%, and 1% levels, respectively.

of observed interpersonal differences. Finally, a large portion of variation is due to nonshared environment. This set of results is consistent with a broad consensus in behavioral genetics, which is now so firmly established that it is often referred to as a law (Turkheimer, 2000).[13]

Our work is also closely related to a number of recent papers that study more conventional aspects of economic preferences and behavior (Wallace et al., 2007; Cesarini et al., 2009a, b; Zhong et al., 2009; Zyphur et al., 2009). In fact, our variance component estimates are quite similar to those reported by Cesarini et al. (2009a, b; 2010) for risk and giving, but somewhat lower than the heritability estimates reported by Zhong et al. (2009). Compared to the consensus estimates in the literature on the heritability of intelligence (Bouchard & McGue, 1981) and the so-called "big five" factors of personality (Jang et al., 1996), it is clear that our estimates are low. Much of this difference likely reflects noise. Indeed, we conjecture that once noise is filtered out, the heritability of most economic preferences will look quite similar to that of the "big five" in personality research.

There is some evidence in the previous literature on the heritability of susceptibility to behavioral anomalies, including the demonstration in Cesarini et al. (2010) that MZ twins are more likely to exhibit concordance in a binary variable proxying for return-chasing behavior. The first paper exclusively devoted to decision making anomalies was Simonson and Sela (2011), who administered a rich set of questions on decision making to a sample of 110 MZ and 70 DZ pairs of twins. They did not find significant heritabilities for a number of judgment heuristics and discounting, and tentatively proposed that decision tasks involving "prudence" have higher heritabilities than those that do not. This paper's findings are broadly consistent with many of the results in Simonson and Sela (2011), though we find no evidence of different heritabilities across domains.[14]

Interestingly, we find that opposite-sex DZ twins tend to be less similar than same-sex DZ twins. There are several potential, and not mutually exclusive, explanations for this finding. One is that different genes, or different features of the family environment, explain variation in men and women. Another is that there are forces not captured by

[13] A common misunderstanding is to equate high heritability with "immutability" (Goldbeger, 1979; Pigliucci, 2001; Beauchamp et al., 2011). This is a mistake for a number of reasons. One is that the causal pathway from molecular genetic variants to complex social outcomes often involve environmental variables that are manipulable. Indeed, behavioral anomalies are sensitive to subtle contextual differences and changes in the instructions (Gigerenzer, 1991). Moreover, as explained in Goldberger (1979), heritability estimates tell us what fraction of phenotypic variance is explained by genes, not whether a particular policy passes the cost benefit test.

[14] A convincing demonstration that heritabilities differ significantly across domains would require very large samples and should ideally also be based on measurement-error adjusted estimates of the variance components.

the simple ACE model, such as sibling interactions, which inflate sibling similarity. If such interactions are more intense in same-sex twin pairs, this may help explain the excess similarity of same-sex DZ twins over opposite-sex pairs without having to invoke explanations based on heterogeneous environmental or genetic effects that vary by sex.

The finding that genes can account for a considerable share of individual differences in behavioral anomalies points to a source of heterogeneity that has traditionally been somewhat overlooked in economics and finance, namely, genetic and biological variation. Obtaining a better understanding of the biological and genetic mechanisms that account for the heritable variation in anomalies is an important next step in the effort to integrate behavioral genetics into economics. There are several complementary approaches that one might take to answer this question. One strategy has been to try to directly identify hormonal and molecular genetic associates of economic preferences (for a review, see Beauchamp et al., 2011). Although this is an exciting area of research, several authors have cautioned that given the small samples typically used in these studies, the false discovery rate is likely to be very high (Benjamin, 2010; Beauchamp et al., 2011). It is as of yet an open question whether biological variables such as genetic markers and hormone levels will prove to be more reliable predictors of individual differences than traditional demographic variables.

Although the search for robust demographic correlates to behavioral anomalies has been largely disappointing, it seems clear that individuals with low cognitive ability are more prone to making economic decisions that defy standard assumptions (Stanovich & West, 1998, 2000; Stanovich, 2003; Frederick, 2005; Benjamin et al., 2006), suggesting that cognitive ability might be a mechanism mediating some of the heritable variation in the behavioral anomalies. Consistent with the findings in this literature, a majority of the anomalies in our data are positively and significantly correlated with performance on the CRT (Frederick, 2005). However, the correlations are rather small with the exception of three variables: illusion of control, insensitivity to sample size, and representativeness. These were also the only variables for which there was a significant genetic correlation between the CRT scores and the anomaly. Given that insensitivity to sample size and representativeness are by construction very similar to the type of questions used in standard tests of cognitive ability, their common origins can hardly be considered surprising. Somewhat more intriguingly, both procrastination and ambiguity aversion correlated negatively with performance on the CRT. For procrastination, the finding may be driven by a nonmonotonic relationship between the CRT score and

the response to the question. For ambiguity aversion, no obvious explanations suggests itself.[15]

There are a number of reasons to expect that a better understanding of the genetic factors that underlie individual differences in behavioral anomalies will benefit economics, finance, and the decision sciences. First, insofar as it is reasonable to classify most of these behaviors as mistakes, knowledge of the forces that generate these mistakes may provide cues about where to look for policy levers that reduce their prevalence and impact. Such information may also turn out to be helpful for understanding how various environmental conditions interact with genetic endowments (Benjamin et al., 2007). A second motivation for studying individual differences is that knowledge about which economic agents are the most vulnerable to behavioral anomalies may have implications for efforts to predict when individual-level biases can be expected to have aggregate effects. An important strand of work in economics has demonstrated that, at least in some economic environments, people who interact in a market frequently tend to exhibit behavior that more closely resembles that predicted by standard economic theory (List, 2003; Levitt & List, 2007). Much remains to be learned about the extent to which these patterns are generated by social and market forces, or if they instead, to a large extent, are driven by selection into economic environments based on individual differences, for example, in genetic endowments.

7. Conclusion

Recent years have witnessed an increasing interest in the sources of individual differences in economic and financial decision making. Many of the papers in this vein of research use behavioral genetic methodologies to separately identify the role of rearing conditions, genetic factors, and other influences. This paper extends the existing literature to an important set of behaviors that has so far received only limited attention, namely, behavioral anomalies. Using survey-based evidence from more than 11,000 Swedish twins, we demonstrate that well-documented departures from expected utility maximization are moderately heritable, with most point estimates suggesting that 20%–30% of individual variation can be accounted for by genes. These results underline the importance of genetic differences as a source of heterogeneity in susceptibility to behavioral anomalies.

[15] The existing literature offers no real clue to the rather puzzling negative correlation, because previous studies found no correlation between ambiguity aversion and cognitive ability, albeit one between ambiguity aversion and orbitofrontal lesions (Hsu et al., 2005; Borghans et al., 2009).

References

Angeletos, G. M., D. Laibson, A. Repetto, J. Tobacman, and S. Weinberg. "The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation," *J. Econom. Perspect.* 15 (2006), 47–68.

Arrow, K., *Essays in the Theory of Risk-Bearing* (Amsterdam, NL: North-Holland, 1971).

Barberis, N., and R. H. Thaler, "A Survey of Behavioral Finance," in G. M. Constantinides, M. Harris, R. Stulz (Eds), *Handbook of the Economics of Finance* (Amsterdam, NL: Elsevier, 2003).

Barnea, A., H. Cronqvist, and S. Siegel, "Nature or nurture: What determines investor behavior?," *J. Financial Econom.* 98 (2010), 583–604.

Beauchamp, J. P., D. Cesarini, M. Johannesson, M. van der Loos, P. Koellinger, P. J. F. Groenen, J. H. Fowler, N. Rosenquist, A. R. Thurik, and N. A. Christakis, "Genetics and Economics," *J. Econom. Perspect.*, forthcoming (2011).

Beauchamp, J. P., D. Cesarini, M. Johannesson, E. Lindqvist, and C. Apicella, "On the sources of the height–intelligence correlation: New insights from a bivariate ACE model with assortative mating," *Behav. Genet.* 41 (2010), 242-252.

Benjamin, D. J., "A white paper," A. Lupia (Ed.), *Genes, Cognition, and Social Behavior: Next Steps for Foundations and Researchers* (Arlington, VA: National Science Foundation, 2010).

Benjamin, D. J., S. Brown, and S. Shapiro, "Who is "Behavioral"? Cognitive Ability and Anomalous Preferences," mimeo (2006).

Benjamin, D. J., C. F. Chabris, E. L. Glaeser, V. Gudnason, T. B. Harris, D. I. Laibson, L. J. Launer, and S. Purcell, "Genoeconomics" in M. Weinstein, J. W. Vaupel, and K. W. Wachter (Eds.), *Biosocial Surveys* (Washington, DC: The National Academies Press, 2007).

Borghans, L., B. H. H. Golsteyn, J. J. Heckman, and H. Meijers, "Gender Differences in Risk Aversion and Ambiguity Aversion," *J. Europ. Econom. Assoc.*, **7** (2009), 649-658.

Bouchard Jr, T. J., and M. McGue, "Familial Studies of Intelligence: A Review," *Science* 212 (1981), 1055–1059.

Bouchard Jr, T. J., and M. McGue, "Genetic and Environmental Influences on Human Psychological Differences," *J. Neurobiol.* 54 (2003), 4–45.

Camerer, C. F., *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton, NJ: Princeton University Press, 2003).

Camerer, C. F., and G. Loewenstein, "Behavioral Economics: Past, Present, Future," in C. F. Camerer, G. Loewenstein, M. Rabin (Eds), *Advances in Behavioral Economics.* (Priceton, NJ: Princeton University Press, 2004).

Camerer, C. F., G. Loewenstein, and D. Prelec, "Neuroeconomics: How Neuroscience Can Inform Economics," *J. Econom. Lit.* 43 (2005), 9-64.

Cesarini, D., C. T. Dawes, J. H. Fowler, M. Johannesson, P. Lichtenstein, and B. Wallace, "Heritability of Cooperative Behavior in the Trust Game," *Proc. Natl. Acad. Sci. USA* 105 (2008), 3721–3726.

Cesarini, D., C. T. Dawes, M. Johannesson, P. Lichtenstein, and B. Wallace, "Experimental Game Theory and Behavior Genetics," *Ann. NY Acad. Sci.* 1167 (2009a), 66–75.

Cesarini, D., C. T. Dawes, M. Johannesson, P. Lichtenstein, B. Wallace, "Genetic Variation in Preferences for Giving and Risk-Taking," *Quart. J. Econom.* 124 (2009b), 809–842.

Cesarini, D., M. Johannesson, P. Lichtenstein, Ö. Sandewall, and B. Wallace, "Genetic Variation in Financial Decision Making," *J. Finance* 65 (2010), 1725–1754.

Choi, J. J., D. Laibson, B. Madrian, and A. Metrick, "Optimal Defaults," *Amer. Econom. Rev. Papers Proc.* 93 (2003), 180–185.

Cohen, J. L., and W. T. Dickens, "A Foundation for Behavioral Economics,". *Amer. Econom. Rev. Papers Proc.* 92 (2002), 335–338.

Cosmides, L., and J. Tooby. "Better than Rational: Evolutionary Psychology and the Invisible Hand," *Amer. Econom. Rev. Papers Proc.* 84 (1994), 327–335.

Cronqvist, H., and S. Siegel, "The origins of savings behavior," *Amer. Finance Assoc. Annual Meeting 2011, Denver* (2011)..

Dickens, W. T., and J. R. Flynn, "Heritability Estimates versus Large Environmental Effects: The IQ Paradox Resolved," *Psychol. Rev.* 108 (2001), 346–369.

Dominicus, A., A. Skrondal, H. K. Gjessing, N. L. Pedersen, and J. Palmgren, "Likelihood Ratio Tests in Behavioral Genetics: Problems and Solutions,". *Behav. Genet.* 36 (2006), 331–340.

Ellsberg, D., "Risk, Ambiguity, and the Savage Axioms," *Quart. J. Econom.* 75 (1961), 643–669.

Fama, E., "Market Efficiency, Long-Term Returns, and Behavioral Finance," *J. Financial Econom.* 49 (1998), 283–306.

Frederick, S., "Cognitive Reflection and Decision Making," *J. Econom. Perspect.* 19 (2005), 25–42.

Frederick, S., G. Loewenstein, and T. O'Donoghue, "Time Discounting and Time Preference: A Critical Review," *J. Econom Lit.* 40 (2001), 351–401.

Fudenberg, D., "Time Discounting and Time Preference,". *J. Econom. Lit.* 44 (2006), 694–711.

Gigerenzer, G., "How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases"," *Europ. Rev. Soc. Psych.* 2 (1991), 83–115.

Gigerenzer, G., R. Hertwig, and U. Hoffrage, "Cognitive Illusions Reconsidered," in C. R. Plott and V. L. Smith (Eds), *Handbook of Experimental Economics Results, vol. 1* (Amsterdam, NL: Elsevier, 2008).

Goldberger, A. S., "Twin Methods: a Skeptical View," in P. Taubman (Ed.), *Kinometrics: Determinants of Socioeconomic Success Within and Between Families* (Amsterdam, NL: North-Holland, 1977).

Goldberger, A. S., "Heritability," *Economica* 46 (1979), 327–347.

Harrison, G. W., and J. List. "Field Experiments," *J. Econom. Lit.* 42 (2004), 1009–1055.

Hsu, M., M. Bhatt, R. Adolphs, D. Tranel, and C. F. Camerer, "Neural Systems Responding to Degrees of Uncertainty in Human Decision-Making," *Science* 310 (2005), 1680–1683.

Isacsson, G, "Estimating the economic return to educational levels using data on twins," *J. App. Econom.* 19 (2004), 99–119.

Jang, K. L., W. J. Livesley, and P. A. Vernon, "Heritability of the Big Five Personality Dimensions and Their Facets: A Twin Study," *J. Pers.*, 64 (1996), 577–592.

Jencks, C., "Heredity, Environment, and Public Policy Reconsidered," *Amer. Sociology Rev.*, 45 (1980), 723–736.

Kahneman, D., and A. Tversky, "Subjective Probability: A Judgment of Representativeness," *Cogn. Psych.* 3 (1972), 430–454.

Kahneman, D., and A. Tversky, "On the psychology of prediction," *Psychol. Rev.* 80 (1973), 237–251.

Kahneman, D., and A. Tversky, "Choices, Values, and Frames," *Amer. Psychol.* 39 (1984), 341–350.

Kahneman, D., and A. Tversky, "On the Reality of Cognitive Illusions," *Psychol. Rev.* 103 (1996), 582–591.

Laibson, D., "Golden Eggs and Hyperbolic Discounting," *Quart. J. Econom.* 112 (1997), 443–477.

Langer, E. J., "The Illusion of Control," *J. Pers. Soc. Psych.* 32 (1975), 311–328.

Levitt, S., and J. List., "What Do Laboratory Experiments Measuring Social Preferences Tell us About the Real World?," *J. Econom. Perspect.* 21 (2007), 153–174.

Lichtenstein, P., P. F. Sullivan, S. Cnattingius, M. Gatz, S. Johansson, E. Carlström, C. Björk, M. Svartengren, A. Volk, L,. Klareskog, U. de Faire, M. Schalling, J. Palmgren, and N. L. Pedersen, "The Swedish Twin Registry in the Third Millennium: An Update," *Twin Res.* 9 (2006), 875–882.

List, J., "Does Market Experience Eliminate Market Anomalies?," *Q. J. Econom.* 118 (2003), 41–71.

Lykken, D. T., M. K. McGue, and A. Tellegen, "Recruitment Bias in Twin Research: The Rule of Two-Thirds Reconsidered," *Behav. Genet.* 17 (1987), 343–362.

Madrian, B., and D. F. Shea, "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior," *Quart. J. Econom.* 116 (2001), 1149–1187.

Martin, N., and L. Eaves, "The Genetic Analysis of Covariance Structure," *Heredity* 38 (1977), 79–95.

Neale, M. C., and M. M. Miller, "The Use of Likelihood-Based Confidence Intervals in Genetic Models," *Behav. Genet.* 27 (1997), 113–120.

Neale, M. C, and H. Maes, *Methodology for Genetic Studies of Twins and Families* (Dordrecht, NL: Kluwer Academic Publishers, 2004).

Oechssler, J., A. Roider, and P. W. Schmitz, "Cognitive abilities and behavioral biases," *J. Econom. Behav. Organ.* 72 (2009), 147–152.

Pigliucci, M., *Phenotypic plasticity: beyond nature and nurture* (Baltimore, MD: Johns Hopkins University Press, 2001).

Plomin, R., J. C. DeFries, and G. E. McClearn, and P. McGuffin, *Behavioral Genetics*, 5th ed. (New York, NY: Worth Publishers, 2009).

Rabin, M., "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica* 68 (2000), 1281–1292.

Samuelson, W., and R. Zeckhauser, "Status-Quo Bias in Decision-Making," *J. Risk Uncertainty* 1 (1988), 7–59.

Simonson, I., and A. Sela, "On the heritability of consumer decision making: An exploratory approach for studying genetic effects on judgment and choice," *J. Consumer Res.* 37 (2011), 951–966.

Sloman, S. A., "The empirical case for two systems of reasoning," *Psych. Bull.* 119 (1996), 3–22.

Stanovich, K. E., *Who Is Rational? Studies of Individual Differences in Reasoning* (Mahwah, NJ: Lawrence Erlbaum Associates, 2003).

Stanovich, K. E., and R. F. West, "Individual Differences in Rational Thought,". *J. Exp. Psych. [Gen.]* 127 (1998), 161–188.

Stanovich, K. E., and R. F. West, 2000. "Individual differences in reasoning: Implications for the rationality debate?," *Behav. Brain Sci.* 23 (2000), 645–726.

Thaler, R. H., "Mental Accounting Matters," *J. Behav. Decision Making* 12 (1999), 183–206.

Turkheimer, E., "Three Laws of Behavior Genetics and What They Mean," *Curr. Dir. Psych. Sci.* 9 (2000), 160–164.

Tversky, A., and D. Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science* 185 (1974), 1124–1131.

Tversky, A., and D. Kahneman, "The Framing of Decisions and the Psychology of Choice," *Science* 211 (1981), 453–458.

Tversky, A., and D. Kahneman, "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgement," *Psychol. Rev.* 90 (1983), 293–315.

Tversky, A., and D. Kahneman, "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *J. Risk Uncertainty* 5 (1992), 297-323.

Wallace, B., D. Cesarini, P. Lichtenstein, and M. Johannesson, "Heritability of Ultimatum Game Responder Behavior," *Proc. Natl. Acad. Sci. USA* 104 (2007), 15631–15634.

Zhong, S., S. H. Chew, E. Set, J. Zhang, H. Xue, P. C. Sham, R. P. Ebstein, and S. Israel, "The Heritability of Attitude Toward Economic Risk," *Twin Res. Human Genet.* 12 (2009), 103–107.

Zyphur, M.J, J. Narayanan, R. D. Arvey, and G. J. Alexander, "The Genetics of Economic Risk Preferences," *J. Behav. Dec. Making*, 22 (2009), 367–377.

Appendix A

In this appendix, we present the English translations of the survey questions that were used in this paper.

7.1. Loss aversion.

Q1 Imagine that you could take part in a lottery where a coin toss determines if you win 2,000 SEK ("heads" in the coin toss) or lose 1,000 SEK ("tail" in the coin toss). Would you choose to participate in the lottery?

Yes
No

Q2 Imagine that you could take part in a lottery where a coin toss determines if you win 1,500 SEK ("heads" in the coin toss) or lose 1,000 SEK ("tail" in the coin toss). Would you choose to participate in the lottery?

Yes
No

Q3 Imagine that you could take part in a lottery where a coin toss determines if you win 2,500 SEK ("heads" in the coin toss) or lose 1,000 SEK ("tail" in the coin toss). Would you choose to participate in the lottery?

Yes
No

7.2. Self-control.

7.2.1. *Short-term time preference.*

Q1 Imagine having to choose between receiving a sum of money today or waiting to receive a larger sum in one week. Which would you choose?

5,000 SEK today
6,000 SEK in a week

Q2 Imagine having to choose between receiving a sum of money today, or waiting to receive a larger sum in one week. Which would you choose?

5,000 SEK today
7,000 SEK in a week

Q3 Imagine having to choose between receiving a sum of money today or waiting to receive a larger sum in one week. Which would you choose?

5,000 SEK today
5,500 SEK in a week

7.2.2. *Procrastination.*

How often does it happen that you do not pay bills on time?

Never
Once a year
Once every six months
Several times every six months
Once a month
Several times a month

7.3. Default bias.

Q1 A few years ago it became possible to switch electricity provider. Have you switched electricity provider since this possibility was introduced?

Yes
No

Q2 A few years ago it became possible to switch telephone operator from Telia. Have you switched telephone operator since this possibility was introduced?

Yes
No

Q3 Have you left the Church of Sweden since the separation from the state?

Yes
No, I am still in the Church of Sweden
No, I did not belong to the Church of Sweden at the separation from the state

7.4. Illusion of control.

Imagine that you could participate in one of the two lotteries below, where the chance of winning is the same.

Lottery 1: You are allocated a lottery ticket and every one in a thousand of the participants will win 10,000 SEK.
Lottery 2: You can pick a lottery ticket yourself and every one in a thousand of the participants will win 9,000 SEK.

Which of these two lotteries would you choose?

Lottery 1
Lottery 2

7.5. Ambiguity aversion.

Imagine that there is an urn with 30 red balls and 60 other balls that are either black or yellow. The number of balls of each color is determined in advance, but you do not know the exact number of balls that are black or yellow, only that the total number is 60. The balls are well mixed so that every ball has the same chance of being drawn. Imagine that you can draw one ball from this urn and that you have to choose between the following three lotteries. Which lottery would you choose?

Lottery A: You receive 900 SEK if a red ball is drawn.
Lottery B: You receive 1,000 SEK if a black ball is drawn.
Lottery C: You receive 1,000 SEK if a yellow ball is drawn.

7.6. Fungibility of money.

Q1 Imagine that you have decided to watch a play that costs 100 SEK. When you enter the theatre to buy the ticket you discover that you have lost a 100 SEK bill. Will you still pay 100 SEK to watch the play?

Yes
No

Q2 Now imagine that you have decided to watch a play and that you have already bought a ticket for 100 SEK. When you enter the theatre you discover that you have

lost the ticket. It is impossible to get a refund for the lost ticket. Would you buy a new ticket for 100 SEK?

Yes
No

7.7. The representativeness heuristic.

7.7.1. *Conjunction fallacy.*

Linda is 31 years old, single, outspoken and very talented. She has a university degree in philosophy. As a student she was very involved in discrimination and social justice issues. She also participated in several anti nuclear demonstrations. Which of the following alternatives is the most likely?

A: Linda works in a bank.
B: Linda works in a bank and is active in the feminist movement.

Alternative A
Alternative B

7.7.2. *Base Rate Fallacy.*

Kalle is attractive, athletic, drives a Mercedes and has a very attractive girlfriend. Which of the following alternatives is the most likely?

A: Kalle is a professional tennis player
B: Kalle works as a nurse.

Alternative A
Alternative B

7.7.3. *Insensitivity to Sample Size.*

There are two hospitals in a city. In the big hospital, 45 children are born every day, and in the small hospital 15 children are born everyday. On average, 50% of the children born are boys, but it varies from day to day. In which hospital do you think that it is most likely that more than 60% of the children born are boys in a specific day?

The big hospital
The small hospital

7.8. Cognitive Reflection.

We used exactly the same three items as in Frederick (2005) and hence do not reproduce them here.

Appendix B

In this appendix, we sketch the assumptions needed to estimate bivariate heritability (Martin & Eaves, 1977). Readers interested in a more comprehensive development are referred to Beauchamp et al. (2010). We write

$$P_1 = a_1 A_1 + c_1 C_1 + e_1 E_1,$$

$$P_2 = a_2 A_2 + c_2 C_2 + e_2 E_2,$$

where P_1 and P_2 are the two traits of interest, and all variables are expressed as deviations from zero with unit variance. In matrix notation, we can write this as

$$P = \tilde{a}\tilde{A} + \tilde{c}\tilde{C} + \tilde{e}\tilde{E},$$

where $\tilde{a}$ is a diagonal two-by-two matrix whose diagonal elements are a_1 and a_2, and $\tilde{A} = [A_1, A_2]'$. Here, $\tilde{c}$, $\tilde{C}$, $\tilde{e}$, and $\tilde{E}$ are defined similarly. We impose the same assumptions as in the univariate case. In addition, we assume that all elements of $\tilde{A}$, $\tilde{C}$, and $\tilde{E}$ are mutually uncorrelated. Stacking the observations in each family, $Y = (P_1, P_2, P_1^*, P_2^*)'$, we have the following symmetric matrix of moment conditions:

$E(YY') =$

$$\begin{bmatrix} a_1^2+c_1^2+e_1^2 & & & \\ \rho_A a_1 a_2+\rho_C c_1 c_2+e_1 e_2 \rho_E & a_2^2+c_2^2+e_2^2 & & \\ R_i a_1^2+c_1^2 & R_i \rho_A a_1 a_2+\rho_C c_1 c_2 & a_1^2+c_1^2+e_1^2 & \\ R_i \rho_A a_1 a_2+\rho_C c_1 c_2 & R_i a_2^2+c_2^2 & \rho_A a_1 a_2+\rho_C c_1 c_2+e_1 e_2 \rho_E & a_2^2+c_2^2+e_2^2 \end{bmatrix}$$

where $R_i = 0.5$ if the observation is from a DZ pair and 1 otherwise. Here, the parameter ρ_A is the correlation between the standardized latent genetic endowments A_1 and A_2, and ρ_C and ρ_E are analogously defined. Given the standardization, it is readily verified that there are nine independently informative equations and hence that there are exactly enough moments to identify the parameters of the model. The genetic correlation squared is the R_i^2 from a hypothetical regression of the genetic endowment of the first trait on the genetic endowment of the second trait. A high R_i^2 is consistent with the hypothesis that genetic variance in cognitive reflection mediates some of the heritable variation in the behavioral anomalies. However, although a high genetic correlation is suggestive, it does not prove mediation. For example, two traits could be explained by entirely different sets of genes and yet be correlated because of linkage disequilibrium.[16]

[16] Linkage disequilibrium refers to the tendency for alleles that are located near each other on a chromosome to be inherited together.

PAPER 5

Is there a long-term effect of Africa's slave trades?

Margherita Bottero Björn Wallace

Abstract. Nunn (2008) found a negative relationship between past slave exports and economic performance within Africa. Here we investigate these findings and the suggested causal pathway in further detail. Extending the sample period back in time we reveal that the coefficient on slave exports did not become significantly negative until 1970, and that it was close to zero in 1960. One potential explanation for this temporal pattern could be decolonization. But, somewhat puzzlingly, we document a similar downward slope for the relationship between slave raids contemporaneous to the African slave trades and economic performance within Italy. In addition, our reading of the historical and anthropological literature differs from that of Nunn. For instance, taking a global rather than African perspective we find that the African slave trades cannot without difficulties explain the patterns of ethnic fractionalization that we observe today.

1. Introduction

A great number of hypotheses have been put forth in order to explain Africa's anomalously poor economic performance (Rodney, 1972; Easterly & Levine, 1997; Sachs & Warner, 1997; Collier & Gunning, 1999; Acemoglu, Johnson & Robinson 2001; Hernández-Catá, Schwab, & Lopez-Claros, 2004). One popular theory claims to have traced its roots back to past slave trades, and in particular, the trans-Atlantic slave trade. A recent proponent of this theory is Nathan Nunn (2008) who has uncovered evidence in favor of a substantial negative relationship between past slave exports and current economic performance within Africa. Nunn argues that his evidence taken together with the historical literature suggests that slave trade had an adverse effect on economic development and that the most likely causal pathway goes via impeded state formation and increased ethnic fractionalization. However, despite its intuitive appeal and supporting data, we argue that the long run effects of slave trade are not necessarily as clear-cut as Nunn's reasoning would lead us to believe. In particular we

We would like to thank Robert A. Foley for kind permission to reprint Figure 1 as well as Tore Ellingsen, Magnus Johannesson, Juanna Joensen, Erik Lindqvist, Tino Sanandaji, Yoichi Sugita and seminar participants at the Stockholm School of Economics for their helpful comments and suggestions. We gratefully acknowledge the Knut and Alice Wallenberg Research Foundation for financial support.

show that a too narrow focus on a single year, 2000, as the outcome variable and a single continent, Africa, as the sample space may be driving his conclusions.

By extending the sample period in Nunn (2008) back to 1960 we demonstrate that the coefficient for past slave exports is declining over time and that it is not significantly below zero before the 1970s. While this finding is in line with Nunn's suggestion that the economic effects of past slave exports did not necessarily manifest themselves until after the decolonization of Africa, we also uncover a number of empirical irregularities in the data. Most notably, the coefficient for past slave exports is often positive instead of negative for those countries that produced oil at any point in time during the sample period. Further, our reading of the historical and anthropological literature is quite different from that of Nunn. In Africa, there is an obvious latitudinal gradient in ethnic fractionalization which is associated with GDP as well as past slave exports (Easterly & Levine, 1997; Alesina et al., 2003; Nunn, 2008). Nunn suggests, presumably in light of this, that the causal mechanism from slave exports to current economic performance goes via ethnic fractionalization. It is argued that the exogenous demand for slaves led to a decrease in trust and an increase in conflict, which in turn impeded state formation and contributed to modern day ethnic fractionalization. That ethnic fractionalization is associated with poor economic performance is well known, although the actual causes remain disputed (Easterly & Levine, 1997; Alesina et al., 2003). What is undisputed, however, is that lower latitudes are strongly associated with higher ethnic and linguistic diversity, not only in Africa, but also globally (Cashdan, 2001; Collard & Foley, 2002). This empirical relationship is perhaps little known outside the rather narrow field of human biogeography, but it mirrors an extensive and earlier literature that documents a latitudinal gradient in present, and past, species diversity, within, as well as across, regions and taxa (Pianka, 1966; MacArthur, 1972; Rosenzweig, 1992, 1995).

While these facts do not disprove, or particularly undermine, the negative association between past slave trade and current economic performance for Africa as a whole, they do suggest that if there is a causal relationship, then it is likely to be more complex, and less straightforward, than what is commonly believed. In fact, to complicate matters further, we demonstrate that there is also a negative relationship between past slave raids and current economic performance within Italy.[1] On the surface this evidence is supportive of a negative relationship between past slave trades and GDP per capita, although such a relationship has admittedly rarely, if ever, been suggested for Italy. However, like for Africa, going back to 1960, the trend of the coefficient on slave raids is downward sloping. Thus, due to the absence of decolonization, the Italian

[1] As measured by regional per capita GDP in 2000.

evidence arguably casts doubt on, rather than support, the hypothesis that past slave exports negatively affected economic development in post-independence Africa.

The remainder of the paper is organized as follows. Section 2 investigates the long run economic consequences of slave trade in more detail. Section 3 reviews the anthropological and historical literature, while section 4 describes the data and the empirical strategy. The empirical results are presented in section 5 and we conclude with a discussion of our findings in section 6.

2. The economic consequences of slave trade

The first and most immediate effect of Africans selling other Africans to non-Africans in return for various commodities and precious metals is obviously an increase of capital and a decrease of labor in Africa. Further, in the presence of decreasing marginal products of capital, land and labor, the marginal and average product of labor for the remaining African population that was not exported as slaves should have increased, *ceteris paribus*. That a large scale catastrophe, like the slave trades were for large parts of Africa, can cause an increase in average and marginal productivity may perhaps seem counter-intuitive, or even perverse, to some, but such outcomes are well documented (Lee, 1973; Findlay & Lundahl, 2002, 2006; Clark, 2005; Pamuk, 2007). In this respect, the economic consequences of the slave trades are clearly analogous to the those of the Black Death in medieval Europe[2], but with the added twist of an influx of capital. However, in contrast to the Black Death the slave trades did not target individuals in an approximately random fashion. Rather, the demand for slaves was concentrated to young and healthy males, in the trans-Atlantic slave trade, or females, in the trans-Saharan slave trade (Phillips, 1985).[3] Thus, since young adult males were arguably the most productive group in pre-industrial societies (Kaplan et al., 2000), selection may potentially have offset other immediate effects of slave exports on GDP per capita, leaving the overall effect unresolved in the absence of more detailed data.

The exact nature and mechanisms behind medium and long run effects of slavery on economic performance are in many ways even less clear than those for short run effects. This pertains in particular to the type of effects that we study here, that is, effects that are observed several decades, or centuries, after the actual trades themselves stopped. Not only does the testing of long run effects involve a Kierkegaardian leap of faith in terms of a large set of untestable exogeneity, homogeneity and *ceteris paribus*

[2] Similar outcomes have been observed also for the Justinian plague in 6^{th} century Egypt and to a lesser extent for the Antonine plague in 2^{nd} century Egypt (see for instance Findlay & Lundahl, 2006; Scheidel 2010a).

[3] Interestingly this suggests that short, and long, term effects of the slave trades may have been heterogenous across Africa. Nunn (2008) implicitly assumes homogenous effects.

assumptions, it also relies on the existence of causal and permanent effects on factors that determine not only aggregate, but also per capita production. These are stringent requirements, in particular the latter one, which asks that any such effects hold also after accounting for population dynamics. Importantly, the method employed by Nunn (2008), regressing current GDP on past slave exports, imply that hundreds of years of history is treated as a black box, an approach that Austin (2008) has critically dubbed "compression of history", and that is particularly sensitive to a number of key assumptions. For instance, in his regressions Nunn (2008) control for colonizer fixed effects, but these are only partially complete and do not cover neither the prolonged Bantu colonization of Eastern and Southern Africa that continued well into the 19^{th} century nor any of its contemporary Arabic incursions into sub-Saharan politics (Gray, 1975; Flint, 1977). Nonetheless, it is not difficult to construct an argument for why an observed association between past slave trade and current economic performance could be causal. As we have seen, the exogenous demand for slaves was targeted to the two groups, young and healthy men and women, that were key to both production and reproduction in pre-industrial societies. The exogenous demand for slaves may also have channelled effort and resources away from productive to destructive and rent-seeking activities such as arms races and slave raiding.

According to Nunn (2008) this is approximately what happened. The African slave trades led to a demographic collapse, the corruption of existing legal systems, increased conflict, an environment of mistrust and smaller and less well functioning states. Slaves were often acquired through raiding, or alternatively sometimes via the legal system. The latter method arguably undermined the rule of law, while slave raiding contributed to create an environment of mistrust and violence. These developments were further fuelled by the influx of European weapons which helped to create an arms race accompanied by a vicious circle of violence that is often referred to as the guns for slave cycle.[4] Compared to other, earlier, episodes of slave trading the African slave trades were unique both in their scale and in how they turned individuals of the same, or similar, ethnicities against each other. The foreign demand for slaves spurred neighboring individuals, groups and villages to raid each other. This in turn lead to the breakdown of existing states and impeded the formation of larger and more well functioning states. Nunn's suggested causal mechanism thus go from past slave trades to current economic

[4] There were arguably precursors. In the trans-Saharan slave trade there was a substantial horses for slaves trade predating the guns for slave trade that is typically associated with the trans-Atlantic slave trade. These horses were often used for military purposes as well as for slave raids and this trade is in many respects analogous to the guns for slave trade. See for instance Phillips (1985) and Fisher (2001).

performance via these phenomena, ethnic fractionalization, weak government, corruption and low levels of trust, which have all been found to associated negatively with per capita GDP and growth.

To test if past slave trades did in fact have a negative effect on current GDP Nunn runs the following regression,

$$\ln y_i = \beta_0 + \beta_1 \ln(\text{exports}_i/\text{area}_i) + \mathbf{C}_i'\delta + \mathbf{X}'\gamma + \varepsilon_i \tag{2.1}$$

where $\ln y_i$ is the natural logarithm of real per capita GDP in country i in year 2000. $\ln(\text{exports}_i/\text{area}_i)$ is the natural logarithm of the total number of slaves exported from country i between 1400 and 1900, normalized by land area. $\mathbf{C}$ and $\mathbf{X}$ are vectors that control respectively for colonizers prior to independence and a set of geographic and climatic variables. The baseline sample consists of all African countries, but a number of these are dropped in robustness checks which also normalize exports by historical population rather than area, along with varying the exact number and composition of the controls. In addition to the standard OLS the author also runs an IV regression where approximate distances from a country to the location of demand for slaves are used as instruments in order to demonstrate causality and overcome measurement error.

The evidence that emerges from this exercise is very much supportive of there being a negative effect of past slave trades on current GDP per capita as the coefficient on exports is typically negative and significant for both the OLS and IV regressions. If correct, the negative effect is not only statistically significant, but also economically so. In Table 1 we report the average GDP per capita across continents in 2000 along with the counterfactual for Africa in the absence of slave exports. Of course, such an exercise is sensitive to a number of assumptions, but as long as it is taken with a pinch of salt it may nonetheless be instructive. As can be seen, according to the OLS estimate for the specification (5) that was used as an example by Nunn (2008) average African GDP per capita would be approximately 50% higher in the absence of slave trade. While this is indeed economically significant it still does not go very far in explaining Africa's comparative economic underdevelopement. In this example, if there were no slave trades Africa's GDP per capita would increase from 38% to 56% of that of the second poorest continent (Asia) and from 5% to 8% of the richest (North America). The calculations thus support the findings from the sub-Saharan countries that were reported in Birchenall (2009a). That is, the OLS estimates reported in Nunn (2008) indicate that past slave exports can explain different economic outcomes within Africa, but only very little of the difference between Africa and the other continents. The IV estimates do however tell a somewhat different story. Using the counterfactual

for the corresponding IV specification we find that in the absence of slave trade African GDP would have been 3.75 times higher than what it is now.[5] Thus, average African GDP per capita would have been 43% higher than the Asian and almost at par with the Latin American.

Table 1. Average GDP across regions in 2000[6]

Region	GDP pc in 2000	Ratio	Ratio (OLS)	Ratio (IV)
North America	27,873	0.05	0.08	0.19
Oceania	20,819	0.07	0.10	0.26
Europe	12,620	0.11	0.17	0.43
Latin America	5,889	0.25	0.36	0.92
Asia	3,791	0.38	0.56	1.43
Africa	1,447	1	1.47	3.75
Africa, counterfactual (OLS)	2,128	0.68	1	2.55
Africa, counterfactual (IV)	5,427	0.27	0.391	1

If we choose to accept the evidence that Nunn (2008) presents, it still raises a number of questions. Slavery is an ancient institution that until recently was widespread across the globe (Phillips, 1985). Today, Africa is by far the poorest continent. Why then, did slavery cause poverty in Africa, but not elsewhere? Though perhaps not central to detecting a negative association between past slave trades within Africa and current economic performance, it is an important motivating question. Many countries that historically faced extensive slave taking, raiding and trading are today much richer than any African country. Thus, there does not seem to exist a deterministic relationship between slave trade and economic performance. By claiming that the African slave trades were unique Nunn (2008) at least partially circumvents this critique.

[5] When interpreting these estimates it is important to keep in mind that in addition to the standard caveats the coefficient on slave exports for the IV regression used in calculating the counterfactual is significant at the 10%, but not at the 5%, level. In addition, of the four instruments the three for the Islamic slave trades are insignificant while that for the trans-Atlantic is significant at the 10% level. The IV estimate that the calculations are based on is the only one in the paper, (3), which uses both the (almost) full set of controls and the whole sample.

[6] Note: GDP pc in 2000 is the population weighted continental real GDP per capita in 2000 based on Maddison (2010). Ratio is the African GDP per capita in 2000 divided by that of the other continents. Ratio (OLS) is ditto for the counterfactual OLS estimates. Ratio (IV) is ditto for the counterfactual IV estimates. Armenia was counted as belonging to Europe along with those countries that have a majority of the population that is of European descent.

3. The historical and anthropological evidence

The conclusion that past slave trades had a negative effect on modern-day per capita GDP hinges critically on a number of factors. Crucially, Nunn (2008) argues that the African slave trades were unique[7] for three reasons i) their volume ii) how they turned individuals of same of similar ethnicities against each other and iii) how they corrupted existing legal structures. These features of the African slave trades were also those that supposedly led to ethnic fractionalization, and in turn lower GDP per capita. Hence, given this mechanism, the uniqueness of the African slave trades and their causal link to ethnic fractionalization becomes central to assessing the validity of the claim that the African slave trades led to depressed economic performance. In what follows we will scrutinize these three aspects of the slave trades in detail, collapsing the latter two into one, and thereafter we proceed to take a closer look at how ethnicity was imputed and the validity of the instruments. Doing this, we find that the African slave trades were neither unique in scale nor scope. In addition, we raise a number of concerns as regards the causal relationship between slave trades and ethnic fractionalization.

But, before proceeding we would like to comment on a couple of important issues that require less detailed attention. The first concerns slavery as an institution, while the second concerns the definition of Africa. Both of these issues relate to selection, but they are also of a more general interest for evaluating the long run effects of slave trade. In particular, if there were some African societies that had slavery prior to the slave trades, and if there were those that did not, then this could be a potential source of selection into the slave trades. Reviewing the literature Nunn (2008) finds, although not explicitly stating so, that at the onset of the trans-Atlantic slave trade there was no indigenous slavery in sub-Saharan Africa and that slavery only existed in those African societies that were part of the older Islamic slave trades. However, this conclusion stands in stark contrast to the opinions of a number of authorities such as Phillips (1985), Thornton (1998), Lovejoy (2000), Fage & Tordoff (2002) and Austin (2008). Most notably, Austin (2008) addresses Nunn's reading of the literature directly:

> "Finally (and contrary to Nunn, 2008: 139), there is no dispute in the specialist Africanist literature today that 'domestic' slavery, while not universal, pre-dated the Atlantic slave trade, and not only in areas involved in the Islamic slave trades. The debate is between those who argue that slavery was widespread in Africa and slave trading routine (Thornton, 1998: 73–97), and those who think that they were as yet common only within a few societies, and thus became widespread

[7] Interestingly, he makes this claim without *any* reference to slave trades outside Africa.

> only as a joint product of the Atlantic slave trade (Lovejoy, 2000)" (p. 1006).

Another, and perhaps more important, concern is the definition of the sample as well as of Africa. First, it is important to remember that Africa is a culturally, economically and genetically diverse continent, and hence not necessarily a natural unit for studying the effects of slave trades (Gray, 1975; Flint, 1977; Oliver, 1978; Cavalli-Sforza, Menozzi & Piazza, 1994; Anon., 2008). North Africa belongs, both culturally and genetically, to the Mediterranean and middle Eastern sphere and is in many respects different from sub-Saharan Africa. Similarly, the island nations in the sample have little in common with mainland Africa. At the onset of the trans-Atlantic slave trade four of these were uninhabited and two were inhabited by a mix of Polynesian, Arab and Bantu people that were primarily of non-sub-Saharan African extraction. In addition, at the time parts of southern Africa was sparsely populated by archaic hunter-gatherers (Oliver, 1978; Lee, 1979; Klein, 2009). Thus, the sample used in the main analysis contains not only exporters and non-exporters of slaves, but also importers as well as countries that were uninhabited. Although, these problems are partially overcome in sensitivity analyses, the issue whether the modern geographic definition of Africa is a natural unit of analysis for studying the economic consequences of slave trades remain. As we shall see these concerns will also be important for evaluating the effect of slavery on ethnic fractionalization as well as for the relationship between selection into slavery and development.

3.1. Volume. Like all slave trades, the African slave trades were unique. While it is possible that the absolute volume of these trades were greater than those of other slave trades, this is not necessarily true. More importantly, the African slave trades were most likely not unmatched in scale. Rather, what makes the African slave trades unique in terms of volume is the comparatively detailed source material. Although there are many gaps, and some debate, we can with reasonable confidence estimate the extent of the trans-Atlantic slave trade. The *Voyages: The Trans-Atlantic Slave Trade Database* which maps the vast majority of the trans-Atlantic slave voyages estimates that this trade alone shipped some 12.5 million slaves from Africa. That is, approximately 34,000 slaves per year during more than 350 years. Although the trades were a persistent feature from the second half of the 15^{th} century and onwards they were in fact highly skewed towards the period 1750-1850 which saw some 60% of aggregate exports. Thus, at its zenith an average of approximately 75,000 slaves were shipped yearly across the Atlantic, with an estimated peak of 117,644 in 1829. This is no doubt a lot of people, and it has been suggested that the probability of being captured as a slave during a lifetime in coastal West and Middle Africa could at times have been as

high as 9.3% (Whatley & Gillezeau, 2009). But, does millions and millions of people make it unique? While there is little actual data for other episodes of slave trading the evidence suggests that this is not the case. Staying in Africa, the trans-Saharan slave trade was also substantial. In his paper Nunn (2008) uses estimates from Austen (1979) who calculated that approximately 7.5 million slaves were exported across the Sahara.[8] The trans-Saharan slave trade was however much more prolonged and less intense than the trans-Atlantic, and it lasted approximately from 650 to 1900 with a peak average of 14,000 slaves a year for the period 1800-1880. While these numbers are both much smaller and less precise than those for the trans-Atlantic trade they illustrate that although the latter is the most studied and well known slave trade, there are many other trades of note.

Perhaps most importantly, late republican and early imperial Rome is often considered to have been the foremost so called slave society in history, arguably surpassing even the New World in the colonial era (Phillips, 1985; Bradley, 1994; Scheidel, 2010b). While there is plenty of evidence as to the nature, and approximate extent, of slavery in Rome, little is known about actual trades and the true number of slaves that were involved. Nonetheless, off the cuffs estimates suggest that in total some 100-200 million people would have been held in slavery by the ancient Romans, and that at its height several millions of slaves lived within the boundaries of the Roman empire (Scheidel, 2010b). All of these slaves had to come from somewhere, and even if we allow for organic growth, and the fact that there were a number of slave reservoirs, it is likely that some areas must have supplied Rome with slaves at the same rate as Africa did for the New World, at least if we adjust for population. For example, Gaul remained an important source of Roman slaves for centuries. It has been suggested that this trade dates back at least to the 6^{th} century BC and that it peaked in the late republic and early empire (Bradley, 1994; Nash Briggs, 2003). Estimates on the extent of this trade are hard to come by, but an often quoted number based on written and archaeological sources puts it at 15,000 per year in peacetime (Tchernia, 1983). Uncertain as it may be, this estimate is still barely half of that for the trans-Atlantic slave trade, and substantially smaller than those for the peak years. However, Caesar himself is said to have taken some 1 million slaves in the Gallic wars, and although these numbers are notoriously unreliable and exaggerated it is not unlikely that the number of slaves taken from Gaul at the height of the Caesar's wars may well have equalled that for Africa in 1829 (see discussions in Tchernia, 1983; Bradley 1994; Scheidel, 2007, 2010b). Importantly, if we adjust for population a different picture emerges. Table 2 reports the estimated number of slaves taken during the Atlantic, Saharan and Gallic slave

[8] This estimate is highly uncertain (Austen, 1979; Austin, 2008).

trades along with the yearly fraction of the estimated population for the Atlantic and Gallic slave trades as well as their ratios.

Table 2. Average slave exports, and exports as a fraction of population[9]

	Atlantic	Saharan	Gaul (I)	Gaul (II)	Gaul (III)	Gaul (IV)
Yearly average	34,211	5,960	15,000	15,000	15,000	15,000
Ratio	1	5.74	2.28	2.28	2.28	2.28
Fraction	0.0005	–	0.0019	0.0014	0.0008	0.0004
Ratio	1	–	0.27	0.37	0.60	1.34

Of course, all estimates except those for the number of slaves shipped across the Atlantic are highly uncertain. A fact that is particularly emphasized by the use of no less than four different population estimates ranging from below 8 million to 40 million for Gaul during the high empire. Interestingly, using the population data from Maddison (2010) the estimated yearly exports as a fraction of the population for the trans-Atlantic slave trade is 0.27-0.37 of that from Gaul. Thus, while we have little actual data, the evidence that is available strongly suggests that the African slave trades were not unique in scale.

Likewise, the discussion on the demographic impacts of the slave trades is somewhat lopsided and builds exclusively on Manning's (1990) work which relies crucially on rather obscure assumptions. Unlike Manning many authors, including Malthus (1817) and Fage & Tordoff (2002), do not believe that the long run demographic impact of the Atlantic slave trades were substantial. In fact, this is the prediction from a parsimonious application of Malthusian theory in contrast to Manning's (1990) "[some] thirty pages of source code, written in the Pascal language" (p. 179). If the carrying capacity of land did not change due to the slave trades, then the population should, *ceteris paribus,* have recovered, but the continuous outflow of men and, in particular, women in reproductive age could at least in principle have caused a delay. It is of course also possible that the slave trades did in fact lead to a more permanent downward shift

[9] Note: Atlantic refers to the average number of slaves exported from Africa in the trans-Atlantic slave trade, as estimated in the *Voyages: The Trans-Atlantic Slave Trade Database*. Fraction of population was calculated using linear extrapolation from aggregate 25-year export data and linear extrapolation of population data from Maddison (2010). Gaul (I) uses a low estimate of 8 million for the population in Gaul during the high empire based on the population data for Europe in Maddison (2010). Gaul (II) uses a high estimate of 11 million based on the same data in Maddison (2010). Gaul (III) uses the lower bound of 15 million for the population of Gaul during the high empire in Ferdière (1988). Gaul (VI) uses the upper bound of 40 million in Ferdière (1988).

in population, but the less sophisticated and complex the economic system, the less likely this is to have happened. Either way, that the slave trade led to a demographic collapse is neither a fact nor the default hypothesis. Rather, it is a position held by some scholars.

3.2. The slave trades and ethnic fractionalization. Today, Africa arguably has a higher ethnic diversity than any other continent (Easterly & Levine, 1997; Fearon, 2003). While diversity is often seen as a blessing, there is also a well established negative association between ethnic fractionalization and economic performance. One explanation for Africa's, and sub-Saharan in particular, ethnic fractionalization could be the slave trades. According to Nunn (2008) the African slave trades were not only unique in their scale, but also "because, unlike previous slave trades, individuals of the same or similar ethnicities enslaved each other" (p. 142) a fact which in turn led to "particularly detrimental consequences, including social and ethnic fragmentation, political instability and a weakening of states, and the corruption of judicial institutions" and, in the extension, impaired economic performance. The idea that evil begets evil is intuitively appealing, and has received substantial support as regards the African slave trades. Several historians and economists argue that the exogenous demand for slaves led to conflicts, destabilized existing states, impeded state building and territorial expansion (Rodney, 1972; Lovejoy, 2000; Whatley, 2008; Whatley & Gillezeau, 2009; Nunn & Wantchekon, 2010). However, the introduction of economic incentives to military might and increased fire power does not necessarily lead to smaller states and ethnic groups in the long run. In fact, many of the larger and most famed precolonial Africa states such as the Asante, Benin, Bornu, Dahomey, Ghana, Kanem, Kano, Kongo, Mali, Songhay and Wolof were intimately linked to slave holding, raiding and trading (Thornton, 1982, 1998; Phillips, 1985; Meillassoux, 1991; Lovejoy, 2000; Fisher, 2001; McIntosh, 2001; McCaskie, 2002), and they most likely, at least in part, owe their modern day reputation to slaving activities. Just like for the *ad hoc* evidence, there are theoretical predictions that go in both directions, and many specialists are much less assertive than Nunn when it comes to drawing conclusions on the effects of the slave trades on the size of past and present states and ethnic fractionalization (Phillips, 1985; Thornton, 1998; Austin, 2008). In fact, there are even those who in contrast to Nunn (2008) favor an interpretation where the slave trades may have contributed, although not necessarily uniformly, to the centralization and strengthening of states (for instance, Fage, 1969; Klein, 1992). Thus, the question of whether slavery caused ethnic fractionalization ultimately becomes an empirical question. Here we do not try to prove or disprove causality, but rather we ask i) if the African slave trades really were the only slave trades that turned individuals of same and similar ethnicities

against each other and ii) whether if looking at the anthropological and biogeographical evidence from a global perspective there are alternative, and more likely, explanations for the pattern of ethnic fractionalization that we observe in Africa today.

3.2.1. *Were the African slave trades unique in enslaving individuals of same or similar ethnicities?* To prove the uniqueness of the African slave trades with respect to how they influenced social, ethnic and political institutions and interactions, Nunn (2008; Nunn & Wantchekon, 2009) provides the reader with a wealth of historical evidence. For those readers, like us, who are not Africanists the information is mostly new, somewhat overwhelming and rather convincing, which makes it all the more easy to loose sight of the fact that existence is a necessary, but not sufficient, condition for uniqueness. The fact that the African slave trades led to the corruption of judicial systems and the enslavement of individuals of same and similar ethnicities does not make them unique. It only makes them unique if similar effects did not occur elsewhere. In fact, returning to Europe and the ancient Roman empire there is strong evidence in favor of the African slave trades not being unique in how they enslaved individuals of same and similar ethnicities. Roman laws made provisions for the enslavement of Romans, primarily via the right to enslave foundlings, although such behavior was typically not encouraged (Phillips, 1985; Bradley, 1994; Scheidel, 2007). Rather, like most states the government tried to counteract the enslavement of their own citizens, but debt and penal slavery meant both scope and incentives for trying to abuse the legal system, and kidnappings, as well as rare captures in civil wars, occurred from time to time. With respect to non-citizens the Romans were less scrupulous. Given the scale of the Gallic slave trade and raiding it is hardly surprising to find that not only Romans, but also those of Gallic ethnicity were involved. A well know anecdote from Diodorus (1939) describes how "many of the Italian traders...believe that the love of wine of these Gauls is their own godsend...for in exchange for a jar of wine they receive a slave, getting a servant in return for the drink" (p. 167). Of course, we do not know where these slaves came from, but they were most likely of Gallic or German origin, a hypothesis that receives strong support both from geography and the prevalence of slaves from these areas in Rome. In fact, it has been suggested that some Gallic tribes and elites were involved in raiding slaves from their neighbors already during the early iron age, slaves that were subsequently sold to the Etruscans (Nash Briggs, 2003).

Like Gaul, Germania was a Roman slave reservoir (Phillips, 1985; Bradley, 1994; Scheidel, 2007), and again like for Gaul not much is known about German slave taking and trading, but we can at least glean some information from Tactitus (1948) who in describing the gambling habits of Germans notes the following:

> "The loser goes into slavery without complaint; younger or stronger he may be, but he suffers himself to be bound. Such is their perverse persistence, or, to use their own word, their honour. Slaves of this sort are sold and passed on, so that the winner may be clear of the shame that even he feels in his victory" (p. 121).

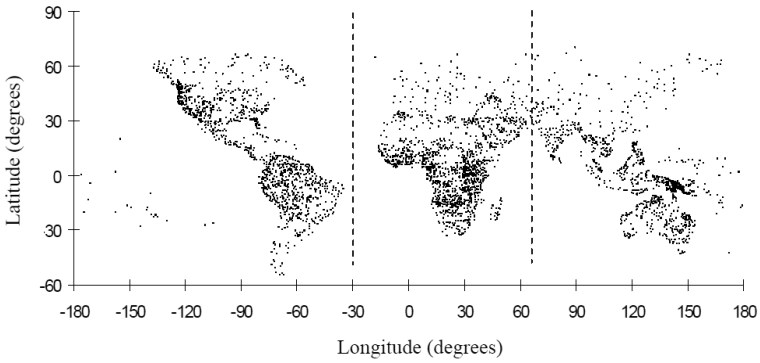
While we most certainly cannot trust every detail of what Diodorus and Tacitus tell us, their writings do illustrate that the enslavement, and subsequent trade, of compatriots most likely happened not only in Africa during the trans-Atlantic slave trade, but also in Europe in Roman days. Similarly, warring Gallic and German tribes were probably involved in both the capture and trading of their enemies as slaves, enemies that most likely were of neighboring and similar ethnicities (Rives, 1999; Nash Briggs, 2003; Scheidel, 2007; Smith, 2009). Interestingly, and in contrast to what we would expect in the presence of negative long term effects of slave exports, only a few hundred years after these slave trades had subsided somewhat Rome itself was overrun by people from Gaul and Germania, and today, the region that was formerly known as Gaul is one of the very richest in the world.

In fact, Europe faced several additional and more recent episodes of slave raiding and trading. Perhaps most notably, from the Muslim invasion of the Iberian peninsula and onwards Spain suffered a great number of wars, skirmishes and slave raids (Phillips, 1985). Christians enslaved neighboring Muslims, and vice versa. These "internal" struggles ended with the end of the Reconquista in 1492, but the Muslim slave raids did not. Pirating activities and coastal raids in the Mediterranean using North Africa as base continued well into the 19^{th} century and plagued in particular Spain and Italy. Although these raids are relatively unknown, it has been estimated that as many as 1,250,000 Europeans were captured as slaves by Saracen pirates between 1530 and 1780 (Davis, 2003). Importantly, while the Saracen captains were Muslim, many were Christian renegades who occasionally even returned to haunt their own home regions. Thus, the experiences of Roman Europe, and early modern Mediterranean stand in sharp contrast to Nunn's (2008) claim that the African slave trades were unique "because, unlike previous slave trades, individuals of the same or similar ethnicities enslaved each other" (p. 142).

3.2.2. *Ethnic fractionalization.* Despite its importance in Nunn's argument little effort is made to disentangle the empirical relationship between past slave trades and ethnic fractionalization, and we have to make do with a sketchy story and a significant correlation. We agree with Nunn that such a correlation exists, however we believe that it is to a large extent spurious. The reason for this is that it is easy to demonstrate that there is not only a significant correlation between past slave exports and

ethnic fractionalization, but also one between past slave exports and latitude, as well as one between latitude and ethnic fractionalization. Based on biogeographical and anthropological evidence we believe that the latter association is of first order and that it temporally precedes the one between slave trades and ethnic fractionalization. If we zoom out of Africa and instead study the world as a whole, we quickly discover that the pattern between ethnic fractionalization and latitude that Nunn observes for Africa holds also globally, despite there being no other continents that were substantial early modern net exporters of slaves. This relationship is illustrated in Figure 1 and has been demonstrated by Nettle (1999), Cashdan (2001) and Collard and Foley (2002) who show that absolute latitude explains a large fraction of diversity for a number of ethnic and linguistic measures both within and between continents. Interestingly, the pattern is very similar for present and past species diversity, for which the latitudinal gradient has been well know since at least the turn of the last century (Pianka, 1966; MacArthur, 1972; Rosenzweig, 1992, 1995).

Figure 1. Distribution of human cultures by longitude and latitude[10]



Two potentially important explanations for why human ethnic and linguistic diversity is correlated with latitude, without having to resort to more fundamental explanations of species diversity itself, are carrying capacity and species diversity (Cashdan, 2001; Foley & Collard, 2002; Moore et al., 2002). Carrying capacity of land for a given technological level is correlated with population density in a Malthusian world and higher population density should reasonably, *ceteris paribus*, lead to higher ethnic

[10] Note: From Collard and Foley (2002). Reprinted with kind permission of Professor Robert A. Foley. Each dot represents a culture in the *Atlas of World Cultures*.

diversity. Similarly, in a Malthusian world carrying capacity is presumably correlated with latitude. High species diversity on the other hand can also be expected to be associated with higher ethnic and linguistic diversity as it increases the number of ecological niches and may consequently contribute both to increased specialization and hamper expansion. Thus, it is hardly surprising that measures of productivity, climatic variability, species and habitat diversity have been found to correlate with linguistic and ethnic diversity on a global as well as African scale.

While it is true that the latitudinal gradient in ethnic fractionalization is global, there are also indications that the relationship is stronger in Africa than elsewhere (Foley & Collard, 2002). Again, the exact reason for this is unclear, and it is indeed possible that there is a second order effect that goes from slave trades to ethnic fractionalization. However, this is not necessarily the most parsimonious explanation. Africa is also the continent that has been inhabited the longest by anatomically modern humans (Klein, 2009). Hence, there are reasons to believe that ethnic differentiation amongst hunter-gatherers may have been higher in Africa than in more recently colonized continents (Ahlerup & Olsson, 2007). Also genetic diversity is substantially higher in Africa than elsewhere and most scholars agree that this is primarily due to the deeper evolutionary history of our species on this continent (Cavalli-Sforza, Menozzi & Piazza, 1994). It is natural to assume that higher genetic diversity, all else equal, contributed to higher ethnic diversity. In addition, the Neolithic revolution begun later in sub-Saharan Africa than elsewhere, and the spread of agriculture is likely to have contributed to decreased ethnic fractionalization (Ammerman & Cavalli-Sforza, 1971; Smith, 1995; Zohary & Hopf, 2000; Diamond, 2002; Ahlerup & Olsson, 2007). In fact, this relationship potentially holds also within Africa, where the comparatively recent Bantu expansion went from west to east and from centre to south, a pattern that mimics that of the spatial distribution of ethnic fractionalization. We consider most of these explanations, again predating the slave trades, more parsimonious and most likely more important in explaining diverging African ethnic diversity than the slave trades themselves. Independently of whether one agrees with us or not, it is safe to say that there are many competing explanations for why there is a correlation between slave exports and latitude. We also note that the causality could in fact not only be spurious, but also reversed. If ethnic fractionalization was higher in tropical Western and Middle Africa, then it is not unlikely that these areas selected into the slave trades, at least in theory.

The idea that the pattern of ethnic fractionalization and comparatively low economic development that we see in Africa today can be explained by events and environmental factors that pre-date the slave trades have received support in a number of

recent papers that are methodologically related to Nunn (2008), but that go further back in time. For instance, Garner (2006) found that African population densities in 1500 are correlated with current economic performance. Similarly, using large and medium sized cities as proxy for development Birchenall (2009a) argues that already pre-modern Africa was lagging behind. In addition, Birchenall documents that pathogenic stress is associated with current economic performance. A finding that is replicated by Bhattacharyya (2008) who augments the approaches of Acemoglu, Johnson and Robinson (2001) and Nunn (2008) with a measure of malaria risk, and finds that this particular environmental factor seems to be more important for determining current economic outcome than either colonial institutions or the slave trades. Although we can of course not be sure, it is not unreasonable to assume that the ecological determinants of disease environments in Africa today at least in part overlap with the ecological determinants of disease environments prior to the slave trades. A strong indication that this was indeed the case is given by Cashdan (2001) who found that disease environment is related to ethnic diversity.

Studying even more fundamental parameters, Olsson and Hibbs (2005), Michalopoulos (2009), Motamed, Florax and Masters (2009), Beck and Sieber (2010) and Fenske (2010) find that standard biogeographic variables are associated with pre-colonial institutions and current economic performance, presumably via their effects on the transition to agricultural production, agricultural practices, urbanization and state development. These findings have received additional support from Putterman (2008) who demonstrates a more direct relationship between the onset of agriculture and current economic performance. In a similar vein, Ahlerup and Olsson (2007), Birchenall (2009b) and Ashraf and Galor (2009) argue that age of human settlement, age of state experience, and distance from the putative home of anatomically modern humans influence ethnic diversity as well as economic performance. More specifically, Ashraf and Galor (2009) find a hump-shaped relationship between genetic diversity and economic performance while Spolaore and Wacziarg (2009) find that genetic distance from the world's technological frontier is associated with GDP per capita. Taken together, these papers suggests that the patterns of ethnic diversity and economic performance observed by Nunn (2008) need not be driven by the slave trades, rather they could with ease be attributed to other, more fundamental, factors that pre-date the slave trades.

3.3. Did really the most advanced African societies select into the slave trades? According to Nunn (2008) "evidence show that it was actually the most developed areas of Africa that tended to select into the slave trades" (p. 140). A finding that the author himself refers to as a "seemingly paradoxical relationship" and that may indeed sound a bit puzzling. But, is it true? Nunn builds his argument on demographic

and ethnographic evidence.[11] It is true that most demographic evidence suggests that West and Middle Africa was more densely populated than Southern Africa. In addition, it is arguably not difficult to show that there were many kingdoms in Middle Africa that were more administratively and technologically advanced than what it is often assumed. This does, however, not amount to proving that these were the most developed African societies. True, large parts of Southern Africa was still inhabited by archaic hunter-gatherer populations that were not particularly technologically advanced (Oliver, 1978; Lee, 1979; Klein, 2009). Even so, there are three important problems with Nunn's (2008) reasoning. The first refers to selection. The ethnographic accounts that he uses to demonstrate the political and technological sophistication of Middle African societies are missing for large parts of Africa, including most parts of Middle Africa (Gray, 1975; Flint, 1977; Oliver, 1978; Thornton, 1998; Austin, 2008). The actual available historical evidence outside a few kingdoms that were in comparatively frequent contact with Europeans and Arabs is typically scant. Hence, what is at most demonstrated is that some societies were at least as, or more, developed that the reader would have expected, not that they were more developed than other contemporary African societies.

Secondly, since net primary productivity has a latitudinal gradient, a high population density does not imply a high level of development in and of itself (Cramer et al., 1999; Galor & Weil, 2000; Austin, 2008; Birchenall, 2009b). Rather, they typically reflect extrapolations based on the carrying capacity of the land and 20^{th} century population densities (see for instance Austin, 2008; Hopkins, 2009).[12] That being said, Middle Africa was probably substantially more developed than most parts of Southern Africa (Gray, 1975; Flint, 1977; Oliver, 1978; Lee, 1979). There are many strong indications that such was the case. It is however not necessarily true that Middle Africa was more developed than Eastern Africa. For instance, the Arabic influence was comparatively strong in the east, and there were complex city structures such as the famed great Zimbabwe.

Thirdly, and perhaps most importantly, there seems to be somewhat of a confusion as regards the definition of Africa in the paper. The statement that tropical Africa was the most developed part of sub-Saharan Africa is rather uncontroversial and it may

[11] Interestingly, a recent paper by Fenske (2010) has demonstrated that for Africa the existence of slavery is associated with agricultural suitability. This finding presumably introduces a link between technological development and slavery, but not necessarily with slave exports. Another indication in this direction is given by Bezemer, Bolt and Lensik (2009) who find that not only export slavery, but also the existence of indigenous slavery is negatively associated with current GDP per capita.

[12] Addressing Nunn's (2008) use of historical populations Austin (2008, p. 1002) writes "Projecting back from already very dubious figures for ca. 1900 or ca. 1920, there is simply no epistemological basis for Nunn's use of the word 'data'— literally, 'things that are given' or granted—to refer to the guesses that have been made about the population of future African countries in 1400 (Nunn, 2008: 158)."

be possible to stretch it as far as to claim that the west was more developed than the east, but that is far from given. However, by claiming that the most advanced African societies selected into the slave trades Nunn (2008) goes one step further. In effect, by doing so he indirectly claims that sub-Saharan Africa was more developed than North Africa. In contrast, laymen and professional historians alike would probably disagree and instead conclude that North Africa was substantially more developed than sub-Saharan Africa at the onset of the slave trades. The eastern part of North Africa neighbors, and is sometimes included in, the fertile crescent where the Neolithic revolution begun. Consequently, it was home to one of the first civilizations known to man, ancient Egypt (Gray, 1975; Flint, 1977; Oliver, 1978). In addition, large parts of North Africa were integrated parts of the Hellenic and Roman empires, and at the onset of the trans-Atlantic slave trade most of North Africa was an integral part of the Muslim and Mediterranean world. In fact, the only way that the areas that make up the North African countries of today may have selected into the slave trades were as slave importers. Thus, we have to conclude that in his empirical strategy Nunn treats the whole of Africa as Africa, but that elsewhere in the paper he seemingly refers only to sub-Saharan Africa. This is somewhat problematic, and once again highlights the issue of whether Africa as a geographical entity is the appropriate sample/unit of analysis.

3.4. The imputation of ethnicity[13]. The key variable in Nunn's (2008) analysis is total slave exports per country. While the number of slaves shipped from each country during the trans-Atlantic slave trade can quite easily be calculated with reasonable confidence the actually number of slaves exported is much more difficult to approximate since trade with the Europeans primarily took place at trading posts along the west coast of Africa. Similarly, both the actual number of slaves exported in the Islamic slave trades and their origins are highly uncertain. This makes the novel and unusually careful estimates of slave exports that Nunn provides one of the main contributions of his paper. To impute the number of slaves exported from each country Nunn extracted the ethnicities of tens of thousands of slaves from administrative records. These ethnicities were then mapped and aggregated onto the modern day countries of Africa so that they could be used to calculate the ratios of slaves exported from each coastal country relative to the land locked countries further inland. While we are sympathetic towards the methodology and the attempt, we still believe that

[13] Throughout this section the calculations are based on the estimates in *Voyages: The Trans-Atlantic Slave Trade Database*. This database covers the period 1514-1866. Nunn (2008) uses data from Elbl (1997) for earlier trades, we do not. However, only one of the 54 samples used by Nunn to impute ethnicity belongs to this earlier period.

there are some problems that may seriously bias the analysis. First, ethnicities are not always stable and there were substantial migrations in early modern African history. In fact, as pointed out by Austin (2008) the idea that you can easily map past ethnicities to modern day countries in Africa goes counter to much of the work that has been done in ethnography. But, even if mapping is in theory possible at least one problem remains. The sample of administrative records that Nunn (2008) uses is not likely to be representative of past slave trades as it is presumably based on what records are readily available today rather than a random sample of all slaves that were exported in the African slave trades.

Indeed, a quick look at Table 3 and how ethnicities were imputed reveals that there is reason for concern. For the trans-Atlantic slave trade 54 different samples were used, however, two single points in time and space, Trinidad 1813 and Sierra Leone 1848, represent almost a third (30.9%) of the total sample. In fact, the 50 year period 1801-1850 is overrepresented by a factor of almost two in Nunn's sample, while the 50 year period 1701-1750 is barely represented at all despite its share of aggregate exports being approximately 20%. Further, if we would instead look at single years within each time period the bias would become even more apparent. Some of this temporal bias is however mitigated by the fact that many of the slaves were presumably shipped to the New World years prior to them being recorded in the censuses, notarial and other administrative records that were used for imputation. Unfortunately, it is not possible to disentangle when the slaves in Nunn's sample arrived to the New World. Equally, and perhaps more, damaging is the spatial bias of the sample. For instance, more than 15% of Nunn sample is made up by slaves registered in Trinidad in 1813. It has been estimated that some 44,000 slaves disembarked in Trinidad and Tobago between 1600 and 1826, representing 4 per mill of the aggregate estimated trans-Atlantic slave trade. That is, slaves that were shipped to Trinidad are arguably more than 37 times overrepresented in the sample that was used to impute ethnicity. Similarly, observations in Sierra Leone in 1848 make up more than 15% of the sample although slaves disembarked in Africa only made up less than 1.5% of total exports, and likewise, early colonial Peru[14] is overrepresented by a factor of about seven and Haiti by a factor greater than two. In contrast, Brazil, which was by far the leading slave importer with more than 45% of total disembarkations, is underrepresented by a factor greater than three and Jamaica with about 9.5% is not represented at all. Thus, Trinidad, Sierra Leone, early colonial Peru and Haiti, which make up almost 60% of Nunn's sample, only represent approximately 10.5% of the aggregate trade while

[14] Although Nunn's sample(s) for Peru start in 1548 and end in 1702 the careful reader may nonetheless want to disregard this estimate with reference to the previous footnote.

Jamaica and Brazil, which imported well over half of the total number of exported slaves, only make up some 13% of Nunn's sample. For the Islamic slave trade, the bias is likely to be worse. The ethnicities of slaves shipped across the Indian ocean is imputed by using six samples and for those shipped across the Sahara and the Red Sea there are two samples. While for two of these trades the samples at least contain several thousand individuals, for the latter they only contains 67 slaves of 32 ethnicities.

Table 3. Examples of bias in data used in the imputation of ethnicity[15]

	1701-50	1801-50	T&T	Jam	Bra	4 Nunn	2 Voyages
Nunn	0.0063	0.5610	0.1545	0.000	0.1287	0.5978	0.1287
Voyages	0.2045	0.2913	0.0041	0.0953	0.4545	0.1052	0.5498
Ratio	0.03	1.93	37.59	–	0.28	5.68	0.23

3.5. The instruments. To deal with measurement error and to establish causality Nunn (2008) develops four instruments, one for each slave trade. The method of instrumental variables provides a general solution to the problem of an endogenous explanatory variable. However, the resulting estimates will be identified only if two conditions are satisfied. First, the instrument must be uncorrelated with the error term, and secondly the instrument has to be partially correlated with the instrumented variable once the other controls have been netted out. While the second condition is testable, the first has to be maintained. In this section we put forward reasons to believe that both requirements are only weakly satisfied, and in particular that (i) the instrument may be correlated with the error term, and (ii) the instrument may be too weakly correlated with the instrumented variable. These concerns, if founded, would mean that the IV estimates could be biased and inconsistent (Wooldridge, 2002; Stock & Yogo, 2010).

The chosen instruments are more or less identical in construction, and they are all derived from proxies for the distance needed to transport slaves from their country of origin to the location of demand. While geographic instruments are popular, they are also often problematic. Here, we in particular worry that the location of demand and supply for slaves could have common and unobserved determinants that are correlated

[15] Note: Nunn is the fraction of slaves in Nunn's (2008) sample that was recorded during the specified time period/location. Voyages is ditto for slaves shipped/disembarked in *Voyages: The Trans-Atlantic Slave Trade Database*. Ratio is Nunn divided by Voyages. T&T is Trinidad and Tobago. Jam is Jamaica. Bra is Brazil. 4 Nunn is T&T, Sierra Leone, Haiti & early colonial Peru. 2 Voyages is Bra & Jam.

with underlying geographic variables, which in turn are correlated with the instruments. An indication that this could indeed be a concern is given by the fact that once the climatic and geographic controls are included in the first stage regression for the full sample none of the instruments are significant at the 5% level and the F-statistic drops below 2.[16] As the trans-Atlantic slave trade was the most important African slave trade, and given that it was the only one for which the instrument was significant at the 10% level for all specifications, we will focus our attention to this instrument.

First of all, the location of demand for slaves in the New World was non-random and primarily driven by the suitability of land for cash crops, and in particular sugar (Fogel, 1989; Engerman & Sokoloff, 2002). Sugar is a crop that requires high temperatures, sunshine, moisture, limited draught and that is most commonly grown in the tropics (Bakker, 1999), a fact which introduces a geographic component into the location of demand for slaves. Given the taxing nature of manual labor in sugar production it would hardly be surprising if plantation owners demanded slaves that were suitable and well adopted for hard physical labor in high temperatures and humidity as well as for life in tropical disease environments. Elsewhere in his paper, Nunn (2008) argues that the ethnicity of a slave was an important and reliable label that had real economic meaning, and this is a reasoning that is in part borne out by the legal preoccupation with slave ethnicities in Rome (Bradley, 1994). While these labels may not always have been accurate, we tend to agree with Nunn (2008), but we suspect that one of the reasons for their importance was that slaves from different areas of Africa may have been differently suited for work in sugar plantations.

Presumably, slaves that were suitable for working the cane fields were to be found in areas with similar climates to those where sugar was grown. If true, this introduces a common geographic component into the location of demand and supply for slaves that could potentially be correlated with the instrument. In fact, two out of Nunn's three measures of climate are negatively correlated with the instrument and positively correlated with slave exports. Moreover, when these are introduced into the first stage of the IV regression along with the other controls the instrument becomes, as mentioned above, insignificant at the 5% level. Likewise, although malaria was most likely introduced to the New World either by Europeans or their African slaves, it soon became indigenous and a major source of sickness (Packard, 2007). Thus, it is hardly surprising to find that a measure of the fraction of the population in African countries that live in areas of high intensity malaria transmission (Rowe et al., 2006) is highly negatively

[16] More specifically, the F-statistic is 1.73 and the p-values for the instruments range from 0.093 (trans-Atlantic slave trade) to 0.507 (Red Sea). In contrast, without controls all instruments but the one for the Red Sea slave trade ($p = 0.998$) are individually significant at the 1% level, and the F-statistic is 15.4. For further details, see page 162 in Nunn (2008).

correlated with the instrument, and positively correlated with slave exports. Indeed, if we add a sugar production dummy based on production data from *FAOSTAT* to the first stage regression the instrument becomes insignificant also at the 10% level for the full sample with geographic controls. Further, adding the malaria index results in none of the coefficients being close to significance.[17] In fact, like for Nunn's (2008) reduced sample, the instrument for the Red Sea trade, becomes positive. While this last result is particularly intriguing, it must be said that it comes at the cost of reducing the sample from 52 to 44 countries.

4. The data and the empirical methodology

Before proceeding with the analysis we present the data we use in greater detail. Throughout, while we share some of the misgivings about the demographic and ethnic variables that Austin (2008) voiced, we will take Nunn's (2008) data as given.

4.1. Data.

4.1.1. *Africa.* In this paper we extend the analysis of Nunn (2008) back in time. In order to do so we take the data made available by the author on his webpage[18] as given and augment it with the variables that change over time, using the same sources as Nunn.[19] Data on per capita GDP for the years 1960 to 2006 was obtained from the most recent update by Maddison (2010). Similarly, like Nunn, we use data on oil, diamond and gold production from the British Geological Survey's *World Mineral Statistics/Production.*[20] In contrast to Nunn who uses 31-year averages we use annual production in our analysis. Since the production figures overlap between editions we always use the most recent available estimate. In a few instance when there were obvious typos with too many, or too few, zeros these were cross-checked with other editions of the said publication, or with data from the *U.S. Energy Information Administration*, and corrected. Unfortunately there is no separate data on gold production in Burundi

[17] If we add the sugar dummy to the fist-stage the p-values range from 0.146 (Atlantic) to 0.818 (Saharan), while they range from 0.349 (Atlantic) to 0.645 (Red Sea) if we add the malaria measure. Finally, if we add both the sugar dummy and the malaria measure the p-values range from 0.318 (Saharan) to 0.729 (Red Sea), with that for the trans-Atlantic distance instrument being 0.491.

[18] http://www.economics.harvard.edu/faculty/nunn

[19] Note that we follow Nunn (2005, 2008) in treating Ethiopia and Eritrea as one country.

[20] The name *World Mineral Statistics* was used in publications printed between 1978 and 2004. From 2005 and onwards the publication is named *World Mineral Production.* Prior to 1978 the publication was known as *Statistical Summary of the Mineral Industry - World Production, Exports and Imports.* All of these publications are available on the web from the British Geological Survey's *World mineral statistics archive.* The data on oil and gold from the 1960s were converted from long tons to tonnes, and ounces to kilos, respectively.

and Rwanda for the years 1960-1962, only information on joint production. To overcome this problem we impute the production for these years using country specific aggregate gold production 1963-2006 as weights.[21]

4.1.2. *Italy.* Our analysis for Italy tries to mimic that for Africa. Data on slave raiding for the years 1530-1780 was collected from Davis (2003).[22] While data on regional per capita GDP in 2000 were obtained from Istat (2011), and data on regional areas from *Nationalencyklopedin.* Coastal length was obtained from *APAT – Servizio Difesa delle Coste*, and climatic data from the *CRU CL 2.0* dataset (New et al., 2002).[23] Note that this data provides average humidity rather than average maximum afternoon humidity. Historical regional GDP was taken from Paci and Saba (1997). All GDP data was converted to 1990 International Geary-Khamis dollars using Maddison (2010) in order to make estimates more readily comparable to those for Africa. Finally, centroids for the regions were obtained from the *NGA GEOnet Names Server.*

4.2. Empirical methodology. In our empirical analysis we have gone to great lengths to be as faithful as possible to the spirit of Nunn (2008). For Africa, we simply extend his econometric analysis from the single year 2000 to every year in the period 1960 to 2006. This is a natural extension, and our motivation for doing so is that we want to be able to assess the stability over time for the relationship between slave trades and economic performance. Following Nunn, we study six different specifications of equation (2.1). Of these, specifications (3) and (6) are evaluated for a reduced sample of 42, rather than 52, countries. The countries dropped in these two specifications are the North African countries Morocco, Algeria, Tunisia, Libya, Egypt and the island nations Seychelles, Mauritius, Comoros, São Tomé and Príncipe, and Cape Verde. Table 4 summarizes the differences between the six specifications. Note that in contrast to Nunn we focus exclusively on the OLS regressions, leaving the more contentious IV regressions aside.

[21] The results are not sensitive to this choice, since joint production was negligible 1960-1962.

[22] We only include observations for which an actual number or estimate was given. If an observation included locations in more than one region, the aggregate number was divided equally among the locations before mapping these into their respective regions.

[23] The dataset provides estimates 10 minute latitudes and longitudes. Our measures are calculated as the weighted, by distance, value using the values for the nearest coordinates to the center of each region.

Table 4. Summary of the controls in specification (1)-(6), Africa

Specification		Controls
(1)		colonizer fixed effects
(2)	controls in (1) +	distance from the equator, longitude, minimum monthly rainfall, average maximum humidity, average minimum temperature, and proximity to the ocean measured by the natural log of coastline divided by land area
(3)	controls in (2) −	island and North African countries
(4)	controls in (2) +	Islam indicator, French legal origin, island and North Africa fixed effects
(5)	controls in (4) +	natural log of the annual average per capita production between 1970 and 2000 of gold, oil, and diamonds
(6)	controls in (5) −	island and North African countries

To study the effects on past slave raids on economic performance within Italy, we replicate the above analysis for Italian regions. Whenever possible, we use the same controls as for Africa, and if unavailable, we use the closest possible substitute. While the economic, geographic and climatic data do not present any additional problems to those described in 4.1.2, the colonizer fixed effects pose a more difficult challenge. Africa has experienced extensive colonization, and accordingly, Nunn includes colonizer fixed effects to "control for the other significant event in Africa's past, colonial rule" (p. 154). The history of Italy is rich in similar type events[24], making it difficult to define what regions were 'colonized', and by whom. In fact, several regions have been subjected to a series of 'colonizers'. Moreover, we are dealing with a relatively small sample of no more than 20 regions, which precludes the possibility of having a large set of fixed effects dummies. Therefore, we include only a dummy indicating if a particular region was under the dominion of the House of Savoy prior to unification. The reason for this is that Italy was united under the auspices of Savoy, and hence theirs is the political entity that stands out most significantly in modern Italian history.

[24] For instance, the Spanish ruled the Kingdom of the two Sicilies for a prolonged period of time and the Pope controlled large fractions of central Italy. In fact, Trentino-Alto Adige and Trieste were not integrated into Italy until the end of the Great War.

Table 5. Summary of the controls in specification (1), (2) and (4), Italy

Specification	Controls
(1)	Savoy fixed effects
(2)	controls in (1) + distance from the equator, longitude, minimum monthly rainfall, average humidity, average minimum temperature, and proximity to the ocean measured by the natural log of coastline divided by land area
(4) & (5)	controls in (2) + island and south of Italy fixed effects

Besides the colonizer dummies, the indicators for North Africa, French legal origin and Islam do not have direct counterparts for Italy. Nunn justifies the use of these three variables with them being of interest as "additional control variables to account for potential differences between islands or North African countries and the rest of Africa. Two core differences between North Africa and the rest of Africa is that North African countries are predominantly Islamic and that they all have legal systems based on French civil law" (p. 156). In other words, the three dummies primarily serve the purpose of capturing what unobservables make North Africa different, and in the extension, more prosperous than the rest of the continent. For Italy, the concern is quite the opposite. Rather than worrying about unobservables explaining the comparative prosperity of the north, there are reasons to be concerned about the anomalously poor performance of the south. For this reason, we add a dummy variable for the south of Italy in our regressions. Finally, we should mention that we do not include the mineral controls in the Italian analysis. Mineral production is a very small part of most modern western economies, and in Italy the quantities produced are negligible. With these changes in mind, we run specification (1), (2) and (4), with controls as summarized in Table 5.

In what follows, we focus on specifications (1) and (5). Note that for Italy specification (5) does not account for mineral production, which would render it equivalent to specification (4). To avoid unnecessary confusion, when talking about Italy we will refer to specification (4) as specification (5). The primary reason for doing so is comparison, as specification (5) seems to be the one preferred by Nunn.

5. Empirical results

Tables 6 and 7 present the results for specifications (1) and (5). Each of the two tables contain four columns. The first two presents the estimates for Africa, using GDP per capita in 1960 and 2000 as dependent variables.[25] The remaining two columns present the estimates from the same regressions for Italy.

Table 6. Relationship between slave exports and income

	Dependent variable is log real GDP , ln$y$				
	Specification (1)				
	Africa 1960	Africa 2000	Nunn 2000	Italy 1960	Italy 2000
ln(exports/area)	−0.028	−0.110**	−0.112**	−0.042	−0.050**
	(0.021)	(0.024)	(0.024)	(0.021)	(0.014)
Colonizer/Savoy	Yes	Yes	Yes	Yes	Yes
Constant	7.557**	7.924**	7.930**	8.498**	9.759**
	(0.552)	(0.629)	(0.634)	(0.081)	(0.053)
R-squared	0.20	0.51	0.51	0.35	0.46
Observations	52	52	52	20	20
Std. errors in parentheses;* sign. 5%; ** sign. at 1%.					

Several interesting results can be gleaned from the tables. First, using yearly data on mineral production, rather than 31 year averages, our estimates are very close to those in Nunn (2008). Second, for 2000 there is not only a negative relationship between slave exports and per capita GDP for Africa, but also one for Italy. However, this relationship is insignificant in specification (5), that is, after introducing climatic and geographic controls.[26] Finally, the relationship between slave exports and per capita GDP is not stable over time. For specification (1) the coefficient on slave exports is insignificant for both Africa and Italy in 1960. In fact, for Africa, the coefficient is *not* significant in any of the six specifications, and for Italy it is positive in specification (5).

To a certain extent our findings are not surprising. The declining coefficient on slave exports for Africa was pre-viewed by the *ad-hoc* analysis in Nunn who split the

[25] The tables in the empirical section present the exact estimates only for 1960 and 2000. We run yearly (1960-2006) regression for all six specifications for Africa. Likewise, for Italy we run specification (1), (2) and (4) for 1960, 1975 and 2000. All betas are reported graphically in the paper, but exact estimates outside 1960 and 2000 are omitted. These are however available from the authors upon request.

[26] The p-value for the coefficient on slave raids is 0.105. Note however that for Italy there are only 20 regions. Interestingly, all coefficients in specification (5) have the same sign as for Africa.

African sample into countries 'high' and 'low' slave exporters and found a negative relationship between economic growth and past slave exports.[27]

Table 7. Relationship between slave exports, income and oil in year 2000

	dependent variable is log real GDP , ln$y$				
	Specification (5)				
	Africa 1960	Africa 2000	Nunn 2000	Italy 1960	Italy 2000
ln(exports/area)	−0.025	−0.097**	−0.103**	0.025	−0.014
	(0.028)	(0.033)	(0.034)	(0.014)	(0.008)
Abs latitude	0.023	0.025	0.023	−0.010	0.015
	(0.014)	(0.017)	(0.017)	(0.042)	(0.023)
Longitude	0.007	−0.004	−0.004	−0.099*	−0.053*
	(0.004)	(0.005)	(0.005)	(0.037)	(0.020)
Min avg rainfall	−0.004	−0.002	−0.001	0.001	−0.001
	(0.005)	(0.006)	(0.006)	(0.003)	(0.001)
Avg max hum.	0.006	0.016	0.015	−0.005	0.003
	(0.009)	(0.011)	(0.011)	(0.022)	(0.012)
Avg min temp.	0.036	−0.005	−0.015	−0.023	−0.012
	(0.022)	(0.026)	(0.026)	(0.021)	(0.011)
ln(coastline/area)	0.076*	0.076	0.082*	0.000	0.006
	(0.034)	(0.039)	(0.040)	(0.008)	(0.004)
Island indicator	−0.298	−0.138	−0.150	−0.253	−0.061
	(0.431)	(0.498)	(0.516)	(0.173)	(0.095)
Percent Islamic	−0.003	−0.007*	−0.006*	–	–
	(0.003)	(0.003)	(0.003)		
French Legal Origin	−0.278	0.609	0.643	–	–
	(0.404)	(0.464)	(0.470)		
ln(oil)	0.080*	0.069**	0.078**	–	–
	(0.035)	(0.024)	(0.027)		
ln(gold)	0.028	0.012	0.011	–	–
	(0.022)	(0.018)	(0.017)		
ln(diamonds)	−0.001	−0.037	−0.039	–	–
	(0.025)	(0.029)	(0.043)		
North Africa/South	−0.015	-0.096	−0.304	−0.289	−0.213*
	(0.434)	(0.488)	(0.517)	(0.149)	(0.082)
Colonizer/Savoy	Yes	Yes	Yes	Yes	Yes
Constant	7.294**	5.308**	6.067**	10.780**	9.840**
	(1.082)	(1.222)	(1.204)	(2.598)	(1.429)
R-squared	0.65	0.78	0.77	0.94	0.96
Observations	52	52	52	20	20

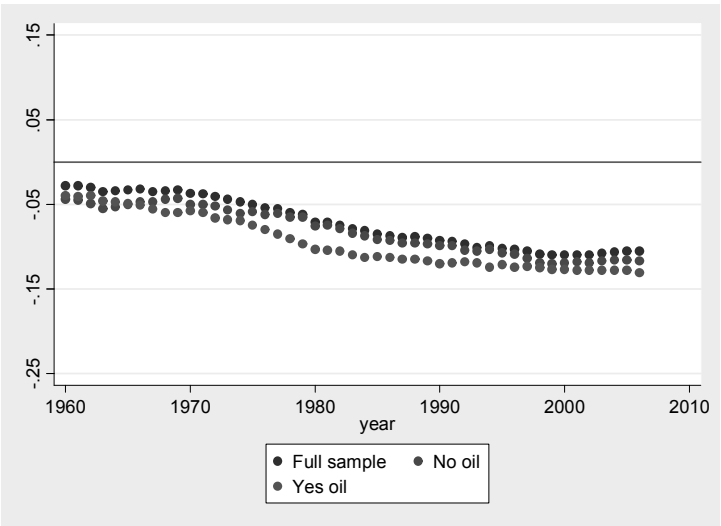
Std. errors in parentheses;* sign. 5%; ** sign. at 1%.

What is perhaps more surprising is that there is also a negative relationship between slave raiding and economic performance for Italy, and that time trend on the coefficient

[27] In his working paper Nunn (2005) also run simple growth regressions. The results from these are suggestive of a negative relationship between growth and slave exports for the years 1960-2000. He does however not study if the relationship holds also in 1960. Further, in the growth regressions oil and mineral production is ignored.

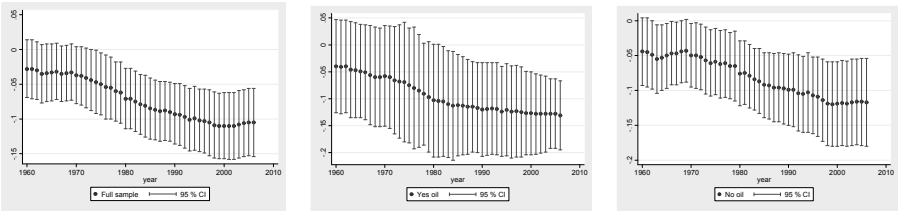
for Italy is similar to that of Africa. To our knowledge the literature does not offer any obvious explanation to why this should be so. In order to investigate the time trends in further detail we plotted the yearly African slave export betas for the years 1960-2006 (Figures 2, 3, 4 & 5) along with those for Italy and the years 1960, 1975 and 2000 (Figures 6 & 7). The declining coefficients on slave exports for both Africa and Italy can be seen clearly in Figures 2, 4 and 6, which confirm the conclusions drawn on the basis of tables 6 and 7.

Figure 2. Development of betas, 1960-2006, Africa specification (1)



In addition, Figures 2 and 4 report the time trend for the African slave export coefficients when the sample is split into those countries who produced oil at any point during the sample period, and those who did not. While these patterns are fairly regular for specification (1) they look highly irregular for specification (5).

Figure 3. Development of betas with confidence intervals, 1960-2006, Africa (1)



For those countries that produced oil, the coefficient makes several jumps of substantial magnitudes, and contrary to the hypothesis, it is positive for approximately half of the years. While the interpretation of this result is not straightforward, it may be indicative of there being some kind of interaction between oil production, GDP per capita and slave trades that we cannot disentangle, and that could potentially affect the analysis. This intuition receives support from the additional specifications that are reported in the Appendix. For specification (2) and (4) the beta for the oil producing countries is always significant, while for specification (3) the beta for the oil producing countries is instead always negative, and sometimes substantially so.

Figure 4. Development of betas, 1960-2006, Africa specification (5)

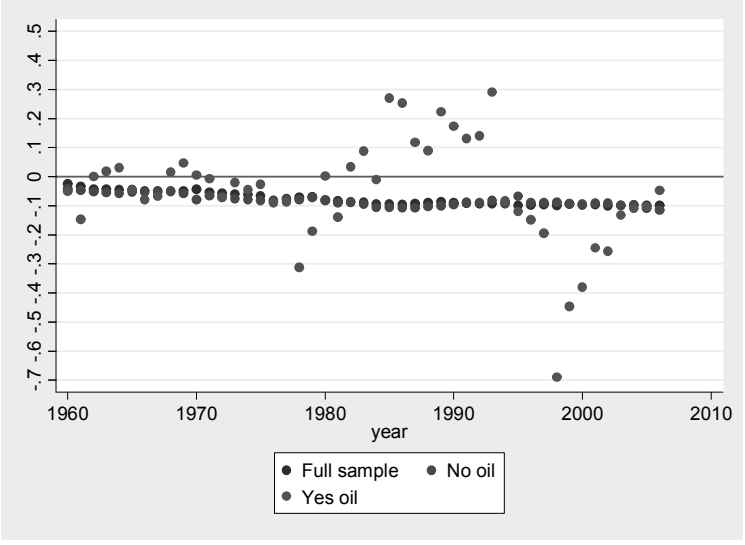
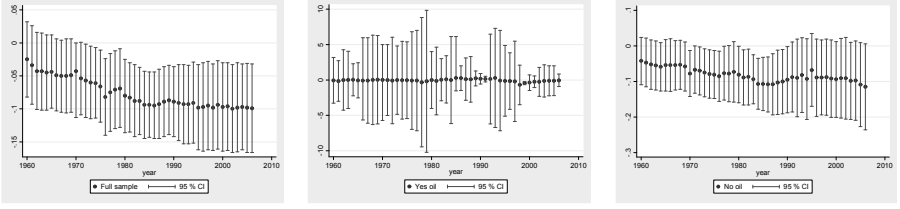


Figure 5. Development of betas with confidence intervals, 1960-2006, Africa (5)



Finally, in Figures 3, 5 and 7 we report the confidence intervals for each yearly beta. As can be seen, neither in specification (1), nor (5), is the beta significantly negative before the 1970s.

Figure 6. Development of betas, 1960-2006, Italy specification (1) & (5)

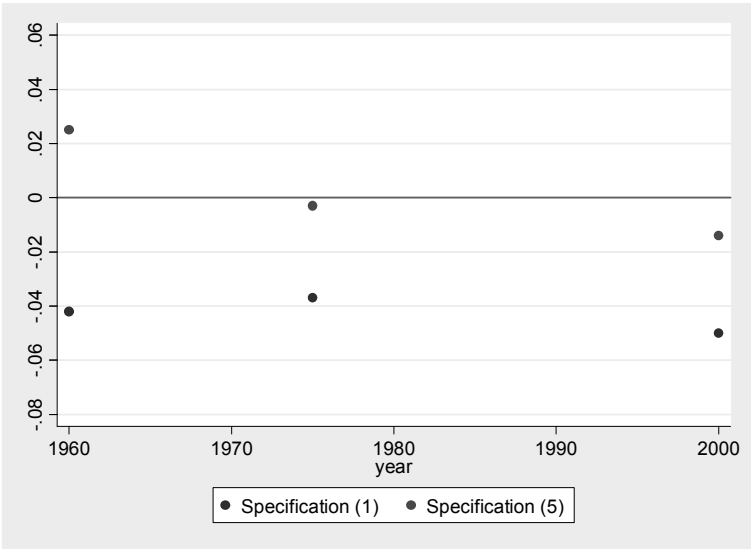
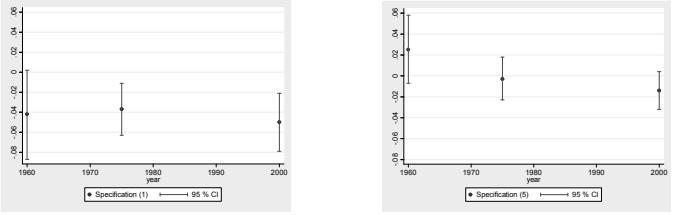


Figure 7. Development of betas with confidence intervals, 1960-2006, Italy specification (1) & (5)



6. Discussion and conclusions

The African slave trades have come to epitomize human tragedy on a grand scale. Millions and millions of people were uprooted, kidnapped, and sold off as slaves to far distant countries. Millions more died in the process. But, did the horrors of the slave trades also contribute to the relative poverty of modern day Africa? In a recent paper Nunn (2008) argues that they did. Here we have argued that although this

is an appealing theory, the slave trades may not have been that influential. Taking Nunn's methodology at face value a researcher studying Africa in the early 2000s would have had to conclude that past slave trades did have a negative effect on GDP per capita. In fact, as we show, there is even some indicative evidence that this could be true not only for Africa, but also for Italy. The Italian evidence does, however, not necessarily support Nunn's conclusions, rather it highlights an important concern with his paper. Slavery was until recently widespread across the globe, and today many formerly slave exporting regions are comparatively rich. Thus, there does not seem to exist a deterministic long run relationship between slave exports and economic performance. We believe that any paper which argues that slave trade had negative long run effects on economic performance needs to address these facts, and look beyond Africa. In his paper, Nunn circumvents the problem by claiming that the African slave trades were unique in scale and scope. Curiously, he establishes the purported uniqueness without a single reference to any other slave trade. In contrast, we study other slave trades, and find that the African slave trades were, at least after adjusting for population, neither unique in scale nor scope.

Further, and in contrast to Nunn, we do not believe that the African slave trades were particularly prominent in causing the latitudinal pattern of ethnic fractionalization that we observe today. Rather, the positive correlation between absolute latitude and ethnic fractionalization in Africa belongs to a global relationship between ethnic, as well as species, diversity and latitude. Since this pattern arguably predates the slave trades, and since ethnic fractionalization is highly, and significantly, correlated with both past slave exports and current GDP per capita this is a potential source of a spurious correlation between the two. Following a similar line of reasoning, we also argue that correlations with climate, disease environments and suitability of land for sugar cane production are potential threats to the validity of the instruments. There is some suggestive evidence to this extent, and when proxies for these variables are included as controls, the instruments become insignificant. In addition, we raise a number of issues concerning potential biases in the imputation of ethnicities, the definition of the sample and selection into the slave trades.

Finally, going back both in time, and to Nunn's methodology, we demonstrate that a researcher studying the African slave trades in the early 1960s would, in contrast to Nunn, have had a difficult time to conclude that past slave trades had a negative impact on GDP per capita. More specifically, we find that the coefficient on slave exports is both small and insignificant in 1960, and that it remains so throughout the 1960s. The trend is however declining and by 1970 the coefficient is typically significant at the 5% level. To a certain extent this result was anticipated by Nunn, who suggested that

the negative effects of the slave trades may have become more pronounced in post-independence Africa. Thus, although our findings raise some doubts concerning the nature and stability of the negative relationship between slave exports and economic performance, they do in part support the hypothesis that they had a negative impact on state development and economic growth after decolonization. However, returning to Italy, we find a weak relationship between slave exports and regional GDP per capita in 1960.

In other words, despite the fact that no decolonization occurred in Italy the coefficient on slave exports shows a similar downward sloping trend, over time, to that for Africa. With this, and the other concerns that we have raised, in mind we conclude that the evidence in favor of the hypothesis that the African slave trades had an adverse effect on modern day economic development is weak.

References

Acemoglu, D., S. Johnson, and J. A. Robinson, "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review* 91 (2001), 1369-1401.

Ahlerup, P., and O. Olsson, "The Roots of Ethnic Diversity," Working Papers in Economics No. 281, School of Business, Economics and Law, Göteborg University (2007).

Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg, "Fractionalization," *Journal of Economic Growth* 8 (2003), 155-194.

Ammerman, A. J., and L. L. Cavalli-Sforza, "Measuring the Rate of Spread of Early Farming in Europe," *Man* (New Series) 6 (1971), 674-688.

Anon., *Africa: Atlas of Our Changing Environment* (Nairobi, KN: UNEP, 2008).

Anon., *eia - U.S. Energy Information Administration*, http://www.eia.doe.gov/countries/country-data.cfm?fips=CG (accessed 2011-03-27).

Anon., *FAOSTAT*, http://faostat.fao.org/ (accessed 2011-04-07).

Anon, *Istat Statistics*, http://dati.istat.it/?lang=en (accessed 2011-03-17).

Anon, *Nationalencyklopedin*, http://www.ne.se (accessed 2011-03-17).

Anon, *NGA GEOnet Names Server*, http://earth-info.nga.mil/gns/html/ (accessed 2011-03-17).

Anon, *World mineral statistics archive*, http://www.bgs.ac.uk/ mineralsuk/statistics/worldArchive.html (accessed 2011-03-16).

Ashraf, Q., and O. Galor, "The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development," mimeo (2009).

Austen, R. A., "The Trans-Saharan Slave Trade: A Tentative Census," in Henry A. Gemery, and Jan S. Hogendorn (Eds.). *The Uncommon Market: Essays in the Economic History of the Atlantic Slave Trade* (New York, NY: Academic Press, 1979).

Austin, G., "The 'Reversal of fortune' thesis and the compression of history: perspectives from African and comparative economic history," *Journal of International Development* 20 (2008), 996-1027.

Bakker, H., *Sugar Cane Cultivation and Management* (Amsterdam, NL: Springer, 1999).

Barbano, A., "SIGC - Sistema Informativo Geografico Costiero: Studi per la caratterizzazione della costa," *APAT - Servizio Difesa delle Coste*, http://www.apat.gov.it/site/_files/Archimede/SIGC_APAT.pdf (accessed 2011-05-02).

Beck, J., and A. Sieber, "Is the Spatial Distribution of Mankind's Most Basic Economic Traits Determined by Climate and Soil Alone?" *PlosOne* 5 (2010): e10416. doi:10.1371/journal.pone.0010416.

Bezemer, D., J. Bolt, and R. Lensink, "Indigenous Slavery in Africa's History: Conditions and Consequences," mimeo (2009).

Bhattacharyya, S., "Root Causes of African Underdevelopment," Working Papers in Trade and Development, No. 2008/16, Australian National University (2008).

Birchenall, J. A., "Quantitative Aspects of Africa's Past Economic Development," mimeo (2009a).

Birchenall, J. A., "Population and Long-Term Economic Development: A Re-examination and Some New Evidence," mimeo (2009b).

Bradley, K., *Slavery and Society at Rome* (Cambridge, UK: Cambridge University Press, 1994).

Cashdan, E., "Ethnic diversity and its environmental determinants: Effects of climate, pathogens, and habitat diversity," *American Anthropologist* 103 (2001), 968-991.

Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza, *The History and Geography of Human Genes* (Princeton, NJ: Princeton University Press, 1994).

Clark, G., "The Condition of the Working Class in England, 1209–2004," *Journal of Political Economy* 113 (2005), 1307-1340.

Collard, I. F., and R. A. Foley, "Latitudinal patterns and environmental determinants of recent human cultural diversity: do humans follow biogeographical rules?" *Evolutionary Ecology Research* 4 (2002), 371-383.

Collier, P., and J. W. Gunning, "Why has Has Africa Grown Slowly?" *Journal of Economic Perspectives* 13 (1999), 3-22.

Cramer, W., D. W. Kicklighter, A. Bondeau, B. Moore III, G. Churkina, B. Nemry, A. Ruimy, A. L. Schloss, and the participants of the Potsdam NPP model intercomparison, "Comparing global models of terrestrial net primary productivity (NPP): overview and key results," *Global Change Biology* 5 (1999), 1-15.

Davis, R. C., *Christian Slaves, Muslim Masters - White Slavery in the Mediterranean, the Barbary coast, and Italy, 1500-1800* (New York, NY: Palgrave Macmillan, 2003).

Diamond, J., "Evolution, consequences and future of plant and animal domestication," *Nature* 418 (2002), 700-707.

Diodorus, S., *The Library of History, vol. III - Loeb Classical Library*, (Cambridge, MA: Harvard University Press, 1939).

Easterly, W., and R. Levine, "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics* 112 (1997), 1203-1250.

Engerman, Stanley L., and K. L. Sokoloff, "Factor endowments, inequality, and paths of development among New World economies," NBER Working Paper Series No 9259 (2002).

Elbl, I., "The Volume of the Early Atlantic Slave Trade, 1450-1521," *Journal of African History*, 38 (1997), 31-75.

Estimates database, *Voyages: The Trans-Atlantic Slave Trade Database,* http://slavevoyages .org/ tast/assessment/estimates.faces?yearFrom=1501& yearTo=1866&flag=2 (accessed 2011-04-02).

Fage, J. D., "Slavery and the Slave Trade in the Context of West African History," *Journal of African History*, 3 (1969), 393-404.

Fage, J. D., and W. Tordoff, *A History of Africa 4th ed.* (Abingdon, UK: Routledge, 2002).

Fearon, J., "Ethnic and Cultural Diversity by Country," *Journal of Economic Growth* 8 (2003), 195-222.

Fenske, J., "Does Land Abundance Explain African Institutions," mimeo (2010).

Ferdère, A., *Les campagnes en Gaule romaine*, 2t (Paris, FR: Errance, 1988).

Findlay, R., and M. Lundahl, "Towards a Factor Proportions Approach to Economic History: Population, Precious Metals and Prices from the Black Death to the Price Revolution," in R. Findlay, L. Jonung, and M. Lundahl (Eds.), *Bertil Ohlin: A Centennial Celebration 1899-1999* (Cambridge, MA: MIT Press, 2002).

Findlay, R., and M. Lundahl, "Demographic Shocks and the Factor Proportions Model: From the Plague of Justinian to the Black Death" in R. Findlay, R. G.H. Henriksson, H. Lindgren and M. Lundahl (Eds), *Eli Heckscher, International Trade and Economic History* (Cambridge, MA: MIT Press, 2002).

Fisher, H. J. *Slavery in the History of Muslim Black Africa.* (New York, NY: New York University Press, 2001).

Flint, J. E. (Ed.), *The Cambridge History of Africa, Volume 5, From c.1790 to c.1870* (Cambridge, UK: Cambridge University Press, 1977).

Fogel, R. W., *Without consent or contract - The rise and fall of American slavery* (New York, NY; W. W. Norton & Company, 1989).

Galor, O., and D. N. Weil, "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond," *American Economic Review* 90 (2000), 806-828.

Garner, P., "Reversal in Africa," mimeo (2006).

Gray, R. (Ed.), *The Cambridge History of Africa, Volume 4, From c.1600 to c.1790* (Cambridge, UK: Cambridge University Press, 1975).

Hernández-Catá, E., K. Scwab, and A. Lopez-Claros (Eds), *The Africa Competitiveness Report 2004* (Geneva: World Economic Forum, 2004).

Hopkins, A. G., "The New Economic History of Africa," *Journal of African History* 50 (2009), 155-177.

Kaplan, H., K. Hill, J. Lancaster, and A. M. Hurtado, "A theory of human life history evolution: Diet, intelligence, and longevity," *Evolutionary Anthropology*, 9 (2000), 156-185.

Klein, M. A. "The Impact of the Atlantic slave trade on the societies of the western Sudan," in J. E. Inikori, S. L. Engerman (Eds), *The Atlantic Slave Trade - Effects on Economies, Societies, and Peoples in Africa, the Americas, and Europe* (Durham, NC: Duke University Press, 1992).

Klein, R. G., *The Human Career - Human Biological and Cultural Origins,* 3rd Ed. (Chicago, IL: The University of Chicago Press, 2009).

Lee, R. B., *The !Kung San: Men, Women and Work in a Foraging Society* (Cambridge, UK: Cambridge University Press, 1979).

Lee, R., "Population in Preindustrial England: An Econometric Analysis," *Quarterly Journal of Economics* 87 (1973), 581-607.

Lovejoy, P. E., *Transformation in Slavery - A history of Slavery in Africa, Second Edition* (Cambridge, UK: Cambridge University Press, 2000).

MacArthur, R. H., *Geographical Ecology: Patterns in the Distribution of Species* (New York, NY: Harper & Row, 1972).

Maddison, A, *Historical Statistics of the World Economy: 1-2008 AD*, http:// www.ggdc.net/ MADDISON/Historical_Statistics/horizontal-file_02-2010.xls (accessed 2011-03-17).

Malthus, T. R., *An Essay on the Principle of Population*, 5th edition (London, UK: John Murray, 1817).

Manning, P., *Slavery and African Life - Occidental, Oriental, and African Slave Trades* (Cambridge,UK: Cambridge University Press. 1990).

McCaskie, T. C. *State and Society in Pre-Colonial Asante* (Cambridge, UK: Cambridge University Press, 2002).

McIntosh, S. K. "Tools for Understanding Transformation and Continuity in Senegambian Society: 1500-1900," in C. R. DeCorse (Ed.), *West Africa during the Atlantic*

Slave Trade - Archaeological Perspectives (London, UK: Leicester University Press, 2001).

Meillassoux, C., *The Anthropology of Slavery - The Womb of Iron and Gold* (Chicago, Il: The University of Chicago Press, 1991).

Michalopoulos, S., "The origins of Ethnolinguistic Diversity," Working Paper No. 110, Collegio Carlo Alberto (2009).

Moore, J. L., L. Manne, T. Brooks, N. D. Burgess, R. Davies, C. Rahbeck, P. Williams, and A. Balmford, "The distribution of cultural and biological diversity in Africa," *Proceedings of Royal Society B* 269 (2002), 1645-1653.

Motamed, M. J., R. J. G. M. Florax, and W. A. Masters, "Geography and Economic Transition: Global Spatial Analysis at the Grid Level," mimeo (2009).

Nash Briggs, D., "Metals, Salt, and Slaves: Economic Links Between Gaul and Italy From the Eigth to the Late Sixth Century," *Oxford Journal of Archaeology* 22 (2003), 243-259.

Nettle, D., *Linguistic Diversity* (Oxford, UK: Oxford University Press, 1999).

New, M., D. Lister, M. Hulme, and I. Makin, "A high-resolution data set of surface climate over global land areas," *Climate Research* 21 (2002), 1-25.

Nunn, N., "Slavery, Institutional Development, and Long-Run Growth in Africa, 1400-2000," mimeo (2005).

Nunn, N., "The Long-Term Effects of Africa's Slave Trades," *Quarterly Journal of Economics* 123 (2008), 139-176.

Nunn, N., and L. Wantchekon, "The Slave Trade and the Origins of Mistrust in Africa," mimeo (2010).

Olsson, O., and D. A. Hibbs, "Biogeography and long-run economic development," *European Economic Review* 49 (2005), 909-938.

Oliver, R. A. (Ed.), *The Cambridge History of Africa, Volume 2, c. 500 B.C. to A.D. 1050* (Cambridge, UK: Cambridge University Press, 1978).

Paci, R., and A. Saba, "The Empirics of Regional Economic Growth in Italy. 1951-1993," Working Paper CRENoS 199701, Centre for North South Economic Research, University of Cagliari and Sassari, Sardinia (1997).

Packard, R., *The Making of a Tropical Disease - A Short History of Malaria* (Baltimore, MD: Johns Hopkins University Press, 2007).

Pamuk, S., "The Black Death and the origins of the 'Great Divergence' across Europe, 1300-1600," *European Review of Economic History* 11 (2007), 289-317.

Phillips, W. D. Jr., *Slavery from Roman Times to the Early Transatlantic Trade* (Manchester, UK: Manchester University Press, 1985).

Pianka, E. R., "Latitudinal Gradients in Species Diversity: A Review of Concepts," *American Naturalist* 100 (1966), 33-46.

Putterman, L., "Agriculture, Diffusion and Development: Ripple Effects of the Neolithic Revolution," *Economica* 75 (2008) 729-748.

Rives, J. B., "Introduction," in Tacitus, *Germania* (Oxford, UK: Clarendon Press, 1999).

Rodney, W., *How Europe Underdeveloped Africa* (London, UK: Bogle-L'Ouverture, 1972).

Rosenzweig, M. L., "Species Diversity Gradients: We Know More and Less Than We Thought," *Journal of Mammology* 73 (1992), 715-730.

Rosenzweig, M. L., *Species diversity in space and time* (Cambridge, UK: Cambridge University Press, 1995).

Rowe, A. K., S. Y. Rowe, R. W. Snow, E. L. Korenromp, J. R. M. Armstrong Schellenberg, C. Stein, B. L. Nahlen, J. Bryce, R. E. Black, and R. W. Steketee, *Estimates of the burden of mortality directly attributable to malaria for children under 5 years of age in Africa for the year 2000 - Final Report For the Child Health Epidemiology Reference Group (CHERG)* (2006).

Sachs, J. D., and A. M. Warner, "Sources of Slow Growth in African Economies," *Journal of African Economies* 6 (1997), 335-376.

Scheidel, W., "The Roman Slave Supply," Princeton/Stanford Working Papers in Classics (2007).

Scheidel, W., "Roman wellbeing and the economic consequences of the 'Antonine Plague'," Princeton/Stanford Working Papers in Classics (2010a).

Scheidel, W., "Slavery in the Roman economy," Princeton/Stanford Working Papers in Classics (2010b).

Smith, B. D. *The Emergence of Agriculture* (New York, NY: Scientific American Library, 1995).

Smith, R. L., *Premodern Trade in World History* (Abingdon, UK: Routledge, 2009).

Spolaore, E., and R. Wacziarg, "The Diffusion of Development," *Quarterly Journal of Economics* 124 (2009), 469-529.

Stock, J. H., and M. Yogo, "Testing for Weak Instruments in Linear IV Regressions," Technical Working Paper 284, National Bureau of Economic Research (2002).

Tacitus, P. C., *Tacitus on Britain and Germany - A New Translation, by H. Mattingly* (West Drayton, UK: Penguin Books, 1948).

Tchernia, A., "Italian wine in Gaul at the end of the Republic," in Peter Garnsey, Keith Hopkins, and C. R. Whittaker (Eds), *Trade in the Ancient Economy* (Berkley, CA: University of California Press, 1983).

Thornton, J., "The Kingdom of Kongo, ca. 1390-1678. The Development of an African Social Formation," *Cahiers d'études africaines* 22 (1982), 325-342.

Thornton, J., *Africa and Africans in the Making of the Atlantic World, 1400-1800, Second Edition* (Cambridge, UK: Cambridge University Press, 1998).

Whatley, W., "Guns-For-Slaves: The 18th Century British Slave Trade in Africa," mimeo (2008).

Whatley, W., and R. Gillezeau, "The Impact of the Slave Trade on African Economies," mimeo (2009).

Wooldridge, J. M., *Econometric Analysis of Cross Section and Panel Data* Cambridge, MA: MIT Press, 2002).

Zohary, D., and M. Hopf, *Domestication of Plants in the Old World* 3rd Ed. (Oxford, UK: Oxford University Press, 2000).

Appendix A

In the Appendix, we report the estimates for all four remaining specifications equation (2.1).

Specification (2)

Table 8. Relationship between slave exports and income

	dependent variable is log real GDP , ln$y$			
	Specification (2)			
	Africa 1960	Africa 2000	Italy 1960	Italy 2000
ln(exports/area)	−0.007	−0.075*	0.026	−0.015
	(0.025)	(0.029)	(0.017)	(0.010)
Abs latitude	0.019	0.017	0.065	0.053*
	(0.014)	(0.016)	(0.034)	(0.020)
Longitude	0.008	−0.000	−0.095*	−0.064**
	(0.005)	(0.005)	(0.033)	(0.019)
Min avg rainfall	−0.004	−0.001	0.001	−0.002
	(0.006)	(0.007)	(0.003)	(0.002)
Avg max humidity	0.007	0.009	0.012	0.015
	(0.010)	(0.012)	(0.025)	(0.014)
Avg min temperature	0.025	−0.017	−0.001	0.000
	(0.024)	(0.028)	(0.023)	(0.013)
ln(coastline/area)	0.055	0.086*	−0.005	0.004
	(0.033)	(0.038)	(0.009)	(0.005)
Colonizer FE/Savoy	Yes	Yes	Yes	Yes
Constant	6.224**	6.685**	6.067**	7.463**
	(1.013)	(1.182)	(2.158)	(1.224)
R-squared	0.39	0.61	0.90	0.93
Observations	52	52	20	20

Std. errors in parentheses;* sign. 5%; ** sign. at 1%

Specification (3)

Table 9. Relationship between slave exports and income

	dependent variable is log real GDP , ln$y$	
	Specification (4)	
	Africa 1960	Africa 2000
ln(exports/area)	−0.054	−0.104**
	(0.035)	(0.036)
Abs latitude	0.010	−0.005
	(0.019)	(0.020)
Longitude	0.004	−0.008
	(0.006)	(0.006)
Min avg rainfall	0.005	0.008
	(0.008)	(0.008)
Avg max humidity	−0.000	0.008
	(0.012)	(0.012)
Avg min temperature	0.029	−0.038
	(0.027)	(0.028)
ln(coastline/area)	0.087*	0.089*
	(0.040)	(0.041)
Colonizer FE	Yes	Yes
Constant	7.554**	7.802**
	(0.750)	(0.769)
R-squared	0.42	0.64
Observations	42	42

Std. errors in parentheses;* sign. 5%; ** sign. at 1%.

Specification (4)

Table 10. Relationship between slave exports, income and oil in year 2000

	dependent variable is log real GDP , ln$y$	
	Specification (4)	
	Africa 1960	Africa 2000
ln(exports/area)	−0.023	−0.080*
	(0.030)	(0.035)
Abs latitude	0.022	0.020
	(0.015)	(0.017)
Longitude	0.008	−0.004
	(0.005)	(0.006)
Min avg rainfall	−0.006	−0.000
	(0.006)	(0.007)
Avg max humidity	0.004	0.009
	(0.010)	(0.011)
Avg min temperature	0.045	−0.004
	(0.024)	(0.027)
ln(coastline/area)	0.086*	0.093*
	(0.037)	(0.042)
Island indicator	−0.546	−0.299
	(0.461)	(0.525)
Percent Islamic	−0.006*	−0.008**
	(0.002)	(0.003)
French Legal Origin	−0.291	0.692
	(0.438)	(0.499)
North Africa indicator	0.530	0.449
	(0.421)	(0.480)
Colonizer FE	Yes	Yes
Constant	6.728**	6.313**
	(1.113)	(1.267)
R-squared	0.53	0.71
Observations	52	52

Std. errors in parentheses;* sign. 5%; ** sign. at 1%.

Specification (6)

Table 11. Relationship between slave exports, income and oil in year 2000

	dependent variable is log real GDP , ln$y$	
	Specification (6)	
	Africa 1960	Africa 2000
ln(exports/area)	−0.029	−0.125**
	(0.027)	(0.032)
Abs latitude	0.018	0.007
	(0.015)	(0.016)
Longitude	0.006	−0.010
	(0.005)	(0.005)
Min avg rainfall	0.007	−0.005
	(0.007)	(0.008)
Avg max humidity	−0.001	0.016
	(0.009)	(0.010)
Avg min temperature	0.023	−0.029
	(0.022)	(0.024)
ln(coastline/area)	0.077*	0.084*
	(0.032)	(0.035)
Percent Islamic	0.002	−0.005
	(0.003)	(0.003)
French Legal Origin	−1.146	−0.531
	(0.634)	(0.728)
ln(oil)	0.093*	0.073**
	(0.036)	(0.021)
ln(gold)	0.048*	0.025
	(0.020)	(0.017)
ln(diamonds)	0.005	−0.052
	(0.023)	(0.027)
Colonizer FE	Yes	Yes
Constant	8.138**	7.328**
	(0.588)	(1.256)
R-squared	0.74	0.81
Observations	42	42

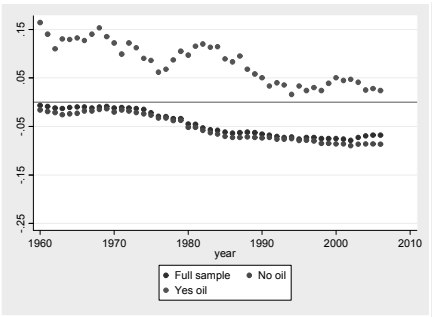
Std. errors in parentheses;* sign. 5%; ** sign. at 1%.

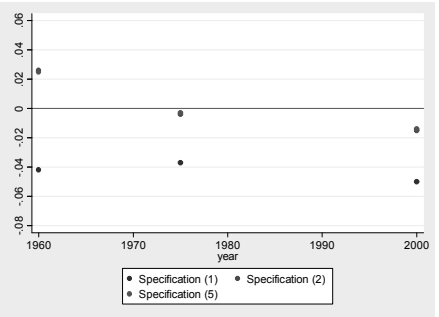
Appendix B

Here follows the development of betas over 1960-2000 for all omitted specifications, with and without the confidence intervals.

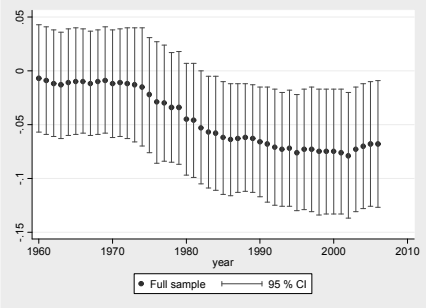
Specification (2)

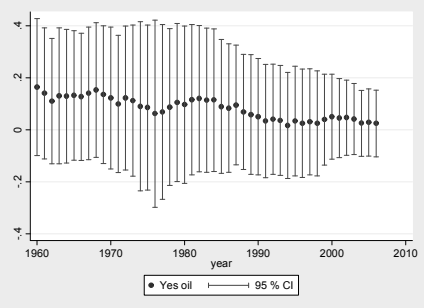
Development of betas 1960-2006, Africa and Italy

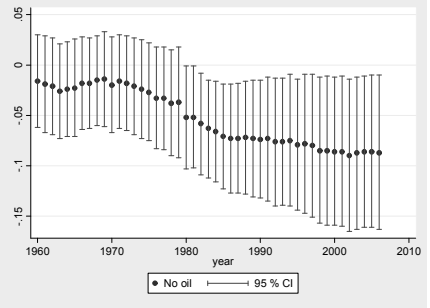


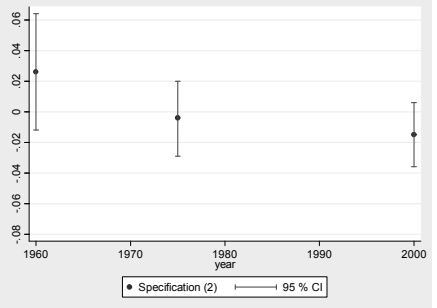


Development of betas with confidence intervals 1960-2006, Africa and Italy



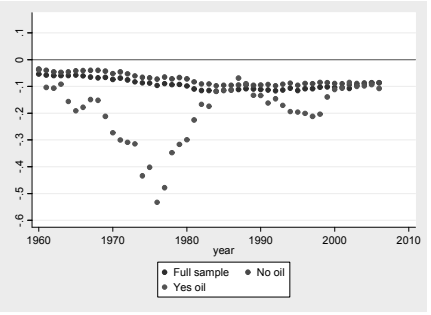




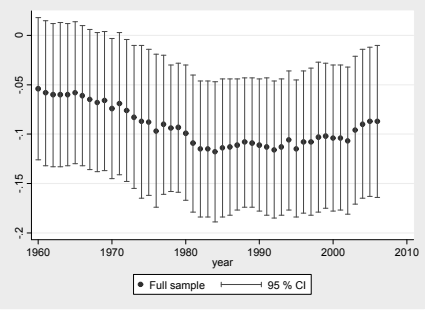


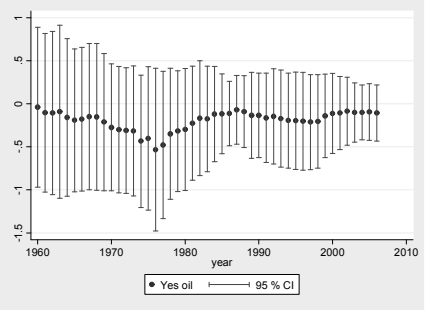
Specification (3)

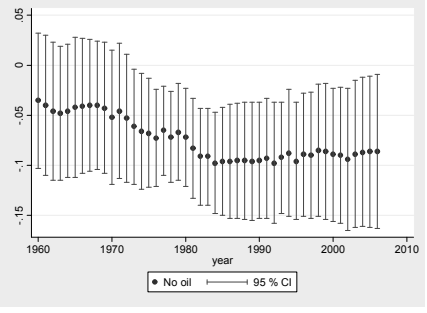
Development of betas 1960-2006, Africa



Development of betas with confidence intervals 1960-2006, Africa

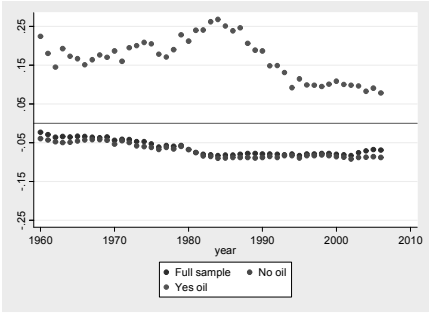




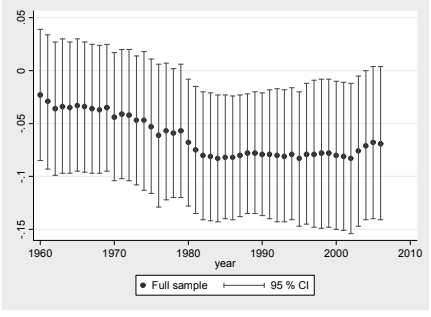


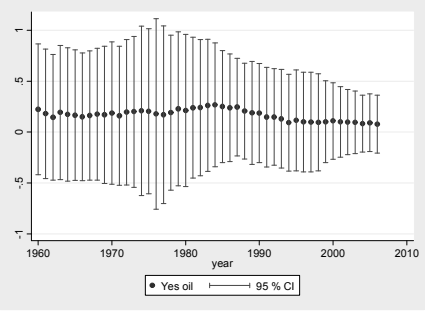
Specification (4)

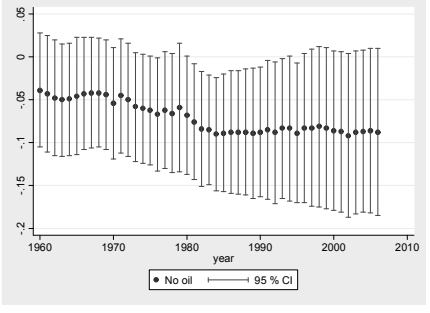
Development of betas 1960-2006, Africa



Development of betas with confidence intervals 1960-2006, Africa

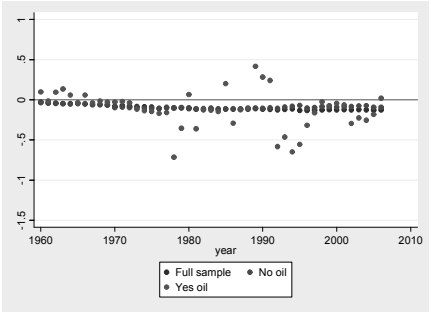




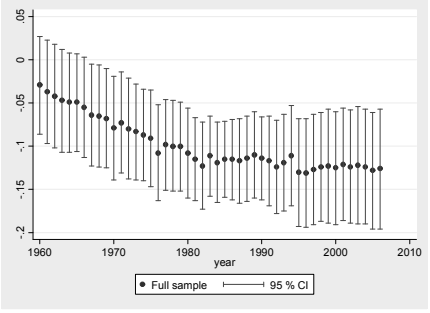


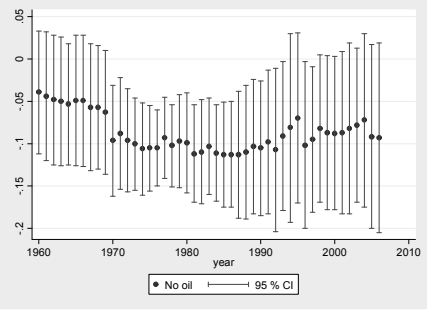
Specification (6)

Development of betas 1960-2006, Africa



Development of betas with confidence intervals 1960-2006, Africa[28]





[28] Note: For this specification, there are not enough observations to plot the case of countries that produce oil.

PAPER 6

Is there an adverse effect of sons on maternal longevity?

David Cesarini Erik Lindqvist Björn Wallace

Abstract. Recent years have witnessed the emergence of a literature examining the effects of giving birth to sons on postmenopausal longevity in pre-industrial mothers. The original paper in this lineage used a sample (n=375) of Sami mothers from northern Finland and found that, relative to daughters, giving birth to sons substantially reduced maternal longevity. We examine this hypothesis using a similar and a much larger sample (n=930) of pre-industrial Sami women from northern Sweden, who in terms of their demographic, sociocultural and biological conditions, closely resemble the original study population. In contrast to the previously reported results for the Sami, we find no evidence of a negative effect of sons on maternal longevity. Thus, we provide the most compelling evidence to date that the leading result in the literature must be approached with scepticism.

1. Introduction

Did giving birth to sons reduce maternal old-age longevity in pre-industrial humans? This question was posed, and answered in the affirmative, by Helle, Lummaa and Jokela (henceforth HLJ; Helle et al., 2002a, b) who used a sample of 375 pre-industrial Sami women from northern Finland. Restricting the sample to mothers who reached an age of at least 50, and regressing the longevity of these mothers on the number of sons and daughters produced, the authors found that maternal longevity decreased by 0.65 years for each son born while it increased by 0.44 years for each daughter.

These results, and their interpretation as evidence in favor of negative 'effects' of sons on maternal longevity, drew criticism almost immediately. In a rejoinder to HLJ, Beise and Voland (2002) reported no evidence for an association between the number of sons born and maternal longevity in neither Canadian nor German pre-industrial populations. Helle et al. (2002b, p. 317) retorted that Beise and Voland had

A slightly different version of this paper was published as Cesarini, D., E. Lindqvist, and B. Wallace, "Is there an adverse effect of sons on maternal longevity?," *Proceedings of Royal Society B* 276 (2009), 2081-2084. We thank Carin Hedlund and Maria Larsson at the Demographic Database, Umeå University, for their help with the data retrieval. We rely on research by Peter Sköld, Per Axelsson and Gabriella Nordin at Umeå University for our identification of the Sami population. Financial support from the Jan Wallander and Tom Hedelius Foundation is gratefully acknowledged.

made a 'disputable attempt to refute our conclusions', citing the unique demographic, sociocultural and ecological conditions of the Sami. Since the original HLJ paper made a general claim about sons reducing longevity in pre-industrial mothers, and not just in their Sami study population, this response is not fully satisfactory. Yet, the HLJ reply did make the valid point that, in principle, it is possible that the relationship between the number of sons born and old-age maternal longevity varies considerably depending on cultural and biological conditions that may be idiosyncratic to each population studied.

Additional fuel to Beise and Voland's scepticism has been offered by a number of recent attempts to replicate HLJ's findings in other populations. For instance, Van de Putte et al. (2004) studied women in a Flemish agricultural village, born between 1700 and 1870. In their baseline regression, there was only weak evidence for a negative correlation between number of sons and maternal longevity, although the association did become statistically significant when the sample was restricted to women born before 1815 married to 'ordinary laborers'. Furthermore, no evidence for differential effects on longevity by the sex of offspring was found by Jasienska et al. (2006) in a study of mothers in rural Poland. Similarly, Hurt et al. (2006) failed to find any association between the number of sons and maternal mortality in a sample from modern-day Bangladesh. Finally, Cesarini et al. (2007) reported null results for a large sample of eighteenth and nineteenth century women from northern Sweden. A recent paper by Harrell et al. (2008) did, however, find some evidence of a relationship between the sex composition of offspring and maternal longevity, but the estimated coefficients were small. Taken together, these papers uniformly find much weaker associations between the number of sons born and maternal longevity than the original HLJ paper did. In fact, a majority of published papers fail to find a stable and significant relationship.

The hypothesis that the sex of offspring might have adverse long-term effects on maternal longevity is not unreasonable, a priori. A distinction can be made between biological and environmental mechanisms through which the sex composition of a mother's children might matter (Van de Putte et al., 2004; Harrell et al., 2008). It is well known that the physiological costs of childbearing are higher for sons, as evidenced by their higher intrauterine growth rates, birth weights (Marsal et al., 1996; de Zegher et al., 1999) and the greater maternal energy intake during male pregnancies (Tamimi et al., 2003). Furthermore, women carrying male fetuses have elevated levels of testosterone (Meulenberg & Hofman, 1991), a known immunosuppressant (Folstad & Karter, 1992). In addition to biological channels, it has been proposed that daughters may relieve the burden on their mothers to carry out domestic tasks (e.g. Turke, 1988; Hames &

Draper, 2004). However, what is at stake is not whether these theoretical channels are plausible, but whether they are consistent with the body of empirical evidence.

In this paper, we ask whether there is merit to HLJ's claim that the unique demographic, sociocultural and biological conditions that characterize their Sami study population can explain the inconsistencies between their findings and the subsequent replication attempts. We test the hypothesis that the sex composition of offspring affects postmenopausal longevity using a sample of Sami women almost three times larger than that in HLJ. While there is some cultural and genetic heterogeneity among different Sami populations, this heterogeneity is comparatively small. As a group, the Sami can be culturally and genetically distinguished from other European populations, and they are often thought of as an outlier (Roung, 1969; Nickul, 1977; Tambets et al., 2004). This cultural and genetic uniqueness shared by our and HLJ's sample populations, combined with the fact that the border between Sweden and Finland is a recent construction that never actually hindered contact, arguably makes the Sami of northern Sweden ideal for replicating HLJ.

As we have indicated elsewhere, we have misgivings about the methodological approach taken by HLJ (Cesarini et al., 2007), but we will suppress these concerns here as far as possible and focus primarily on the issue of replicability. To preview our results, we find no evidence in favor of the hypothesis that sons had a negative impact, relative to daughters, on maternal longevity. That is, holding parity constant, there is no evidence that mothers with more sons had a shorter postmenopausal longevity in our Sami population.

2. Materials and methods

The regression run by HLJ is of the form,

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 d_i + \epsilon_i \tag{2.1}$$

where Y_i is the longevity of mother i; x_i is the longevity of the husband of mother i; z_i is the number of sons of mother i; and d_i is the number of daughters of mother i. Note that if parity $(z_i + d_i)$ is correlated with unobserved variables that affect old-age longevity, such as health or wealth, then the estimates will be biased due to omitted variable bias. Indeed, evidence from both hunter–gatherer and agro-pastoralist, as well modern industrial, societies suggests that family size is correlated with health, wealth and other indicators of longevity, although the strength and direction of this correlation may vary (see, for instance, Kaplan, 1996; Dolbhammer & Oeppen, 2003). If this is true also in our population, the results in HLJ tell us only whether giving birth to sons is a predictor of longevity, but do not speak about the issue of causality, contrary

to what is implied by the title of HLJ's paper. Under less restrictive assumptions, however, any significant difference in the regression coefficients such that $\hat{\beta}_2 - \hat{\beta}_3 \lesseqgtr 0$ can be interpreted as reflecting a difference in the relative cost of producing a son or a daughter (Cesarini et al., 2007). Consider a simple rearrangement of equation (2.1) as follows:

$$Y_i = \alpha_0 + \alpha_1 x_i + \alpha_2 z_i + \alpha_3 n_i + \epsilon_i \tag{2.2}$$

where n_i is the parity of mother i (i.e. $n_i = z_i + d_i$). This is equivalent to the regression in HLJ with $\alpha_0 = \beta_0$, $\alpha_1 = \beta_1$ and $\alpha_2 = \beta_2 - \beta_3$. If α_2 is negative, this means that giving birth to sons is costlier relative to daughters. Note that α_2 can be negative even if both boys and girls have a positive causal effect on longevity (i.e. if $\beta_3 > \beta_2 > 0$). The more general issue of how parity is associated with parental longevity is reviewed in Le Bourg (2007).

The rearrangement in equation (2.2) is only for expositional convenience. The estimated α_2 in equation (2.2) is mathematically equivalent to the estimate of $\beta_2 - \beta_3$ in equation (2.1). Cesarini et al. (2007) provided a more thorough discussion on this point.

3. Data

Our dataset was constructed by the staff at the Demographic Database at Umeå University, Sweden specifically for the purpose of this paper. The sample consists of Sami women whose reproductive history is known, and who were born between 1698 and 1840 in either one of the 12 studied parishes (Table 1) belonging to the counties Jämtland and Norrbotten. The sample is restricted to women who reached the age of 50 and had at least one child. These selection criteria are equivalent to those used by HLJ.

The data sources are separate registers of catechetical examinations, births and baptisms, banns and marriages, migration, deaths and burials. There are, of course, uncertainties concerning the completeness of the data. However, there is no reason to believe that faulty or missing information has affected the data in any way that will introduce a systematic bias. Importantly, our original sources are fully comparable to those used by HLJ as the laws governing the keeping of parochial records were identical in the two countries for the vast majority of the sample period.

Our classification of Sami ethnicity is based on unpublished research at Umeå University by Peter Sköld, Per Axelsson and Gabriella Nordin. The staff at the Demographic Database created six different indicator variables for Sami ethnicity. The two most informative indicators are based on an individual's place of residence and notes

about relatives' (siblings and parents) ethnic affiliation. Several villages and areas are known to have been populated almost exclusively by the Sami during our sample period. Additionally, there are Sami indicator variables based on miscellaneous remarks made by the priest, occupational status, surnames or whether a death occurred in the Alpine region. All six indicator variables take the value 1 if a piece of information indicates Sami ethnicity, and 0 if this is not the case. For example, the occupational variable takes the value 1 if the parish record indicates an occupation specific to the Sami population (e.g. reindeer herder), and 0 if the registered occupation does not indicate Sami ethnicity or if there is no record on occupation at all. Hence, the fact that an indicator variable is 0 does not imply that the person is not of Sami ethnicity.

Table 1. Sample parishes

Parish	Number of women
Frostvikens lappförsamling	13
Föllinge	4
Föllinge lappförsamling	21
Gällivare	286
Hede lappförsamling	3
Hotagen	12
Hotagens lappförsampling	7
Jokkmokk	153
Jukkasjärvi	164
Karesuando	212
Kvikkjokk	41
Undersåkers lappförsamling	14
$\sum$	930

The Pearson correlation between the number of Sami indicators of a wife and the number of Sami indicators for her husband is 0.81. To select our sample, we calculate the number of Sami indicators (i.e. indicator variables equal to 1) for the wife and the husband of each couple. In our basic sample, we consider couples where there is at least one Sami indicator for the mother. This sample consists of 930 married couples and their 5,741 children. We then perform robustness checks with subsamples selected under more restrictive criteria.

4. Results

The results from regression (2.2) are displayed in Table 2. In contrast to HLJ, our estimates imply a small positive, but statistically insignificant, association between maternal old-age longevity and giving birth to a son instead of a daughter. The coefficient on parity is close to zero and statistically insignificant, implying that there is no association between parity and postmenopausal longevity. Our data allow us to reject much smaller relative effects of boys on maternal old-age longevity than those

implied by HLJ. The point estimate of α_2 implied by HLJ's results (-1.09) is thus firmly rejected in our data (F=20.41; p<0.00001). The 95% confidence interval for the coefficient α_2 ranges from -0.32 to 0.78, hence the null hypothesis that the two coefficients β_2 and β_3 are equal cannot be rejected. These results remain essentially unchanged if women with a longevity shorter than 50 are also included in the sample (column 2). The only difference is that the coefficient on parity is now large and statistically significant, reflecting the reverse causality from (premenopausal) longevity to parity. The results do not change appreciably if we use more restrictive definitions of Sami ethnicity (columns 3–6).

In additional analyses not reported here, we find that the results are similar also when stillborn children are excluded, when survival models (Cox and Weibull) are estimated instead of least squares, or when standard errors are clustered at the parish level (available on request from the authors).

5. Conclusions

This paper has attempted to replicate the findings in HLJ for a sample that strongly resembles the original study population in terms of their genetic, demographic, sociocultural and ecological conditions. Our failure to find any evidence in support of the hypothesis that sons reduce old-age maternal longevity, in a sample approximately three times larger than that used by HLJ, provides the most compelling evidence to date that the results reported in HLJ ought to be interpreted with great caution. In our previous work, we concluded our review of the cumulative evidence by noting that, on balance, 'it is the Sami population (in HLJ) which is an outlier in need of an explanation' (Cesarini et al., 2007, p. 544). The results reported here reinforce this conclusion. The original HLJ paper made the strong general claim that sons reduced longevity in pre-industrial humans. This general claim has received little support in subsequent research.

Modest adverse relative effects of sons cannot be ruled out, but the suggestion that giving birth to a son as opposed to a daughter reduced life expectancy by over a year, an extraordinary finding if true, seems to be an anomaly found in only one small sample.

Table 2. Sons and maternal longevity

	Dependent Variable: Maternal Longevity					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	66.999**	41.051**	66.178**	66.707**	66.117**	67.035**
	(1.984)	(2.600)	(2.359)	(3.276)	(2.091)	(2.507)
Number of sons	0.230	0.410	−0.048	0.091	0.263	−0.015
	(0.278)	(0.383)	(0.348)	(0.482)	(0.313)	(0.391)
Parity	0.009	0.974**	0.373	−0.113	0.097	0.259
	(0.184)	(0.258)	(0.231)	(0.351)	(0.207)	(0.268)
Husband's longevity	0.068*	0.247**	0.053	0.095	0.067*	0.049
	(0.028)	(0.038)	(0.035)	(0.050)	(0.030)	(0.037)
Observations	930	1221	582	304	802	497
Sami indicator: mother	One or more	One or more	Two or more	Three or more	One or more	Two or more
Sami indicator: husband	-	-	-	-	One or more	Two or more
Longevity restriction	≥ 50	-	≥ 50	≥ 50	≥ 50	≥ 50
R^2	0.009	0.092	0.015	0.014	0.013	0.010

Heteroskedasticity robust standard errors (White, 1980) in parentheses.
* significant at 5%; ** significant at 1%.

References

Beise, J., and E. Voland, "Effect of producing sons on maternal longevity in premodern populations," *Science* 298 (2002), 317.

Cesarini, D., E. Lindqvist, and B. Wallace, "Maternal longevity and the sex of offspring in pre-industrial Sweden," *Ann. Hum. Biol.* 34 (2007), 535-546.

de Zegher, F. H. Devlieger, and R. Eeckels, "Fetal growth: boys before girls," *Horm. Res.* 51 (1999), 258-259.

Dolbhammer, G., and J. Oeppen, "Reproduction and longevity among the British peerage: the effect of frailty and health selection," *Proc. R. Soc. B* 270 (2003), 1541-1547.

Folstad, I., A. J. Karter, "Parasites, bright males and the immunocompetence handicap," *Am. Nat.* 139 (1992), 603-622.

Hames, R., and P. Draper, "Women's work, child care, and helpers-at-the-nest in a hunter-gatherer society," *Hum. Nature* 15 (2004), 319-341.

Harrell, C.J., K.R. Smith, and G.P. Mineau, "Are girls good and boys bad for parental longevity?," *Hum. Nature* 19 (2008), 56-69.

Helle, S., V. Lummaa, and J. Jokela, "Sons reduced maternal longevity in preindustrial humans," *Science* 296 (2002a), 1085.

Helle, S., V. Lummaa, and J. Jokela, "Effect of producing sons on maternal longevity in premodern populations - Response," *Science* 298 (2002b), 317.

Hurt, L. S., C. Ronsmans, and M. Quigley, "Does the number of sons born affect long-term mortality of parents? A cohort study of rural Bangladesh," *Proc. R. Soc. B* 273 (2006), 149-155.

Jasienska, G., I. Nenko, and M. Jasienski, "Daughters increase longevity of fathers, but daughters and sons equally reduce longevity of mothers," *Am. J. Hum. Biol.* 18 (2006), 422-425.

Kaplan, H., "A theory of fertility and parental investment in traditional and modern human societies," *Yearbook of Physical Anthropology* 101 (1996), 91-135.

Le Bourg, E., "Does reproduction decrease longevity in human beings?," *Ageing Res Rev* 6 (2007), 141-149.

Marsal, K., P. H. Persson, T. Larsen, H. Lilja, A. Selbing, and B. Sultan, "Intrauterine growth curves based on ultrasonically estimated foetal weights," *Acta Paediatr.* 85 (1996), 843-848.

Meulenberg, P. M., and J. A. Hofman, "Maternal testosterone and fetal sex," *J. Steroid Biochem. Mol. Biol.* 39 (1991), 51-54.

Nickul, K., *The Lappish Nation-Citizens of Four Countries* (Bloomington, IN: Indiana University, 1977).

Roung, I., *Samerna* (Stockholm, Swe: Aldus/Bonniers, 1969).

Tambets, K., S. Rootsi, T. Kivisild, H. Help, P. Serk, et al., "The Western and Eastern roots of the Saami—the story of genetic "outliers" told by mitochondrial DNA and Y chromosomes," *Am. J. Hum. Genet.* 74 (2004), 661-682.

Tamimi, R., P. Lagiou, L. Mucci, C. C. Hsieh, H. O. Adami, and D. Trichopoulos, "Average energy intake among pregnant women carrying a boy compared to a girl," *BM*J 326 (2003), 1245-1246.

Turke, P., "Helpers-at-the-nest: childcare networks on ifaluk,". in L. Betzig, M. Borgerhoff Mulder, and P. Turke (Eds), *Human reproductive behavior: A Darwinian perspective* (Cambridge, UK: Cambridge University Press,.1988).

Van de Putte, B., K. Matthijs, and R. Vlietinck, "A social component in the negative effect of sons on maternal longevity in pre-industrial humans," *J. Biosoc. Sci.* 36 (2004), 289-197.

White, H., "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica* 48 (1980), 817–838.

The Stockholm School of Economics

A complete publication list can be found at www.hhs.se/research/publications.
Books and dissertations are published in the language indicated by the title and can be ordered via e-mail: efi.publications@hhs.se.

A selection of recent publications

Books

Barinaga, Ester (2010). *Powerful dichotomies.*
Ericsson, Daniel (2010). *Den odöda musiken.*
Ericsson, Daniel (2010). *Scripting creativity.*
Holmquist, Carin (2011). *Kvinnors företagande – kan och bör det öka?*
Lundeberg, Mats (2011). *Improving business performance: a first introduction.*
Melén, Sara (2010). *Globala från start. Småföretag med världen som marknad.* Forskning i Fickformat.
Mårtensson, Pär och Mähring, Magnus (2010). *Mönster som ger avtryck: Perspektiv på verksamhets-utveckling.*
Sjöström, Emma (2010). *Ansiktslösa men ansvarsfulla.* Forskning i fickformat.
Wijkström, Filip (2010). *Civilsamhällets många ansikten.*
Engwall, Lars (2009). *Mercury meets Minerva: business studies and higher education: the Swedish case.*

Dissertations

Alexandersson, Gunnar (2010). *The accidental deregulation: essays on reforms in the Swedish bus and railway industries 1979-2009*.
Bohman, Claes (2010). *Attraction: a new driver of learning and innovation.*
Buturak, Gökhan (2011). *Choice deferral, status quo bias, and matching.*
Ejenäs, Markus (2010). *Ledning av kunskapsintegration - förutsättningar och hinder: en studie av en fusion mellan IT- och managementkonsulter.*
Engvall, Anders (2010). *Poverty and conflict in Southeast Asia.*
Glassér, Charlotte (2010). *The fountainhead of innovation health: a conceptualization & investigation.*
Hemrit, Maetinee (2011). *Beyond the Bamboo network: the internationalization process of Thai family business groups.*
Juks, Reimo (2010). *Corporate governance and the firm's behaviour towards stakeholders.*
Lundvall, Henrik (2010). *Poverty and the dynamics of equilibrium unemployment: [essays on the economics of job search, skills, and savings].*
Lychnell, Lars-Olof (2010). *IT-relaterad verksamhetsförändring: processer som formar växelspelet mellan utveckling och användning.*
Magnusson Bernard, Kristin (2010). *Remittances, regions and risk sharing.*
Mohlin, Erik (2010). *Essays on belief formation and pro-sociality.*
Monsenego, Jérôme (2011). *Taxation of foreign business income within the European internal market: an analysis of the conflict between the objective of achievement of the European internal market and the principles of territoriality and worldwide taxation.*
Nakatani, Tomoaki (2010). *Four essays on building conditional correlation GARCH models.*
Nelson, Maria. *Utflyttning av aktiebolag: en analys i ljuset av den internationella skatterätten och EU-rätten.*
Siming, Linus (2010). *Private equity and advisors in mergers and acquisitions.*
Sjöquist Rafiqui, Pernilla (2010). *Evolving economic landscapes: institutions and localized economies in time and space.*

Stepanok, Ignat (2011). *Essays on international trade and foreign direct investment.*
Strid, Ingvar (2010). *Computational methods for Bayesian inference in macroeconomic models.*
Sunesson, T. Daniel (2010). *School networks and active investors.*
Tolstoy, Daniel (2010). *International entrepreneurship in networks: the impact of network knowledge combination on SMEs' business creation in foreign markets.*
Öhman, Niclas (2010). *Considering intentions.*