Slave trades, credit records and strategic reasoning

Slave trades, credit records and strategic reasoning

Four essays in microeconomics

Margherita Bottero

Stockholm School of Economics

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

KEYWORDS: Cognitive hierarchy; Centipede game; History; Slave trades; Limited records; Bounded memory; Reputation; Credit markets, Credit registers.

Slave trades, credit records and strategic reasoning.
Four essays in microeconomics.

ISBN 978-91-7258-850-9

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/her research in his/her own ways as an expression of his/her 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

E quindi uscimmo a riveder le stelle

Dante, *La Divina Commedia*
Inferno, XXXIV, 139

Ai miei genitori

Acknowledgement

I am extremely grateful to my supervisor, Jörgen Weibull, for having helped me thoroughly in the process of writing this thesis. I am equally indebted to my co-supervisor Juanna Joensen for her much appreciated support. Very warm thanks go to my co-author Giancarlo Spagnolo, who has shared with me many interesting research ideas. The second part of this dissertation has truly benefited from his suggestions. I am thankful to Björn Wallace for having let me take advantage of his experience and of his haphazardly vast knowledge.

I wish to thank the Faculty and the Ph.D. students at the Department of Economics for providing such a stimulating research environment. In particular, my work has been noticeably improved from the discussions I have had with Tore Ellingsen and Stefano de Michelis. Axel Bernergård, Tobias Laun, Erik Mohlin, Ettore Panetti and Björn Wallace have patiently read through my papers and provided very constructive criticisms and suggestions. Special thanks go to Ritva Kiviharju, Carin Blanksvärd and Lilian Öberg for their extremely efficient and overwhelmingly nice assistance with the many administrative matters.

During the five years of my Ph.D., I had the pleasure of spending most of my time with the *dude of all dudes* Björn Wallace, the *Germans* Tobias Laun and Andrè "Pipito" Romahn, and Ettore Panetti. Thanks for being such good friends.

Special thanks go to Linus Siming for having coached me, along with David Domeij, marvelously well for the job market, and for having provided much exhilarating entertainment along the way. Sara Formai and Elena Mattana have livened up my days in Stockholm, together with Agatha Murgoci and Arvid Malm.

I would like to acknowledge the encouragement and love I received from my friends and family, especially from my lovely grandparents.

Finally, I wish to thank my parents Saula and Matteo. For me, you are the best parents *in the world*.

Contents

Introduction

This thesis consists of four independent chapters, in which well-known economic theories are employed to investigate, and better understand, data and facts from the real world. Although in fairly distant topics, each paper is an example of how economics, and more precisely microeconomics, offers a rigorous and effective framework to reason about what happens around us. In this sense, my dissertation fully represents what I have learnt in these five years.

The first paper addresses the experimental behavior of subjects that interact with each other, non-cooperatively, in a laboratory setup. The experimental evidence is found to be at odds with the predictions of classical game-theory, and I explore whether a model of bounded rationality can instead succeed in explaining the data. The second paper looks at another type of data, historical rather than experimental. Together with Björn Wallace, we raise doubts, methodological and interpretational, regarding the validity of a recent finding that documents a sizeable effect of Africa's past slave trades on current economic performance.

The last two papers investigate the phenomenon of limited records, understood as the limited availability of past public data regarding a transacting partner. The former is a survey, written jointly with Giancarlo Spagnolo, wherein we discuss the literatures that have independently studied whether limited records may actually prompt beneficial reputation effects. We argue that what is known about this type of informational arrangement is little and scattered, and that this is problematic given the large number of real-life situations featuring limited records. These conclusions prepare the ground for the last paper of this dissertation, which presents a model of limited credit records. The model aims at providing a framework for evaluating the current privacy provisions in the credit market which mandate the removal of information about borrowers' past performance from public registers after a finite number of years.

1. Cognitive Hierarchies and the centipede game

There is by now ample evidence demonstrating that when individuals are presented with game theoretic interactions their behavior frequently departs from the equilibrium predictions, especially during the first rounds of play (see for instance Costa-Gomes, Crawford, 2006 and references therein). This empirical anomaly motivates the ongoing search for new, non-equilibrium, models that are capable of explaining the behavioral patterns observed in laboratory and field experiments. Among the many contributions in this literature, the Level k (Nagel, 1995; Stahl, Wilson, 1994, 1995) and cognitive hierarchy models (Camerer et al., 2004) stand out for their success in predicting and explaining actual behavior in a number of normal-form games by relying only on a parsimonious set of assumptions. How these models perform when applied to extensive-form games is, however, still an open question. In the first chapter of my dissertation, I address this issue by suggesting and testing a possible extension of the cognitive hierarchy model that is directly applicable to the centipede game (Rosenthal, 1982).

The centipede game is a two-player game of perfect information that represents an interactive situation where the players play sequentially and non-cooperatively for a limited number of rounds. The players take turns in deciding whether to terminate the game or to continue it. Continuing increases the value of the joint payoffs, but is not individually profitable if the game is terminated in the next round. If so, the player who passed receives a payoff lower than what he could have secured himself by terminating at the preceding round. If players seek to maximize their monetary gains, the centipede game features a unique (subgame perfect) equilibrium, in which the player who moves first terminates the game. This sharp theoretical prediction is at odds with the behavior of actual players who typically continue until the central rounds, and occasionally until the end. Empirically, no, or very few, participants have been observed to take the pot at the first round (see for instance Fey et al., 1996; McKelvey, Palfrey, 1992; Nagel, Tang, 1998; Rapoport et al., 2003).

This divergence between the equilibrium predictions and the experimental evidence appears analogous to that which has been successfully addressed by non-equilibrium models in normal-form games (Costa-Gomez et al., 2001; Costa-Gomez, Crawford, 2006; Ho et al., 1998; Nagel, 1993, 1995; Stahl, Wilson 1994, 1995). The key novelty of these models is that players are allowed to have beliefs regarding the behavior of their opponents that are much simpler than those assumed in traditional game theory.

Building on this approach, and more precisely on the work of Camerer et al.'s (2004), I present and test an extensive-form version of the normal-form cognitive hierarchy model. Further, I evaluate empirically the resulting predictions using laboratory data borrowed from a previously published experiment (McKelvey, Palfrey, 1992). My paper features two main contributions. First, it extends the experimental literature on the cognitive hierarchy model, which is here directly applied to the extensive-form centipede game. Secondly, it demonstrates that in the centipede game the cognitive hierarchy approach leads to predictions which are not fully backwardly inductive and that can help to explain some key features of the experimental data.

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

A great number of hypotheses have been put forth in order to explain Africa's anomalously poor economic performance (Acemoglu et al., 2001; Easterly, Levine, 1997; Rodney, 1982). 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, Björn Wallace and I 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 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 (Alesina et al., 2003; Easterly, Levine, 1997; 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 (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. 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 evidence arguably casts doubt on, rather than support, the hypothesis that past slave exports negatively affected economic development in post-independence Africa.

3. Privacy, reputation and limited records: a survey

Reputation is a potentially valuable asset because it helps individuals to signal their trustworthiness and their value as a contractual counterpart (Bar-Isaac, Tadelis, 2008). Typically, reputations are accumulated over time, and are backed by relevant information on past behavior and performances. For this reason, the process of reputation formation is likely to benefit from the collection, storage and diffusion of personal data, and to be at odds with privacy concerns that call for some of these data to remain undisclosed. In a neoclassical market, in fact, the participants will themselves optimally decide which, and how much, information is needed to assess the value of a transacting partner. Accordingly, a privacy regulation that mandates the removal of data from the public domain is at best unjustified and, at worst, it opens the door for "fraud" (from Posner's blog, 8th May 2005, see also Posner, 1983, Nock, 1993).

However, besides the canonical argument that we are not (always) capable of optimal information-processing in the neoclassical sense, there are instances in which privacy regulations can alleviate market failures that arise even when assuming rational agents (Acquisti, 2010; Acquisti, Varian, 2005; Calzolari, Pavan, 2006; Hermalin, Katz, 2006; Hirshleifer, 1971; Taylor, 2004a, 2004b). Moreover, in many situations, reputation effects have been demonstrated to disappear over time (Cripps et al., 2004) when access to information is unrestricted.

Originally, in response to this last concern, limited records were put forward as a device to help sustain reputation incentives over time (Ekmekci, 2011; Liu, 2010; Liu, Skrzypacz, 2011). However, limited records can also be seen as an arrangement that protects privacy, for they are compatible with the principle of data retention. These two considerations highlight the key appeal of this informational arrangement: it can simultaneously fulfill privacy needs and increase efficiency via strengthening reputation concerns. Being also relatively easy to implement, limited records appear to be an ideal policy for a wide range of real-life situations, where either privacy is socially desirable, or reputation is needed to curb informational asymmetries.

Motivated by this link between limited records, reputation and privacy, Giancarlo Spagnolo and I review the theoretical literature up to date, and find that the state of the art is unsatisfactory. In particular, the relevant game theoretic works do not offer any indication on how the length of records should vary with the model's fundamentals, although limited records have been demonstrated capable of supporting reputation incentives over time. For what concerns the privacy debate, although privacy per se is a topic not unknown to the economics literature (Acquisti, 2010, and references therein), the study of limited records has been left almost neglected.

The lack of a cohesive theoretical foundation guiding the design of limited records is highlighted by observing how they have been implemented in real-life situations. Reviewing the evidence, we uncover a tremendous amount of heterogeneity, which we believe stresses once more that a unified framework of reference is both missing and needed.

4. Limited credit records and market outcomes

In most modern credit markets, credit registers collect and share data on borrowers' past behavior with other market participants, in an effort to reduce the well-known informational asymmetries that characterize lending relationships (Jappelli, Pagano, 2002). These asymmetries take the form of adverse selection (Akerlof, 1970) or moral hazard. The former arises before the loan is contracted, because lenders can at best only partially observe a borrower's ability and other relevant characteristics. Moral hazard instead occurs after the loan is contracted and stems from the imperfect observability of the borrower's actions. Consequently, reliable information on a borrower's past behavior serves both as a screening device for lenders, and as a "reputational" collateral for borrowers (Vogel, Burkett, 1986), which they can use to signal to potential lenders their private information and intended choices.

Interestingly, from a legal point of view credit data fall under the umbrella of personal data and are thus subject to the privacy protection provisions that regulate the extent to which personal information can be handled, shared and stored. In particular, an important by-product of privacy protection is the principle of data retention, which prescribes that personal data, as collected by any data user, should be retained for a limited period of time only, and by any means for no longer than what is necessary for the intended purpose. Although such a principle may look intuitively appealing, it is not immediately clear whether it is desirable from a strictly economic standpoint. In fact, data retention directly mandates the removal from the market of the very data that have a key role in disciplining the informational asymmetries described above. Indeed, its application has stirred an intense and so far unresolved debate about the implications of limited records for privacy, efficiency and reputation (Bottero, Spagnolo, 2011; Sartor, 2006).

The inconclusiveness of the academic and political debate is reflected in the choices made by policy makers. As of 2007, about 90 percent of the countries in which a credit register was operating mandated the removal of (at least some) credit data after a limited period of time (Elul, Gottardi, 2008). Although suggestive of a certain degree of homogeneity across countries, this is only superficial. In practice, there is no consensus on how to set the length of credit registers' records.

This paper intends to develop a formal analysis of the effects of data retention provisions on borrowers' behavior and market outcomes. To this end, I study a stylized model of a credit market for consumers and small businesses where a credit register decides on the amount of information to be disclosed to the lenders. I directly compare three informational arrangements, which I call full withholding, full disclosure and limited records. Under full withholding, privacy is enforced and no information about the borrowers is disclosed. Under full disclosure, all the relevant information is made available to the lenders. Finally, with limited records, the credit register implements a data retention policy.

My analysis demonstrates that it is possible to characterize sufficient conditions under which each of the three arrangements leads to a lower (ex ante) default rate than the other two. Interestingly, when the moral hazard in the market is high, limited records outperform the two other arrangements, supporting the case for privacy protection legislations. In a situation of high moral hazard, in fact, the two extreme policies of full disclosure and withholding hamper the performance of low-ability borrowers while not eliciting a sufficiently higher repayment rate from the high-ability group. Limited records, instead, provide borrowers of all types with the possibility to alleviate, at least in part, the moral hazard problem by providing them with a credible device, the records, to build or restore a reputation for being a trustworthy borrower.

By moderating the moral hazard problem, this last arrangement leads to the lowest default rate.

Further, it turns out that limited records may also lead to higher aggregate welfare, albeit under fairly restrictive conditions. More precisely, when moral hazard is high and borrowers are not too different in their abilities, limited records are more efficient because the resulting interest rates succeed in increasing borrowers' aggregate welfare, while keeping lenders' welfare constant. Yet, under the model's assumptions, if these conditions are not met, it is preferable to disclose all information about the borrowers.

Summing up, the main contribution of this analysis is to show that data-protection laws endogenously create scope for the borrowers to link their current choices of effort to better contractual terms in the future, by building a reputation for being of high ability and thus securing themselves contracts at lower interest rates. Such a possibility elicits higher effort, which, under the assumption of the model, is beneficial for the overall economy, by prompting higher repayment rates and lower interest rates.

References

Acemoglu, D. , Johnson, S. , Robinson, J. A. , 2001. The colonial origins of comparative development: an empirical investigation. Amer. Econ. Rev. : 91, 1369-1401.

Acquisti, A. , 2010. The economics of personal data and the economics of privacy. OECD Committee for Information, Computer and communication policy, DSTI/ ICCP /IE/ REG (2010)4.

Acquisti, A. , Varian, H. R. , 2005. Conditioning Prices on Purchase History. Marketing Sci. : 24, 1-15.

Alesina, A. , Devleeschauwer, A. , Easterly, W. , Kurlat, S. , Wacziarg, R. , 2003. Fractionalization. J. Econ. Growth : 8, 155-194.

Akerlof, G. A. , 1970. The market for "lemons": quality uncertainty and the market mechanism. Quart. J. Econ. : 84, 488-500.

Bar-Isaac, H. , Tadelis, S. , 2008. Seller reputation. Foundations and Trends in Microeconomics: 4, 273-351.

Calzolari, G. , Pavan, A. , 2006. On the optimality of privacy in sequential contracting. J. Econ. Theory :130, 168-204.

Camerer, C. , Ho, T. H. , Chong, J. K. , 2004. A cognitive hierarchy theory of one-shot games. Quart. J. Econ. 119, 861-898.

Cashdan, E. , 2001. Ethnic diversity and its environmental determinants: effects of climate, pathogens, and habitat diversity. American Anthropologist: 103, 968-991.

Collard, I. F., Foley, R. A. , 2002. Latitudinal patterns and environmental determinants of recent human cultural diversity: do humans follow biogeographical rules? Evol. Ecol. Res. : 4, 371-383.

Costa-Gomes, M. , Crawford, V. , Broseta, B. , 2001. Cognition and behavior in normal-form games: An experimental study. Econometrica. 69, 1193-1235.

Costa-Gomes, M. , Crawford, V. , 2006. Cognition and behavior in two-person guessing games: An experimental study. Amer. Econ. Rev. 96, 1737-1768.

Cripps, M. W. , Mailath, G. J. , Samuelson, L. , 2004. Imperfect monitoring and impermanent reputation. Econometrica: 72, 407-432.

Easterly, W., Levine, R. , 1997. Africa's growth tragedy: policies and ethnic divisions. Quart. J. Econ. : 112, 1203-1250.

Ekmekci, M. , 2011. Sustainable reputations with rating systems. J. Econ. Theory : 146, 479-503.

Fey M. , McKelvey, R. , Palfrey, T. , 1996. An experimental study of constant sum Centipede games. Int. J. Game Theory. 25, 269-287.

Hermalin, B. E. , Katz, M. L. , 2006. Privacy, property rights and efficiency: The economics of privacy as secrecy. Quant. Market. Econ. : 4, 209-239.

Hirshleifer, J. , 1971. The private and social value of information and the reward to inventive activity. Amer. Econ. Rev. : 61, 561-574.

Ho, T. H. , Camerer, C. , Weigelt, K. , 1998. Iterated dominance and iterated best response in experimental 'p-beauty contests'. Amer. Econ. Rev. 39, 649-660.

Jappelli, T. , Pagano, M. , 2002. Information Sharing, Lending and Defaults: Cross-Country Evidence. J. Banking Finance : 26, 2017-2045.

Liu, Q. , Skrzypacz, A. , 2011. Limited records and reputation. Mimeo.

Liu, Q. , 2010. Information Acquisition and Reputation Dynamics. forthcoming, Rev. Econ. Stud. .

McKelvey, R. , Palfrey, T. , 1992. An experimental study of the centipede game. Econometrica: 60, 803-836.

Nagel, R. , 1993. Experimental results on interactive competitive guessing. Discussion paper No. B-236, University of Bonn.

Nagel, R. , 1995. Unraveling in guessing games: An experimental study. Amer. Econ. Rev. 85, 1313-1326.

Nagel, R. , Tang, F. , 1998. Experimental Results on the Centipede Game in Normal Form: An Investigation on Learning. J. Math. Psychol. 42, 356-384.

Nock, S. L. , 1993. The costs of privacy: surveillance and reputation in America. New York: Aldine.

Posner, R. A. , 1983. Privacy and related interests. In: The economics of Justice. Harvard University Press.

Nunn, N. , 2008. The long-term effects of Africa's Slave Trades. Quart. J. Econ.:123, 139-176.

Posner, R. A. , 1983. Privacy and related interests. In: The economics of Justice. Harvard University Press.

Rapoport, A. , Stein, W. E. , Parco, J. E. , Nicholas, T. E. , 2003. Equilibrium Play and Adaptive Learning in a Three-Person Centipede Game. Games Econ. Behav. 43, 239-265.

Rodney, W. , 1982. How Europe underdeveloped Africa. London, UK: Bogle-L'Ouverture, 1972.

Rosen, J. , 2000. The unwanted gaze: The destruction of privacy in America. Random House, New York.

Rosenthal, R. , 1982. Games of perfect information, predatory pricing, and chainstore paradox. J. Econ. Theory. 25, 92-100.

Rosenzweig, M. L. , 1992. Species diversity gradients: we know more and less than we thought. J. Mammal: 73 , 715-730.

Rosenzweig, M. L. , 1995. Species diversity in space and time. Cambridge, UK: Cambridge University Press.

Sartor, G. , 2006. Privacy, Reputation, and Trust: Some Implications for Data Protection. Mimeo.

Stahl, D. , Wilson, P. W. , 1994. Experimental evidence on players' models of other players. J. Econ. Behav. Organ. 25, 309-327.

Stahl, D. , Wilson, P. W. , 1995. On Players' Models of Other Players: Theory and Experimental Evidence. Games Econ. Behav. 10, 218-254.

Taylor, C. , 2004a. Privacy and information acquisition in competitive markets. W.P. 13271, Olin Program in Law & Economics, UC Berkeley.

Taylor, C. , 2004b. Consumer privacy and the market for customer information. RAND: 35, 631-650.

Vogel, R. C. , Burkett, P. , 1986. Deposit mobilization in developing countries: the importance of reciprocity in lending. J. Develop. Area: 20, 425-438.

PAPER 1

Cognitive Hierarchies and the centipede game

Margherita Bottero

Abstract. Here, I adopt the cognitive hierarchy approach to analyze the centipede game. To this end, I present and study an extensive-form version of Camerer et al.'s (2004) original normal-form model. The resulting predictions are evaluated empirically using laboratory data borrowed from a previously published experiment. The paper features two main contributions. First, it extends the experimental literature on the cognitive hierarchy model, which is here directly applied to the extensive-form centipede game. Secondly, it demonstrates that in the centipede game the cognitive hierarchy approach leads to predictions which are not fully backwardly inductive and that can help to explain some key features of the experimental data.

1. Introduction

There is by now ample evidence demonstrating that when individuals are presented with game theoretic interactions their behavior frequently departs from the equilibrium predictions, especially during the first rounds of play (see for instance Costa-Gomes and Crawford, 2006, and references therein). This empirical anomaly motivates the ongoing search for new, non-equilibrium, models that are capable of explaining the behavioral patterns observed in laboratory and field experiments. Among the many contributions in this literature, the Level k (Nagel, 1995; Stahl, Wilson, 1994, 1995) and cognitive hierarchy models (Camerer et al., 2004) stand out for their success in predicting and explaining actual behavior in a number of normal-form games by relying only on a parsimonious set of assumptions[1]. How these models perform when applied

I would like to thank Tore Ellingsen, Erik Mohlin, Juanna Joensen, Björn Wallace, Jörgen Weibull and participants at the AISC VI meeting in Naples, at the SSE lunch seminar series and at the 3rd Maastricht Behavioral and Experimental Economics Symposium for their helpful comments and suggestions. I gratefully acknowledge the Knut and Alice Wallenberg Research Foundation for financial support.

[1] See for instance Costa-Gomes et al. (2009) for applications of the Level k and cognitive hierarchy models to various normal-form games, based on laboratory data. Crawford and Iriberri (2007a) and Brocas et al. (2009) study games with private information using the Level k model. Östling et al. (2009) and Brown et al. (2009) test the Level k model's predictions using field data. Ellingsen and

to extensive-form games is, however, still an open question. Here, I address this issue by suggesting and testing a possible extension of the cognitive hierarchy model that is directly applicable to the centipede game (Rosenthal, 1982).

The centipede game is a two-player game of perfect information that represents an interactive situation where the players play sequentially and non-cooperatively for a limited number of rounds. The players take turns in deciding whether to terminate the game or to continue it. Continuing increases the value of the joint payoffs, the pot, but is not individually profitable if the game is terminated in the next round. If so, the player who passed receives a payoff lower than what he could have secured himself by terminating at the preceding round. If players seek to maximize their monetary gains, the centipede game features a unique (subgame perfect) equilibrium, in which the player who moves first terminates the game. This sharp theoretical prediction is at odds with the behavior of actual players who typically continue until the central rounds, and occasionally until the end. Empirically, no, or very few, participants have been observed to take the pot at the first round (see for instance Fey et al., 1996; McKelvey and Palfrey, 1992; Nagel and Tang, 1998; Rapoport at al., 2003).

This divergence between the equilibrium predictions and the experimental evidence appears analogous to that which has been successfully addressed by non-equilibrium models in normal-form games (Costa-Gomez et al., 2001; Costa-Gomez and Crawford, 2006; Ho et al., 1998; Nagel, 1993, 1995; Stahl and Wilson 1994, 1995). The key novelty of these models is that players are allowed to hold beliefs regarding the behavior of their opponents that are much simpler than those assumed in traditional game theory. More precisely, the Level k model assumes that players' behavior can be described by their (behavioral) type which they draw at the beginning of the game from a hierarchy of possible types. At the bottom of this hierarchy lies a non-strategic type 0, who uniformly randomizes among his actions. Next, there is type 1, who believes that the opponent will act as a type 0, and who plays a best response based on this conjecture. Higher types adjust their beliefs via recursive thought experiments: type 2 best responds to type 1, type 3 to type 2 etcetera. This assumption is appealing because it grants rationality to players of all types but type 0, while allowing them to adopt non-equilibrium behaviors thanks to the simplified belief structure.

Camerer et al.'s (2004) cognitive hierarchy model builds on the Level k approach by allowing for a slightly more complex belief structure. More precisely, the model allows a type k player, $k > 0$, to envision the possibility that his opponent may be of any of the lower types. Thus, a type k not only best responds to type $k-1$, but to a mixture

Östling, (2006) study several games with pre-play communication, where the communication phase and the actual play of the game can be thought of as two separate rounds of an extensive-form game.

of lower-level types. The types' frequencies may be treated either as a parametrized distribution, usually Poisson, or as an unconstrained types distribution. This richer set of assumptions has been successfully employed to explain players' behavior in a number of two-person finite dominance-solvable games, such as guessing games (Costa-Gomez and Crawford, 2006), zero-sum hide-and-seek games (Crawford and Iriberri, 2007b) and even the normal-form centipede game (Kawagoe and Takizawa, 2010).

In this paper, I present a version of the cognitive hierarchy model that is directly applicable to the extensive-form six-step centipede protocol that was studied by McKelvey and Palfrey (1992). I call this model the *EF-CH* model, where the acronym stands for extensive-form cognitive hierarchy. Possibly, the *EF-CH* model is the most faithful translation of Camerer's model to this extensive-form game. The choice to apply the model to McKelvey and Palfrey's six-step centipede game is functional[2] to producing predictions that are readily testable, thanks to the experimental data collected and made available by the authors.

The *EF-CH* model rests on four assumptions. First, players can be of one of three types, 0, 1 and 2. Second, type 0 players randomize between actions with a constant probability, while players of type 1 and 2 use backward induction. Third, players of type 1 and 2 think that the opponent is of any of the types below their own, and they share the same belief on type 0s' randomization probability. Note that type 1s and type 2s do not envision the possibility of types equal to, or higher than, their own. The fourth assumption is methodological, players maximize their monetary payoffs. Together, these hypotheses allow me to derive the payoff-maximizing behavior, in expectation, for each type and thus to lay down verifiable predictions on the play of the game and the frequency of types in the experimental population.

In addition, I also study two extensions of such a model. First, I allow type 1 and type 2 players to have different beliefs on the behavior of type 0s. This special case, which I call *heterogenous EF-CH* (*HEF-CH*), features a more "sophisticated" type 2, that not only forms his own belief on the behavior of type 0s, but also holds a (second-order) belief on type 1s' beliefs on type 0. This novel assumption departs from the original spirit of the cognitive hierarchy model by undermining its tight recursive structure, but it allows for a larger degree of flexibility that can better accommodate a variety of situations. For example, imagine that the players in a population can be divided into three groups. Players in the first group randomize, while those in the second group expect, erroneously, to meet altruistic opponents and best respond to this belief. Players in the third group instead realize that there is the possibility of

[2] After having presented the model, I will discuss whether it can be generalized to any two-player, finite, extensive-form game of perfect information and the challenges that such a task would face.

meeting players from either of the two previous groups, and correctly anticipate the mixture of their behaviors and best respond to it. For instance, it is possible to think of situations involving different sets of social norms that have come to coexist. This and other similar situations can be captured by the *HEF-CH* specification. Continuing with the example, the model would let type 1 players believe that type 0s are altruists, and allow type 2 players to correctly recognize type 1s' mistake.

The second extension I consider is to allow for Bayesian updating in the players' decision processes. There is evidence that when playing repeatedly in the same interaction players revise their prior beliefs regarding the opponent's type using Bayes' Law, although this topic is still open to debate (Cosmides and Tooby, 1996; Fudenberg and Levine, 1998, and references therein). Updating is a cognitively advanced task, which requires the use of costly cognitive resources to improve upon the priors. For this reason, and to be consistent with the spirit of the cognitive hierarchy model, I allow only type 2 players to update their belief on the type of the opponent.

The *EF-CH* model produces a large number of theoretical results, that describe the behaviors that maximize the players' expected payoffs for any belief they may have. Among these optimal behaviors, there is the backward induction prescription to terminate at the first round, but it is far from being the only one. In particular, there are other beliefs such that it is payoff maximizing for players of type 1 and 2 to continue until the last node. Bayesian updating does not have any observable effect on the predictions of the model, neither by introducing new payoff-maximizing behaviors nor by eliminating existing ones. However, it alters the set of beliefs for which the prescribed behaviors are payoff-maximizing. Using the assumptions of the *HEF-CH* model, instead, I can characterize new payoff-maximizing behaviors, including type 2 players terminating at the central round. Borrowing data from McKelvey and Palfrey's (1992) experiment, I test whether the predictions of the model can actually explain experimental behavior. The *EF-CH* model fits the data fairly well, but the heterogenous extension outperforms it in terms of empirical fit. In general, all specifications fall short of fully fitting the data, by failing to predict enough terminations at the central nodes of the game. I attribute this shortcoming to the assumption of constant randomization from part of the type 0s, and I will discuss this issue further in section 6.

To my knowledge, this paper is the first that applies and tests the cognitive hierarchy framework directly to the extensive-form centipede game. Camerer (2003) mentions that the Level k model could explain the experimental evidence for this game, but he does not provide a formal analysis. Ho et al. (2008) study a normal-form centipede game assuming a cognitive hierarchy structure that allows for learning. Similarly, Nagel and Tang (1998) look at a normal-form centipede game to evaluate different learning models.

Kawagoe and Takizawa (2010) compare the empirical fit of the Level k model and the Agent quantal response model[3], without and with altruistic types (Fey et al., 1996; McKelvey and Palfrey, 1992, 1998), when applied to the reduced normal-form[4] of the centipede game. The results of their analysis show that the Level k model outperforms the other two. My work differs from Kawagoe and Takizawa principally in that I study the cognitive hierarchy, rather than the Level k, model. Further, I look directly at the extensive-form centipede, and introduce a model that is tailored to translate the normal-form cognitive hierarchy model into a proper extensive-form model. Although Kawagoe and Takizawa present evidence that supports the equivalence of studying the reduced normal-form and the extensive-form of the centipede, I believe that there are substantial gains from applying non-equilibrium models directly to extensive-form games. First, this permits us to reason about non-equilibrium assumptions in a dynamic framework and about their interaction with mechanisms such as belief updating. Secondly, this exercise can contribute to the search for a general extensive-form non-equilibrium model. Besides these and some other minor divergences, however, Kawagoe and Takizawa's and my paper are very close in spirit and they are inspired by the same question: how would non-equilibrium models fare when applied to extensive-form games.

Palacios-Huerta and Volij (2009) recently presented interesting experimental results on the centipede game (for a discussion, see also List et al., 2010). They documented that chess players rarely fail to terminate at the first round of the game (the subgame perfect equilibrium) when playing with other chess players, while they typically continue to later rounds when playing with college students. In line of principle, this result could be given a cognitive hierarchy interpretation. It is possible that chess players have beliefs on the behavior of other chess players and college students such that it is payoff maximizing, in expectation, to terminate as soon as possible with the first group and to continue for a while with the latter[5]. Finally, in a recent paper, Gerber and Wichardt (2010) investigate the efficacy of bonuses and insurances as deferring devices

[3] The agent quantal response model assumes that each player treats his future self as an independent player with a known probability distribution over actions.

[4] To corroborate this assumption, the authors present the result of an experiment showing that players do not behave differently when the same game is presented to them in extensive or normal form.

[5] Interestingly, also college students are documented to terminate earlier when playing with chess players rather than with other students.

for postponing termination of the play of the centipede game[6]. In explaining their experimental results, the authors discuss a model that has a flavor of non-equilibrium thinking but is neither the Level k nor the cognitive hierarchy model.

The remainder of the paper is organized as follows. Section 2 discusses the centipede game and the relevant literature in ampler detail. Section 3 introduces the assumptions of the model and section 4 analyzes the results. Section 5 presents the empirical analysis. Section 6 discusses the results and section 7 concludes.

2. The centipede game

In game theory, the centipede is a finite and sequential game of perfect information with two player roles, here labelled A and B. Following McKelvey and Palfrey (1992), I here consider the 6-node version of the centipede game that is displayed in Figure 1. At each node j, the player who has to move can choose either to continue to play (c_j) or to terminate the game (t_j). Upon the decision to terminate the game at a node j, the players receive the payoffs appearing at the end of branch t_j, where the first amount pertains to player (role) A. At the first node, the total pot sums up to \$0.5, thereafter each time it is passed on its total value is multiplied by two, until it reaches a total of \$32 in the last node.

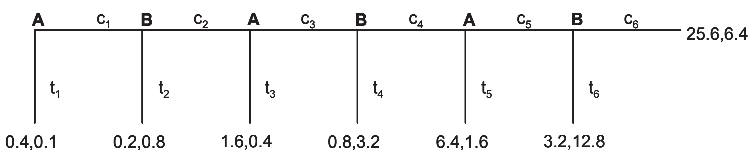


Fig. 1. The Centipede game

Since the game is sequential and there is perfect information about the history of play, a new subgame begins at each node, for a total of six (proper) subgames. Subgames are important in the search for the equilibrium because a strategy profile such that the players do not want to deviate at any subgame is also a subgame perfect strategy profile (Selten, 1965). Following standard practice in game theory, it is assumed that the players are rational and have von Neumann-Morgenstern utility functions. Kuhn (1950, 1953) demonstrated that subgame perfect equilibria can be found working *backwardly* from the end of the tree, replacing the starting node of each subgame by the vector of payoffs that corresponds to the optimal choice at the subsequent nodes.

[6] The authors are motivated by the fact that, in the standard centipede game, longer interactions result in increased aggregate welfare for both players.

Applied to the game in Figure 1, this procedure begins with considering player B's choice at node 6, the last subgame of the game. There, player B's optimal choice is t_6 because, being rational, he chooses the action that delivers him the higher payoff. Then, the players can replace the subgame with the payoff vector corresponding to the choice t_6, obtaining a new, artificial centipede. Applying the same reasoning, it is possible to establish that player A, being rational, will choose t_5 at node 5, and so on. The collection of these choices, one for each personal decision node of the game tree, defines the unique[7] subgame perfect equilibrium of the game, that is, for each player to terminate as soon as he gets the move.

Despite being straightforward, the backward induction result is uncomfortable, for it clashes with experimental evidence and arguably with introspection. To improve on this front, the game theoretic literature has proposed several equilibrium refinements. In particular, these suggested refinements all deal with the issue of what players should believe after observing a deviation from the equilibrium strategy. The backward induction procedure, in fact, rests on the assumption that both players know that they both will keep the pot if given the chance, and that they both know that they both know this fact, and so on. Modifying this assumption allows us to obtain other equilibria (see, for instance, Asheim and Dufwenberg, 2003; Ben-Porath, 1997; Kreps and Wilson, 1982; Kreps et al., 1982; Reny, 1998; Selten, 1975).

In this paper, I depart from the equilibrium analysis to assume that players are rational and that they hold mutual and subjective conjectures on the play of the opponent. To describe how these payoff-maximizing players play the game it is sufficient to know which conjecture they have about the opponent (a conjecture is here the node where the opponent is believed to terminate). In the next section, I will show that these simple requirements go a long way in explaining the laboratory evidence on the centipede.

As mentioned in the Introduction, the reason for focusing on the game in Figure 1 is that it is part of a series[8] of experiments conducted by McKelvey and Palfrey (1992). Thanks to the availability of their laboratory data, it is possible to directly test the model's predictions. The experiment consists of three[9] sessions, labeled 5, 6 and 7, two with twenty participants, and one with eighteen, none of whom had played the game before. At the beginning of each session, participants were divided into two groups

[7] Thanks to the following proposition: every finite, generic, extensive form game of perfect information has a unique subgame perfect equilibrium.

[8] In the same paper, the authors consider also two four-step games, one with a "high payoff" treatment.

[9] Sessions 1, 2 and 3 involved other treatments. Specifically, these consisted of a shorter (4-step) centipede game, and a "high payoff" condition with stakes substantially higher than those considered here.

of ten (or nine in the case of session 6) that were randomly allocated to roles A or B. Each subject was then matched with an individual from the other group to play the game ten (nine) times. It was made common knowledge at the beginning of each session that no individual was going to be matched with the same opponent twice. In this sense, there is no learning, and matching can be treated as random.

The laboratory data that was collected in this experiment uncovered two main features that characterize the subjects' behavior. Table 1 shows the frequency f_j of plays terminating at node j. The backward induction outcome, which would require $f_1 = 1$, never occurs in the data. What instead appears is that terminations occur mostly after the third node, and that they are concentrated around nodes 3, 4 and 5. Few plays continue without interruption to the end of the centipede.

Table 1: Frequency of Observations at each Terminal Node

Session	N	f_1	f_2	f_3	f_4	f_5	f_6	f_7
5	100	.02	.09	.39	.28	.20	.01	.01
6	81	.00	.02	.04	.46	.35	.11	.02
7	100	.00	.07	.14	.43	.23	.12	.01
Total 5-7	281	.007	.064	.199	.384	.253	.078	.014

It is possible to use the f_js to back out the implied probability p_j of terminating at any given node j, conditional on the event that node j is reached, $p_j = f_j / \left(f_j + ... + f_7\right)$. These are shown in Table 2.

Table 2: Implied Take Probabilities

Session	p_1	p_2	p_3	p_4	p_5	p_6
5	0.02	0.09	0.44	0.56	0.91	0.50
6	0.00	0.02	0.04	0.49	0.72	0.82
7	0.00	0.07	0.15	0.54	0.64	0.92
Total 5-7	0.01	0.06	0.21	0.53	0.73	0.85

For instance, in Session 5, if a play reaches node 5, the ex ante probability that it will end there is $p_5 = f_5 / \left(f_5 + f_6 + f_7\right) = 0.2 / \left(0.2 + 0.01 + 0.01\right) = 0.91$. The backward induction outcome would imply $p_1 = 1$, but this is not documented in the data. The most evident and consistent empirical pattern is that the probability of taking the pot increases as the play approaches the last nodes[10].

Summing up, there are two important empirical features that a theory aiming to explain the centipede game should replicate. The first is that the distribution of terminations is concentrated around the mean, and the second is that the pattern of the conditional take probability is increasing in the length of the centipede.

[10] In Session 5, the figure at the last move (0.50) is based on two observations only.

3. The *EF-CH* model

Consistent with experimental evidence, I assume that players can be of one of three types, 0, 1 and 2, which are recursively defined as follows. Players of type 0 are at the bottom of the types' hierarchy and behave as non-strategic randomizers. At the beginning of the game, they observe a realization q of the random variable $\tilde{q}$ and continue at each node with that probability. Let $\tilde{q}$ be distributed with density f over the support $[0, 1]$. Note that there is no heterogeneity among type 0s, in the sense that they all randomize in the same way. Also note that when q is equal to one half we have the uniform randomization case. Next in the hierarchy, is type 1. Players of this type are Savage rational decision-makers that maximize their expected monetary payoffs through reasoning via backward induction. In doing so, they assume that *all* the other players are of type 0, and they assign a point belief λ to q. The belief λ is the realization of a random variable $\tilde{\lambda}$ defined on the same support as q, which is observed by all type 1 players at the beginning of the game. This implies that, as was the case for type 0s, all players of type 1 are homogeneous in their beliefs (and hence in their behavior). The players of type 2 are at the top of the type-hierarchy. They are similar to type 1s in three respects. First, they are Savage rational maximizers that reason via backward induction. Secondly, they observe the same realization λ and use it as point belief on q. Thirdly, they do not consider the possibility that other players may be of the same type as their own. However, differently from type 1s, they can guess correctly that a fraction s_0 of the opponents is of type 0 and a fraction s_1 is of type 1. Since the two fractions do not sum up to one[11], type 2s normalize the probability of playing with a type 0 and a type 1 to $\sigma_0 = \frac{s_0}{s_0+s_1}$ and to $\sigma_1 = 1 - \sigma_0 = \frac{s_1}{s_0+s_1}$ respectively.

I will extend this model in two ways. First, I will assume that type 2 players observe λ, and thus that they are able to correctly predict the behavior of type 1 players, but assign to q a point belief μ, which is the realization of a random variable $\tilde{\mu}$ different from $\tilde{\lambda}$[12]. I call this case *HEF-CH*, which stands for *heterogenous EF-CH*, because type 1 and type 2 players have heterogeneous beliefs. The second extension that I consider is to allow for Bayesian updating during the play of the centipede game. In particular, I study an extension of the model above where I allow type 2 players to engage in Bayesian updating on the prior σ_0, after having observed the opponent's move. Given that type 2s correctly anticipate the node at which type 1 players terminate, they update on σ_0 by making use of the fact that *before that node* only players of type 0 drop off the game. Upon observing a continuation decision

[11] Following Camerer at al. (2004), type 2 players do not envision the possibility of players of their own or higher type, and accordingly they normalize the true proportions of type 0s and 1s in the population (that do not sum up to one) by dividing them by their sum.

[12] But defined on the same support $[0, 1]$.

before the type 1s' expected terminal node, type 2 players re-assess downwardly the probability of being playing with a type 0. More precisely, after having observed the first pass from the opponent, a type 2 player can revise his prior to be playing with a type 0 player downward to $\sigma_0^{R1} = \frac{\sigma_0 \rho_2}{1-\sigma_0(1-\rho_2)}$, where the superscript $R1$ stands for first revision. Accordingly, the updated probability that the opponent is instead of type 1 becomes $\sigma_1^{R1} = 1 - \sigma_0^{R1} = \frac{1-\sigma_0}{1-\sigma_0(1-\rho_2)}$. If the opponent chooses to pass a second time, such a belief can be updated again, this time to $\sigma_0^{R2} = \frac{\sigma_0 \rho_2^2}{1-\sigma_0(1-\rho_2^2)}$.

Given the length of the centipede game in Figure 1, players revise only twice (at nodes 3 and 5 for player role A and nodes 2 and 4 otherwise[13]). However, the updating process can be generalized to any number n of revisions: $\sigma_0^{Rn} = \frac{\sigma_0 \rho_2^n}{1-\sigma_0(1-\rho_2^n)}$. As expected, the revised belief goes to 0 as the number of revisions goes to infinity,

$$\lim_{n\to\infty} \sigma_0^{Rn} = \frac{\sigma_0 \rho_2^n}{1-\sigma_0(1-\rho_2^n)} = 0$$

Similarly, σ_1^{Rn} goes to 1, $\lim\limits_{n\to\infty} \sigma_1^{Rn} = 1$.

Applied to the centipede game, these assumptions will result in a set of outcomes, where by outcome I mean a collection of terminal nodes and beliefs for type 1s and 2s, such that, given the beliefs, it is payoff maximizing to stop at the prescribed nodes. Note that *ex ante* players' beliefs can fall anywhere on the support, so that the model is descriptive of a variety of behaviors. To increase the precision of its predictions, it is possible to specify more accurate hypotheses on the players' beliefs. Alternatively, if one wants, or has to, remain agnostic about the players' beliefs, an "empirical approach" can be adopted. This consists in taking all the outcomes to the data and judge empirically which one fits the evidence better. In this paper, I will follow the latter approach. First, however, the next section presents the outcomes of the model.

4. Analyzing the model

In this section, I first study the *EF-CH*'s predictions. Then, I consider in turn the *Bayesian* case, in which type 2 players undertake Bayesian updating, and the *HEF-CH* case, which allows type 2 players to form a belief μ on q that is different from λ. In general, the assumptions lead to a partition of the belief-space Ω in a number of non-overlapping components o such that $\bigcup\limits_{o\in\Omega} = \Omega$. Each component o, which I call *outcome*, is associated with a set of strategies (r_A^*, r_B^*) for type 1s and (t_A^*, t_B^*) for type 2s, such that if a player's belief(s) fall in that outcome it is payoff maximizing for him to follow r^* or t^*. For each outcome, or, alternatively, for each belief the players may have, it is possible to derive the probability $\bar{q}_n$, $n = 1, 2, ..., 7$, that a randomly drawn

[13] Revision at node 6 is possible but useless, given that the game terminates there.

pair of players terminates at node n. The probability $\bar{q}_n$ is a function of the shares of the players' types in the population, s_0, s_1 and s_2, and of q, the type 0's randomization probability, and, as we will see in the next section, it will serve as the basis for the maximum likelihood analysis.

Let us begin with the *EF-CH* model, and assume that type 1s and type 2s have observed a realization λ of the belief on q. Consider a type 1 individual in role A that maximizes the expected payoff $E[\pi(\cdot)]$ of his actions. At the last node he compares the expected value of c_5, $E\pi(c_5) = \lambda\pi(c_6) + (1-\lambda)\pi(t_6)$, and that of t_5, $\pi(t_5)$, and plays the action with highest expected value. Then, he takes one step back and proceeds in the same way at node 3 and 1. Reasoning in a similar fashion for the other player role, the expected payoff maximizing strategies for a type 1 are,

$$r_A^* = \begin{cases} c_1c_3c_5 & \text{if } \lambda \geq \frac{1}{7} \\ t_1t_3t_5 & \text{otherwise} \end{cases} \tag{4.1}$$

$$r_B^* = \begin{cases} c_2c_4t_6 & \text{if } \lambda \geq \frac{1}{7} \\ t_2t_4t_6 & \text{otherwise} \end{cases}$$

By assumption of the *EF-CH* model, type 2 players assign to q the same λ as type 1s, hence they can anticipate the behavior described in expression 4.1. Assigning belief σ_0 to the event that the opponent is of type 0, type 2 players reason backwardly to find their payoff maximizing behavior. After some computations, it turns out that the payoff maximizing strategies t_A^*, t_B^* are,

$$t_A^* = \begin{cases} c_1c_3c_5 & \text{if } \sigma_0 \geq \bar{\lambda} \text{ and } \lambda \geq \frac{1}{7} \\ c_1c_3t_5 & \text{if } \sigma_0 < \bar{\lambda} \text{ and } \lambda \geq \frac{1}{7} \\ t_1t_3t_5 & \text{otherwise} \end{cases} \tag{4.2}$$

$$t_B^* = \begin{cases} c_2c_4t_6 & \text{if } \lambda \geq \frac{1}{7} \\ t_2t_4t_6 & \text{otherwise} \end{cases}$$

where $\bar{\lambda} = \frac{6.4}{22.4\lambda+3.2}$. The interpretation of expression 4.2 is intuitive. Type 2s' strategy depends both on λ, which is the parameter describing the behavior of type 0s, and type 1s, and on the belief σ_0 on the shares of type 0s in the population. As long as the type 0 players are believed to randomize with a probability that is high, $\lambda \geq \frac{1}{7}$, it is optimal for the type 2s to terminate at the last round. In particular, if type 2s believe that there are many type 0s that continue with a high probability, then it is payoff-maximizing, in expectation, to choose *pass* even at node 5. Note however that the rationality assumption makes sure that, when playing in position B, type 2s always take at node 6. When $\lambda < \frac{1}{7}$, instead, both type 1s and type 2s terminate at the first node.

To sum up, the *EF-CH* assumptions can give rise to three outcomes, only one of which prescribes the backward induction behavior. The outcomes can be represented on the (λ, σ_0) plane, as shown in Figure 2,

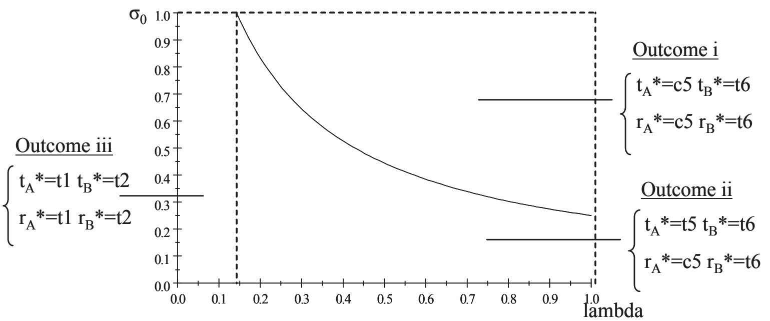


Fig. 2. Outcomes of the EF-CH model

Outcome *i* consists of those beliefs such that both types 1 and 2 continue to the end of the game. The beliefs that form outcome *ii*, instead, are such that it is payoff-maximizing for type 2 players to terminate at the last round and for type 1s to continue to the end of the game. Finally, the beliefs in outcomes *iii* are such that both types terminate at the first round. Note that it is possible to read the behavior of type 1 players along $\sigma_0 = 1$[14]. Figure 2 also portrays the uniform randomization hypothesis, $\lambda = \frac{1}{2}$. In this case, type 1 players play until the last round, while type 2 players may either continue to the last round or terminate at node 5 depending on their belief σ_0. The termination probabilities are reported in Appendix B.

4.1. Introducing Bayesian updating. As mentioned in the Introduction and in section 3, a natural extension of the *EF-CH* model is to allow for Bayesian updating. Here I assume that type 2 players update their prior σ_0 on the share of type 0 players in the population. It turns out that if type 2 players make use of the updated beliefs σ_0^{R1}, σ_0^{R2} during the play of the game the outcomes of the Bayesian model are behaviorally indistinguishable from the non-Bayesian. To see why, consider first the case $\lambda < \frac{1}{7}$. Regardless of the belief σ_0, it is optimal for type 2s to terminate at the first round. In this case, in fact, type 1s terminate at the first round (see expression 4.1) and type 0s are also believed to terminate with a high probability (λ is low). If $\lambda \geq \frac{1}{7}$, instead, the type 2s' payoff maximizing strategies remain the same as in expression 4.2, with the difference that the set of beliefs for which they are optimal changes slightly. In particular, $\bar{\lambda}$ is replaced by $\tilde{\lambda} = \frac{1}{7\lambda^3 - \lambda^2 + 1}$. The Bayesian outcomes *Bi*, *Bii* and *Biii*

[14] In fact, the assumptions of the model are such that it is "as if" players of type 1 had a belief $\sigma_0 = 1$ regarding the population composition.

are represented on the (λ, σ_0) plane in Figure 3,

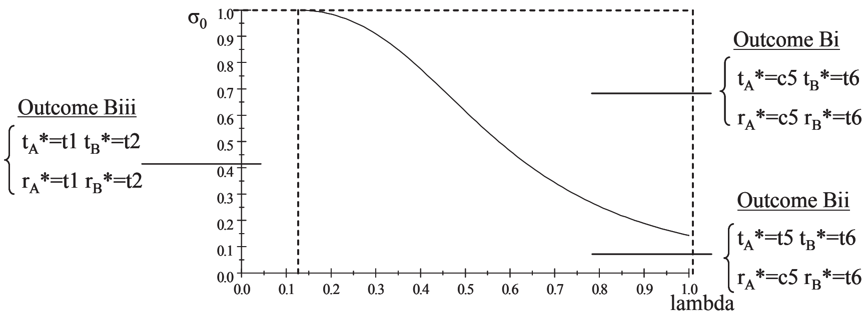


Fig. 3. Outcomes of the Bayesian EF-CH model

Figure 3 highlights two main differences with respect to the non-Bayesian model. First, for player role A, continuing at the last round is optimal for a smaller set of beliefs. This happens because when node 5 is reached a player has to believe in a high σ_0 to continue, since c_5 is profitable only with a type 0 opponent. However, during the play of the game, σ_0 has been steadily revised downwardly. Thus, it is payoff maximizing to play c_5 only for high values of the prior σ_0. It is also worth noting that when Bayesian updating is possible, playing in role B gives a strategic advantage over playing in role A. Starting to play at node 2 means, in fact, that the player can revise his prior *before* having to play the first time, by observing the opponent's move at node 1. While this does not affect the outcomes here, it indicates a strategic advantage associated with playing in position B.

4.2. The *HEF-CH* case. In the *HEF-CH* case, I relax the assumption that type 2s and type 1s hold the same belief on q. In particular, I assume that type 2 players assign a belief μ to q and correctly guess type 1s' belief λ. Thanks to this last assumption, we can consider separately the cases when $\lambda \geq \frac{1}{7}$, when type 1 players continue to the last round, and $\lambda < \frac{1}{7}$, when type 1s terminate at the first round instead (see expression 4.1). If we begin with the case $\lambda \geq \frac{1}{7}$, which I call *HEF-CH*$^+$ case, then type 2s' payoff maximizing strategies are given by,

$$t_A^* = \begin{cases} c_1c_3c_5 & \text{if } \sigma_0 > \bar{\mu} \text{ and } \mu \geq \frac{1}{7} \\ c_1c_3t_5 & \text{if } \sigma_0 < \bar{\mu} \text{ and } \mu \geq \frac{1}{7} \text{ or } \mu < \frac{1}{7} \text{ and } \sigma_0 \leq \frac{2}{3} \\ c_1t_3t_5 & \text{if } \frac{2}{3} < \sigma_0 \leq 0.93 \text{ and } \mu < \frac{1}{7} \\ t_1t_3t_5 & \text{otherwise} \end{cases} \tag{4.3}$$

$$t_B^* = \begin{cases} c_2c_4t_6 & \text{if } \mu \geq \frac{1}{7} \text{ or } \mu < \frac{1}{7} \text{ with } \sigma_0 \leq \frac{2}{3} \\ c_2t_4t_6 & \text{if } \frac{2}{3} < \sigma_0 \leq 0.93 \text{ or } \mu < \frac{1}{7} \\ t_2t_4t_6 & \text{otherwise} \end{cases}$$

where $\bar{\mu} = \frac{6.4}{22.4\mu+3.2}$. Compared to the *EF-CH* model, the *HEF-CH*$^+$ differs in that there are now beliefs such that it is optimal for type 2 players to terminate at the central round (outcome III). Outcome III arises because termination at the central round is payoff maximizing for type 2s when $\lambda > \frac{1}{7}$, $\mu < \frac{1}{7}$ and σ_0 is high, but not very high. However, these conditions cannot be met in the *EF-CH* model, where by definition $\lambda = \mu$. Graphically, the four outcomes of the *HEF-CH*$^+$ can be represented as in Figure 4,

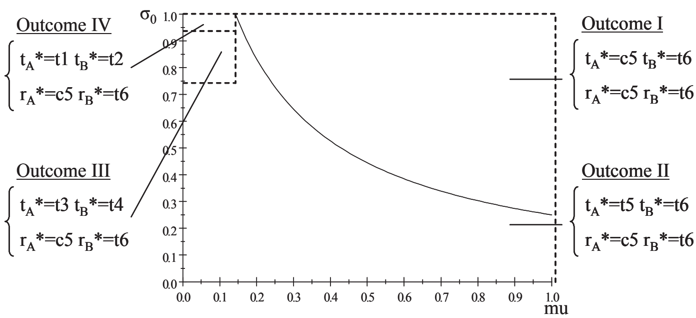


Fig. 4. Outcomes of the HEF-CH$^+$ case

Outcome I is identical to outcome i in the *EF-CH* model. Outcome II is behaviorally identical to outcome ii for type 2 players, but note that it is here optimal for a larger set of beliefs. For instance, even when type 2s have a low belief on q, $\mu < \frac{1}{7}$, they may continue to the last round provided that they believe that there are many type 1s in the population. As the share of type 1s decreases, with $\mu < \frac{1}{7}$, type 2 players terminate increasingly early, that is, either at the central round, outcome III, or at the first round, outcome IV.

The second case to consider, which I call *HEF-CH*$^-$, arises when $\lambda < \frac{1}{7}$. Here type 1 players terminate at the first round, as described in expression 4.1. Note that this implies a strategic advantage of being allocated to role B. Type 2 players in that role will get to know, before taking the first decision, the opponent's type. If the game is still ongoing at node 2, then they are for sure[15] playing with type 0. It follows that type 2s' payoff maximizing strategy for position B, t_B^* , is the same as r_B^* (see expression 4.1). However, when playing in role A, a type 2 player has no information on the type

[15] It is worth highlighting here the role which is played by the assumption that players do not make errors in implementing their preferred strategy.

of his opponent. Thus, if he believes $\mu < \frac{1}{7}$ it is optimal for him to play t_1, because even if there were only types 0s in the population, such a μ_0 would be too low to make it profitable to play c_2. If instead $\mu \geq \frac{1}{7}$, it turns out that it is payoff maximizing to play,

$$t_A^* = \begin{cases} c_1c_3c_5 & \text{if } \sigma_0 \geq \tilde{\mu} \\ t_1t_3t_5 & \text{otherwise} \end{cases} \tag{4.4}$$

$$t_B^* = \begin{cases} c_2c_4t_6 & \text{if } \mu \leq \frac{1}{7} \\ t_2t_4t_6 & \text{otherwise} \end{cases}$$

where $\tilde{\mu} = \frac{0.2}{0.6\mu + 2.4\mu^2 + 22.4\mu}$. As for the *HEF-CH*$^+$ model, we can represent graphically the possible outcomes on the (μ, σ_0) plane, holding $\lambda \geq \frac{1}{7}$, as shown in Figure 5,

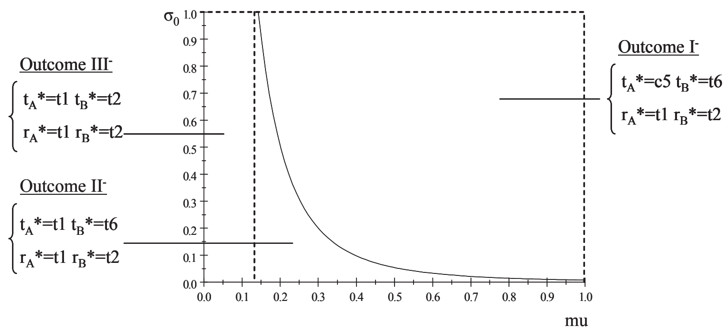


Fig. 5. Outcomes of the HEF-CH$^-$ case

Outcome I^- consists of those beliefs that make it optimal to terminate at the last round. Outcome II^- collects those beliefs such that $t_A^* = t_1t_3t_5$ and $t_B^* = c_2c_4t_6$. Finally, in outcome III^-, the payoff-maximizing strategies are $t_A^* = t_1t_3t_5$ and $t_B^* = t_2t_4t_6$. The termination probabilities can be found in Appendix B.

It is possible to allow for Bayesian updating also in the *HEF-CH* case. The computation are shown in Appendix A. This extension produces three outcomes, of which only one, $BIII$, is behaviorally different from those already studied. According to $BIII$, type 1 players continue to the last round (terminal nodes c_5 and t_6). Type 2 players, instead, terminate at the first node when in position A and at node 6 when in position B.

5. Taking the model to the data

In this section, I use the data collected by McKelvey and Palfrey (1992) in their experiment to test the predictions of the *EF-CH* model and its extensions. The 58 experimental subjects were first divided into three groups, which were later split in two halves that were randomly assigned to role A or B. Then, each player in role A (B) played the game with all the participants in his group that were allocated to role B (A), generating a total of 281 observations. To address potential concerns with learning during the ten repetitions, I also separately analyze first response data.

Before taking the theoretical results to the data, there are several considerations to be made. I estimate the model assuming that type 0 players actually exist in the population. Unreasonably high estimated shares of type 0s could suggest that the model is not well specified. The probability with which type 0s terminate, q, will be estimated from the data. This will allow me to obtain more information on the way players randomize in the game. Moreover, there are instances in which the behaviors prescribed by the different outcomes are indistinguishable. For instance, outcomes i and Bi prescribe continuation to the end of the game. Outcomes ii and Bii prescribe that type 2 players terminate at the last round, and that type 1 players continue to the last round. Outcomes iii and III^- prescribe that type 1s and 2s terminate at the first round. The impossibility to distinguish among different outcomes by looking at players' behavior implies that I cannot always infer the players' beliefs. The final consideration is that, given that I did not make any assumption on the players' beliefs, *ex ante* all the outcomes are equally likely. For this reason, here I will try to discriminate between (distinguishable) outcomes by empirically searching for the one that predicts and describes the experimental evidence best. To this end, I first employ a maximum likelihood analysis to find the parameter estimates for each outcome[16]. Then, I compare the different outcomes in two ways. First, I look at the information criteria AIC and BIC to shed some light on which specification is preferable from an econometric point of view. Secondly, I use the maximum likelihood estimates to derive each outcomes' predicted frequencies of termination, and the implied take probabilities, to compare them with the actual data.

Suppose now that type 1 and type 2 players have observed beliefs λ (as well as μ in the *HEF-CH* case) and σ_0 according to the model's hypotheses and that type 0 players follow the deterministic behavior described in section 3. I make two further assumptions that will facilitate the empirical analysis. First, I assume that type 2s' beliefs on the population composition are correct in expectation. Secondly, I assume that there is the same distribution of types for each player role. In other words, regardless of whether a player is drawn to play in role A or B, he is equally likely to be of a certain type. Under these assumptions it follows that each repetition of the game can be viewed as

[16] In the estimation I will not allow for errors in actions. This is primarily because McKelvey and Palfrey (1992) do not find any support for errors in action when analyzing the same dataset.

an experiment with seven mutually exclusive ways to terminate, $t_1, ..., t_7$, which have probabilities $\bar{q}_1, ..., \bar{q}_7 = 1 - \sum_{j=1}^{6} \bar{q}_j$. These are derived for each outcome in the previous section. Repeating the experiment n independent times, and assuming that the $\bar{q}_j$s remain the same from trial to trial, the probability of observing a random sample $\mathbf{t} = [T_1, ..., T_7 = n - \sum_{j=1}^{6} T_j]$ is distributed according to a multivariate binomial,

$$f(\mathbf{t}) = \frac{n!}{t_1!...t_6!\left(n - \sum_{j=1}^{6} t_j\right)!} \bar{q}_1^{t_1}...\bar{q}_6^{t_6} \left(1 - \sum_{j=1}^{6} \bar{q}_j\right)^{n - \sum_{j=1}^{6} t_j} \tag{5.1}$$

The density function (5.1) depends on the parameters s_0, s_1, q, which are the population shares of players of type 0 and of type 1, and the probability q with which type 0s continue at each node. For any outcome, the likelihood function for a sample of N independent observations is the product of N densities,

$$L(s_0, s_1, q) = \prod_{i=1}^{N+1} f(\mathbf{t}^i; s_0, s_1, q) \tag{5.2}$$

and the log-likelihood $\lambda = \log [L(s_0, s_1, q)]$ becomes,

$$\lambda = \sum_{i=1}^{N+1} \ln A + \sum_{i=1}^{N+1} t_1^i \ln \bar{q}_1 + ... + \sum_{i=1}^{N+1} t_6^i \ln \bar{q}_6 + \sum_{i=1}^{N+1} \left(n - \sum_{j=1}^{6} t_j\right) \ln \left(1 - \sum_{j=1}^{6} \bar{q}_j\right) \tag{5.3}$$

where $A := \frac{n!}{t_1!...t_6!\left(n - \sum_{j=1}^{6} t_j\right)!}$.

By substituting the $\bar{q}_j$s implied by each outcome into expression (5.3), and maximizing with respect to the parameters, I obtain the maximum likelihood estimates for the population composition and q. This procedure will produce consistent estimates, since the model is identified[17], the parameter space (s_0, s_1, q) is compact and λ is continuous in (s_0, s_1, q) for almost all t[18].

5.1. Estimation results for the *EF-CH* model. Table 3 presents the maximum likelihood estimates of the shares of type 0s, s_0, and type 1s, s_1, together with the estimated continuation probability q, for each possible outcome of the *EF-CH* model.

[17] There are no observationally equivalent parameters, provided that we accept the indeterminacy between behaviorally equivalent outcomes.

[18] Since the observations are not i.i.d., the convergence in probability of the average log-likelihood to the objective function is guaranteed because the sequence of the average log-likelihood is stochastically equicontinuous.

Table 3: Maximum Likelihood Estimation
EF-CH model

	i / Bi	*ii / Bii*	*iii / Biii*
-LogL	555.82	620.15	490.06
s_0	0.747**	0.361**	0.852**
	(0.004)	(0.015)	(0.014)
s_1	–	0.360**	–
		(0.015)	
q	0.936**	0.597**	0.283**
	(0.029)	(0.022)	(0.004)
Obs.	281	281	281

Std. errors in par. ;* sign. 5%; ** sign. at 1%

The estimated fraction of type 0s in the population is very high in outcomes *i / Bi* and *iii / Biii*, while it is around thirty per cent in outcome *ii / Bii*. Note also that the estimates of q vary substantially across outcomes, but in outcome *ii / Bii* they are close to the uniform randomization value. This last result suggest the presence of the a truly randomizing type 0 in the population.

The results in Table 3 suggest that all the outcomes can be estimated using the data. However, these estimates are not informative as regards which of the outcomes is preferable in the sense of being the correct model of behavior and in terms of parsimony of parameters. Regarding the last point, note that outcome *ii / Bii* needs two parameters to describe the population, s_0 and s_1, while the other two outcomes use only s_0. This follows from the fact that in outcomes *i / Bi* and *iii / Biii*, type 1s and type 2s behave in the same way, and hence they are estimated to be the remaining $1 - s_0$ share of the population. To address the issue of which model is preferable, the simplest way is to look at the informational criteria, AIC and BIC, computed for each outcome as shown in Table 4. The information criteria help identifying the correct, most parsimonious model specification. The two criteria are very similar, with the difference that the BIC penalizes additional parameters more severely. In general, for both criteria, the lower the value, the more preferable the model, in term of parsimony of parameters and explanatory power.

Table 4: Information Criteria for Outcomes' Comparisons

Outcome	Obs.	LogL(null)	LogL(model)	dg. freedom	AIC	BIC
i / Bi	281	.	−555.82	2	1115.64	1122.91
ii / Bii	281	.	−620.15	3	1246.30	1257.22
iii / Biii	281	.	−490.06	2	984.13	991.41

For both criteria outcome *iii / Biii* have the lowest value. However, this result may be driven by the fact that the log-likelihood function of outcome *iii / Biii* is much simpler than those for the other outcomes (see Appendix B). Yet, the information criteria allow for model comparison only and do not provide any diagnostic on the fit of the model. To get a rough measure of how well the outcomes fit the data, I use the maximum likelihood estimates to predict the frequency of termination f_j and the implied take probability p_j for each outcome. First, however, I re-run the analysis considering only first-response data to gather some more information on the reliability of the estimates.

5.1.1. *First-response data.* The estimates in Table 3 are obtained by using the full sample of observations. To address possible concerns regarding learning during the game, I here restrict the analysis to first response data, in order to verify if this leads to different results. Table 5 shows the maximum likelihood estimates for the *EF-CH* model,

Table 5: Maximum Likelihood Estimation
EF-CH model, first response data

	i / Bi	*ii / Bii*	*iii / Biii*
-LogL	50.82	58.51	56.06
s_0	0.890**	0.317**	0.756**
	(0.128)	(0.052)	(0.014)
s_1	–	0.291**	–
		(0.049)	
q	0.759**	0.586**	0.313**
	(0.032)	(0.099)	(0.005)
Obs.	29	29	29

Std. errors in par. * sign. 5%; ** sign. 1%

The estimates do not change substantially with respect to those in Table 3, suggesting that the data from the full sample can be treated as independent. For the outcomes that appear in Table 5, it is possible to compute the AIC and the BIC, as shown in Table 6

Table 6: Information Criteria for Outcomes' Comparisons

Outcome	Obs.	LogL(null)	LogL(model)	dg. freedom	AIC	BIC
i / Bi	29	.	−50.82	2	105.65	108.39
ii / Bii	29	.	−58.51	3	123.03	127.13
iii / Biii	29	.	−56.06	2	116.13	118.87

Now outcome *i / Bi* scores the lowest on both criteria. Such a change in the ranking of the three outcomes seems to confirm the suspicion that the low values scored by outcome *iii / Biii* were an artifact of its extremely simple log-likelihood function.

5.1.2. *Fit.* I can obtain the distribution of terminations as predicted by the outcomes of the *EF-CH* model by substituting the parameters s_0, s_1 and q into each outcome's $\bar{q}_n$ using the estimates in Table 3. The result of this exercise can be found in Table 7, wherein the actual f_js are reported as well.

Table 7: Proportion of Observations at each Terminal Node
Empirical and Predicted Distributions

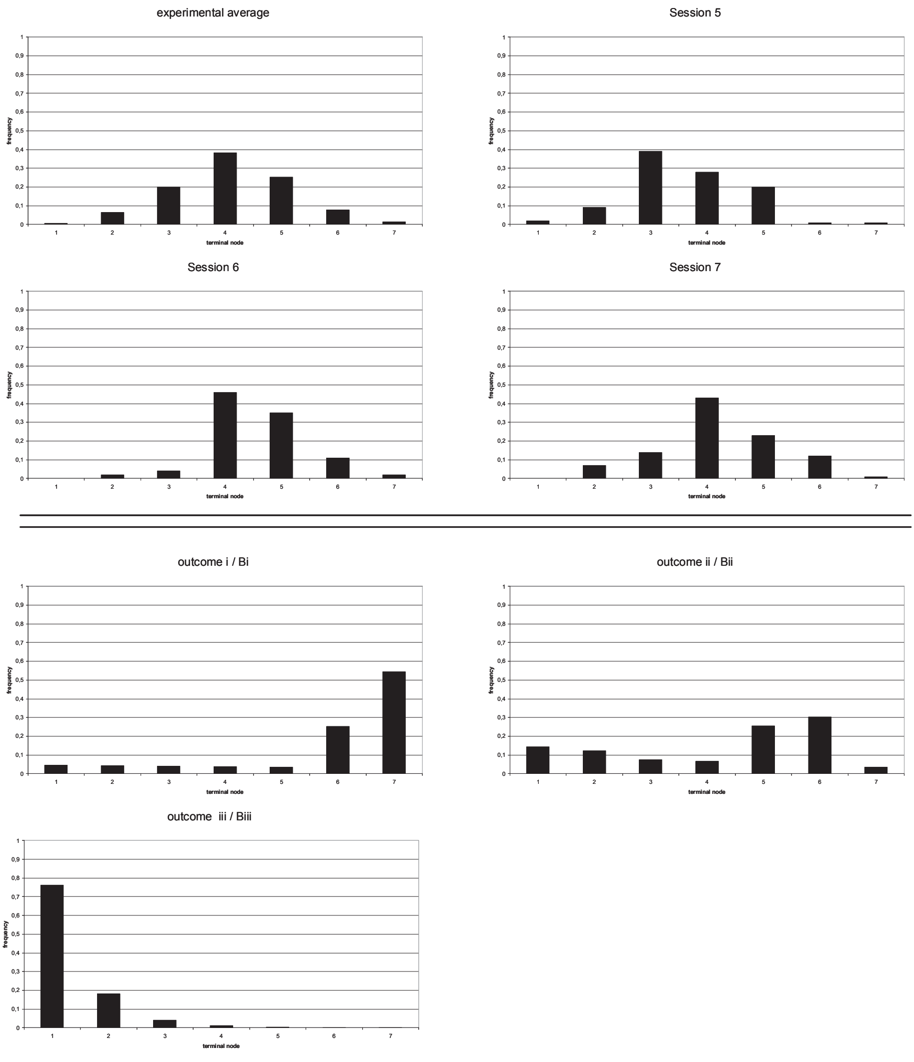
Session	N	f_1	f_2	f_3	f_4	f_5	f_6	f_7
5	100	.02	.09	.39	.28	.20	.01	.01
6	81	.00	.02	.04	.46	.35	.11	.02
7	100	.00	.07	.14	.43	.23	.12	.01
Total 5-7	281	.007	.064	.199	.384	.253	.078	.014
Outcome		$\hat{f}_1$	$\hat{f}_2$	$\hat{f}_3$	$\hat{f}_4$	$\hat{f}_5$	$\hat{f}_6$	$\hat{f}_7$
i / Bi	–	.045	.043	.040	.039	.036	.253	.544
ii / Bi	–	.144	.123	.074	.066	.255	.303	.035
iii / Biii	–	.762	.181	.04	.011	.003	.002	.001

It is easy to see that the predicted terminations occur either very early or very late in the game. Outcome *iii* / *Biii*, which is the one with the lowest informational criteria, predicts an unreasonably high share of terminations at the first node. In contrast, outcome *i* / *Bi* predicts too many terminations at the last two nodes. On the other hand, outcome *ii* / *Bii* seems to fit the empirical termination frequencies comparatively well.

The results in Table 7 can be plotted using an histogram as shown in Figure 6. The first four panels present the aggregate empirical distribution of terminations, together with those for the individual sessions 5, 6 and 7. As can be seen, the laboratory data show a marked tendency for plays to terminate at nodes 3, 4 and 5. The average distribution is almost normal around the central round, while the individual sessions show slightly different patterns. Terminations do, however, very rarely occur before node 2. The remaining three panels of Figure 6 show the predicted distribution of terminations according to the outcomes of the *EF-CH* model. Clearly, the outcomes differ substantially in their predictions, but none really fits the data. Outcome *i* / *Bi* puts too much weight on late terminations with respect to the sample data. Equally far from replicating the empirical pattern, outcome *iii* / *Biii* predicts that the largest share of terminations takes place at node 1. This leaves outcome *ii* / *Bii*, which, despite having the highest information criteria among the three, predicts that terminations may take place at all nodes, although the largest share of terminations is predicted at nodes 5 and 6.

To further investigate the predictive power of the *EF-CH* outcomes, we can use the data in Table 7 to derive the predicted implied take probabilities. Table 8 presents these results. As discussed in section 2, this probability, p_j, is higher in the later nodes of the game. Again, outcome *ii* / *Bii* is the one that follows the empirical evidence the closest, which is hardly surprising given that this outcome is the one that better fits the data in terms of predicted distribution of terminations, f_js.

Figure 6. Empirical and predicted distributions of terminations[19]



[19] Note: On the horizontal axis are the seven terminal nodes at which the game can end, where the seventh node represents the case in which the player in position B passes at the last node. On the vertical axis is the frequency of terminations at each node.

In general, the empirical analysis demonstrated that the *EF-CH* model outperforms the backward induction prediction, but falls short of replicating the high number of terminations at the central nodes of the game. Among the possible outcomes, outcome *ii* / *Bii* is the one that better fits the data. In other words, we can represent the experimental evidence fairly well if we assume a population of players composed by the three types in equal shares, and if we let the type 0 players randomize with a probability of sixty percent on the choice of continuation, while type 1s and 2s continue to the last round (with type 2s terminating there).

Table 8: Implied take probabilities
Empirical and Predicted Distributions

Session	p_1	p_2	p_3	p_4	p_5	p_6
5	0.02	0.09	0.44	0.56	0.91	0.50
6	0.00	0.02	0.04	0.49	0.72	0.82
7	0.00	0.07	0.15	0.54	0.64	0.92
Total 5-7	0.01	0.06	0.21	0.53	0.73	0.85
	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$	$\hat{p}_5$	$\hat{p}_6$
outcome *i* / *Bi*	0.05	0.32	0.04	0.05	0.04	0.31
outcome *ii* / *Bii*	0.14	0.14	0.10	0.10	0.43	0.89
outcome *iii* / *Biii*	0.76	0.76	0.70	0.65	0.50	0.66

If anything, the cognitive hierarchy approach misses a mechanism to generate terminations at the central round.

5.2. Estimation results for the *HEF-CH* case. In what follows, I present the estimation results for the *HEF-CH* outcomes. Note that outcome I and II are behaviorally indistinguishable from outcome i and outcome ii respectively, and that the same holds true for outcome III^- and outcome iii. With regard to the Bayesian extension of the *HEF-CH* case (see Appendix A), only outcome $BIII$ prescribes a new behavior. To avoid repetitions, I do not present the estimates for the outcomes that are behaviorally indistinguishable from those already discussed. The maximum likelihood estimates are shown in Table 9.

Table 9 : Maximum Likelihood Estimation
HEF-CH case

	III	IV	$BIII$	I^-	II^-
-LogL	576.65	650.29	589.81	635.61	654.22
s_0	0.181**	0.077**	0.128**	0.447**	0.479**
	(0.036)	(0.012)	(0.007)	(0.000)	(0.000)
s_1	0.569**	0.975**	0.516**	0.004**	0.000
	(0.053)	(0.000)	(0.003)	(0.000)	(0.000)
q	0.792**	0.742**	0.354**	0.370**	0.343**
	(0.016)	(0.028)	(0.003)	(0.001)	(0.000)
Obs.	281	281	281	281	281

Std. errors in parentheses;* sign. 5%; ** sign. at 1%

Recall that in outcomes III and IV type 1 players continue to the last round, while type 2 players terminate at the central and the first nodes, respectively. In outcomes I^- and II^- type 1 players terminate at the first round, while type 2 players continue to the last round in outcome I^- while in outcome II^- they terminate at the first node if playing in position A and at the last one if in position B. This stark contrast in the prescribed behavior between the two player-roles for type 2 persists in outcome $BIII$, with the difference that type 1 players continue to the last round. With respect to the *EF-CH* case, it is worth noting that the estimates point towards a more realistic population composition. On average type 0 players constitute one third of the population. As for q, this varies between thirty and eighty per cent. All the estimates, besides those for outcome II^-, are significant. Table 10 presents the information criteria for these outcomes.

Table 10: Information Criteria for Outcomes' Comparisons

Outcome	Obs.	LogL(null)	LogL(model)	dg. freedom	AIC	BIC
III	281	.	−563.21	3	1159.30	1170.21
IV	281	.	−650.29	3	1304.59	1311.87
$BIII$	281	.	−589.81	3	1185.63	1196.54
I^-	281	.	−635.61	2	1275.22	1282.50
II^-	281	.	−638.09	2	1314.45	1325.37

Outcomes III and IV score the lowest information criteria, noticeably lower than those computed for outcomes i and ii. Before looking at how well these outcomes fit the experimental evidence, I run the same analysis on first-response data only, in order to check if the results are consistent with the ones just discussed.

5.2.1. *First response data.* As for the *EF-CH* model, it is possible to use first response data to carry out the maximum likelihood estimation. The results appear in Table 11.

Table 11: Maximum Likelihood Estimation
HEF-CH case, first response data

	III	IV	$BIII$
-LogL	64.58	69.60	63.51
s_0	0.166**	0.935**	0.447
	(0.030)	(0.028)	(0.000)
s_1	0.723**	0.098**	0.370**
	(0.071)	(0.024)	(0.000)
q	0.776**	0.769**	0.169**
	(0.043)	(0.040)	(0.000)
Obs.	29	29	29

Std. errors in par. ;* sign. 5%; ** sign. at 1%

The estimates for outcomes III, IV and $BIII$ do not differ substantially from those in Table 9 that were obtained by using the full sample. However, the estimation was impossible to carry out for outcomes I^- and II^-, which may be suggestive of a problem either with the outcomes or with the data. Neither of these explanations is problematic. First, the first response sample consists of 29 observations, which is a small sample to perform maximum likelihood analysis on. Secondly, outcome I^- and II^- are those that in the full sample scored the lowest informational criteria, which already then hinted that they were not performing particularly well in explaining the data. For the outcomes in Table 11 we can also look at the information criteria.

Table 12: Information Criteria for Outcomes' Comparisons

Outcome	Obs.	LogL(null)	LogL(model)	dg. freedom	AIC	BIC
III	29	.	−64.58	3	135.16	139.26
IV	29	.	−69.60	3	145.21	149.31
$BIII$	29	.	−63.51	2	131.03	133.77

The AIC and the BIC for outcomes III and $BIII$ maintain the same ranking as in the full sample, and outcome IV continues to be the least preferable among the alternatives. This confirms the results obtained above that point at outcome III as the preferable among the alternative ones.

5.2.2. *Fit.* Using the estimates in Table 9, I can obtain the predicted distribution of terminations for each outcome. The results of this exercise are presented in Table 13.

Table 13: Proportion of Observations at each Terminal Node
Predicted Distributions

Outcome	$\hat{f}_1$	$\hat{f}_2$	$\hat{f}_3$	$\hat{f}_4$	$\hat{f}_5$	$\hat{f}_6$	$\hat{f}_7$
III	.120	.105	.303	.185	.041	.117	.129
$BIII$	.760	.150	.010	.070	.000	.010	.000
IV	.234	.180	.132	.102	.077	.106	.169
I^-	.280	.200	.080	.060	.020	.330	.030
II^-	.830	.050	.070	.000	.020	.010	.020

Looking at Table 13 it is easy to see that the problem of the *EF-CH* model largely persists. In fact, the outcomes fail to predict the large number of terminations at the central round. However, the *HEF-CH* outcomes fare somewhat better that than those of the *EF-CH* model. To better understand why, I plot the predicted frequency of termination in a histogram, as can be seen in Figure 7.

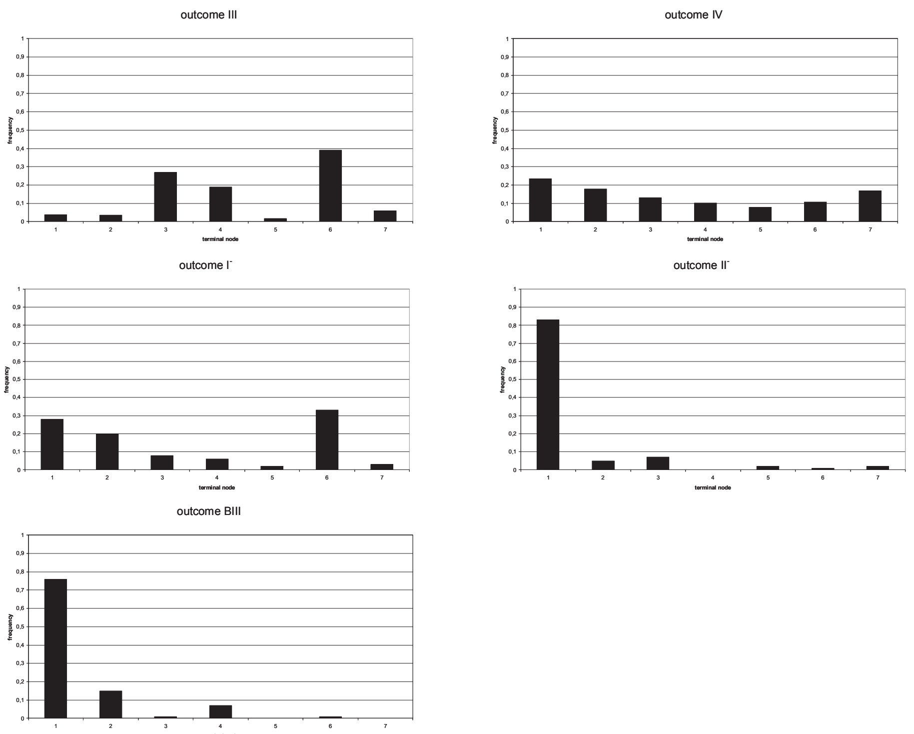
The patter of the distributions in Figure 7 are more diverse than those in Figure 6. In particular, outcome III, outcome IV and outcome I^- are doing comparatively well. As before, we can use the predicted frequency of terminations to back out the predicted implied take probabilities, which are shown in Table 14.

Table 14: Implied take probabilities
Predicted Distributions

	$\hat{p}_1$	$\hat{p}_2$	$\hat{p}_3$	$\hat{p}_4$	$\hat{p}_5$	$\hat{p}_6$
outcome III	0.12	0.12	0.39	0.39	0.14	0.48
outcome $BIII$	0.76	0.63	0.11	0.88	0.00	1.00
outcome IV	0.83	0.33	0.58	0.00	0.40	0.33
outcome I^-	0.28	0.28	0.15	0.14	0.05	0.92
outcome II^-	0.83	0.29	0.58	0.00	0.40	0.33

The figures in Table 14 are closer to the actual values than those in Table 6, for the *EF-CH* model. In particular, outcome III is the closest to the average empirical data, besides a noticeable discrepancy at node 5.

Figure 7. Predicted distributions of terminations[20]



[20] Note: On the horizontal axis are the seven terminal nodes at which the game can end, where the seventh node represents the case in which the player in position B passes at the last node. On the vertical axis is the frequency of terminations at each node.

Summing up, the empirical analysis of the *HEF-CH* models suggest that, among the alternatives, outcome III is the outcome that best explains the subjects' behavior when playing the game. Outcome III describes a situation in which type 1 players continue to the last round, assigning a high probability to the event that the opponent passes the pot back ($\lambda \geq \frac{1}{7}$). Type 2 players, instead, assign to such an event a probability μ strictly lower than that of type 1, and believe that the population consists mostly of type 0 players. Note that, according to the estimates in Table 9, type 2 players have a correct belief on the population composition, but they are mistaken in assigning a low continuation probability. Since the actual q is estimated at 0.79, instead, it is type 1 players that have the correct belief on q.

6. Discussion

The *EF-CH* model[21] described in section 3 has brought about a variety of findings that deserve discussion. Most importantly, the extensive-form cognitive hierarchy assumptions prescribe optimal behaviors that differ from the backward induction equilibrium. In particular, the *HEF-CH* model is such that it is possible for type 2 players to terminate at any node of the game. This result can be interpreted as a simple way out of the backward induction paradox, as defined by the discrepancy between equilibrium predictions and laboratory behavior. By relaxing the equilibrium requirements on the players' beliefs it is possible to generate payoff maximizing behaviors that are, in theory, consistent with the experimental data.

However, when brought to the data, the model partially fails the empirical test. While it succeeds in consistently replicating the pattern of implied take probabilities, it falls short of predicting the almost normal distribution of the frequency of terminations[22]. The experimental evidence seems to suggest that the central round features some aspects that make it different from the initial and the last one. At least in part, this is likely to be driven by behavioral factors, such as regret, risk or loss aversions.

The present model can be improved along a number of dimensions. The most straightforward possibility is to allow players to be noisy best responders to their beliefs, and blend into the current model some form of quantal response. McKelvey and Palfrey (1992, 1998) have applied a static agent quantal response equilibrium model to precisely the centipede game, with satisfactory results. The explanatory power of such a model, however, has been demonstrated by the same authors to significantly

[21] It is worth stressing that a similar model of players' behavior has been put forward and later empirically tested by Jehiel (Jehiel, 2005; Huck et al. 2009; Huck and Jehiel; 2004) in his analogy-based expectation equilibrium concept. The *EF-CH* model however, differs in that it is inspired by the cognitive hierarchy model and is only concerned with the centipede game.

[22] By assigning a too high probability to the event that the play reaches the final round.

improve by allowing for altruism. More precisely, the authors put forward a model of incomplete information in which there is the possibility to be matched with an altruistic opponent that always chooses to pass. Allowing for both errors in actions and in beliefs, it can be shown that in equilibrium players adopt mixed strategies that prescribe to terminate with a probability that increases in the length of the play. Interestingly, it turns out that, according to the estimates, about five per cent of the experimental subjects behaved as altruists. Compared to McKelvey and Palfrey's (1992) model, the *EF-CH* has the advantage of not requiring players to be able to derive an equilibrium in mixed strategies. However, it would certainly be possible to enrich the current model with both the error structure and the altruism assumption studied therein.

Besides quantal response and behavioral concerns, my work could be improved by modifying the non-equilibrium assumptions described in section 3. For instance, it would be possible to increase the number of types in the population. Since higher types tend typically[23] to terminate earlier, such an assumption could in part improve the currently poor empirical fit of the *EF-CH* and *HEF-CH* models. Alternatively, the mismatch between the predicted and empirical distributions of terminations could be addressed by relaxing the assumption of constant randomization from part of type 0 players.

In part, these possibilities have been indirectly investigated by Kawagoe and Takizawa (2010) who compare the explanatory power of the Level k model with that of quantal response models when explaining a number of centipede protocols. Besides assuming a Level k belief structure, while I assume that beliefs follow the cognitive hierarchy assumptions, the authors allow for four instead of three types and assume that the type 0 players only exist in the mind of the type 1s. The authors find that in most setups, including the six-step centipede protocol studied here, the Level k model[24] outperforms both the agent quantal response model and the agent quantal response model with altruistic types. In general, their estimates show a population composed predominantly of type 1s, roughly seventy per cent. This is fairly close to the estimates obtained here when looking at outcome III, where the type 1 players are 57 per cent of the population, type 0s 18 per cent and type 2s are 25 per cent. Such a similarity is encouraging, suggesting that the assumption of a hierarchy composed by three types only is not off the mark, at least in the case of a six-step centipede game. However,

[23] Higher types, in fact, have by definition increasingly sophisticated beliefs about the opponent's behavior. As these beliefs approach the common knowledge assumption, the resulting payoff-maximizing behavior tends to the Nash equilibrium, which, in the context of the Centipede game, is to terminate at the first round.

[24] More precisely, they assumed that players can be of four types, $1 \leq k \leq 4$, and they assume that type 0 players do not exist in the population, but only in the mind of type 1 players.

the authors show that for other centipede protocols the fraction of type 2s and 3s may be sizeable.

Kawagoe and Takizawa's results suggest that allowing for more types is an important extension of this model, which is likely to improve on the empirical fit. In doing so, however, it is important not to fail the spirit of the original research question, namely how far would a realistic model of players' behavior take us into explaining the experimental data on the centipede game. Admitting type 3 players, in fact, means to assume that in the experimental population some individuals are prepared to undertake high cognitive costs to decide how to play. These costs, together with the empirical finding that players' behavior in most games can be described assuming types 0, 1 and 2, speak somewhat against allowing for higher types of players.

When compared with other models that address the backward induction paradox, my approach appears substantially different from those put forward in the behavioral literature, which typically invokes factors like altruism, fairness or similar concepts. While the behavioral approach most certainly has some explanatory power for the experimental results that document late termination in the game, it fails to explain the mechanism driving the paradox. This is embedded in the payoff structure, that simultaneously makes it attractive to be at the later nodes in the game, where payoffs are higher, and unattractive to reach them, since for every two subsequent nodes the first mover prefers to terminate rather than continue. Altruism, fairness or reciprocity are channels that may explain the preferences to terminate at later stages of the game, where both players are better off, rather than terminate at the beginning of it. Yet, at the last node, reciprocity cannot be enforced, since there are no more moves, and fairness is impossible, since the division of the pot is bound to be uneven by the payoff structure. Altruism is also problematic, since a player, to be altruistic, would have to forego at least half of his payoff at the last node. The behavioral solutions appear to be corollary to the explanation of the paradox, but fail to directly address it. On the contrary, the cognitive hierarchy assumptions offer a mechanism that may reconcile rational play with the continuation patterns found in the data, thus avoiding the paradox and allowing for behavioral refinements.

Interestingly, the ancillary finding that in this model Bayesian updating is ineffective may be relevant to the discussion regarding the extent to which Bayesian reasoning is an appropriate description of human behavior (see, for instance, Cosmides and Tooby, 1996). However, the most likely explanation of such a result is the very simplified model of type 1s' beliefs about types 0. As it stands now, the constant randomization assumption clashes with the increase in payoffs across nodes. Moreover, it is arguably restrictive, even granted that keeping simple the beliefs of type 1s' is in the spirit of the model . A model that allows type 1s to undertake some Bayesian updating might very well fare better than the *EF-CH* specification: to improve the model in this respect is surely another important extension.

Finally, I would like to comment on whether the model put forward in this paper could be generalized to a model applicable to any game in the class of two-player, extensive-form, finite game of perfect information. To this end, two major challenges arise. First, in arbitrary extensive-form games it may be difficult to the depart from the assumption of uniform randomization regarding the behavior of type 0 players. In the centipede game, the players at each node face only two decisions, and behavioral considerations provide a fairly solid background to justify randomization different from the uniform case. However, were the player to face three or more decision nodes, it is not a priori clear how to depart from the uniform randomization assumption. The second challenge that a general *EF-CH* model would have to face is how to address the possibility of signalling. It is well-known, in fact, that in multistage games players may adopt certain strategies in order to influence their opponent's beliefs in the earlier stage of the game, to build a reputation which can be exploited at later stages (Kreps et al., 1982). A generalized *EF-CH* model may find it difficult to account for this possibility, especially because it could make it difficult to solve recursively for the players' strategies. Although these issues are not straightforward to address, an extensive-form cognitive hierarchy model appears highly desirable to improve our understanding of players' behavior in this important class of games.

7. Conclusions

In this paper, I have applied the cognitive hierarchy model (Camerer et al., 2004) to the extensive-form centipede game. I assumed three types of players: a non-strategic randomizer, type 0, and types 1 and 2, that reason by backward induction according to their beliefs. These were structured following the conventional cognitive hierarchy hypotheses according to which a type k player assumes that the opponent belongs to a mixture of lower-level types, and that type k players, $k \neq 0$, assign the same belief on the behavior of type 0s. Besides this model, which I called *EF-CH*, I also studied two extensions. The first allowed type 2s to undertake Bayesian updating on their prior about the opponent's type, and the second, which I called *HEF-CH*, relaxed the assumption that type 1s and 2s have the same belief on the behavior of type 0s.

In both cases, the assumptions resulted in outcomes that improve upon the backward induction result, prescribing, to a different extent, terminations at nodes later than the first. To discriminate which of the possible specifications explains the data best, I estimated and compared them using standard maximum likelihood tools. The empirical analysis produced several results. First, the outcomes of the *EF-CH* model predict terminations either too early or too late in the game to match the data. The best fit is obtained assuming a population composition where the three types are present in equal shares, and such that type 0s are more likely to continue than to terminate, while type 1s and 2s continue to the last round. This result can be improved by one of the *HEF-CH* outcomes, which assumes a population composed mostly by type 1 players and to a lesser extent by type 0 players that continue with a high probability. Despite that the fit of this outcome is closer to the data, it still falls short of predicting the large share of terminations that occur during the experiment at nodes 3 and 4. Interestingly, allowing for Bayesian updating does not seem to have any noticeable effect on the model's predictions or empirical performance.

To conclude, the extensive-form cognitive hierarchy model has proven to be reasonably effective and parsimonious in explaining the backward induction paradox. However, the empirical analysis suggests that there is more to the plays of the centipede game than different cognitive levels among the players. Most likely, the experimental behavior is the result of the combined actions of players of different cognitive types and the effect of relevant behavioral factors such as altruism and fairness.

References

Aumann, R. , 1995. Backward induction and common knowledge of rationality. Games Econ. Behav. 8, 6-19.

Battigalli, P. , 1997. On rationalizability in extensive form games. J. Econ. Theory.74, 40-61.

Ben-Porath, E. , 1997. Rationality, Nash Equilibrium and Backwards Induction in Perfect- Information Games. Rev. Econ. Stud. 64, 23-46.

Brandenburger, A. , Friedenberg, A. , 2007. Axioms for backward induction. unpublished manuscript.

Camerer, C. , 2003. Behavioral Game Theory: Experiments in strategic interaction. Princeton: Princeton University Press.

Camerer, C. , Ho, T. H. , Chong, J. K. , 2004. A cognitive hierarchy theory of one-shot games. Quart. J. Econ. 119, 861-898.

Cosmides, L. , Tooby, J. , 1996. Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. Cognition. 58: 1-73.

Costa-Gomes, M. , Crawford, V. , Broseta, B. , 2001. Cognition and behavior in normal form games: An experimental study. Econometrica. 69, 1193-1235.

Costa-Gomes, M. , Crawford, V. , 2006, Cognition and behavior in two-person guessing games: An experimental study. Amer. Econ. Rev. 96, 1737-1768.

Crawford, V. , Iriberri, N. , 2007. Fatal attraction: Salience, naivete, and sophistication in experimental hide-and-seek games. Amer. Econ. Rev. 97, 1731-1750.

Ellingsen, T. , Östling, R. , 2006. Organizational Structure as the Channeling of Boundedly Rational Pre-play Communication. SSE/EFI Working Paper Series 634.

Fey, M. , McKelvey, R. , Palfrey, T. , 1996. An experimental study of constant sum Centipede games. Int. J. Game Theory. 25: 269-287.

Gerber, A. , Wichardt, P. C. , 2007. Finite-order beliefs and welfare-enhancing instruments in the centipede game. Working Paper Series, ISSN 1424-0459, University of Zurich.

Hart, S. , 2002. Evolutionary dynamics and backward induction. Games Econ. Behav. 41, 227–264.

Ho, T. H. , Camerer, C. , Weigelt, K. , 1998. Iterated dominance and iterated best response in experimental 'p-beauty contests'. Amer. Econ. Rev. 39, 649-660.

Ho, T. H. , Wang, X. , Camerer, C. , 2008. Individual differences in EWA learning with partial payoff information. Econ. J. 118, 37-59.

Kawagoe, T. , Takizawa, H. , 2008, Level-k Analysis of Experimental Centipede Games. Available at SSRN: http://ssrn.com/abstract=1289514.

Kreps D. M. , Wilson, R. , 1982, Sequential Equilibria. Econometrica. 50, 863-894.

Kreps D. M. , Milgrom, P. , Roberts, J. , Wilson, R. , 1982. Rational Cooperation in the finitely repeated Prisoner's Dilemma. J. Econ. Theory. 27, 245-52.

Kuhn, H. W. , 1950. Extensive games. Proc. Natl. Acad. Sci. U.S.A. 36, 570–576.

Kuhn, H. W. , 1953. Classics in Game Theory. Princeton: Princeton University Press.

McKelvey, R. , Palfrey, T. , 1992. An experimental study of the centipede game. Econometrica: 60, 803-836.

McKelvey, R. , Palfrey, T. , 1995. Quantal response equilibria for normal form games. Games Econ. Behav. 10, 6-38.

McKelvey, R. , Palfrey, T. , 1996. A Statistical Theory of Equilibrium in Games. Japanese Econ. Rev. 47: 2.

McKelvey, R. , Palfrey, T. , 1998. Quantal response equilibria for extensive form games. Exper. Econ. 1, 9-41.

Nagel, R. , 1995, Unraveling in guessing games: An experimental study. Amer. Econ. Rev. 85, 1313-1326.

Nagel, R. , Tang, F. ,1998. Experimental Results on the Centipede Game in Normal Form: An Investigation on Learning. J. Math. Psychol. 42, 356-384.

Nöldeke G. , Samulesson, L. , 1997. A Dynamic Model of Equilibrium Selection in Signaling Markets. J. Econ. Theory. 73: 118-156.

Rapoport, A. , Stein, W. E. , Parco, J. E. , Nicholas, T. E. , 2003. Equilibrium Play and Adaptive Learning in a Three-Person Centipede Game. Games Econ. Behav. 43, 239-265.

Reny, P. , 1998. Rationality, Common Knowledge, and the Theory of Games. Technical Report, University of Western Ontario.

Rosenthal, R. , 1982. Games of perfect information, predatory pricing, and chainstore paradox. J. Econ. Theory. 25: 92-100.

Selten, R. , 1965. Spieltheoretische Behandlung eines Oligopmodells mit nachfrage-tragheit. Z. Staatswissenschaft 121: 301- 324.

Selten, R. , 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. Internat. J. Game Theory. 4, 25-55.

Stahl, D. , Wilson, P.W. , 1994. Experimental evidence on players' models of other players. J. Econ. Behav. Organ. 25, 309-327.

Stahl, D. , Wilson, P.W. , 1995. On Players' Models of Other Players: Theory and Experimental Evidence. Games Econ. Behav. 10, 218-254.

Appendix A: Bayesian updating

Expression A.1 shows the payoff-maximizing strategies for type 2 players in the Bayesian *EF-CH* model

$$t_A^* = \begin{cases} c_1c_3c_5 & \text{if } \sigma_0 \geq \tilde{\lambda} \text{ and } \lambda \geq \frac{1}{7} \\ c_1c_3t_5 & \text{if } \sigma_0 < \tilde{\lambda} \text{ and } \lambda \geq \frac{1}{7} \\ t_1t_3t_5 & \text{otherwise} \end{cases} \tag{A.1}$$

$$t_B^* = \begin{cases} c_2c_4t_6 & \text{if } \lambda \geq \frac{1}{7} \\ t_2t_4t_6 & \text{otherwise} \end{cases}$$

where $\tilde{\lambda} = \frac{1}{7\lambda^3 - \lambda^2 + 1}$.

We can now look at the effect of allowing for Bayesian updating in the *HEF-CH* model. An informal argument may illustrate how Bayesian updating would affect the outcomes in the game. Consider a player i of type 2 at the beginning of the game. Suppose his beliefs fall in outcome III, meaning that at the beginning of the game it is optimal for i to terminate at node 3 (or 4 if he plays in role B). With Bayesian updating, however, once the play reached the second round, i may decide to continue further. This happens because i knows that only type 0s terminate before the last round, since by assumption he knows λ. Thus, upon receiving the pot back, i revises downward the probability of being playing with a type 0. In particular, player i may revise σ_0 downward enough to decide to continue rather than stop at node 3 (or 4). Looking at Figure 2, this would correspond to a drop in σ_0 from outcome III to outcome II.

As before, I will consider separately the two cases $\lambda \geq \frac{1}{7}$ and $\lambda < \frac{1}{7}$, and start with the former. The behavior of type 1 players' remains the same as that described in expression 4.1, while the payoff maximizing strategies for type 2s do not. The strategy t^* can still be derived by backward induction, now making use at each node of the updated beliefs

$$t_A^* = \begin{cases} c_1c_3c_5 & \text{if } \sigma_0 > \frac{1}{7\mu_0^3 - \mu_0^2 + 1} \\ c_1c_3t_5 & \text{if } \sigma_0 < \frac{1}{7\mu_0^3 - \mu_0^2 + 1} \\ t_1t_3t_5 & \text{if } \sigma_0 > 0.93 \text{ and } \mu_0 < \frac{1}{7} \end{cases} \tag{A.2}$$

$$t_B^* = c_2c_4t_6$$

A type 2 player in role B has only one payoff maximizing strategy, continuing to the last node: the fact that the opponent passes at node one conveys enough information to make it optimal for the type 2 players to continue the game to the last node. Figure A1 illustrates the (t_A^*, t_B^*) strategies in the $(\mu_0; \sigma_0)$ plane.

Outcome BI is defined by the beliefs that make it payoff maximizing to play $t_A^* = c_1c_3c_5$ and $t_B^* = c_2c_4t_6$. In outcome BII it is $t_A^* = c_1c_3t_5$ $t_B^* = c_2c_4t_6$ and in outcome BIII $t_A^* = t_1t_3t_5$ and $t_B^* = c_2c_4t_6$. The termination probabilities are given in Appendix B.

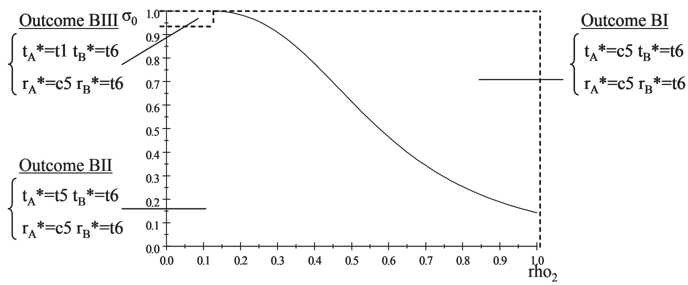


Fig. A1. Outcomes of the HEF-CH Bayesian case

This case presents two main differences with respect to the non-Bayesian *EF-CH*. First, the behavior $t_A^* = c_1t_3t_5$, $t_B^* = c_2t_4t_6$ is no longer payoff maximizing (in expectation) for any beliefs. In fact, a player starting off with the beliefs that made it payoff maximizing to stop at node 3 (outcome III), once at node 3 revises σ_0 downward enough to keep playing at least one more node. Secondly, continuation at node 5 is now payoff maximizing for a smaller set of beliefs than for the *EF-CH* case. When node 5 is reached, two opposing considerations come into play. On the one hand, a player at node 5 has to believe in a high σ_0 to choose to continue, because c_5 is profitable only with a type 0 opponent. On the other hand, the likelihood of being playing with a type 0 has been revised downwardly, given that the game has not been interrupted. The beliefs for which it is optimal to play c_5, then, shrink with respect to the *EF-CH* model.

In case $\lambda < \frac{1}{7}$, the Bayesian updating takes place only in the first round and only for those type 2s that are playing in position B, since those in the other role terminate at node 1, as shown in expression 4.4. Knowing that all type 1s go out after the first round, those type 2s that receive the pot at node 2 revise σ_0 to $\sigma_0^{R1} = 1$. This revision affects type 2s' behavior differently for different μ_0. If $\mu_0 \geq \frac{1}{7}$, then type 2s terminate at the last round, otherwise they take the pot immediately.

Appendix B: termination probabilities $\bar{q}_n$

Remark 1. *Procedure to find the payoff maximizing strategies in the case* $\rho_1 = \rho_2 = \frac{1}{2}$.

Consider a type 1 player in role A. If $\rho_1 = \frac{1}{2}$, at the last node $E\pi(c_5) = \frac{1}{2}\pi(c_6) + \frac{1}{2}\pi(t_6) = 14.4$ and $\pi(t_5) = 6.4$, meaning that his optimal choice is c_5. Reasoning backwardly, one step ahead, the expected value of playing c_3 is $E\pi(c_3) = \frac{1}{2}\pi(c_4) + \frac{1}{2}\pi(t_4) = 7.6$, greater than $\pi(t_3) = 1.6$. At the initial node, again it is the case that $E\pi(c_1) = \frac{1}{2}\pi(c_2) + \frac{1}{2}\pi(t_2) = 4 > 0.4 = \pi(t_1)$. Then, type 1's optimal strategy is $r_A^* = c_1c_3c_5$. When in role B, a type 1 plays t_6 at the last node. At the fourth, $E\pi(c_4) = \frac{1}{2}\pi(c_5) + \frac{1}{2}\pi(t_5) = 7.2 > 3.2 = E(t_4)$. At node two, the value of terminating is $\pi(t_2) = 0.8$ and that of continuing is $E\pi(c_2) = \frac{1}{2}\pi(c_3) + \frac{1}{2}\pi(t_3) = 3.8$. Summing up, the payoff maximizing strategies for a type 1 type are: $r_A^* = c_1c_3c_5$ and $r_B^* = c_2c_4t_6$. Reasoning in the same way for a type 2 player, we have that $t_A^* = c_1c_3t_5$ and $t_B^* = c_2c_4t_6$, which is the best reply to (r_A^*, r_B^*).

Termination probabilities outcomes i / Bi / I / BI

$\bar{q}_1 = s_0(1-q)$
$\bar{q}_2 = s_0^2\, q(1-q) + (s_0 - s_0^2)(1-q)$
$\bar{q}_3 = s_0^2q^2(1-q) + (s_0 - s_0^2)q(1-q)$
$\bar{q}_4 = s_0^2q^3(1-q) + s_0q - s_0^2q - s_0q^2 + s_0^2q^2$
$\bar{q}_5 = s_0^2q^4(1-q) + s_0q^2 - s_0q^3 - s_0^2q^2 + s_0^2q^3$
$\bar{q}_6 = s_0^2q^5(1-q) + s_0q^2 - s_0^2q^2 + 1 - 2s_0 + s_0^2$
$\bar{q}_7 = s_0^2q^6 + s_0q^3 - s_0^2q^3$

Termination probabilities outcomes ii / Bii / II / BII

$\bar{q}_1 = s_0(1-q)$
$\bar{q}_2 = s_0(1-q) - (1-q)^2s_0^2$
$\bar{q}_3 = s_0^2q^2 - s_0^2q^3 + s_0q - s_0q^2 - s_0^2q + s_0^2q^2$
$\bar{q}_4 = s_0^2q^3 - s_0^2q^4 + s_0q - s_0q^2 - s_0^2q + s_0^2q^2$
$\bar{q}_5 = s_0^2q^4 - s_0^2q^5 + s_0q^2 - s_0q^3 - s_0^2q^2 + s_0^2q^3 + (1 - s_0 - s_1)(s_0q^2 + 1 - s_0)$
$\bar{q}_6 = s_0^2q^5 - s_0^2q^6 + s_0q^3 - s_0^2q^3 + s_0s_1q^2(1-q) + s_1 - s_0s_1$
$\bar{q}_7 = s_0q^3(s_0q^3 + s_1)$

Termination probabilities outcomes iii / Biii / III$^-$

$\bar{q}_1 = 1 - s_0q$
$\bar{q}_2 = s_0q - s_0^2\, q^2$
$\bar{q}_3 = s_0^2q^2(1-q)$

$\bar{q}_4 = s_0^2 q^3(1-q)$
$\bar{q}_5 = s_0^2 q^4(1-q)$
$\bar{q}_6 = s_0^2 q^5(1-q)$
$\bar{q}_7 = s_0^2 q^6$

Termination probabilities outcome III

$\bar{q}_1 = s_0(1-q)$
$\bar{q}_2 = s_0^2 q(1-q) + (s_0 - s_0^2)(1-q)$
$\bar{q}_3 = s_0^2 q^2(1-q) + (s_0 - s_0^2)q(1-q) + (1-s_0-s_1)s_0 q + (1-s_0-s_1)(1-s_0)$
$\bar{q}_4 = s_0^2 q^3(1-q) + s_0 q^2 - s_0^2 q^2 - s_0 s_1 q^2 + s_0 s_1 q - s_0 s_1 q^2 + s_1 - s_1^2 - s_1 s_0$
$\bar{q}_5 = s_0^2 q^4(1-q) + s_1 s_0 q^2 - s_1 s_0 q^3$
$\bar{q}_6 = s_0^2 q^5(1-q) + s_1 s_0 q^2 + s_1^2$
$\bar{q}_7 = s_0^2 q^6 + s_1 s_0 q^3$

Termination probabilities outcome IV

$\bar{q}_1 = s_0(1-q) + 1 - s_1 - s_0$
$\bar{q}_2 = s_0^2 q(1-q) + s_0(1-s_1-s_0)q + s_0 s_1(1-q) + (1-s_0-s_1)s_1$
$\bar{q}_3 = s_0^2 q^2(1-q) + s_0 s_1 q(1-q)$
$\bar{q}_4 = s_0^2 q^3(1-q) + s_0 s_1 q(1-q)$
$\bar{q}_5 = s_0^2 q^4(1-q) + s_0 s_1 q^2(1-q)$
$\bar{q}_6 = s_0^2 q^5(1-q) + s_0 s_1 q^2 + s_1^2$
$\bar{q}_7 = s_0^2 q^6 + s_1 s_0 q^3$

Termination probabilities outcome BIII

$\bar{q}_1 = s_0(1-q) + s_1$
$\bar{q}_2 = s_0^2 q(1-q) + s_0 s_1 q + s_0(1-s_0-s_1)(1-q) + s_1(1-s_0-s_1)$
$\bar{q}_3 = s_0^2 q^2(1-q) + s_0(1-s_0-s_1)q(1-q)$
$\bar{q}_4 = s_0^2 q^3(1-q) + s_0(1-s_0-s_1)q(1-q)$
$\bar{q}_5 = s_0^2 q^4(1-q) + s_0(1-s_0-s_1)q^2(1-q)$
$\bar{q}_6 = s_0^2 q^5(1-q) + s_0(1-s_0-s_1)q^2(1-q) + s_0(1-s_0-s_1)q^3 + (1-s_0-s_1)^2$
$\bar{q}_7 = s_0^2 q^6 + s_0(1-s_0-s_1)q^3$

Termination probabilities outcome I^-

$\bar{q}_1 = s_1 + s_0(1-q)$
$\bar{q}_2 = s_0^2 q(1-q) + s_0 s_1 q + s_0(1-s_0-s_1)(1-q) + s_1(1-s_0-s_1)$
$\bar{q}_3 = s_0^2 q^2(1-q) + s_0(1-s_0-s_1)q(1-q)$
$\bar{q}_4 = s_0^2 q^3(1-q) + s_0(1-s_0-s_1)q(1-q)$

$\bar{q}_5 = s_0^2 q^4(1-q) + s_0(1-s_0-s_1)q^2(1-q)$
$\bar{q}_6 = s_0^2 q^5(1-q) + s_0(1-s_0-s_1)q^3 + s_0(1-s_0-s_1)q^2(1-q) + (1-s_0-s_1)^2$
$\bar{q}_7 = s_0^2 q^6 \ + s_0(1-s_0-s_1)q^3$

Termination probabilities outcome II^{-}

$\bar{q}_1 = 1 - s_0 q$
$\bar{q}_2 = s_0^2\ q(1-q) + s_0 s_1 q$
$\bar{q}_3 = s_0^2 q^2(1-q) + s_0(1-s_0-s_1)q(1-q)$
$\bar{q}_4 = s_0^2 q^3(1-q)$
$\bar{q}_5 = s_0^2 q^4(1-q) + s_0(1-s_0-s_1)q^2(1-q)$
$\bar{q}_6 = s_0^2 q^5(1-q) + s_0(1-s_0-s_1)q^3$
$\bar{q}_7 = s_0^2 q^6$

PAPER 2

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

Margherita Bottero and 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 I 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 I. Average GDP across regions in 2000[6]

Region	GDP pc in 2000	Ratio	Ratio (OLS)	Ratio (IV)
North America	27873	0.05	0.08	0.19
Oceania	20819	0.07	0.10	0.26
Europe	12620	0.11	0.17	0.43
Latin America	5889	0.25	0.36	0.92
Asia	3791	0.38	0.56	1.43
Africa	1447	1	1.47	3.75
Africa, counterfactual (OLS)	2128	0.68	1	2.55
Africa, counterfactual (IV)	5427	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 II 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 II. Average slave exports, and exports as a fraction of population[9]

	Atlantic	Saharan	Gaul (I)	Gaul (II)	Gaul (III)	Gaul (IV)
Yearly average	34211	5960	15000	15000	15000	15000
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:

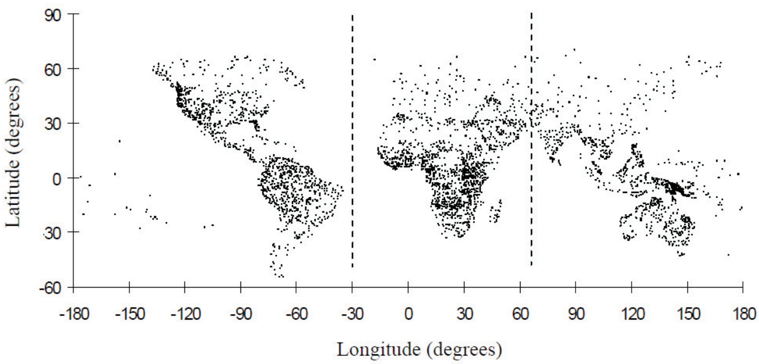
> "The loser goes into slavery without complaint; younger or stronger he may be, but he suffers himself to be bound. Such is their perverse persistence, or, to use their own word, their honour. Slaves of this sort are sold and passed on, so that the winner may be clear of the shame that even he feels 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.

[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)."

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 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

[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.

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 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 III 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

[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.

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 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 III. 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 (Stock & Yogo, 2010; Wooldridge, 2002).

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

[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.

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 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

[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).

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 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*,

[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 corrected. Unfortunately there is no separate data on gold production in Burundi 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 (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 IV 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 IV: 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 V. 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 V.

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 VI and VII 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 VI. 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 VII: 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

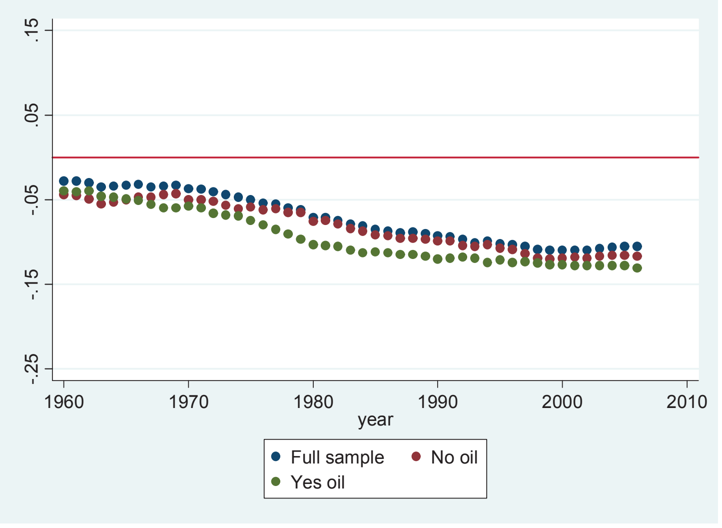
Std. errors in parentheses;* sign. 5%; ** sign. at 1%.

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.

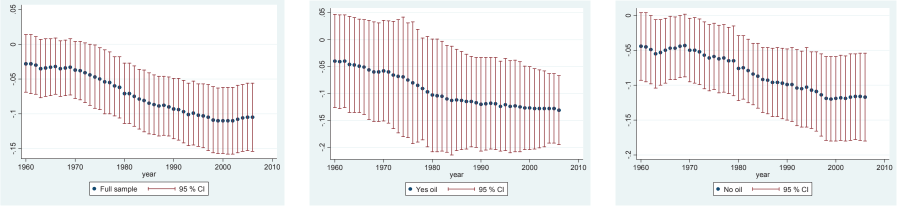
for Italy is similar to that of Africa. To our knowledge the literature does not offer any obvious explanation to why this should be so. In order to investigate the time trends in further detail we plotted the yearly African slave export betas for the years 1960-2006 (Figures 2, 3, 4 & 5) along with 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 VI and VII.

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)

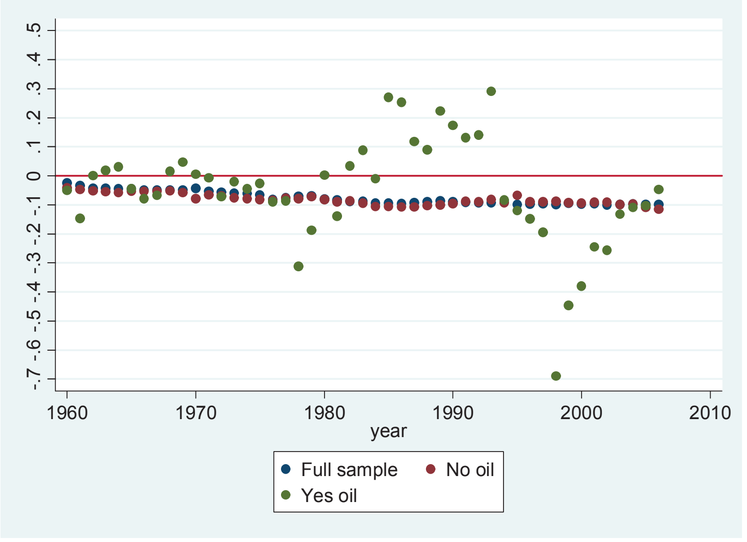
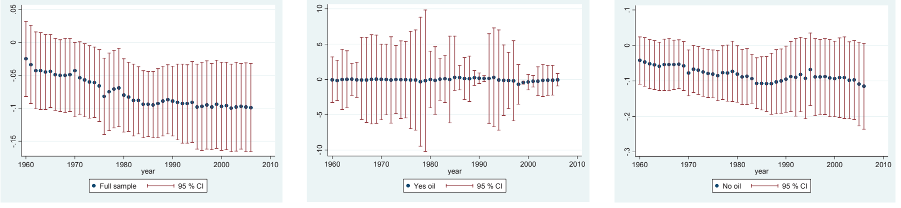


Figure 5. Development of betas with confidence intervals, 1960-2006, Africa (5)



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)

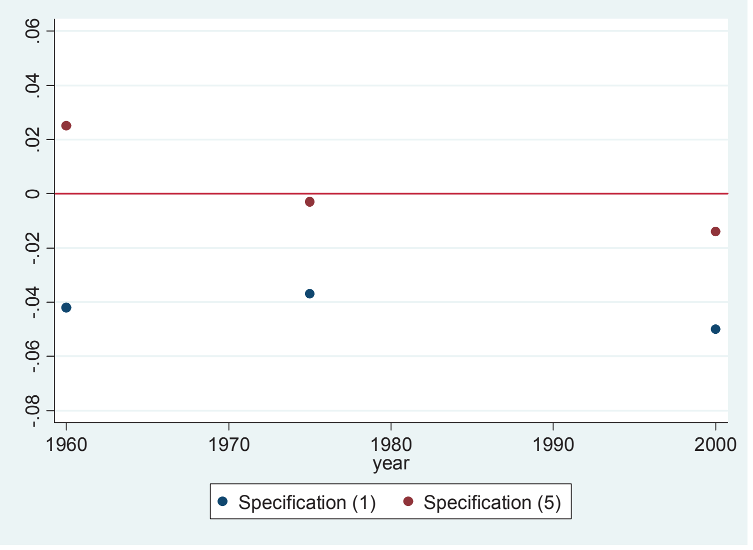
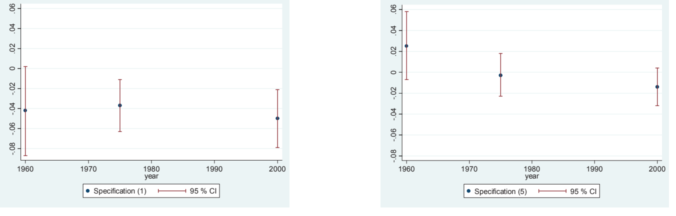


Figure 7. Development of betas with confidence intervals, 1960-2006, Italy specification (1) & (5)



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, Daron, Simon Johnson, and James A. Robinson, "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review* 91 (2001), 1369-1401.

Ahlerup, Pelle, and Ola Olsson, "The Roots of Ethnic Diversity," Working Papers in Economics No. 281, School of Business, Economics and Law, Göteborg University (2007).

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg, "Fractionalization," *Journal of Economic Growth* 8 (2003), 155-194.

Ammerman, Albert J., and L. Luca 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, Quamrul, and Oded Galor, "The "Out of Africa" Hypothesis, Human Genetic Diversity, and Comparative Economic Development," mimeo (2009).

Austen, Ralph 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, Gareth, "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, Henk, *Sugar Cane Cultivation and Management* (Amsterdam, NL: Springer, 1999).

Barbano, Angela, "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, Jan, and Andrea 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, Dirk, Jutta Bolt, and Robert Lensink, "Indigenous Slavery in Africa's History: Conditions and Consequences," mimeo (2009).

Bhattacharyya, Sambit, "Root Causes of African Underdevelopment," Working Papers in Trade and Development, No. 2008/16, Australian National University (2008).

Birchenall, Javier A., "Quantitative Aspects of Africa's Past Economic Development," mimeo (2009a).

Birchenall, Javier A., "Population and Long-Term Economic Development: A Reexamination and Some New Evidence," mimeo (2009b).

Bradley, Keith., *Slavery and Society at Rome* (Cambridge, UK: Cambridge University Press, 1994).

Cashdan, Elisabeth, "Ethnic diversity and its environmental determinants: Effects of climate, pathogens, and habitat diversity," *American Anthropologist* 103 (2001), 968-991.

Cavalli-Sforza, L. Luca, Paolo Menozzi, and Alberto Piazza, *The History and Geography of Human Genes* (Princeton, NJ: Princeton University Press, 1994).

Clark, Gregory, "The Condition of the Working Class in England, 1209–2004," *Journal of Political Economy* 113 (2005), 1307-1340.

Collard, Ian F., and Rober 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, Paul, and Jan Willem 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, Robert C., *Christian Slaves, Muslim Masters - White Slavery in the Mediterranean, the Barbary coast, and Italy, 1500-1800* (New York, NY: Palgrave Macmillan, 2003).

Diamond, Jared, "Evolution, consequences and future of plant and animal domestication," *Nature* 418 (2002), 700-707.

Diodorus, Siculus, *The Library of History, vol. III - Loeb Classical Library*, (Cambridge, MA: Harvard University Press, 1939).

Easterly, William, and Ross Levine, "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics* 112 (1997), 1203-1250.

Engerman, Stanley L., and KennethL. Sokoloff, "Factor endowments, inequality, and paths of development among New World economies," NBER Working Paper Series No 9259 (2002).

Elbl, Ivana, "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://slave voyages .org/ tast/assessment/estimates.faces?yearFrom=1501& yearTo=1866&flag=2 (accessed 2011-04-02).

Fage, John D., "Slavery and the Slave Trade in the Context of West African History," *Journal of African History*, 3 (1969), 393-404.

Fage, John D., and William Tordoff, *A History of Africa 4th ed.* (Abingdon, UK: Routledge, 2002).

Fearon, James D., "Ethnic and Cultural Diversity by Country," *Journal of Economic Growth* 8 (2003), 195-222.

Fenske, James, "Does Land Abundance Explain African Institutions," mimeo (2010).

Ferdère, Alan, *Les campagnes en Gaule romaine*, 2t (Paris, FR: Errance, 1988).

Findlay, R., and Mats Lundahl, "Towards a Factor Proportions Approach to Economic History: Population, Precious Metals and Prices from the Black Death to the Price Revolution," in Ronald Findlay, Lars Jonung, and Mats Lundahl (Eds.), *Bertil Ohlin: A Centennial Celebration 1899-1999* (Cambridge, MA: MIT Press, 2002).

Findlay, R., and Mats Lundahl, "Demographic Shocks and the Factor Proportions Model: From the Plague of Justinian to the Black Death" in Ronald Findlay, Rolf G.H. Henriksson, Håkan Lindgren and Mats Lundahl (Eds.), *Eli Heckscher, International Trade and Economic History* (Cambridge, MA: MIT Press, 2002).

Fisher, Humphrey J. *Slavery in the History of Muslim Black Africa.* (New York, NY: New York University Press, 2001).

Flint, John E. (Ed.), *The Cambridge History of Africa, Volume 5, From c.1790 to c.1870* (Cambridge, UK: Cambridge University Press, 1977).

Fogel, Robert W., *Without consent or contract - The rise and fall of American slavery* (New York, NY; W. W. Norton & Company, 1989).

Galor, Oded, and David N. Weil, "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond," *American Economic Review* 90 (2000), 806-828.

Garner, Phillip, "Reversal in Africa," mimeo (2006).

Gray, Richard (Ed.), *The Cambridge History of Africa, Volume 4, From c.1600 to c.1790* (Cambridge, UK: Cambridge University Press, 1975).

Hernández-Catá, Ernesto, Klaus Scwab, and Augusto Lopez-Claros (Eds.), *The Africa Competitiveness Report 2004* (Geneva: World Economic Forum, 2004).

Hopkins, Antony G., "The New Economic History of Africa," *Journal of African History* 50 (2009), 155-177.

Kaplan, Hillard, Kim Hill, Jane Lancaster, and A. Magdalena Hurtado, "A theory of human life history evolution: Diet, intelligence, and longevity," *Evolutionary Anthropology*, 9 (2000), 156-185.

Klein, Martin A. "The Impact of the Atlantic slave trade on the societies of the western Sudan," in Joseph E. Inikori, Stanley 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, Richard G., *The Human Career - Human Biological and Cultural Origins,* 3rd Ed. (Chicago, IL: The University of Chicago Press, 2009).

Lee, Richard B., *The !Kung San: Men, Women and Work in a Foraging Society* (Cambridge, UK: Cambridge University Press, 1979).

Lee, Ronald, "Population in Preindustrial England: An Econometric Analysis," *Quarterly Journal of Economics* 87 (1973), 581-607.

Lovejoy, Paul E. *Transformation in Slavery - A history of Slavery in Africa, Second Edition* (Cambridge, UK: Cambridge University Press, 2000).

MacArthur, Robert H., *Geographical Ecology: Patterns in the Distribution of Species* (New York, NY: Harper & Row, 1972).

Maddison, Angus. *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, Thomas R., *An Essay on the Principle of Population*, 5th edition (London, UK: John Murray, 1817).

Manning, Patrick, *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, Susan K. "Tools for Understanding Transformation and Continuity in Senegambian Society: 1500-1900," in Christopher R. DeCorse (Ed.), *West Africa during the Atlantic Slave Trade - Archaeological Perspectives* (London, UK: Leicester University Press, 2001).

Meillassoux, Claude, *The Anthropology of Slavery - The Womb of Iron and Gold* (Chicago, Il: The University of Chicago Press, 1991).

Michalopoulos, Stellios, "The origins of Ethnolinguistic Diversity," Working Paper No. 110, Collegio Carlo Alberto (2009).

Moore, Joslin L., Lisa Manne, Thomas Brooks, Neil D. Burgess, Robert Davies, Carsten Rahbeck, Paul Williams, and Andrew Balmford, "The distribution of cultural and biological diversity in Africa," *Proceedings of Royal Society B* 269 (2002), 1645-1653.

Motamed, Mesbah J., Raymond J. G. M. Florax, and William A. Masters, "Geography and Economic Transition: Global Spatial Analysis at the Grid Level," mimeo (2009).

Nash Briggs, Daphne, "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, Daniel. *Linguistic Diversity* (Oxford, UK: Oxford University Press, 1999).

New, Mark, David Lister, Mike Hulme, and Ian Makin, "A high-resolution data set of surface climate over global land areas," *Climate Research* 21 (2002), 1-25.

Nunn, Nathan, "Slavery, Institutional Development, and Long-Run Growth in Africa, 1400-2000," mimeo (2005).

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

Nunn, Nathan, and Leonard Wantchekon, "The Slave Trade and the Origins of Mistrust in Africa," mimeo (2010).

Olsson, Ola, and Douglas A. Hibbs, "Biogeography and long-run economic development," *European Economic Review* 49 (2005), 909-938.

Oliver, Roland A. (Ed.), *The Cambridge History of Africa, Volume 2, c. 500 B.C. to A.D. 1050* (Cambridge, UK: Cambridge University Press, 1978).

Paci, Raffaele, and Andrea 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, Randall, *The Making of a Tropical Disease - A Short History of Malaria* (Baltimore, MD: Johns Hopkins University Press, 2007).

Pamuk, Sevket, "The Black Death and the origins of the 'Great Divergence' across Europe, 1300-1600," *European Review of Economic History* 11 (2007), 289-317.

Phillips, William D. Jr., *Slavery from Roman Times to the Early Transatlantic Trade* (Manchester, UK: Manchester University Press, 1985).

Pianka, Eric R., "Latitudinal Gradients in Species Diversity: A Review of Concepts," *American Naturalist* 100 (1966), 33-46.

Putterman, Louis, "Agriculture, Diffusion and Development: Ripple Effects of the Neolithic Revolution," *Economica* 75 (2008) 729-748.

Rives, James B., "Introduction," in Tacitus, *Germania* (Oxford, UK: Clarendon Press, 1999).

Rodney, Walter, *How Europe Underdeveloped Africa* (London, UK: Bogle-L'Ouverture, 1972).

Rosenzweig, Michael L., "Species Diversity Gradients: We Know More and Less Than We Thought," *Journal of Mammology* 73 (1992), 715-730.

Rosenzweig, Michael L., *Species diversity in space and time* (Cambridge, UK: Cambridge University Press, 1995).

Rowe, Alexander K., Samantha Y. Rowe, Robert W. Snow, Eline L. Korenromp, Joanna R. M. Armstrong Schellenberg, Claudia Stein, Bernard L. Nahlen, Jennifer Bryce, Robert E. Black, and Richard 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, Jeffrey D., and Andrew M. Warner, "Sources of Slow Growth in African Economies," *Journal of African Economies* 6 (1997), 335-376.

Scheidel, Walter, "The Roman Slave Supply," Princeton/Stanford Working Papers in Classics (2007).

Scheidel, Walter, "Roman wellbeing and the economic consequences of the 'Antonine Plague'," Princeton/Stanford Working Papers in Classics (2010a).

Scheidel, Walter, "Slavery in the Roman economy," Princeton/Stanford Working Papers in Classics (2010b).

Smith, Bruce D. *The Emergence of Agriculture* (New York, NY: Scientific American Library, 1995).

Smith, Richard L., *Premodern Trade in World History* (Abingdon, UK: Routledge, 2009).

Spolaore, Enrico, and Romain Wacziarg, "The Diffusion of Development," *Quarterly Journal of Economics* 124 (2009), 469-529.

Stock, James H., and Motohiro Yogo, "Testing for Weak Instruments in Linear IV Regressions," Technical Working Paper 284, National Bureau of Economic Research (2002).

Tacitus, Publius Cornelius, *Tacitus on Britain and Germany - A New Translation, by H. Mattingly* (West Drayton, UK: Penguin Books, 1948).

Tchernia, André, "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, John, "The Kingdom of Kongo, ca. 1390-1678. The Development of an African Social Formation," *Cahiers d'études africaines* 22 (1982), 325-342.

Thornton, John, *Africa and Africans in the Making of the Atlantic World, 1400-1800, Second Edition* (Cambridge, UK: Cambridge University Press, 1998).

Whatley, Warren, "Guns-For-Slaves: The 18th Century British Slave Trade in Africa," mimeo (2008).

Whatley, Warren, and Rob Gillezeau, "The Impact of the Slave Trade on African Economies," mimeo (2009).

Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data* Cambridge, MA: MIT Press, 2002).

Zohary, Daniel, and Maria 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 (1).

Specification (2)

Table VIII: 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 IX: 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 X: 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 XI: 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

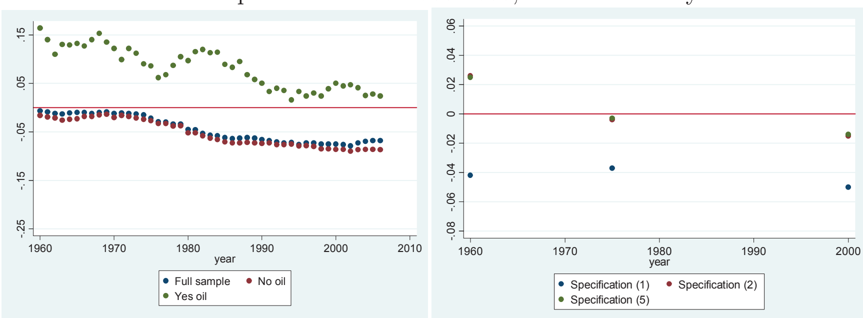
Std. errors in parentheses;* sign. 5%; ** sign. at 1%.

Appendix B

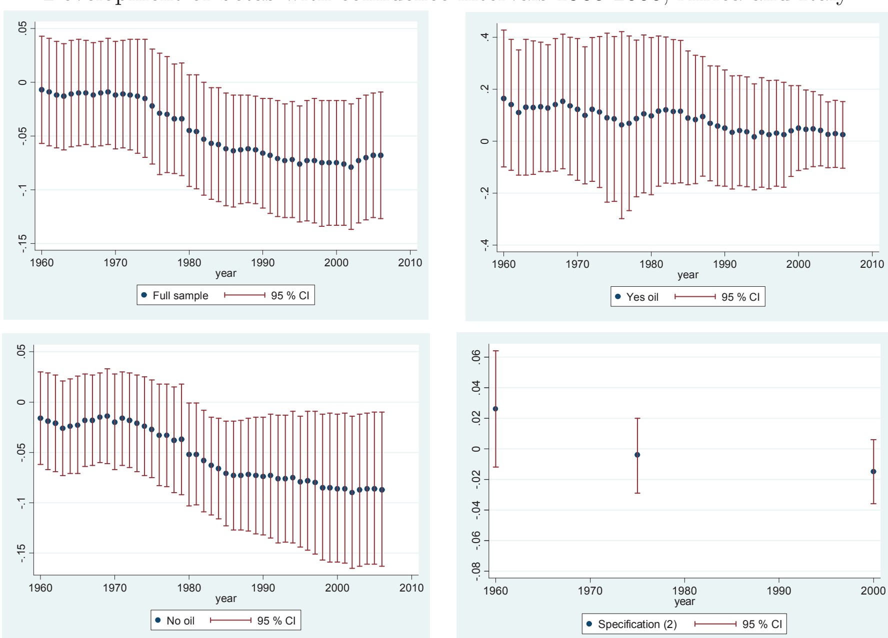
Here follows the development of betas over 1960-2000 for all omitted specifications, with and without the confidence intervals.

Specification (2)

Development of betas 1960-2006, Africa and Italy

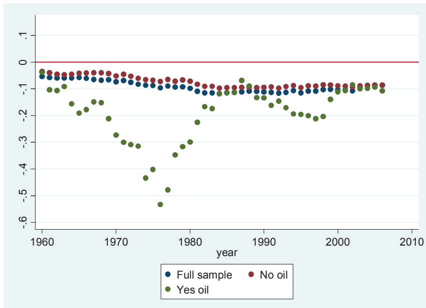


Development of betas with confidence intervals 1960-2006, Africa and Italy



Specification (3)

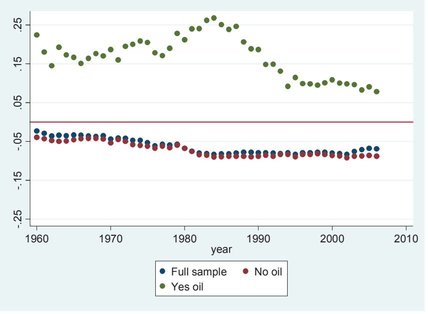
Development of betas 1960-2006, Africa



Development of betas with confidence intervals 1960-2006, Africa

Specification (4)

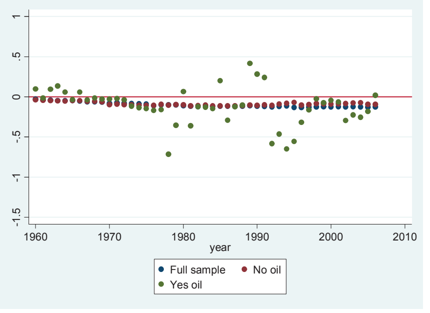
Development of betas 1960-2006, Africa



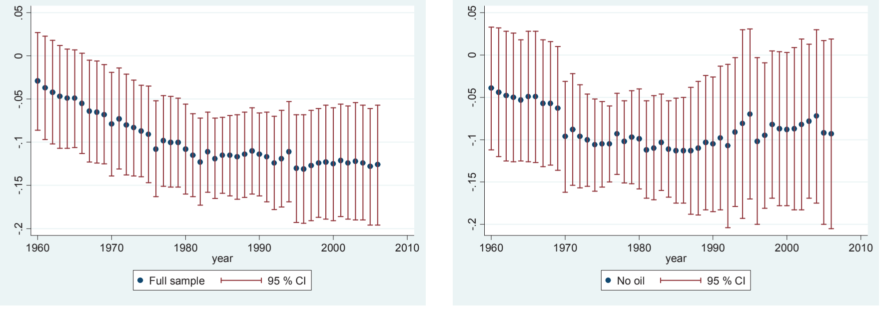
Development of betas with confidence intervals 1960-2006, Africa

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 3

Reputation, privacy and limited records: a survey

Margherita Bottero and Giancarlo Spagnolo

Abstract. We investigate the relationship between reputation, privacy and limited records. Limited records have been demonstrated capable of supporting efficient reputation incentives in a large class of games, where these would otherwise dissipate with time. Further, limited records are compatible with privacy concerns. In particular, they conform to the principle of data retention, which mandates the removal of personal data from public registers after a finite number of years. It is the opportunity of coupling efficient reputation incentives and privacy needs that makes limited records attractive, and motivates our survey. We find that little is known about the design and implementation of limited records, and that what is know is scattered. Given the links between this informational arrangement and valuable goals such as efficiency and privacy, the state of the art appears to us highly unsatisfactory. We conclude that a general model of limited records is greatly needed.

1. Privacy, limited records and reputation

Reputation is a potentially valuable asset because it helps individuals to signal their trustworthiness and their value as a contractual counterpart (Bar-Isaac, Tadelis, 2008). Typically, reputations are accumulated over time, and are backed by relevant information on past behavior and performance. For this reason, the process of reputation formation is likely to benefit from the collection, storage and diffusion of personal data, and to be at odds with privacy concerns that call for some of these data to remain undisclosed. In a neoclassical market, in fact, the participants will themselves optimally decide which, and how much, information is needed to assess the value of a transacting partner. Accordingly, a privacy regulation that mandates the removal of data from the public domain is at best unjustified and, at worst, it opens the door for "fraud[1]" (see also Posner, 1983; Nock, 1993).

We would like to thank Alessandro Acquisti, Marieke Bos, Kasper Roszbach, Andy Skrzypacz, Björn Wallace and Jörgen Weibull for helpful comments and suggestions, as well as Marco Genovese for his assistance with data collection. Margherita Bottero gratefully acknowledges the Knut and Alice Wallenberg Research Foundation for financial support.

[1] From Posner's blog, 8th May 2005. http://www.becker-posner-blog.com/2005/05/index.html

However, besides the canonical argument that we are not (always) capable of optimal information-processing in the neoclassic sense, there are instances in which privacy regulations can alleviate market failures that arise even when assuming rational agents (Acquisti, 2010; Acquisti, Varian, 2005; Calzolari, Pavan, 2006; Hermalin, Katz, 2006; Hirshleifer, 1971; Taylor, 2004a, 2004b). Moreover, in many situations, reputation effects have been demonstrated to disappear over time (Cripps et al., 2004) when access to information is unrestricted.

Originally, in response to this last concern, limited records were put forward as a device to help sustain reputation incentives over time (Ekmekci, 2011; Liu, 2010; Liu, Skrzypacz, 2011). However, limited records can also be seen as enforcing privacy concerns, for they are compatible with the principle of data retention[2]. These two considerations highlight the key appeal of this informational arrangement: it can simultaneously fulfill privacy needs and increase efficiency via strengthening reputation concerns. Being also relatively easy to implement, limited records appear to be an ideal policy for a wide range of real-life situations, where either privacy is socially desirable, or reputation is needed to curb informational asymmetries.

Motivated by this link between limited records, reputation and privacy, we review in depth the theoretical literature up to date, and find that the state of the art is unsatisfactory. In particular, the relevant game theoretic works do not offer any indication on how the length of records should vary with the model's fundamentals[3], although limited records have been demonstrated capable of supporting reputation incentives over time. For what concerns the privacy debate, although privacy *per se* is a topic not unknown to the economics literature (Acquisti, 2010, and references therein), the study of limited records has been left almost neglected.

The lack of a cohesive theoretical foundation guiding the design of limited records is highlighted by observing how they have been implemented in real-life situations. Reviewing the evidence, we uncover a tremendous amount of heterogeneity, which we believe stresses once more that a framework of reference is both missing and needed.

Privacy legislations are possibly the most obvious case of limited records in everyday life. Albeit very pervasive, however, these laws are often formulated in strikingly vague terms. For instance, consider the EU Directive on Data Protection (95/46/EC) which mandates that "[M]ember States shall provide that personal data must be [...] adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed". This excerpt illustrates the problems connected to the implementation of limited records. First, personal data is a surprisingly wide

[2] Which mandates the removal of information collected prior a certain date.
[3] Such as the relative gravity of moral hazard, adverse selection or the players' patience.

umbrella that encompasses personal information, such as sex, race, religious affiliation, but that also includes information on past credit remarks, arrears, bankruptcy as well as criminal convictions. Besides this, the translation of the concept "adequate, relevant and not excessive" into a well-defined retention window has turned out to be challenging. In the case of the credit market, this has resulted in a puzzle that is so emblematic that we will discuss it in an independent section (section 2). As regards the data on criminal convictions, the data-protection principle contemplates their removal with a practice often referred to as expungement[4]. It turns out that in the UK, less serious offences are expungeable after a retention period ranging from three to ten years[5]. In the United States, retention periods vary across states. In Washington state, convictions for misdemeanors are expungeable after three years but DUI (driving under the influence) convictions must remain on the records for at least seven years. In New Jersey, the retention period is ten years for indictable convictions, five years for disorderly offenses and two years for breaking municipal ordinances. Interestingly, in most countries child offenders are typically granted a blank slate upon reaching the age of maturity.

Data removal is also mandated by other types legislations besides privacy laws, usually in an effort to providing "fresh starts". A relevant example is that of breaches of the driving code. The common driver-licence *point systems* that punish infractions of the driving code work by removing points for each past infringement, eventually up to suspending the licence when all points are lost. While these systems are widespread, the time-span that they use to calculate a driver's current score varies wildly across countries[6]. In Australia the system considers infractions realized in the last three years for new licence holders, but it only takes the last years into consideration when it comes experienced drivers. Similarly, in Italy the system begins with considering the infringements in the last three years, but as the driver becomes experienced the memory is shortened to two years. In other countries, instead, there is no such discount for experience. In the UK, for example, data are retained up to eleven years, in the state of New York for one and a half year, in Colorado for three years and in California for ten years.

Surprisingly, there are also situations for which a limited records arrangement appears to have been chosen "endogenously" by the market, without being mandated by law. One such case is the *bonus-malus* systems based on records of past claims in

[4] According to the Black's Law dictionary (1999), the expungement process is a "process by which record of criminal conviction is destroyed or sealed from the state or Federal repository." (p. p. 582).

[5] For instance, the former applies to an order of detention in a detention centre while the latter a sentence of imprisonment or youth custody for between 6 and 30 months.

[6] The information we present has been collected from the homepages of national authorities. The "memory" is the minimum number of years of past information necessary to calculate the score.

automobile third-party insurance. In most countries automobile insurance providers implement bonus-malus systems that link the insurance premium to the number and severity of the insured's claims or accidents during the past years. Though very similar in aim and scope, these systems differ considerably across countries in terms of the rating scale and the scoring rules which determine the ranking of gains and losses. More interestingly, they differ in how many years of information on claims, or accidents, they use to calculate the scores. In 1995 the *minimum* data retention period required to be inputted in these scoring systems was ten years in Norway, six in Sweden, four in Belgium, three in France, two in Germany and one in many other countries (Lemaire, 1995).

Another situation in which limited records have been chosen by the market designers, without legal intervention, is the case of rating systems in electronic marketplaces. For instance, eBay enforced for five years, from 2003 to 2008, a reputation mechanism which compounded into a seller's public score the whole history of his ratings. In 2008, however, this mechanism was modified to include only the ratings received in the last twelve months.

From the examples above, it transpires that although limited records is perceived to be an attractive informational arrangement, little is known about how to implement it, with the result that existing practices look inspired by common sense. Given that a general theory of limited records is needed, it seems to us that it would have to account for, and build on, the existing results, and to this end we review the current state of the art. The survey is organized as follows. First, we describe the case of privacy legislations in the credit market, and argue that it demonstrates in a very transparent way why we need to know more about the relationship between privacy, limited records and reputation. In section 3, we address the theoretical backbone of this survey. In the first part of the section, we will give a more precise meaning to the notion of reputation effects and show why they often are found to disappear over time. Then, we will survey three models that suggest limited records as a potential remedy to this impermanence. Section 4 will look at the applied works that study the effects of limited records on credit market outcomes. In section 5, we will consider the case of electronic marketplaces, and section 6 concludes.

2. The puzzle of privacy laws in the credit market

Privacy protection entails benefits and costs. Benefits appear both in non-economic and economic terms. The former accrue because the endorsement of such a policy fosters individual autonomy, dignity and freedom. Yet, further beneficial effects can arise in terms of genuine economic outcomes, by providing economic actors with a new, clean reputation and, hopefully, with the incentive to maintain it clean by taking efficient actions. Erasing data from the public domain, however, prevents the market from settling on the complete-information equilibrium, so that data retention may result in distorted equilibrium prices and quantities.

The tension between the costs and benefits of privacy is particularly critical for credit markets, which are well-known to suffer from severe adverse selection and moral hazard. In this context, data on borrowers' past behavior, such as total loan exposure, number of credit lines, defaults and filed bankruptcies, are particularly needed to screen out of the market intrinsically bad borrowers. Thus, their removal from public records after a finite number of years, in accordance with data removal, is not straightforward to justify.

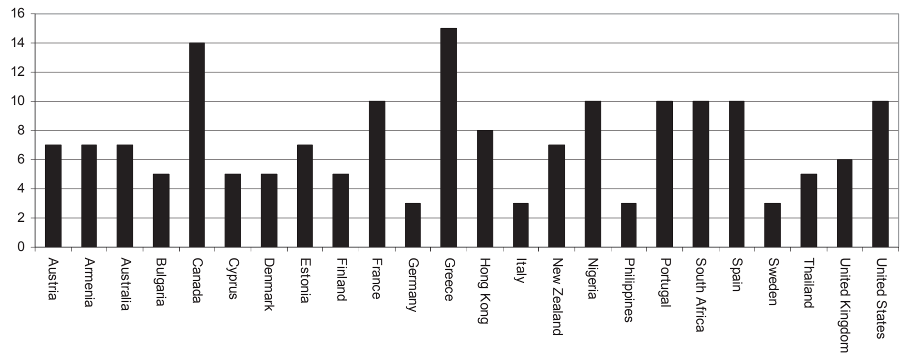
Restating the trade-off between costs and benefits of data retention with reference to the credit market, removing past credit data provides a fresh start to those borrowers and entrepreneurs that were hit by bad luck (Armour, Cumming, 2005). Yet, if the information deleted is a signal of borrower's intrinsic characteristics, such as low skills or fraudulent behavior, limited records effectively impede lenders from screening out of the market those participants that are not credit-worthy. Thus, rather than providing welfare-improving "fresh starts", the legislation can end up clogging the access to, and the functioning of, the credit market as a whole.

In principle, it would seem natural to expect legislators to account for, or at least anticipate, this trade-off. If this is really the case, we should be able to detect a pattern in how data retention has been implemented across the world, and to observe similar retention periods in credit markets characterized by similar features. The evidence, however, shows that there is no such consensus. Figure 1 plots a few retention periods as mandated in different countries, for bankruptcy data regarding the market for credit to household and small entrepreneurs[7]. Although as of 2007, up to 90 percent of the countries wherein a credit register was operating (113 in total) mandated the removal of past credit data after a limited period of time (Elul, Gottardi, 2008), we can see from Figure 1 that there is a large degree of heterogeneity in the length of the data

[7] For a description of credit registers in Europe see the Report of Expert Group on Credit Histories (2009). For an accurate comparison between the United Stated and Europe, see Jentzsch (2001). For more information about South America see Galindo and Miller (2001).

retention window, even for the narrow class of bankruptcy data.

Figure 1. Retention periods for bankruptcy data



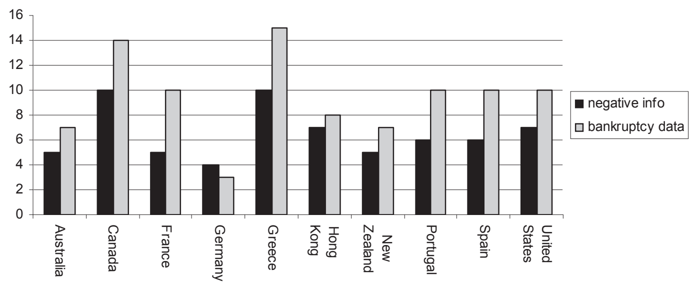
The retention periods range from three years (Germany, Italy, Sweden) up to fourteen and fifteen years (Canada and Greece respectively). Even within continental Europe, we can observe removal after three years in Germany and after seven in neighboring Austria.

Moreover, national laws may mandate different retention periods depending on the type of data (fraudulent bankruptcies, credit remarks, etcetera) or on the type of person filing them (entrepreneurs may be treated differently from households). Figure 2 shows the different retentions periods for two types of negative information, bankruptcy and other adverse information (such as credit remarks, arrears, etcetera).

This puzzling diversity has recently become particularly problematic because of the sharp increase in cross-border data sharing, so that many organizations have highlighted "their concerns about the current patchwork of different data retention periods and expressed their wish for a pan-European data retention period, in order to boost legal security"[8]. However, before addressing the issue of the homogenization of retention periods, it is worth discussing whether these provisions have ever been found to have a positive effect on credit markets' outcomes.

[8] Summary of replies to the public consultation about the future legal framework for protecting personal data, 2009. See also the Communication COM (2010) 609 Final, 2010 http://ec.europa.eu/justice/news/consulting_public/0006/com_2010_609_en.pdf.

Figure 2. Retention bankruptcy and negative data



Elul and Gottardi (2008) look at the empirical relationship between credit provision, as measured by the average ratio of private credit to GDP, at the country level, and current privacy legislations for a sample of more than a hundred countries. This exercise shows that private credit to GDP is almost four times higher in countries where defaults are erased after a finite number of years, than in countries where credit records are not kept (because there is no credit register). More importantly, in countries that mandate data removal, private credit to GDP is more than twice as high than in countries where such data are never erased. Although it is not possible to argue for causality, these findings suggest that whatever the relation between volume of lending and length of retention period may be, it is most likely not linear.

These findings reinforce our intuition that limited records should not only be considered as a way to implement privacy, but also as a way to restore genuinely beneficial reputation concerns. From this perspective, being able to rely on a sound theory is of great importance to design laws and rules in an optimal way.

3. Limited records and reputation

In game theory, reputation effects arise in repeated games with incomplete information. We restrict our discussion to the class of two-player games, where a long-lived informed player repeatedly plays against a sequence of short-lived uninformed opponents, who play the game once and then leave. In such games, incomplete information is typically modelled by allowing for the possibility that the player with private information is either of the normal type or of one of finitely many *commitment* types. The normal type is the familiar payoff-maximizer that chooses his strategy in order to maximize the average discounted payoffs of the repeated stage-game. A commitment type, instead, plays a pre-specified strategy regardless of it being an equilibrium or not. If the probability of meeting a commitment type is small enough, the incomplete-information game can be thought of as a small perturbation of the complete-information game, which the uninformed player expects to be played with a high probability (for a discussion, see Mailath, Samuleson, 2006).

When the short-lived players can observe the whole history of game, and thus the actions taken by the long-lived opponent, it can be demonstrated that a (sufficiently patient) normal player can build a reputation for being of his preferred commitment type, by taking repeatedly the commitment action. Eventually, the long-lived player's reputation[9], given by the opponent's belief on him being of the commitment type, becomes so strong that the opponent begins to best-reply to the commitment strategy. This is a key finding, which demonstrates how allowing for a small amount of incomplete information allows for the past play of the game to affect the expectations regarding its future play.

Clearly, reputation is desirable for the player who is investing in it, as a way to secure himself payoffs better than those he could achieve with complete information. However, reputation can also benefit the short-lived players. This occurs when, in order to signal that he is of a certain commitment type, the normal player takes actions that benefit the aggregate welfare, and that would not be an equilibrium action otherwise. In particular, reputation effects are to be welcomed in repeated games where the only equilibrium in the complete-information game is Pareto-dominated by other strategies, which are however not individually rational in the stage game. Allowing for incomplete information, reputation effects arise endogenously, and induce the normal player to take actions different from the Pareto-dominated equilibrium strategy in anticipation of the gains that will accrue to him from being treated as if being of the commitment type. Reputation effects, then, allow for a play of the game that delivers Pareto-improved payoffs, that could not be reached in the static framework.

With perfect monitoring (and incomplete information), the payoff that can be achieved by the long-lived player can be made arbitrarily close to the maximum equilibrium payoff he can obtain with complete information. Importantly, in repeated games of complete information, desirable payoffs are usually achieved in equilibria that involve actions which are not Nash equilibria of the stage game. Such equilibria are sustained by attaching a negative continuation payoff to deviating players, that is, by prescribing that a deviation "triggers" a punishment phase. Sometimes players complying to the equilibrium strategy are said to maintain a reputation for doing so, and a player's reputation is lost upon entering the punishment phase. This interpretation of reputation,

[9] We can say that reputation takes up a *signalling* role by allowing players to signal that they are, or are not, of a certain type by playing repeatedly certain actions.

however, is substantially different from that in the incomplete information approach. In particular, the latter introduces a meaningful connection between the past and the future play of the game, while with complete information there is no reason to expect the continuation of the game to be different from the original game, after any possible history.

With imperfect monitoring, instead, the "adverse selection" approach to reputation has even starker implications. In particular, it has been found that the long-lived player can secure himself payoffs larger than the maximum that could be achieved without incomplete information. Importantly, it is the case that in building his reputation, the player usually mimics a commitment type that does not suffer from moral hazard, thus taking actions that lead to a Pareto-improvement with respect to the complete information case.

In this survey we will be concerned with the mainstream notion of reputation, which requires incomplete information. In particular, we are interested in those situations in which the players' own desire to build a good reputation leads to an increase in aggregate welfare. Unfortunately, when the uninformed players can observe and recall the whole play of the game (a case which we call complete records) reputation is fragile. Absent imperfect monitoring, a single action that is not compatible with the commitment strategy is enough for the uninformed player to discover the opponent's true type. With imperfect monitoring, the informed player himself finds it increasingly tempting to deviate from the commitment strategy, expecting the uninformed player to attribute such deviations to noise. Indeed, imperfect monitoring preserves the player's reputation for a while, but eventually the player's true type will be learnt, and reputation will cease to play a role (Cripps et al., 2004; see also Özdogan, 2010).

The impermanence of desirable reputation effects has prompted the literature to focus on their rescue, looking for a way to maintain over time strong incentives to build and not to deplete a good reputation. Intuitively, reputation concerns could be kept in check by preventing the informed player from accumulating a too good reputation as well as by granting him the possibility to rebuild a lost reputation. Since a player's reputation corresponds to the belief that the short-lived player has in him being of the commitment type, the intuition above can be restated by saying that to sustain reputation effects over time we need to "continuously replenish" the uncertainty about players' type (Cripps et al., 2004).

In this sense, censoring the short-lived players' access to the information regarding their opponent's past behavior is a viable solution. Limited records appear capable of preventing the informed player from building a too solid reputation, which, as argued above, would provide him with a credible commitment device to keep investing in his reputation. Moreover, actions not compatible with the commitment strategy would be forgotten with the passage of time, allowing the long-run player to rebuild a lost reputation.

For these reasons, and for their evident simplicity, limited records appear to us an ideal device to "continuously replenish" the uncertainty regarding the type of the long-lived player, sustaining over time plays of the game that are preferable from a Pareto point of view. The remainder of this section reviews in the detail the game-theoretic works that investigate whether limited records can effectively succeed in doing so.

3.1. Three models of limited records and incomplete information. To fix ideas, it is useful to lay down some more precise assumptions regarding the interaction under study. Following the literature (Mailath, Samuleson, 2006), we assume a two-player, simultaneous-move, infinitely repeated game of perfect information, played by a long-lived player (a "seller") and a sequence of short-lived opponents (the "buyers"). The latter enter the game, play it once, and then leave. The seller decides whether or not to exert effort in the production of a good, and each buyer decides whether to buy it or not[10]. Alternatively, we can say that the buyer decides on how much to trust the seller, while the seller decides on how much to exploit the buyer's trust.

Payoffs are such that although both the seller and the buyer would benefit from taking the cooperative actions (high effort/do not exploit and buy/trust), these are not an equilibrium of the static game. The only equilibrium of the one-shot, complete information game is for the seller to exert no effort (or to exploit the buyer's trust) and for the buyer not to buy (not to trust). To allow for incomplete information, we perturb this game by introducing the possibility that, besides being of the normal (rational) type, the seller can be of a commitment type. This type always plays the cooperative action[11] (high effort/not exploiting) with a probability greater than zero.

These assumptions describe a repeated game of incomplete information of the kind discussed above, so that, if we allow the short-lived players to observe all the past actions of the seller (a case that we call complete records), as well as those of the previous buyers, we can expect reputation effects to arise. A sufficiently patient seller, in fact, can be demonstrated to optimally choose to mimic the commitment type, so that buyers eventually begin to best respond to the commitment strategy by buying the good/trusting the seller.

[10] The action space need not necessarily to be finite, in which case we would think of a seller that decides how much effort to exert and the buyer how much of the good to buy.

[11] The upcoming results would also hold if there were more than one commitment type. Also, note that the commitment type need not to play the action to which the rational player would like to commit, if he could so (the so-called Stackelberg action, see Mailath, Samuleson, 2006), but this type is only assumed to play the cooperative action with a positive probability at all stages of the game.

Liu and Skrzypacz (2011) study the effects of substituting, in the setup just described, the assumption of complete records with that of (exogenously) *limited* records. In doing so, they are motivated by the observation that limited records is an informational arrangement much more widespread in real-life than complete records. In an effort to capture the features of actual interactions, each short-lived buyers is assumed to observe, upon entering the game, the K most recent actions taken by the seller, and to not be able to observe the actions taken by previous buyers[12]. Further, the seller's payoff is assumed to be submodular in the buyers' likelihood to best reply to the commitment action (by buying/trusting), implying that the seller's temptation to deviate from mimicking the commitment type's strategy increases in the likelihood that the buyer will trust the seller.

When these two assumptions, limited records and submodularity, hold, it is possible to demonstrate that any of the perfect Bayesian equilibria (hereafter, PBE) of the game displays stochastic *cycles* of reputation building and exploitation. More precisely, when the records show any positive number $k \leq K$ of low-effort/exploitation choices (deviations), the seller mixes between exerting high and low effort, and the buyer mixes between buying and not buying (trusting/not trusting), with a higher probability on the trusting action the cleaner is the seller's records. When his records displays no deviations in the previous K periods, the seller is indistinguishable from the commitment type and he deviates for sure, exploiting the current buyer's trust. After this, he starts building his reputation anew.

The reputation building phase terminates for sure when the seller has achieved the clean record with only cooperative actions. At that point, he is indistinguishable from the commitment type and he exploits for sure. Buyers, in fact, upon seeing a clean record, cannot distinguish between a rational and a commitment seller and best reply to the latter, which by definition displays the clean record more often. Importantly, the seller does not further cooperate in equilibrium because an additional cooperative action would not be seen by future buyers. Such an action would only replace the cooperative action recorded K periods ago that is going to be erased by the passage of time. Thus, once he has the clean records, the seller cannot increase the buyer's trust any further, and he optimally exploits it for sure.

The reputation cycles just described arise from the joint action of the submodularity and limited records assumptions. The former creates a tension between the benefit of deviating today and the discounted benefits of deviating tomorrow on a more trusting

[12] The authors argue that these modelling assumptions represent buyers on electronic marketplaces, lenders that access a borrower's credit records or customers that access the registers of previous customers' opinions and complaints. In these situations, the current buyer/lender/customer effectively cannot observe the action taken by his predecessors.

buyer. Such a tension is solved by mixing between mimicking the commitment type and exploiting when the records display any number of deviations. The submodularity assumption also explains why in equilibrium, during the reputation building phase, each individual buyer decides to mix between best replying to the normal type's and to the commitment type's action, even knowing that the seller is not of the commitment type because he has deviations in his records. Each buyer, in fact, knows that during the reputation building phase the seller will not take the exploitative action for sure, foreseeing the gains from exploiting a more trusting buyer in the future. Anticipating this and knowing that some future buyers will believe the seller to be of the commitment type for sure (at the clean record with no deviations), a buyer trusts the seller, to a certain extent, even when he knows his true type. Again because of submodularity, a buyer trusts the seller the more the cleaner his records[13].

Naturally, the assumption of limited records plays as well a key role in allowing for reputation cycles to arise. Were records complete, a current short-run buyer would know that future buyers will be able to access (at least) the same history of the seller's action as he can. Accordingly, once a buyer sees a deviation in the records, he knows that future buyers will see it as well, and that the seller has no incentives to postpone exploitation, since he will never be able to "clean up" his records. Thus, with complete records, the seller would exploit at most once along the equilibrium path, and after that reputation effects would disappear forever.

Summing up, the key finding of Liu and Skrzypacz consists of having demonstrated that in repeated games of incomplete information, limited records together with the, fairly realistic, assumption of submodularity of the seller's payoffs allow for rich reputation dynamics to arise in equilibrium. Importantly, these do not require each individual buyer to believe that the seller is of the commitment type to allow him to invest in reputation building. Anticipating that the rational seller prefers to deviate on future, more trusting, buyers (submodularity), and knowing that a future buyer will indeed see clean records and believe himself to be playing with the commitment type (limited records), buyers that observe a records with any positive number of deviations optimally mix between buying and not buying the good. Accordingly, even a (known) rational seller is in the position to build up a reputation for being of the commitment type. Such a result is very innovative because, to our knowledge, it contrasts with most of the existing results, that document reputation dynamics only when the uninformed player entertains a genuine belief to be playing with the commitment type.

---

[13] The authors comment that "as a result, [the seller and the buyers] collectively exploit future short run players" (Liu, Skrzypacz, 2011, p. 14).

Reputation cycles similar to the ones just described arise also when the assumption of (exogenously) limited records is replaced with that of costly information acquisition, which effectively makes records endogenously limited. Liu (2010) studies this possibility in the game above and assumes that, upon entering the game, a buyer has the option to pay a monetary cost to access, unobserved by the seller, k "bits" of information, which correspond to the records of the seller's actions[14] in the last k periods. The cost of buying information is assumed to increase in the number of periods acquired, so that buyers never buy the full history[15].

With costly information acquisition, in all the PBE of the game the seller builds his reputation by mixing between high and low effort until he has K^* cooperative actions in a row, where K^* corresponds to the maximum number of information bits that each short-run buyer would want to acquire in equilibrium[16]. Each buyer decides whether and how much information to acquire as well as whether to buy the good or not.

Liu demonstrates that, in any PBE of the game, the buyers randomize on how much information to buy. Upon acquiring $k^* \leq K^*$ bits of information, buyers buy the good as long as they do not see any opportunistic action in the acquired records, and otherwise refrain from buying. When they do not acquire any information, buyers mix between buying and not buying with a probability on the first action that is higher the lower the costs of information acquisition, thus free-riding on the previous buyers' information acquisition decisions. Note that a situation in which each buyer acquires (lots of) information for sure is not an equilibrium. Future buyers, in fact, would find it optimal to acquire less information, in order to free-ride on the strict monitoring exerted by their predecessors. But this eventually would weaken the seller's incentives not to cheat, thus undermining the disciplinary effect of information acquisition.

Differently from the case in which records are limited exogenously, here short-run players only trust the long-run player when they genuinely believe he may be of the commitment type. This follows from the fact that records are now monitored randomly, sometimes not at all and other times to different extent. Then, even when maintaining the submodularity assumption, short-run buyers that discover that the seller is of the rational type do not expect him to postpone the exploitative action, because it is not clear whether the future buyers will be more or less trusting, or not trusting at all. Thus, once replaced the assumption of exogenously limited records with that of costly

14 As in Liu, Skrzypacz (2011), short-lived buyers are not assumed to be able to observe each other's actions.

15 In characterizing the equilibrium, Liu disregards the initial periods in which buyers could afford to buy the whole history.

16 Acquiring more than K^* bits of information leads to costs higher than the expected benefits. More precisely, K^* follows from the assumption that the "marginal value of information decreases exponentially in the amount of information" (p. 5).

monitoring, the seller is trusted only when the acquired information is consistent with him being of the commitment type.

In Liu's model, then, reputation building is driven by the assumption of costly information acquisition. Since the buyers that do not acquire any information mix between best replying to the commitment and to the rational action, the seller has an incentive to mix between exploiting and cooperating, in an effort to build up a reputation for being of the commitment type. The reputation building phase ends once the seller achieves a record that makes him indistinguishable from the commitment type, at which point he takes exploitative action for sure[17].

In the two models discussed so far, monitoring has been assumed to be perfect, although limited to the last K periods. Ekmekci (2011), instead, allows for imperfect monitoring, and adapts the incomplete-information game that we have been looking at so far to represent a seller/buyer interaction in an electronic marketplace. The game is similar to the one discussed above, although submodularity is not assumed, so that the seller's temptation to cheat does not increase in the buyers' likelihood to buy[18].

Moreover, because of imperfect monitoring the short-lived buyers can only observe signals about the past play. In particular, we maintain the assumption that buyers cannot see each other's actions, so, without information censoring, they would access the (whole) history of signals about the seller's actions. Were this the case, reputation effects would arise and then disappear, as argued in Cripps et al. (2004). To prevent this from happening, Ekmekci investigates the possibility of limiting the buyers' access to past information. Differently from the limited records we have discussed so far, rather than assuming that buyers can access only part of the history of past signals, Ekmekci assumes a rating system, that produces a noisy signal regarding the past play of the seller[19]. In the first period of the game, the seller's rating is selected at random among the set of possible ratings, which we can think of as a finite set of natural numbers. After that, the rating evolves according to a transition rule chosen by the rating system, which maps the seller's past actions, the system's old ratings and the realized signal[20] into a probability distribution over the possible ratings.

---

[17] As in Liu, Skrzypacz (2011), since the seller has K^* cooperative actions in a row, further cooperation would not be rewarded because no buyer would acquire $K^* + 1$ bits of information.

[18] More precisely, the cost of effort is constant across all actions of the short-run players. That is, taking the opportunistic actions does not deliver a higher payoff when the short-run buyer is (more likely to be) buying/trusting.

[19] The rating system is inspired by the model in Wilson (2002) that studies a non Bayesian decision maker information processing with a finite automata, that, as the rating system, process the information that will be disclosed to the decision maker.

[20] That Ekmekci interprets as rating of the seller's performance

Under this informational arrangement, Ekmekci shows that it is possible to construct a rating system that sustain reputation effects over time. In particular, under such system, it is possible to find a PBE of the game in which the buyers buy at all the ratings but the lowest one and the effort exerted by the seller depends on his rating. At the lowest rating, the seller behaves opportunistically, and the current-period buyer does not buy the good. As the rating improves, the seller mixes between cooperating and taking the opportunistic action, while buyers always buy the good. Finally, at the best rating the seller exploits the buyer who, observing the best rating, buys the good for sure, expecting to be dealing with the commitment type[21].

The importance of Ekmekci's results is best appraised by relating them to the impermanent reputation finding[22] in Cripps et al. (2004). Ekmekci shows, in contrast to Cripps et al.'s, that it is possible to design a rating system such that (i) the buyers never learn fully the seller's type, (ii) the long-lived seller is provided with the right incentives for taking the commitment action. Note that, although not all rating systems which grant (i) also achieve (ii), Ekmecki demonstrates that some exist.

Although the buyers' equilibrium strategy described in Ekmekci's model has a flavor of the reputation cycles documented by Liu, Skrypacz (2011) and Liu (2010), there are a few caveats to keep in mind when comparing the three models. First, in the Ekmekci's model, the seller postpones exploitation by being indifferent between exploiting and not exploiting in each period, rather than because he expects higher benefits from postponing exploitation tomorrow. This follows from the fact that the benefits from exploiting are the same at any rating, since there is no submodularity in the seller's payoff, and buyers do not buy only at the lowest rating. Secondly, similar to Liu (2010) and differently from Liu and Skrzypacz (2011), reputation can be sustained only when the buyers genuinely believe the seller is of the commitment type. This is because in the current model there is no reason for a buyer to expect a rational seller to behave cooperatively, since neither exogenously limited records nor payoff submodularity are assumed. However, thanks to the assumptions on the rating system, and because of imperfect monitoring, ratings different from the top one are also consistent with the commitment strategy. Then, buyers optimally buy at all ratings that are compatible with the seller being a commitment type, and they do not buy the good at the only rating that unequivocally identifies the rational seller (the lowest). Finally, Ekmekci's findings are true for one equilibrium that he constructs, while the results in the two previous models apply to any PBE of the game studied.

[21] This is an equilibrium because the rating system is constructed such that the commitment type is at the best rating much more frequently than the rational type.

[22] Cripps et al. (2004) show how under unlimited records the payoff of the long-run player will eventually approach the equilibrium payoff of the repeated game without uncertainty.

The three models surveyed so far share a number other similarities worth mentioning. First, all the three models assume that the long-lived player is sufficiently patient. Were this not to be the case, the seller's impatience would have to be taken into account explicitly, in designing effective limited records.

Moreover, in all three models the limited records arrangement is not preferable to complete record if incomplete information is assumed away (i.e. allowing the seller to be only of the rational type). In Liu, Skrzypacz (2011) and Liu (2010), the assumption of unobservability of previous short-lived buyers' actions prevents a grim-trigger strategy to be an equilibrium both in the case of limited and complete records. Thus, even with perfect monitoring, both arrangements unfolds in the repetition of the static Nash equilibrium. When the information is instead transmitted via the rating system (Ekmekci, 2011), it turns out that with complete information all the PBEs that can be constructed by transmitting to the buyers a single message that encapsulates the whole history can be replicated by a rating system that shows the current message as well as the complete history of past signals.

Summing up, in games with incomplete information[23] and possibly imperfect monitoring, limited records allow players to build up a reputation for being of the commitment type, which is socially desirable because this type does not suffer from the moral hazard problem. In particular, limited records are effective if the mimicking player is sufficiently patient, so he will find it optimal to invest in reputation building, and if the short-lived players cannot coordinate[24], implying that other desirable equilibria, such as grim-trigger ones, are not feasible. Limited records, however, give rise to cycles that consist of a reputation building phase, which is desirable from a social perspective because the seller often takes an action that Pareto-improves the payoffs, followed by a period of reputation exploitation.

Importantly for our survey, in these models exploitation occurs more often the shorter the length of the records, because with shorter records the rational player succeeds more often in mimicking fully the commitment type, at which point exploitation follows for the reasons outlined above. This result suggests that limited records can sustain beneficial reputation effects, but, to avoid inefficiencies, they have not to be set too short.

4. Limited records and credit markets

As discussed in section 2, privacy laws apply to credit markets, where they prescribe the removal of borrowers' credit data from public registers after a finite number of

[23] And a desirable commitment type.

[24] By not being able to observe each other's actions

years. From a strictly economic perspective, such a provision looks inefficient because it ordains the removal of possibly relevant information from the public domain[25]. Yet, the results in section 3 challenge this intuition, and instead suggest that data retention may be needed to sustain beneficial reputation effects over time. Although this possibility has not been very extensively studied[26], in this section we survey the relevant literature and present the existing results

The interaction between a borrower and a lender is similar to that between a seller and a buyer, since both often feature adverse selection and moral hazard. Adverse selection arises at the pre-contractual stage, whenever borrowers are allowed to have private information about their own characteristics (*type*). Moral hazard occurs instead after the contract has been negotiated, because borrowers can affect the returns to the lenders by taking unobservable and costly actions. Importantly, adverse selection and moral hazard may be, and often are, interrelated. When this is the case, the adverse selection (a borrower's private information) concerns the extent to which he suffers from the moral hazard problem. For instance, this is the case when borrowers differ in their cost of effort. Some borrowers, then, are preferable as transacting partners because they have intrinsic characteristics which induce them to take the opportunistic action less frequently.

If we allow borrowers and lenders to interact repeatedly over time, letting the information on the past play of the game be available at any point in time (complete records), we can analyze the resulting credit market interaction as a repeated game of incomplete information. Thus, according to the results discussed up until now, we can expect reputation effects to arise. More precisely, we can expect borrowers of the less desirable type ("bad" borrowers) to try to build a reputation for being "good" borrowers, in an effort to obtain from the lenders the better contractual terms offered to this latter group. In particular, when adverse selection and moral hazard are interrelated, it may be the case that bad borrowers end up by taking actions that alleviate the moral hazard problem.

Complete records, however, despite creating scope for reputation effects to arise, are also likely to result in their disappearance, as lenders learn each borrower's true type by Bayesian updating. Yet, the results reviewed in section 3 suggest that reputation incentives may be sustained by implementing a limited records policy. If this can be demonstrated to be viable, it would be particularly valuable because, besides increasing efficiency, it would also be in agreement with privacy goals.

[25] Rather than allowing the agents to be free to optimally decide whether to use it or not.

[26] There are however several works that try to assess empirically the impact of privacy provisions on market outcomes (Armour, Cumming, 2005; Bos, Nakamura, 2010; Cohen-Cole et al., 2009; Musto, 2004). The results are mixed.

Interestingly, the works that have explored the above line of reasoning, have found mixed results (Elul, Gottardi, 2008; Vercammen, 1995). In particular, it is not always possible to argue that limited records are preferable to complete records. We believe that this result follows from typical characteristics of the credit markets', at least two of which raise concerns that were not found in the models surveyed thus far. The first feature that pertains specifically to the market for credit is that the severity of adverse selection and moral hazard may differ, and this variation affects the desirability of the limited records arrangement. For instance, when some borrowers are intrinsically not credit-worthy, complete records, that allow the lenders to learn fully the borrowers type, are arguably more efficient than limited records, which would instead allow borrowers of all types to remain in the market. Secondly, and differently from the setup studied in the previous section, in this section we will be looking at a market setup where different informational arrangements affect the contractual terms offered to all market participants. Thus, the composition of borrowers of different types in the market is also an important dimension for assessing the desirability of limited records.

The remainder of this section surveys in depth two papers that try to assess the impact of limited records on credit market outcomes, and it briefly discusses a third work that addresses the same question but from a different perspective. Unless otherwise stated, the setup is assumed to be an unsecured borrowing-lending market for entrepreneurs and small business, where long-lived borrowers apply for credit from short-lived lenders. Borrowers are of different types, some of which are preferable in the eyes of the lenders. Types, however, are private information, so an adverse selection problem arises. Furthermore, once the loan is contracted, borrowers take an unobservable action, which introduces moral hazard into the market. Borrowers are assumed to be long-lived so that we can study the effects that reputation building has on their current behavior. Lenders, instead, are assumed to be short-lived, mainly for two reasons. First, this assumption makes sure that lenders cannot accumulate private information, and have to base their strategy on public information only. This is functional to isolate the effects on information censoring on market outcomes. Secondly, the assumption befits the unsecured credit markets, where lenders typically do not write long-term contracts. Free-entry into the supply side is assumed, so that the credit-industry is competitive, and the zero-profit interest rate prevails. Finally, we can think of a costless technology that enforces the contract, so that borrowers cannot embezzle the realized revenues. We will discuss each model's assumption in more depth below.

4.1. Three models of borrowers' behavior with limited records. While they were not the first to address the topic, Elul and Gottardi (2008) provide the most thorough analysis of the effects that data retention policies have on credit market outcomes. The authors study a borrowing-lending market for entrepreneurs and small business, in which short-lived lenders compete on the contractual terms to offer to long-lived borrowers. Borrowers, who do not have any endowment of their own, need such a loan to undertake a project. For simplicity, it is assumed that the project can result either in $R > 0$ revenues, or in zero revenues. Once revenues are realized, they are used to repay the lenders, thus a loan is either repaid in full or defaulted upon[27]. Whatever remains after the debt is repaid, is used by the borrower for his own consumption[28].

Borrowers can be of one of two types, either safe or risky. The safe type is such that his projects always succeeds, implying that loans to these borrowers are always repaid in full. The loan to a risky agents, instead, may or may not be repaid. More precisely, risky agents can decide whether to exert high or low effort. With low effort, the project fails for sure, while with high effort it results in the project succeeding with a positive probability, strictly lower than one. High effort, however, is costly[29], and the effort choice is unobservable to the lenders. Suitable assumptions on the payoff structure guarantee that in a static framework risky borrowers do not choose high effort, and that lenders do not fund them, because failed projects are assumed to have negative present discounted value.

If the game is repeated, however, the history of past project outcomes allows intertemporal incentives to arise for the risky borrowers, who can build themselves a reputation for being of the safe type. Lenders, in fact, use a borrower's credit history to update their belief regarding his type. As long as a credit history shows no defaults, the lenders' belief in dealing with a safe borrower becomes increasingly strong. By exerting high effort, then, a risky borrower can appear to be of the safe type, thereby securing himself continued access to credit, and at increasingly better terms.

Since lenders are by assumption short-lived, a (long-lived) credit bureau is assumed, that updates and makes available[30] to the lenders the credit history σ_t^i of each borrower i. A credit history consists of the sequence of the outcomes (repayment/default) realized by a borrower in all previous periods up to t. Since, the information contained in σ_t is the only available to the lenders, their strategies, and those of the borrowers',

[27] Borrowers, in fact, do not have any endowment, and cannot write long-term contracts with the short-lived lenders.

[28] It is assumed that saving (negative consumption) is not possible.

[29] This is different from what happens in the Diamond (1989) model, where the moral hazard choice of the "normal" entrepreneurs only involve costlessy choosing between the safe and the risky project.

[30] Lenders are also assumed to have access to the set of contracts offered to borrowers in each past period, although they do not know by which borrower each contract was accepted. The authors justify this assumption by noting that "the set of contracts generally offered to borrowers is available from databases such as "Comperemedia"" (Gottardi, Elul, 2008, p. 8).

are based on it[31]. More precisely, lenders use the information in σ_t to compute the belief $p(\sigma_t)$ that a borrower is of the safe type. This belief is then used to decide on which interest rate to ask for lending their funds[32]. According to the assumptions of the model, upon seeing a single default, the lenders correctly identify the borrower as risky. Borrowers of all types decide which contract to accept[33], and risky borrowers in addition decide on which effort-level to exert. In doing so, they weigh the marginal cost and the marginal benefits of effort. The latter account both for higher current utility, as well as for the higher future payoffs that accrue via reputation building, by lowering the probability of a default.

To begin with, the authors characterize the equilibrium with complete records, in which the credit bureau transmits to current-period lenders the complete history of each borrower's past project outcomes. The authors analyze the Markov equilibrium of the game between the borrowers and the lenders, using $p(\sigma_t)$ as the state. This is a fairly natural restriction[34], because both the borrowers' and the lenders' decisions can only be based on σ_t, and thus on $p(\sigma_t)$. In such an equilibrium, credit records are indeed found to create incentives for the risky borrowers to choose high effort, in order to build a reputation for being of the good type. Moreover, the strength of these incentives is demonstrated to vary with the parameter that measures the cost of effort[35]. Note that reputation is valuable because, in equilibrium, only safe borrowers are offered financing. When a risky borrower is identified, he cannot any longer hope to affect lenders' beliefs by using the outcome of his projects, since there is not, and there

[31] This environment appears inspired by that studied in by Diamond (1989). Diamond was the first to demonstrate that when reputation concerns arise in the credit markets, they have welfare-increasing effects. Reputation, in fact, thanks to its signalling effect, allows only good borrowers to remain in the market, leading to lowers interest rates in equilibrium. In Elul and Gottardi's model, the reputation dynamics work in a similar way, by allowing lenders to screen bad borrowers out of the market (with complete records).

[32] Loans are restricted to be of a single size, normalized to one, and contracts are assumed to be debt contracts, fully characterized by the rate of interest.

[33] In particular, they choose the contract offering the lowest interest rate, since the decision of which contract to accept is unavailable to future lenders.

[34] The authors argue that "non-Markov" equilibria require a degree of coordination among lenders that seems implausible in a real credit markets. However, they also consider the case in which lenders live forever and can write long term contracts, and argue that the findings of the paper would still hold.

[35] When the cost of effort is high, the incentives to exert high effort are weak, and risky entrepreneurs optimally exert low effort whenever financed. Lenders, however, may still decide to finance borrowers as long as they expect enough borrowers to be of the safe type. When instead the cost of effort is low, risky entrepreneurs exert high effort and financing is profitable for any composition of safe and risky types. For intermediate values of the cost of effort, the incentives depend on the belief p that a borrower is of the safe type. If this belief is high, then the equilibrium interest rate is low (as in Diamond, 1989), and high effort can be sustained. As p decreases the incentive to exert low effort weakens, and financing decreases, until we reach a state with no financing.

will never be, any uncertainty about his type. Current lenders anticipate that, for this reason, the borrower is unable to internalize the returns from high effort (that would have otherwise accrued via building good reputation), and, accordingly, they expect him to exert low effort. Due to the assumption on the static moral hazard problem, lenders do not provide a borrower known to be risky with funds.

Having characterized the equilibrium with complete records, the authors move on to discuss the conditions under which a policy of data retention is welfare-enhancing, in terms of total generated surplus. Rather than via "properly" limited records, data retention is here modelled by assuming that once a borrower realizes a default, the credit bureau registers it as a repayment with a strictly positive probability q. Thus, data retention is framed as a "forgetting" policy. The authors demonstrate that this policy is welfare-increasing when the risky borrowers' incentive to exert high effort is strong (i. e. when the cost of effort is not too high), when the borrowers' average quality in the market is not too risky and when the borrowers are sufficiently patient.

To interpret the conditions just outlined, one has to consider that the forgetting policy effectively reintroduces risky borrowers into the market alongside good borrowers. Accordingly, its overall effect on welfare depends on how these "redeemed" borrowers behave once back in the market. In accordance with intuition, total surplus will increase if the borrowers do not slack off a second time, but rather exert high effort. Having clarified this, it appears that the role of the conditions above is precisely to ensure that (reintroduced) risky borrowers will actually behave in a way that increases the realized surplus. Note that, in this model, the forgetting policy hurts the safe borrowers, because it slows down the process of screening risky borrowers out of the market. In turn, this prevents the interest rates from settling on the levels that would prevail with complete information. However, were the safe types also allowed to default, although less frequently than the risky types, forgetting would promote the welfare of both types[36], albeit to a different extent.

We have two main comments on Elul and Gottardi's results. The first concerns the conditions under which forgetting is welfare-enhancing. These describe a market characterized by low adverse selection and low cost of effort, where defaults are likely to happen more frequently out of bad luck rather than out of fraudulent behavior. Under these circumstances, punishing a single default with exclusion forever from the market appears too severe, even from an aggregate welfare perspective. Arguably, in fact, welfare would benefit from reintroducing relatively hard-working but unlucky borrowers. Thus, it seems to us that the result on the efficiency of forgetting is, at

[36] Because now both would benefit from the possibility of being reintroduced in the market.

least in part, driven by the assumption of a too harsh punishment for the defaulting event.

Secondly, Elul and Gottardi's setup is reminiscent of the model studied by Liu and Skrzypacz (2011). In both models, there is a short-lived player, that can be called buyer or lender, who conditions his strategy only upon the information transmitted by a central authority. Further, both models feature a long-lived player, seller or borrower, that can be either of the rational or of a commitment type[37]. In both models, the rational type would want to take the opportunistic action in the static-equilibrium, either exploiting the buyer's trust or exerting low effort. Differently from Liu and Skrzypacz, in Elul and Gottardi's model there is imperfect monitoring of the long-lived player's actions, which cannot be directly[38] observed. However, Liu and Skrzypacz argue that, based on preliminary work, imperfect monitoring would not alter the nature of their results.

Despite these similarities between the two models, they are strikingly different in how information censoring is modelled. While in Liu and Skrzypacz records are effectively limited to the last K periods, Elul and Gottardi assume a forgetting device that transmits complete but "incorrect" records.

Elul and Gottardi argue that the reputation effects that arise under the forgetting specification are comparable to those that would arise under a proper limited records assumption. However, from the discussion in section 3, it seems instead that substituting limited records with forgetting may lead us to overlook richer reputation dynamics. In Elul and Gottardi's model, in fact, once a default is realized and not forgotten, the borrower is uniquely identified as being of low ability, and thus excluded from the market[39]. This is very similar to what happens in Liu and Skrzypacz's model with complete, rather than limited, records. In their model, when information is not censored, exploitation (low effort) can happen at most once, after which reputation is lost forever, just as in Elul and Gottardi, although in this latter model the forgetting policy may postpone the reputation loss for a while.

The effects of forgetting, then, contrast starkly with the dynamics found to arise under proper limited records, which we know to be as rich as to generating cycles of reputation building and exploitation. Thus, we believe it is not unlikely that limited records would produce, also in the context of the credit market, dynamics different

[37] In Elul and Gottardi's model the commitment type is the safe borrower that never fails.

[38] Rather, what is observed is only a noisy signal of the borrower's action, given by the project's outcome. Furthermore, in Elul and Gottardi's model current-period lenders know something about the action taken by the previous period lenders, since they can observe the set of previously offered contracts.

[39] Low ability borrowers have no static incentive to exert high effort, and with low effort their projects are not worth funding.

from those documented under the forgetting assumption. For instance, anticipating that a default will be erased by the passage of time, borrowers may find it optimal to keep building their reputation even when they have such a default in their records. Accordingly, lenders could find it optimal to keep providing funds to these borrowers[40], instead of excluding them from the market altogether. Although Elul and Gottardi's model is informative *per se*, and has the merit of having directly addressed the issue of data retention in the credit market in an comprehensive way, we believe that the issue we just raised deserves further investigation.

The second paper we survey (Vercammen, 1995) studies a credit market similar to that analyzed above, where short-lived lenders compete in offering loans to long-lived borrowers, but it differs in how moral hazard and adverse selection are modeled. More precisely, Vercammen assumes that borrowers can be of one of several, rather than two, types. A borrower's type is thought of as his ability, and it is assumed to only affect the borrower's marginal cost of effort. Intuitively, the higher a borrower's ability, the lower his marginal cost of effort. Borrowers of all types, upon having obtained a loan, have to decide how much effort to exert. As in Elul and Gottardi, the project's revenues are used to repay the loan and, if anything is left, for own (non storable) consumption. Revenues are jointly determined by the borrower's effort choice and the realization of a shock. The disturbance term is assumed to be such that borrowers of any type, and exerting any level of effort, may default[41]. Differently from the previous model, revenues are not a binary variable, and a default occurs when the generated revenues, which may, but need not to, be zero, fall short of repaying the pre-contracted interest rate. When this is the case, the lenders appropriate any revenue that was realized.

Under these assumptions, the interaction between borrowers and lenders features both adverse selection and moral hazard. Since both the effort choice and the shock realization are unobservable, a moral hazard problem arises from the impossibility for borrowers to internalize the returns on effort in the event of a default. In turn, this prompts them to exert less effort than in the perfect information case. Adverse selection, instead, follows from the fact that lenders would prefer to contract with borrowers of high ability. Since a borrower's type is his private information, however, it cannot be credibly communicated, and the well-known market for lemon problem arises.

Note that adverse selection is here modelled differently from what we have seen in Elul and Gottardi. While they assumed a proper commitment type, whose behavior always resulted in a certain outcome (repayment), Vercammen lets borrowers of all

[40] Although of known type, a risky borrower that is known to be committed to high effort is credit-worthy.

[41] Although the higher the effort the higher the probability of repaying.

types to be in the position to repay or default. Furthermore, note how in Vercammen's setup the adverse selection is on the degree of moral hazard: lenders would want to know the borrowers' type to lend their funds to the applicants that suffer less from this distortion. Finally, note that in Vercammen's model all borrowers are assumed to be credit-worthy.

Vercammen studies the PBE of the repeated interaction between lenders and borrowers, and documents that with complete records reputation effects arise in equilibrium. Lenders, in fact, are aware that the borrowers may be of different types, and, accordingly, offer different contracts to borrowers with different credit histories. As before, this information is made available to them by a credit bureau, and consists of the history of a borrower's project outcomes (default/repayment).

Just like in Elul and Gottardi's model, the intertemporal transmission of information creates the incentive for borrowers not only to maximize their current-period utility, but their future payoffs as well. Future payoffs depend on the lenders' belief (on the borrower's type) because lenders offer lower interest rates to the applicants believed to be of high ability, expecting them to default less frequently than their low-ability counterparts. Anticipating these lower interest rates, a borrower has an incentive to build himself a reputation for being of the high type, which he can do by repaying rather than defaulting. Note that the possibility of building a reputation, allowed by the presence of incomplete information, increases the total marginal value of today's effort for a borrower, by connecting it to future payoffs, and, thus, it prompts him to exert higher effort than in the complete information case. In turn, this alleviates the moral hazard problem.

Complete records allow for reputation effects to arise, but unfortunately they also imply their disappearance over time, as lenders learn the borrowers' types. As learning becomes increasingly precise, the equilibrium interest rates will approach those prevailing in the complete information game. Reputation effects also vanish, because as a borrower's type becomes known, the lenders' beliefs can be influenced to a lesser extent. Accordingly, the return on effort that was accruing to borrowers via reputation building disappears, because there is no reason to expect lenders to condition contractual terms on anything else than a borrower's known type.

The impermanence of reputation effects, although swiping adverse selection away from the market, can be welfare-decreasing, at least for some range[42] of the parameters. Motivated by this result, Vercammen investigates the effects on welfare of a policy of

42 In general, the effects of fading adverse selection are ambiguous. However, with fixed investment size, adverse selection does not cause any direct welfare loss, and moral hazard is less severe when it appears jointly with adverse selection than without it. As a result, welfare is higher when both distortions are present. For a discussion, see Vercammen (2002).

data retention. To do so, he restricts lenders to observe only (limited) records of the T most recent outcomes realized by each borrower. It can be demonstrated that such a policy, by effectively preventing the lenders from learning a borrower's true type, allows borrowers to build themselves, and maintain over time, a reputation for being of high ability, with results that are beneficial for aggregate welfare.

Although Vercammen does not explicitly study the determinants of the optimal length of the window T, he suggests that intermediate values are preferable. Too long retention periods, in fact, may risk weakening the reputation incentives, by allowing the lenders to screen borrowers in a rather accurate way. Too short periods, on the other hand, may not allow borrowers enough room to build themselves a solid reputation. Vercammen also highlights that, in designing limited records, it is important to account for the effect that different retention periods have on the welfare of borrowers of different types. In particular, like in Elul and Gottardi's model, high-ability borrowers benefit relatively more from longer retention periods, which allow them to signal credibly their types. Low-ability borrowers, instead, prefer relatively shorter records. Under this arrangement, in fact, possible defaults are forgotten more frequently, and the prevailing interest rates involve a large extent of cross-subsidization, which is attractive for low-ability borrowers (for a discussion, see Vercammen, 2002).

For the purpose of our survey, Vercammen's model is interesting because it illustrates how limited records can be designed in a way that maximize efficiency and respects privacy. This result rests on the assumption that moral hazard and adverse selection are related, since the adverse selection (borrowers are of different, privately known types) is "on" the severity of moral hazard (worse types suffer from a more severe moral hazard problem). Credit records exploit the joint presence of these two informational asymmetries by allowing borrowers the possibility to build a reputation that is beneficial for aggregate welfare. Reputation building, in fact, prompts borrowers of all types to exert an effort higher than in the complete-information case, thereby alleviating the distortions that would otherwise arise. Limited credit records, instead, allow reputation to maintain its disciplining power over time, by providing borrowers (of all types) with a credible commitment device to exert high effort .

It is also worth pointing out, however, that Vercammen's clear-cut results on the desirability of limited records is in part driven by the assumption that all borrowers are credit-worthy. This relaxes the need to screen bad borrowers out of the market, which instead may be an important objective for many market regulators, as, for instance, it was in Elul and Gottardi (2008).

Finally, we want to mention the work of Chatterjee and al. (2010), who study the effects of limited records on households', rather than small entrepreneurs', borrowing and defaulting decisions. The authors offer a fully-fledged model with the objective to provide a theory of credit scoring. This is the widespread practice, often carried out by credit bureaus, of assessing a borrower's credit-worthiness, and collapsing it into a credit score[43], available to lenders and other interested contractual partners upon demand.

Chatterjee et al. provide a theoretical foundation to such practice, and put forward a model showing that credit scores can be interpreted as encapsulating the lenders' beliefs regarding a borrower's expected default probability. Moreover, they succeed in demonstrating within the model's framework a variety of empirical regularities concerning credit scores and access to credit[44]. More importantly for our survey, the authors also use their model to assess the effects of data retention provisions on consumers' welfare. In particular, they focus on the current data removal regulation in the United States, according to which bankruptcy events should be removed from the borrowers' files after ten years. Calibrating their model for the American market, it turns out that the current legislation is not welfare increasing compared to a legislation that does not mandate any removal. The difference, however, is found to be small.

Although Chatterjee et al. do not take a "reputation" perspective, presumably because they focus on household consumption and default decisions, their model offers a different and equally important perspective on limited records. We believe that the perspective they take may provide important insights for assessing the efficiency of data retention provisions.

5. Limited records and electronic marketplaces

In this section, we survey the applied literature that combines economics, business studies and computer science to investigates information censoring in electronic marketplaces. Differently from what we have seen so far, these models typically assume complete but imperfect information, arguably in order to focus on moral hazard, which is the gravest source of distortion in electronic markets. For this reason, the models we are going to survey do not qualify as canonical models of reputation. Yet, we have argued elsewhere that even in the economics literature the term reputation is sometimes extended to indicate certain grim-trigger equilibrium strategies in situations characterized by pure moral hazard (for a discussion, see Mailath, Samuelson, 2006). For this reason, and because they deal with information censoring and efficiency, we

---

43 The FICO score is an example one of such credit-score.

44 These are (i) interest rates are lower for better credit scores, (ii) the removal of adverse event from a borrower's credit score substantially improves his access to credit (iii) opening more credit lines decreases a borrower's score. For more details, see Musto (2004).

have decided to include two such works in our survey, although we are aware that, by doing so, we are abandoning the traditional approach to reputation[45].

The defining feature of electronic marketplaces, such as Amazon and eBay, is their large degree of anonymity. Anonymity arises because buyers typically have no previous experience of transacting with a particular seller, and cannot easily access the experience of other buyers. Sellers, moreover, cannot be prevented from leaving and reentering the market under different identities. Further, buyers often can only partially observe the item they are buying before paying for it. These features effectively constrain buyers' access to relevant information, and, worryingly, allow ample scope for opportunistic behavior.

In physical markets, opportunistic behavior is usually disciplined by legally enforceable contracts, as well as by reputation concerns[46]. The latter usually arise spontaneously and carry the threat of a decrease in, or even of the exclusion from, future transactions. In the electronic marketplaces, however, contracts are cumbersome to write and difficult to enforce, and reputation concerns do not arise spontaneously, due to the large degree of anonymity surrounding the sellers and the buyers[47]. For these reasons, market designers have introduced[48] *online reputation mechanisms*, also referred as rating/scoring systems, which have the explicit goal of promoting trust and inducing cooperation. These mechanisms work by asking the buyers, after any transaction, to submit a rating on the experience they had with the seller. The central system receives the evaluation and makes it accessible to the other potential buyers by compounding it in an aggregate score displayed on the seller's public profile.

Rating systems have very effectively succeeded in eliciting from sellers a concern for reputation, which, by disciplining their opportunistic behavior, works in a welfare-increasing way for the whole market[49]. Many empirical studies document a sizeable

[45] To a certain extent, in economics the theoretical backbone of this literature is found in the works that investigate repeated games with bounded recall. See for instance, Bhaskar (1998), Cole and Kocherlakota (2004), Renault et al. (2006), Barlo et al. (2008) and Doraszelski and Escobar (2011).

[46] In physical markets, reputation arises spontaneously from the repeated interaction of sellers and buyers, because the latter realize the benefits of pooling and sharing their own experience, and the former understand that a good reputation means profits as much as a bad reputation means real economic losses. In electronic marketplaces, however, the interactions are neither sufficiently frequent nor personal for reputation to arise spontaneously.

[47] Sellers may abuse anonymity by reentering the market with different identities, while buyers may either be unable to share their experience, or not find it optimal to share it without being able to benefiting from reciprocation.

[48] See Bakos and Dellarocas (2003) and Dellarocas (2006) for a survey of the various reputation mechanism.

[49] This success can be attributed to the mechanism's ability of coupling two goals. First, rating systems allow for buyers to internalize the return of sharing their experience, with the expectation to benefit from accessing the experience of the others in the future. Secondly, and most importantly, the

price premium to reputation, which can be up to eight per cent of the price (Resnik et al., 2006). Similarly, a decrease in a seller's reputation has been found to be associated with a considerable fall in sales price (Cabral, Hortacsu, 2010).

In this section, we focus on the design-problem of how many of the past ratings received by a seller should be compounded into his score, in order to elicit the maximum attainable efficiency and welfare in the market[50]. Absent incomplete information, the problem of designing the systems' *memory* principally lies in finding a mechanism capable of deterring the sellers from taking the opportunistic action, by threatening them with a punishment high enough to make cooperation preferable. Note how this differs from the problem on which we have been focusing so far, that consisted of how to endogenously create, via the elicitation of reputation concerns, the conditions for players to behave cooperatively.

Before proceeding, to fix the ideas, we briefly describe the theoretical setup that we will be dealing with. An electronic marketplace is often represented by a repeated (second-price) auction between a long-lived seller, who has an item for sale in every period, and many short-lived potential buyers. Buyers observe the information transmitted by the rating system, then bid[51] truthfully their valuation of the good, and leave the market before the next auction starts. The value of the item for sale may either be fixed over time or a random variable. In both cases, the seller can take unobservable actions, possibly but not necessarily costly, that matter for the buyers' realized utility, such as advertising the item honestly or exerting high effort in delivering the good on time. Payoffs are assumed to be such that in the one-shot repetition of the auction the only equilibrium is for the seller to take the opportunistic action and for the buyers to best respond to it, either by not buying the good or by bidding a low price. To represent the reputation mechanism, it is assumed that after any transaction, the buyer that bought the good submits to the central system a rating s. Once the rating is submitted, the system compounds it into the seller's score R, which the future buyers will take into account when deciding on their bids.

Without loss of generality, ratings can be assumed to be a binary variable. Buyers, then, submit a positive rating if the quality of the delivered good is compatible with the seller having taken the cooperative action, and a negative rating otherwise. Imperfect monitoring may also be assumed by letting the relationship between the seller's true

possibility of receiving publicly a bad rating prevents the seller from acting opportunistically, at least when the future losses accruing from bad reputation are as high as the gains from deviating.

[50] Evidently there are many features of a rating system that need to be designed, such as the scale and the format, of the possible ratings, however here we will concentrate uniquely on the problem of how many of the past ratings to compound into a seller's score.

[51] Buyers' valuations of the good are i.i.d. within and across periods.

actions and the received rating be noisy, and thus a seller cannot credibly precommit to cooperation.

5.1. Online reputation mechanisms and limited records. The remainder of this section addresses the question of whether, in designing online reputation mechanisms, limited records are preferable to complete records, from an efficiency perspective. Reputation mechanisms that aggregate all of a seller's past ratings into his public score are said to make use of complete records. Mechanisms in this class are also known as unweighted aggregation mechanisms, for they weight all past ratings equally. Instead, a mechanism is said to make use of limited records when it employs only the ratings received during a predefined window of time, and, for this reason, it can also be referred to as window aggregation mechanism[52].

It can be easily proven that the introduction of a rating system in the repeated auction setup described above is unambiguously welfare-enhancing[53]. In its absence, in fact, the auctions reduce to effectively disconnected one-shot interactions, during which the seller takes the opportunistic action[54] by assumption. A rating system, instead, creates scope for cooperation to take place in equilibrium, by providing the sellers with a credible device to commit to exerting the non-opportunistic action. Via the rating system, in fact, the seller's current behavior becomes tightly linked to his future payoffs, and, in turn, this link is often sufficient to deter him from taking the opportunistic action.

As regards how much of a seller's past ratings to aggregate in his public score, the simplest option is to use all of the available history. Examples of such unweighted aggregation mechanisms are the accumulative score mechanism, that produces a score by summing up all previous ratings, or the average score mechanism, according to which a score is computed as the average of all ratings.

In order to analyze the effect of unweighted mechanisms on efficiency and trade, we begin with looking at a simplified interaction between the seller and the buyers, where the strategic behavior of the latter players is replaced with a heuristic bidding behavior. This simplification allows us to reduce the original strategic interaction to a decision problem for the seller (non-equilibrium approach). In particular, the seller's

[52] Window aggregation mechanisms, that give the same positive weight to all ratings received within a time window and zero weight to all the others, are a subset of the weighted aggregation mechanisms, that includes all the mechanisms that assign a decay weight to past information (as, for instance, the exponential smoothing mechanism studied in Fan et al., 2005).

[53] When a scoring system is introduced, the seller takes the opportunistic action much less frequently or even does not take it at all.

[54] The seller lacks any means to commit to the cooperation strategy, because future short-lived players will not be able to reward (or punish) the seller for his current cooperation.

problem consists of choosing an action that maximizes a *premium function*, describing his expected payment as a function of his current actions and his score. The premium function is meant to represent the non-strategic buyers' behavior. Such a non-equilibrium approach has been taken in this literature on the grounds that equilibrium concepts require a degree of coordination among buyers that is incompatible the anonymity in the market, at least in the short-run (Aperjis, Johari, 2010b).

Analyzing this simplified setup, Aperjis and Johari (2010a) demonstrate that when the premium function[55] is concave in reputation, unweighted aggregation mechanisms are not capable of providing the seller with strong enough incentives to undertake the cooperative action in every period[56]. This result follows from the assumption of concavity of the premium function, which implies that the marginal effect of each additional positive rating on the seller's future profits becomes smaller as the total number of positive ratings increases. As a seller's score increases, in fact, so does his reputation for not taking the opportunistic action, as well as the price he expects to receive from future transactions. Due to the concavity assumption, however, the returns on reputation are decreasing, and so is the opportunity cost of cheating, which corresponds to the foregone increase in expected price that would follow from an increase in the score. In other words, once a seller has accumulated a history of good ratings, the decrease in profit accruing from receiving a bad rating becomes smaller, while the payoff from cheating remains the same. Eventually, this imbalance will induce the seller to take the opportunistic action.

Thus, if the goal of the market designer is to induce the seller to take the cooperative action in every period, unweighted aggregation mechanisms are to be discarded. Aperjis and Johari (2010a, b) suggest instead that this objective could be achieved by the window aggregation mechanism. When limited records are implemented, in fact, it turns out that it is possible to find a window T such that the seller never takes the opportunistic action. More precisely, such a window can be found by solving the seller's maximization problem (in every period) under the constraint that cooperating delivers a higher value than deviating.

Interestingly, it can be demonstrated that if the seller has a discount factor δ and provides on average items of quality q, the optimal window is increasing in δ and decreasing in q. Patient sellers (high δ), in fact, are by definition quite concerned about their future payoff. Accordingly, the threat of the permanence of a bad rating in

---

[55] And assuming perfect monitoring, in the sense that the buyers can correctly infer the seller's action when they consummate the good after purchase.

[56] Ancillary assumptions are mild, such as that the payment to a seller with a maximum reputation is bounded away from zero. This result extends the previous work of Fan et al. (2005) to a larger class of premium functions.

their score for a large number of periods (and the associated payoff losses), as prescribed by a long retention window, is effective in disciplining their behavior today. On the other hand, a long window would only reinforce the incentive to cheat for a seller that has already accumulated a good reputation. This is especially true for the sellers that on average deliver higher quality q. These sellers are likely to obtain high scores rather easily, and they may find it optimal to deviate every once in a while, expecting to receive additional good ratings in the future. When this is the case, only a short window can induce a punishment severe enough to deter the seller from deviating[57].

Similar results to those documented by taking the non-equilibrium approach also obtain when the repeated auction setup is analyzed as a proper game between the seller and the (current and future) buyers. Dellarocas (2005) studies such a case, assuming that the seller, who, as usual, has an item for sale every period, has to decide, unobserved by the buyers, whether to exert costly high effort or shirk (low effort) at zero cost. His effort choice affects the value of the item as it will be perceived by the buyer. Buyers are assumed to derive an expected positive value from consuming the good only when the seller cooperates. Accordingly, if the seller is expected to shirk, no buyer would want to buy the good. The buyers live for one period only, during which they participate in a Vickrey action. Without a rating system, the setup above features a unique Nash equilibrium, where the seller does not exert costly effort and the buyers do not buy.

To remedy such a market failure, a rating system is introduced. Winning bidders share their satisfaction with the purchased item by submitting a rating[58] to the central system, which collects it and makes it available to future buyers. Dellarocas allows for imperfect monitoring, assuming that the buyers' ratings are related to the seller's true actions only probabilistically: both high and low effort may result in the rating reporting high and low item-value, although with different probabilities[59]. Despite imperfect monitoring, with a rating system, it is possible to find an equilibrium of the repeated game there is trade. The maximum efficiency[60] that can be achieved, however, is strictly lower than what could be attained with perfect monitoring. This

---

[57] It is useful to restate the results above in terms of intensity and duration of the effect of reputation mechanisms. A mechanism that chooses a larger aggregation window brings about a stronger duration effect, because a bad rating will affect the seller's future payoffs over a relatively long horizon. A mechanism that instead aggregates ratings over a shorter window has a strong intensity effect, by increasing the incidence of the most recent rating on the incoming future payoffs.

[58] Without loss of generality, ratings can be assumed to be binary, reporting a positive or a negative value.

[59] Although buyers are "making mistakes" in their rating decisions, it is assumed that they do not act strategically in doing this.

[60] In this setup, both the seller's profit and the buyer surplus are proportional to the level of seller's cooperation. Accordingly, efficiency is maximized when the seller (lifetime discounted) payoff are.

second-best result follows from the possibility of fully cooperating sellers receiving bad ratings, and, in fact, the wedge in efficiency between the two scenarios is proportional to the amount of noise in detecting the seller's cooperative action.

It can be demonstrated that the upper bound on the seller's payoffs can be attained by an unweighted mechanism, although under these systems the market-interaction eventually unfolds in no-trade. To see why, consider implementing the complete records arrangement by a two-score system, according to which the seller's public score can be either "good" or "bad". The transition between the two scores depends on the received ratings and on a "forgetting" parameter, chosen by the market designer. This is a probability with which negative ratings are ignored, in an effort to counteract the distortions accruing from imperfect monitoring. As long as the seller receives positive ratings, and the negative ratings are forgotten, the system keeps assigning him the good score. When the seller receives a negative rating that is not forgotten, the system assigns him the bad score. Dellarocas demonstrates that it is possible to choose the forgetting parameter in order to construct an equilibrium[61] where the seller always cooperates while he has the good score, and never does when his score is bad. Buyers, accordingly, always buy the item when the seller's score is good, and never buy it otherwise. In this equilibrium, the rating system is effectively implementing a trigger strategy which fosters cooperation by threatening the seller with exclusion from the market. Although the seller's payoff is maximized by this equilibrium strategy, it is also true that due to imperfect monitoring cooperating sellers will eventually be excluded from the transactions.

Thus, although complete records are efficient, Dellarocas argues that a rating system that allows the seller not to be expelled from the market may be preferable, and he shows that continued trade can be implemented by a mechanism employing limited records. More precisely, a rating system that shows the sum of ratings received over a window of a (publicly known) finite length N induces an equilibrium similar to that arising with complete records. The seller always cooperates whenever he has no negative ratings, and, during these periods, he receives the second-highest payment[62] the buyers are willing to bid. As he accumulates negative ratings, the seller begins to mix between cooperating and cheating, and the buyers, that anticipate correctly this strategy, bid lower and lower prices.

[61] A strategy for the long-run player is public if, at each period, it depends only on the public information and not on the private information. A perfect public equilibrium is a collection of public strategies such that at every period and for every public history the strategy is a Nash equilibrium from that date on.

[62] Recall that a second-price auction is assumed.

This limited-record mechanism can be shown to achieve the same seller's payoff, and efficiency, as a mechanism that publishes complete records. However, while this latter system punishes the seller, as soon as he transits to the bad score, with the exclusion from the market, the current mechanism smooths the necessary punishment over the N periods during which the negative rating transits through the seller's records (before being forgotten). During each of these periods trades occurs, and the seller's punishment takes the form of lower payoffs, that accrue from the fact that the buyers anticipate his mixing strategy.

Having shown that limited-records mechanisms can induce the same efficiency as those employing complete-records, and, as opposed to these, never lead to no-trade, Dellarocas also provides more specific results regarding the length of the system's memory. He notes that, in equilibrium, when the seller accumulates the first negative rating, buyers expect him to mix between cooperating and cheating, and accordingly they bid a lower price than what they were bidding when the seller had no negative ratings (and was expected to cooperate fully). Further trade can happen only if the decrease in the buyers' expectation of high effort, that is, if the decrease in their bidden prices is still enough for the seller to prefer to exert high effort instead of cheating. As negative ratings start accumulating on a seller's profile, the buyers expect him to cooperate with lower and lower probabilities, and bid lower and lower prices, making it increasingly less profitable for the seller to actually exert high effort. Eventually, when the buyers' expectations are low enough that the seller is indifferent between cooperating and cheating, trade does not occur any longer.

Consistently with this reasoning, the longer N, the more difficult is to sustain trade for all possible amount of negative ratings that the seller may have. In particular, as N grows large, the higher the initial buyers' valuations of the good should be to induce a seller to cooperate at all negative ratings he may have. For this reason, mechanisms that are endowed with long windows sustain continued trade in a smaller set of situations (those characterized by borrowers with high initial valuations) than mechanisms that publish only a limited amount of past ratings. Moreover, since in this setup the maximum attainable payoff is independent of N, it is also possible to state that the most efficient, and most widely applicable, reputation mechanism reports only the outcome of the seller's most recent transaction.

It is worth relating Dellarocas' work with that of Ekmekci (2011) presented in section 3. Both Ekmekci and Dellarocas study the efficiency of rating systems in a repeated game with moral hazard, although the former allows for the possibility of the seller being of a commitment type that always cooperate, while Dellarocas studies a pure moral hazard setup. However, Ekmekci also considers the case of complete information, and argues that limited records cannot lead to a higher efficiency than complete records (Ekmecki, 2011). This result is consistent with what Dellarocas finds, because, in Dellarocas, a rating system that publishes only the last period rating is *as* efficient as one that uses complete records, and not more efficient than that. Besides this connection, it is not trivial to compare the two papers, for they rely on very different setups.

Summarizing the results surveyed here, it seems that compounding the whole history of a seller's behavior into his public score may not be the most efficient solution. Limited records, instead, have been found to allow to design rating systems that elicit the highest achievable efficiency.

6. Conclusions

In this survey, we have discussed how limited records can be designed to increase the efficiency in repeated interactions characterized by both adverse selection and moral hazard. In doing so, we have been motivated by the intuition that limited records may simultaneously strengthen reputation concerns and fulfill privacy needs. Data retention is, in fact, one of the regulatory principles put forward by modern-day legal systems to safeguard individual privacy.

Throughout the paper, we have been dealing with a number of different informational arrangements, all of which prevent agents from accessing in full the available information regarding the past behavior of a (potential) transacting partner. This heterogeneity arises in part because limited records have as of yet not been studied in a unified way, and in part because of the need to capture relevant characteristics. Table 1 summarizes the various interpretations given.

Table 1. Informational arrangements considered in the survey

Paper	Informational arrangement
Liu, Skrzypacz, 2011	exogenously limited records of player's actions
Liu, 2010	costly information acquisition of player's actions
Ekmekci, 2011	rating from a finite set
Aperjis and Johari, 2010b	weighted aggregation mechanisms
Dellarocas, 2005	summary statistic most recent K ratings
Elul and Gottardi, 2008	history of outcomes, with forgetting
Vercammen, 1995	history of outcomes, in the last T periods

All of the above definitions share two important features. First, players are restricted to access either a subset of the existing information (Aperjis and Johari, 2010b; Dellarocas, 2005; Liu, 2010; Liu, Skrzypacz, 2011; Vercammen, 1995), or a non-transparent version of it (Ekmekci, 2011; Elul and Gottardi, 2008). Secondly, records contain only information on the past behavior of the player from whom we want to elicit reputations concerns.

Although the works surveyed here are evidently concerned with the same topic, it seems to us that their results rely somehow on *ad hoc* assumptions. For this reason, it is difficult to see how the existing works can provide general guidelines regarding how to implement limited records in an arbitrary market. Moreover, there are several important aspects regarding the design of informational arrangements that are yet to be addressed. Among them, is the identification of the trade-off between the costs and benefits associated with choosing records of different lengths.

A solid theoretical framework is especially needed today since the debate on the effectiveness of data retention is back on the political agenda[63] due to the substantial increase in data collection, and storing possibilities, that the technological development allows. Although all democratic parties express a concern for privacy, the operationalization of data retention remains highly disputed. While the political debate takes shape, the general public appears increasingly worried about the problem, to the point that it has taken its own measures to protect personal data. The expansion of the Internet as a marketplace and the diffusion of Web 2.0 websites have directly confronted a vast number of individuals with the problem of how to protect their privacy, and solutions ranging from the removal of cookies to non-traceable credit cards[64] have arisen spontaneously.

We believe that this growing concern for privacy and the empirical motivations that were discussed in the Introduction are delicate issues that deserve to be dealt within a unified framework. A theory of limited records would benefit a wide variety of markets, most notably those in which privacy needs are indispensable. From the existing works, in fact, we know that limited records can sustain welfare-increasing reputation effects. What we do not know, however, is how to practically design the many features of limited records, among which, most importantly, their length.

[63] For instance, in May 2009 the EU internal Market and Services Directorate General (DG) released its “Report on credit histories”, discussing the harmonization of credit reporting systems across Europe. More recently, in May 2010 the U.S. Representatives Rick Boucher and Cliff Stearns announced a draft legislation tackling online privacy issues, including data retention. For more information, see http://www.boucher.house.gov/images/stories/Privacy_Draft_5-10.pdf.

[64] PrivateBuy.com(TM), for instance, is a service that provides devices to allow consumers to buy on the Web without revealing personal or credit card information to the sellers

References

Acquisti, A. , 2010. The economics of personal data and the economics of privacy. OECD Committee for Information, Computer and communication policy, DSTI/ ICCP /IE/ REG (2010)4.

Acquisti, A. ,Varian, H. R. , 2005. Conditioning Prices on Purchase History. Marketing Sci. : 24, 1-15.

Aperjis, C. , Johari, R. , 2010a. Designing reputation mechanisms for efficient trade. *Mimeo.*

Aperjis, C. , Johari, R. , 2010b. Optimal windows for aggregating ratings in electronic marketplaces. Manag. Sci. : 56, 864 -880.

Armour, J. , Cumming, D. , 2005. Bankruptcy law and entrepreneurship. Amer. Law Econ. Rev. :10, 303-350.

Bar-Isaac, H. , Tadelis, S. , 2008. Seller reputation. Foundations and Trends in Microeconomics: 4, 273-351.

Bakos, Y. , Dellarocas, C. , 2003. Cooperation without enforcement? A comparative analysis of litigation and online reputation as quality assurance mechanisms. MIT Sloan Working Paper No. 4295-03.

Barlo, M. , Carmona, G. , Sabourian, H. , 2008. Repeated games with one-memory. *Manuscript.*

Bhaskar, V. , 1997. Information constraints and the overlapping generations model: Folk and anti-Folk theorem. Rev. Econ. Stud. : 65, 135-149.

Black's Law Dictionary, 1999. B. A. Garner, ED. ,7th ed.. West Group.

Bos, M. , Nakamura, L. , 2010. Loan applications and credit scores: evidence from legally mandated credit remark removal. *Manuscript.*

Cabral, L. , Hortacsu, A. , 2010. The dynamics of seller reputation: theory and evidence from eBay. J. Ind. Econ. : 58, 54-78.

Calzolari, G. , Pavan, A. , 2006. On the optimality of privacy in sequential contracting. J. Econ. Theory :130, 168-204.

Chatterjee, S. , Corbae, D. , Rios-Rull, V. , 2010. A theory of credit scoring and competitive pricing of default risk. *Mimeo.*

Cohen-Cole, E. , Duygan-Bump, D. , Montoriol-Garriga, J. , 2009. Forgive and Forget: Who Gets Credit after Bankruptcy and Why?. W. P. Federal Reserve Bank of Boston.

Cole, H. , Kocherlakota, N. , 2005. Finite memory and imperfect monitoring. Games Econ. Behav. : 53, 59-72.

Cripps, M. W. , Mailath, G. J. , Samuelson, L. , 2004. Imperfect monitoring and impermanent reputation. Econometrica: 72, 407-432.

Dellarocas, C. , 2005. Reputation mechanism design in online trading environments with pure moral hazard. Inform. Syst. Res. : 16, 209 -230.

Dellarocas, C. , 2006. Reputation Mechanisms. In the Handbook on Economics and Information Systems, edited by T. Hendershott. Elsevier Publishing.

Diamond, D. W. , 1989. Reputation acquisition in debt markets. J. Polit. Economy : 97, 828-862.

Doraszelski, U. , Escobar, F. ,2011. Restricted feedback in long term relationships. *Manuscript.*

Elul , R. , Gottardi , P. , 2008. Bankruptcy: is it enough to forgive or must we also forget? Federal Reserve Bank of Philadelphia W.P. 07-10

Ekmekci, M. , 2011. Sustainable reputations with rating systems. J. Econ. Theory : 146, 479-503.

Fan, M. , Tan, Y. , Whinston, A. B. , 2005. Evaluation and design of online cooperative feedback mechanisms for reputation management. IEEE Transactions on knowledge and data engineering: 17, 244- 254.

Galindo, A. , Miller, M. , 2001. Can credit registries reduce credit constraints? Empirical evidence on the role of credit registries in firm investment decisions. Annual Meetings of the Board of Governors, Inter-American Development Bank and Inter-American Investment Corporation.

Hermalin, B. E. , Katz, M. L. , 2006. Privacy, property rights and efficiency: The economics of privacy as secrecy. Quant. Market. Econ. : 4, 209-239.

Hirshleifer, J. , 1971. The private and social value of information and the reward to inventive activity. Amer. Econ. Rev. : 61, 561-574.

Jentzsch, N. , 2001. The economics and regulation of financial privacy - A comparative analysis of the United States and Europe. W.P. 128, Freie Universität Berlin, John F. Kennedy Institute for North American Studies.

Lemaire, J. , 1995. Bonus-Malus Systems in Automobile Insurance. Boston: Kluwer.

Liu, Q. , Skrzypacz, A. , 2011. Limited records and reputation. *Mimeo.*

Liu, Q. , 2010. Information Acquisition and Reputation Dynamics. forthcoming, Rev. Econ. Stud..

Mailath, G. J. , Samuelson, L. , 2006. Repeated games and reputations. Oxford University Press.

Musto, D. K. , 2004. What happens when information leaves a market? Evidence from postbankruptcy consumers. J. Bus. : 77, 725-748.

Nock, S. L. , 1993. The costs of privacy: surveillance and reputation in America. New York: Aldine.

Posner, R. A. , 1983. Privacy and related interests. In: The economics of Justice. Harvard University Press.

Renault, J. , Scarsini, M. , Tomala, T. , 2006. A minority game with bounded recall. *Mimeo.*

Resnick, P. , Zeckhauser, R. , Swanson, J. , Lockwood, K. , The value of reputation on eBay: a controlled experiment. Exper. Econ. : 9, 79-101.

Taylor, C. , 2004a. Privacy and information acquisition in competitive markets. W.P. 13271, Olin Program in Law & Economics, UC Berkeley.

Taylor, C. , 2004b. Consumer privacy and the market for customer information. RAND: 35, 631-650.

Özdogan, A. , 2010. Sequential disappearance of reputations in two-sided incomplete-information games. *Mimeo.*

Vercammen, J. , 2002. Welfare improving adverse selection in the credit markets. Int. Econ. Rev.: 43, 1017-1033.

Vercammen, J. , 1995. Credit bureau policy and sustainable reputation effects in credit markets. Economica: 62, 461-478.

Wilson, A. , 2002. Bounded memory and biases in information processing. NAJ Economics: 5.

PAPER 4

Limited credit records and market outcomes

Margherita Bottero

Abstract. In most modern credit markets, credit registers store and share information regarding borrowers' past behavior. Such information is often removed after a finite number of years as mandated by privacy protection laws. While these laws are motivated by privacy concerns and by the desire to provide a fresh start to unlucky borrowers, they are difficult to justify from an economic perspective, since they effectively prescribe the removal of possibly relevant public information. This paper tries to assess the impact of different lengths of credit records on borrowers' behavior and market outcomes. To do so, I study a stylized credit market for small loans, and demonstrate that limited records endogenously give rise to beneficial reputation effects, which can alleviate the distortions caused by asymmetric information. More precisely, I show that when moral hazard is high, limited records can achieve higher welfare and lead to a lower default rate than records that show all, or nothing, of the past history.

1. Introduction

In most modern credit markets, credit registers collect and share data on borrowers' past behavior with other market participants, in an effort to reduce the well-known informational asymmetries that characterize lending relationships (Jappelli, Pagano, 2002). These asymmetries take the form of adverse selection (Akerlof, 1970) or moral hazard. The former arises before the loan is contracted, because lenders can at best only partially observe a borrower's ability and other relevant characteristics. Moral hazard instead occurs after the loan is contracted and stems from the imperfect observability of the borrower's actions. Consequently, reliable information on a borrower's past behavior serves both as a screening device for lenders, and as a "reputational" collateral for borrowers (Vogel, Burkett, 1986), which they can use to signal to potential lenders their private information and intended choices.

I would like to thank Axel Bernergård, David Domeij, Tore Ellingsen, Tobias Laun, Lars Ljungqvist, Ettore Panetti, Giancarlo Spagnolo, Mark Voorneveld, Jörgen Weibull and seminar participants at the Stockholm School of Economics and at the Bank of Italy for their helpful comments and suggestions. This paper is part of a joint project with Giancarlo Spagnolo on limited records and credit markets. I gratefully acknowledge the Knut and Alice Wallenberg Research Foundation for financial support.

Interestingly, from a legal point of view credit data fall under the umbrella of personal data and are thus subject to the privacy protection provisions that regulate the extent to which personal information can be handled, shared and stored. In particular, an important by-product of privacy protection is the principle of data retention[1], which prescribes that personal data, as collected by any data user, should be retained for a limited period of time only, and by any means for no longer than what is necessary for the intended purpose. Although such a principle may look intuitively appealing, it is not immediately clear whether it is desirable from a strictly economic standpoint. In fact, data retention directly mandates the removal from the market of the very data that have a key role in disciplining the market asymmetries as argued above. Indeed, the issue has stirred an intense and so far unresolved debate[2] about the implications of limited records for privacy, efficiency and reputation (Bottero, Spagnolo, 2011; Sartor, 2006).

The inconclusiveness of the academic and political debate is reflected in the choices made by policy makers. As of 2007, about 90 percent of the countries in which a credit register was operating[3] mandated the removal of (at least some) credit data after a limited period of time (Elul, Gottardi, 2008). Although suggestive of a certain degree of homogeneity across countries, this is only superficial. In fact, there is no consensus on how long the retention period of credit registers' records should be. Table 1 provides a snapshot of a few of the current legislations that apply to bankruptcy information[4].

---

[1] Data retention was first mentioned by the Council of Europe in 1973, when they declared that "[T]hose responsible of these [personal data] files [...] should refrain from storing information which is not necessary for the given purposes". This principle was later on incorporated into the 1995 EU Directive (95/46/EC) on Data Protection, which states that "[M]ember States shall provide that personal data must be [...] adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed".

[2] According to the anti-privacy position, privacy gives the possibility to manipulate one's reputation and to impede the autonomous function of the impersonal market mechanism that allocates credit at better rates to borrowers with better reputations (Posner, 1983). From Posner's blog (May 8^{th} 2005): "All that privacy means in the informational context [...] is that people want to control what is known about them by other people. They want to conceal facts about themselves that are embarrassing or discreditable [...] Such concealment is a species of fraud". Yet, other scholars argue that privacy may be needed when agents are boundedly rational, or even in cases of market failures with rational agents (Taylor 2004a, 2004b). With the former concern in mind, Jeffrey Rosen (2000) writes that "[P]rivacy protects us from being misdefined and judged out of context in a world of short attention spans, a world in which information can be easily confused with knowledge". For a reference to the debate in the context of the credit market, the acts of the Congressional debate on the adoption of the Fair Credit Reporting Act, U.S. Senate 1969 and U.S. House 1970.

[3] The total number of countries in which a register was operating as of 2007 was 113.

[4] For a description of credit registers in Europe, see the Report of Expert Group on Credit Histories (2009). For a detailed comparison between the United Stated and Europe, see Jentzsch (2001). For more information about South America, see Galindo and Miller (2001).

As can be seen, countries have implemented the principle of data retention in a strikingly haphazard way, mandating data removal from after one year (Belgium) up to after fifteen years (Greece). Indeed, even geographically and culturally neighboring countries such as Germany and Austria prescribe surprisingly different retention periods.

Table 1: Withholding periods for bankruptcy data (in years)[5]

Country	Years	Country	Years
United States	10	Finland	5
Brazil	5	Greece	15
Chile	3	Hungary	-
Argentina	2	Italy	3*
Austria	7	Netherlands	-
Denmark	5	Portugal	10*
Bulgaria	5	Germany	3
Sweden	3	Slovenia	5
Czech Republic	-	UK	6

To complicate matters further, national laws mandate different retention periods depending on the type of data (fraudulent bankruptcies, credit remarks, etcetera) as well as on the type of person (entrepreneurs may be treated differently from non entrepreneurs). In recent years, this heterogeneity has become particularly problematic due to the sharp increase in cross-border data sharing. In the Euro-zone, several organizations have already highlighted "their concerns about the current patchwork of different data retention periods and expressed their wish for a pan-European data retention period, in order to boost legal security"[6].

This paper intends to develop a formal analysis of the effects of data retention provisions on borrowers' behavior and market outcomes. To this end, I study a stylized model of a credit market for consumers and small businesses where a credit register decides on the amount of information to be disclosed to the lenders. I directly compare three informational arrangements, which I call full withholding, full disclosure and limited records. Under full withholding, privacy is enforced and no information about the borrowers is disclosed. Under full disclosure, all the relevant information is made available to the lenders. Finally, with limited records, the credit register implements a

[5] Note: Data are taken from Matuszyk, Thomas, (2008), & Report of the Expert Group on Credit Histories (EGCH), May 2009. The entries refer to private credit bureaus, except those marked with an asterisk that refer to a public register.

[6] Summary of replies to the public consultation about the future legal framework for protecting personal data, 2009. See also the Communication COM (2010) 609 Final, 2010 http://ec.europa.eu/justice/news/consulting_public/0006/com_2010_609_en.pdf.

data retention policy, by storing and sharing the outcome of the N most recent projects undertaken by a borrower.

It turns out that it is possible to characterize sufficient conditions under which each of the three arrangements leads to a lower (ex ante) default rate than the other two. Interestingly, when the moral hazard in the market is high, limited records outperform the two other arrangements, supporting the case for privacy protection legislations. In a situation of high moral hazard, in fact, the two extreme policies of full disclosure and withholding hamper the performance of low-ability borrowers while not eliciting a sufficiently higher repayment rate from the high-ability group. Limited records, instead, provide borrowers of all types with the possibility to alleviate, at least in part, the moral hazard problem by equipping them with a credible device, the records, to build or restore a reputation for being a trustworthy borrower. By moderating the moral hazard problem, this last arrangement leads to the lowest default rate.

Further, it turns out that limited records may also lead to higher aggregate welfare, albeit under fairly restrictive conditions. More precisely, when moral hazard is high and borrowers are not too different in their abilities, limited records are more efficient because the resulting interest rates succeed in increasing borrowers' aggregate welfare, while keeping lenders' welfare constant. Yet, under the model's assumptions, if these conditions are not met, it is preferable to disclose all information about the borrowers.

The main result of my analysis is that data-protection laws endogenously create scope for borrowers to link their current choices of effort to better contractual terms in the future, by building a reputation for being of high ability and thus securing themselves contracts at lower interest rates. Reputation concerns elicit higher effort, which under the assumptions of the model is beneficial for the overall economy because it prompts higher repayment rates and lower interest rates. Importantly, the removal of past data also circumvents the problem of disappearing reputation effects (Cripps et al. 2004) by continuously replenishing a borrower's incentives to sustain his reputation (for a game-theoretic discussion, Ekmekci, 2010; Liu, Skrzypacz, 2011).

The present work is connected to the literature on reputation and data retention. For what concerns the credit markets, the effects of credit registers and borrowers' reputation were first addressed by Diamond (1989), who documented positive and persistent effects of reputation on the equilibrium interest rate and the quality of the pool of borrowers. However, Diamond does not address the effects of data retention policies. Rather, the model I put forward builds on Vercammen's (1995) and (2002) companion papers. In his 2002 paper, Vercammen shows in a game theoretic setup that adding adverse selection to a market already characterized by moral hazard may be welfare improving. In the other paper, he demonstrates that transmitting limited records on borrowers' past behavior may give rise to a sequential equilibrium in which borrowers are concerned about their reputation. My model generalizes and extends Vercammen's results. First, I apply his framework to a market setup, in which I can study and compare, within the same model, the policies of full disclosure (moral hazard), full withholding (moral hazard and adverse selection) and of limited records. By explicitly studying the equilibrium that arises under limited records, I investigate Vercammen's suggestion that such a policy may be welfare-enhancing[7]. Further, I relate my findings on expected default rates to Vercammen's results on welfare.

The effects of data retention provisions are also studied by Elul and Gottardi (2008). These authors provide a game theoretic analysis of the effects of forgetting some of the past information on (long-lived) borrowers' past behavior in a credit market where such information is transmitted to (short-lived) lenders by a credit register. Besides a few minor differences[8], our models focus on the same question and they both assume that the information shared by the register is a binary variable that reports if the previous loans have been repaid in full or defaulted upon. Elul and Gottardi, however, model data retention by assuming that the credit register discloses the whole history of borrowers' past behavior, but with a positive probability it "forgets" a default and records a repayments instead. The equilibrium they study, then, revolves around the lenders' posterior belief on the type of the borrowers, which is formed and updated by looking at these modified records. In my model, instead, the information is always correctly registered and fully censored after N periods. Although the authors argue at length that the two formulations lead to similar results, mine allows for a more straightforward discussion on the optimal length of the register's memory, and in particular it allows me to identify the trade-offs that have to be solved by the optimal memory design. My model is also more suitable for accommodating relevant extensions such as whether, and how, borrowers strategically respond to limited records. The two models, however, address complementary aspects of the issue at hand. More precisely, Elul and Gottardi look at the conditions under which a policy of limited records, as opposed to full information, improves the access to credit as well as welfare, while I seek to characterize the conditions under which different informational arrangements, among which limited records, improve on default rates and welfare. Moreover, my model is more closely connected with the literature that studies how limited records can provide permanent reputation effects in repeated interactions, and how this could possibly vary with the relative importance of adverse selection and moral hazard in the market.

[7] Vercammen (1995) only provides a numerical example.

[8] In my model both types of borrowers can fail, while in Elul and Gottardi's model only low-ability borrowers can fail. The authors do however consider the possibility that both types fail in a numerical example.

Chatterjee et al. (2010) present a model of consumer behavior that aims at explaining the fact that defaulted households typically face limited access to credit, and at a higher cost, and that these constraints are lifted with the passage of time. The authors demonstrate that such a pattern can arise in a recursive equilibrium due to borrowers' stochastic income process and lenders' limited information about the borrowers. In the last section of their paper, they also address the issue of data retention with a numerical example calibrated for the current legislation in the United States[9]. The results of this exercise show that data removal leads to lower welfare than a policy of no removal. Differently from the present paper, however, Chatterjee et al. focus on households and consumers rather than on small entrepreneurs. Furthermore, in my paper data removal is the primary object of study.

While the present work is theoretical, it can be related to the empirical literature that tries to estimate the effect of data retention in the credit market, that originated with Musto's (2004) seminal contribution. Musto demonstrates that data retention policies seem to take the market away from the efficient equilibrium and he criticizes data retention provisions for mandating "efficiency in reverse" (p. 725). His findings have, however, been challenged in several subsequent papers, that instead document a positive effect of data protection laws (Armour, Cumming, 2005; Bos, Nakamura, 2010; Cohen-Cole et al., 2009). I will discuss the relationship between my model and the empirical literature in Section 8.

The remainder of the paper is organized as follows. Section 2 describes the model. Section 3 describes the equilibrium under full disclosure and full withholding, and section 4 characterizes the equilibrium with limited records. Section 5 goes through the effects of different choices of memory on borrowers' behavior. Section 6 and 7 present some comparative statics for the three informational arrangements, in terms of the ex ante default rate and aggregate welfare. Section 8 discusses the findings and section 9 concludes.

2. The model

The present model builds on the work of Vercammen (1995, 2002) and provides a stylized representation of an unsecured credit market for small businesses, where borrowers may switch between lenders and where the relevant information on a borrower's past behavior is transmitted via a credit register. The market features adverse selection, because borrowers' ability is private information, and moral hazard, because the effort choice is a hidden action taken after the contract has been negotiated. Time is

[9] The Fair Credit Reporting Act.

infinite and discrete, and at the beginning of each period a competitive credit market opens, where borrowers and lenders interact as described below.

2.1. The borrowers. There is a large number of infinitely-lived, risk-neutral borrowers, who discount the future at rate $\beta < 1$. Borrowers are credit-worthy and identical in everything but in a privately known, fixed[10] ability parameter θ that can take two values, high, H, or low, L, and that affects the disutility of effort. More precisely, in supplying effort $e \in E = [0, 1]$, a borrower pays $c(e, \theta)$, but high-ability borrowers face a lower marginal cost, $c'_e(e, H) < c'_e(e, L)$. Otherwise, the cost function is continuous, twice differentiable and strictly convex, $c'_e > 0$, $c''_{ee} > 0$, and zero effort entails no cost for either type, $c(0, \theta) = 0$. High-ability borrowers are present in a fraction $p \in (0, 1)$ of the population and the remainder is of low ability.

At the beginning of each period, a borrower seeks an indivisible loan and combines it with his own effort in a project that generates non-storable revenues $Y(e, \omega)$ that are used for consumption. Revenues are also a function of a non-observable positive random shock ω, i.i.d. across borrowers and periods, drawn from a distribution F defined on $\Omega = [0, 1]$. The revenue function is continuous, twice differentiable and concave, so that revenues increase in effort, $Y'_e > 0$, and in the realization of the shock, $Y'_\omega > 0$. Further, the marginal productivity of effort decreases in effort and increases in the realization of the shock, $Y''_{ee} < 0$ and $Y''_{e\omega} > 0$. To simplify the analysis, it is also assumed that $Y''_{\omega\omega} > 0$, implying that Y is a strictly concave function[11]. Finally, it is assumed that zero effort and the lowest shock realization both result in zero revenues, $Y(e, 0) = Y(0, \omega) = 0$, regardless of the realized shock and the exerted effort.

A borrower's expected utility is given by expected revenues net of the repayment of the loan and the promised interests, $1 + r$, minus the cost of effort. When the realized revenues are not enough to repay the contracted amount, the lenders appropriate anything that has been generated, and the borrower receives zero utility. To write the expected utility of a borrower, it turns out to be useful to follow Vercammen (1995) and define the "default threshold", which is the realization of the shock such that a borrower that has exerted effort e and has agreed to repay interest rate r generates just enough revenues to honor his debt, that is, $Y(e, \bar{\omega}(e, r)) = 1 + r$. Totally differentiating $\bar{\omega}(e, r)$, it turns out that given an interest rate, higher effort reduces the probability of default, $\bar{\omega}_e(e, r) < 0$, while, given a level of effort, a higher interest rate increases

[10] As in Sharpe, 1990, ability is fixed over time.

[11] Because in this case the determinant of the Hessian, $\det H = R_{ee}R_{\omega\omega} - R_{e\omega}R_{\omega e}$, is by assumption negative.

the probability of default[12], $\bar{\omega}_r(e,r) > 0$. I can now write a type θ borrower's life-time utility from partaking in the credit market as,

$$U(e_t, r_t, \theta) = \sum_{t=0}^{\infty} \beta^t \left[\int_{\bar{\omega}(e_t, r_t)}^{1} [Y(e_t, \omega_t) - (1 + \tilde{r}_t)]\, dF(\omega) - c(e_t, \theta) \right] \tag{2.1}$$

where $\theta = L, H$ and $\tilde{r}_t$ is the anticipated interest rate. Borrowers have perfect foresight, so the anticipated interest rate is equal to the equilibrium interest rate, $\tilde{r}_t = r^*$. Given that the perfect foresight assumption will be maintained throughout the analysis, from now on I will no longer distinguish between the two. For all θ, the function U is continuous, twice differentiable and strictly concave, because the cost function c is strictly convex. The reservation utility is zero for borrowers of both types.

Under the limited records arrangement, borrowers' expected utility will also depend on their past performance as it appears in the credit report c_N. For each loan obtained by the borrower, the report will show a binary variable specifying whether the loan was successfully repaid, S, or defaulted upon, D. Accordingly, in every period t a type-θ borrower faces the following recursive dynamic programming problem,

$$\begin{aligned} V(c_N, \theta) = \max_e & U(e, c_N, \theta) \\ & + \beta \left[\Pr(c'_N = c_{NS} \mid e, c_N) V(c_{NS}, \theta) + \Pr(c'_N = c_{ND} \mid e, c_N) V(c_{ND}, \theta) \right] \end{aligned} \tag{2.2}$$

where the utility function U is defined in (2.1), and c_{NS} (c_{ND}) is the credit report obtained from c_N by deleting the outcome realized N periods ago, which is the last displayed by the record, and adding a repayment (default) in front of the *most recent* position, and finally $\Pr(c'_N = c_{NS} \mid e, c_N) = 1 - F(\bar{\omega}(e, r(c_N)))$, while $\Pr(c'_N = c_{ND} \mid e, c_N) = F(\bar{\omega}(e, r(c_N)))$. Note that, thanks to the recursive nature of the problem, it is possible to substitute the time subscripts with the apostrophe notation when denoting a variable in the next period. The shock ω does not appear as an argument of the value function because $V(c_N, \theta)$ gives the value of being in state c_N in a certain period, *before* the shock is realized. Tomorrow's credit report, c'_N, is a state under the borrower's control that evolves according to the constraint, which describes the impact of the control e on c'_N. The current-period report, c_N, is a proper state and the borrower's type θ acts as a state chosen by nature and fixed over time.

[12] Precisely, $\omega^*_e(e,r) = -\frac{R_e}{R_w} < 0$ and $\omega^*_r(e,r) = \frac{1}{R_w} > 0$ for $e \in \left[e^0(r), 1\right]$, where $e^0(r)$ is such that $R(e^0(r), 1) \equiv 1 + r$.

2.2. The credit industry. There is a competitive credit industry that makes loans of size one to borrowers. Financial contracts come only in the form of one-period debt contracts, which are fully characterized by the rate of interest r. The opportunity costs of these loans is given by $\rho > 0$. There is limited liability, thus, by accepting the contract, a borrower agrees to repay $1+r$ if the revenues from the project are large enough and otherwise to transfer to the lender all realized revenues. Then, $r \in R = [0, \bar{Y}]$ where $\bar{Y} = Y(1,1)$. A costless monitoring technology guarantees that borrowers cannot embezzle the realized revenues.

Lenders know that borrowers differ in their ability, and share a common prior p_0 on the fraction of high-ability borrowers. When they can, they use the information provided by the credit register to update their prior. Since the lenders' prior is common and since they have access to the same information, their posterior beliefs are common as well.

2.3. The credit register. I assume a long-lived credit register, which knows each borrower's type and records, for all previous loans, a binary variable that indicates whether the loan has been repaid or defaulted upon. The register can choose among three informational policies: full disclosure, full withholding and limited records. In the first case, the register discloses the borrowers' type once and for all, leading to a situation in which adverse selection is eliminated from the market. The policy of full withholding instead consists of the register shutting down and not disclosing any information that the lenders could use to update their beliefs. In this case both moral hazard and adverse selection persist in the market. Finally, when the register chooses to disclose limited records, it makes available to the current-period lenders a credit report c_N, one for each borrower, which displays the outcomes he realized during the N most recent periods. Since the outcome of a project can be either a success, S, when the loan is repaid, or a default, D, we let $C^N = \{S, D\}^N$ be the set of all possible credit reports of length N.

2.4. Time-line of the events and equilibrium notions. Time is discrete and infinite and in each period the events unfold as follows. At the beginning of the period, the credit register discloses the information on borrowers according to the chosen policy. Lenders simultaneously offer contracts to the borrowers, specifying the rate of interest that they ask in exchange for the loan. Borrowers choose which contract to accept and which effort level to exert. Next, the outcome of the project is realized and the loan is either repaid or defaulted. The credit register observes the projects' outcomes and updates the records. Then, a new period begins. Figure 1 displays the relevant time-line.

Under the policies of full disclosure and full withholding, the market equilibrium can be described by the infinite repetition[13] of the (Bayesian) Nash equilibrium, with complete and incomplete information respectively. More precisely this equilibrium, if it exists, will consist of an interest rate such that lenders make zero profit in expectation and an effort choice such that, given the equilibrium interest rate, borrowers maximize their revenues.

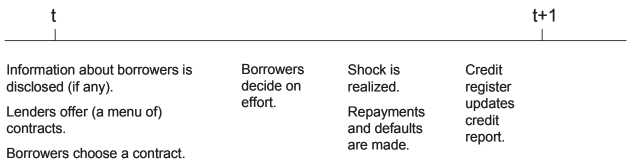


Fig. 1. The time-line of the events.

When the register opts for a limited records policy, it is appropriate to make use of the following rational-expectations, recursive, competitive equilibrium notion, which I call N-period equilibrium.

Definition 1. *A N-period equilibrium consists of a policy function $e^* : C^N \times \{H, L\} \to [0, 1]$, an interest rate function $r^* : C^N \to [0, \bar{Y}]$, a belief function $p^* : C^N \to [0, 1]$ and two distribution functions (G_H, G_L) over the state space C^N such that*

- borrowers behave optimally: given the interest rate function r^* and the belief function p^*, e^* maximizes the borrower's utility

- the credit industry is competitive: for each state c_N, given that the proportion of high type is $p^*_{c_N}$ and given that type-θ borrowers exert $e^*_{\theta c_N}$, the interest rate $r^*_{c_N}$ ensures zero profits

- consistency: If the current distributions of high and low types is given by (G_H, G_L) and borrowers use the policy function e^*, the next period distribution of high and low types is also given by (G_H, G_L). The belief function p^* is consistent with this steady state distribution.

In all three informational arrangements, I will focus only on the steady-state and dispense of the dynamics outside the stationary path.

2.5. Review of the main assumptions. Before proceeding, I briefly review the main assumptions to highlight the role they will play in the model.

- The assumptions of non-storability of revenues and credit-worthiness of both types of borrowers ensures that all borrowers always apply for credit and that

[13] In both information arrangements, since borrowers cannot affect future payoffs with current actions, the equilibrium of the repeated game will boil down to the repetition of the static equilibrium.

credit will not be rationed. The fixed investment assumption is instead key to obtaining a unique equilibrium in the market with full withholding, because lenders cannot price-discriminate by using the size of the loan. Moreover, together with the absence of collaterals, such an assumption makes sure that adverse selection does not result in direct net losses due to distorted investments. Collaterals are not allowed because the model aims at representing an unsecured credit market and because their absence highlights the importance of the information transmitted by the credit register.

- The supply side of the market is assumed to rely only on the information disclosed by the credit register, meaning that lenders cannot accumulate private information on borrowers. Together with the assumption that only one-period, history-independent (non relational) contracts may be offered, this renders the competitive market equivalent to a market populated by short-lived lenders who interact with borrowers only once and then leave the market. The assumption of short-lived lenders is standard in the literature on reputation and credit markets (see Diamond, 1989), for at least two reasons. First, it is adequate to describe the bank-card credit market (as in Musto, 2004). Secondly, it prevents lenders from accumulating private information, which they could compound in "internal ratings" (Jacobson et al., 2006) and use alongside the public credit score to decide the contractual terms.
- The register is assumed to know each borrower's type and history of outcomes. Knowing a borrower's type can be interpreted as knowing those private characteristics that are correlated with borrowers' performance, such as sex, marital status, ethnic background, etcetera. The assumption that only the outcomes of past transactions are recorded is admittedly restrictive. Future work should allow the register to transmit other relevant information, such as total debt exposure or outstanding debt.

The following table summarizes the notation used in the paper,

Table 2. Notation used in the paper

Notation		Definition
$c(e,\theta)$	cost of effort	$c: E \times \{H, L\} \to \mathbb{R}_+$
$Y(e,\omega)$	revenues	$Y: E \times \Omega \to \mathbb{R}_+$
$U(e,r,\theta)$	utility	$U: E \times R \times \{H, L\} \to \mathbb{R}_+$
$\bar{\omega}(e,r)$	default threshold	$\bar{\omega}: E \times R \to [0,1]$
$\Pi(e,r,\theta)$	profits	$\Pi: E \times R \times \{H, L\} \to \mathbb{R}_+$
$e^*_\theta, e^*_{0\theta}, e^*_{\theta c_N}$	equilibrium effort, in full disclosure, full withholding and lim. records	
$r^*_\theta, r^*_0, r^*_{c_N}$	equilibrium int. rate, in full disclosure, full withholding and lim. records	
$p, p_0, p^*_{c_N}$	equilibrium beliefs, in full disclosure, full withholding and lim. records	
c_N	credit report of length N	

Note that first and second derivatives are indicated with one and two apostrophes respectively, and that the subscript(s) indicate the variable(s) with respect to which the derivative is taken. Subscript notation is also used to identify the equilibrium variables under different informational arrangements. This should not create confusion because the latter will always be defined in the text and are marked with an asterisk.

3. The cases of full disclosure and full withholding

In the following section, I describe the equilibria that arise under the policies of full disclosure and full withholding. Both equilibria are characterized by the fact that borrowers cannot influence the credit industry's belief on their type by taking non-equilibrium actions. Under the first arrangement, this follows because information is not transmitted at all, while under the second the lenders already know the borrowers' type for sure. Because of this, different periods are effectively disconnected, so I suppress the time index in this section, with the understanding that all variables refer to period t.

3.1. Full disclosure. The full disclosure case corresponds to a situation in which there is no incomplete information in market. Since the problem is stationary, in every period the borrowers choose effort in order to maximize their period-0 utility U, and the first order condition of a type-θ borrower's problem is given by $U'_e(e,r,\theta) = 0$, or,

$$\int_{\bar{\omega}(e,r)}^{1} Y'_e(e,\omega)dF(\omega) - c'_e(e,\theta) = 0 \tag{3.1}$$

which states that borrowers supply effort until its marginal return equals its marginal cost. Note that the discount factor disappears due to the absence of any intertemporal link, $\beta^0 = 1$. The lower bound of the integral depends on the default threshold $\bar{\omega}(e, r)$, and since $\bar{\omega}'_r > 0$ the marginal value of effort decreases in the promised interest rate r. The second order conditions are satisfied, so that equation (3.1) implicitly defines each type's conditional effort function $e_\theta = e(r, \theta)$. This solution to the maximization of (2.1) is unique and interior[14].

Lenders expect a return $\Pi(e, r, \theta)$ from dealing with a type θ borrower that depends on the borrower's anticipated effort choice e_θ and on the charged interest rate r. Expected returns are given by,

$$\Pi(e, r, \theta) = \int_0^{\bar{\omega}(e^*_\theta, r)} Y(e_\theta, \omega) dF(\omega) + (1 - F(\bar{\omega}(e_\theta, r)))(1 + r) \tag{3.2}$$

and consist of the sum of what the lender will appropriate in the case of a default, and the interest rate in the case of a repayment weighted by the appropriate probabilities. Since the borrowers' types are known, the lenders are able to charge to the high- and low-ability types different interest rates, which I call r_H and r_L respectively. Then, under full disclosure there exists a unique stationary equilibrium in which high-ability borrowers exert more effort and face a lower interest rate than low-ability borrowers.

PROPOSITION 1. *The pairs (e^*_H, r^*_H) and (e^*_L, r^*_L) are the unique stationary equilibrium of the market under full disclosure. Further, $e^*_H > e^*_L$ and $r^*_H < r^*_L$.*

PROOF. This and all the following proofs are relegated to the Appendix. □

Note that the uniqueness of the equilibrium follows also from the assumption that lenders cannot write long-term contracts or condition current contracts on past performance in a relational fashion. Were this not the case, more complex dynamics could arise.

Despite the fact that the full disclosure policy eliminates the adverse selection problem from the market, moral hazard persists and the resulting equilibrium effort is a second best with respect to the case of perfect monitoring[15]. By preventing lenders to make the interest rate contingent on the level of effort, moral hazard implies that when a default occurs the marginal value of effort to a borrower is zero. In the full

[14] Corner solutions are not maxima because zero effort result in zero or negative utility, depending on the shock, while maximum effort, $e = 1$, is too expensive.

[15] The intuition behind such a distortion goes as follows. With moral hazard, the interest rate cannot be conditioned on the effort choice, meaning that the expected value of marginal effort is lower than in the perfect information case. Additional effort, in fact, benefits the lender, but such benefits are not internalized by the borrower because the interest rate cannot be made dependent on the effort choice. Besides reducing the expected value of marginal effort, moral hazard also results in higher interest rates, because these are lower for higher levels of effort (for a more detailed discussion, see Brander and Spencer, 1989).

disclosure equilibrium this type of distortion affects both types' choice of effort, as can be seen in equation (3.1), but it is particularly severe for the low-ability borrowers. Due to their higher marginal cost of effort, in fact, low-ability borrowers are charged a higher interest rate. In turn, this increases the likelihood of defaults and further reduces the marginal value of effort for borrowers in this group.

Before moving on to analyze the full withholding case, it is worth pointing out an alternative interpretation of the full disclosure arrangement. So far, I have maintained the working hypothesis that complete information was granted by the credit register, which, according to the chosen policy, was disclosing the borrowers' type. However complete information may arise also from the lenders's updating process, that allows them to learn the borrowers' type, provided that the market has been operating for a long enough time, and that the register discloses all of the past outcomes (a case that could be labeled "public records"). Under public records, in the long run we would observe the market equilibrium of the game approaching the repetition of the (unique) full disclosure equilibrium described above.

Remark 2. *After sufficiently many periods, the equilibrium with public records is the same as the equilibrium with full disclosure.*

The idea behind remark 1 is intuitive. By observing the history of play, lenders are able to update more and more precisely their common prior. Eventually, their belief will be correct and from that point on[16] the complete information equilibrium will prevail.

3.2. Full withholding. Under the alternative policy of full withholding, lenders know nothing about the borrowers and the equilibrium interest rate will be a function of the prior only[17]. The borrowers' optimization problem does not change with respect to the full disclosure case and it results in a conditional effort choice $e_{\theta 0} = e(r_0, \theta)$ which depends on the anticipated interest rate r_0 and the borrower's own type. The expected return on lending is now given by,

$$\Pi(e_{H0}, e_{L0}, r_0, p_0) = 1 + \rho \tag{3.3}$$

[16] This result reproduces in the market setting a result first stated in Vercammen's (1995) game theoretical analysis of borrowers and lenders. Further, the result in Remark 1 is reminiscent of the Cripps et al. (2004) impermanent reputation result. However in this setup there is no commitment type and reputations arise from the fact that the credit register allows borrowers to care for their own future payoffs. Similarly, reputations are impermanent because once lenders learn correctly each borrowers' type, these are not any longer in the position to affect future contractual terms, and the complete information equilibrium arise.

[17] Borrowers can do nothing to credibly signal their type. In particular, they cannot do so by accepting higher interest rates, because lenders are short-lived and multi-period contracts are not available.

where Π results from averaging the return from lending to borrowers of both types, $\Pi(e_{H0}, e_{L0}, r_0, p_0) = p_0\Pi(e, r_0, H) + (1 - p_0)\Pi(e, r_0, H)$ using the prior p_0 as a weight. Then, by solving together the equations (3.1) and (3.3), it is possible to find the borrowers' optimal effort choices, e^*_{H0} and e^*_{L0}, and the corresponding equilibrium interest rate, r^*_0. The next proposition grants that under full withholding, there exists a unique stationary equilibrium in which borrowers of both types are charged the same interest rate but high-ability borrowers exert more effort than low-ability borrowers.

PROPOSITION 2. *The pairs* (e^*_{H0}, r^*_0) *and* (e^*_{L0}, r^*_0) *are the unique stationary equilibrium of the market under full withholding. Further,* $e^*_{H0} > e^*_{L0}$.

The proof is straightforward so it is omitted. Due to the assumption of fixed investment, in equilibrium lenders offer exactly one interest rate. If they were trying to separate borrowers by offering a menu of two contracts, one with a higher interest rate for the low-ability type and viceversa for high-ability borrowers, both types of borrowers would choose the lower interest rate, which would result in negative expected returns for the lenders.

As I will discuss more extensively in section 8, under this informational arrangement high-ability types are cross-subsidizing low-ability types. This follows because they are charged an interest rate higher than the one in the perfect information case while low-ability types are charged a comparatively lower interest rate.

Before moving on to study the equilibrium with limited records, it is worth pointing out that it is always the case that high-ability borrowers exert more effort under full disclosure, while low-ability borrowers exert more effort under full withholding.

REMARK 3. *It is always true that* $e^*_L < e^*_{L0} < e^*_{H0} < e^*_H$.

The remark above follows from the assumption on marginal costs and the structure of the interest rates.

4. The case of limited records

This section studies the market equilibrium that arises when the credit register opts for a policy of limited records. Under this arrangement, lenders will have access to *credit reports* containing information regarding the outcomes of a borrower's projects in the last N periods. Lenders use the reports to update their beliefs, which in equilibrium results in separate contracts being offered to borrowers with different records. Borrowers, on the other hand, anticipate that their current actions will affect future contractual terms for the incoming N periods and choose their effort accordingly.

Limited records introduce a dynamic programming problem for the borrowers that is the same at each point in time, since in any t there is an infinite number of periods left. The environment is thus stationary and the appropriate solution concept is a version of the recursive competitive equilibrium concept, as defined in section 2.

Differently from the two policies previously analyzed, limited records allow the borrowers to link their current choice of effort to future benefits in terms of lower cost of capital. The main result in this section is that limited records prompt endogenously beneficial reputation effects from part of the borrower, that translate into higher effort and lower cost of capital for the low-ability borrowers with respect to the two previous informational arrangements. This follows because the lenders' beliefs in high ability are increasing in the number of past repayments, meaning that better contractual terms are offered to borrowers with "cleaner" scores. A credit report thus provides a credible commitment device for a borrower to exert higher effort, counteracting the distortionary effects of moral hazard.

4.1. The equilibrium with one-period records. A natural starting point to analyze limited records is to consider the case of $N = 1$, when the register discloses to the credit industry only the outcome of the most recent project for each borrower. More precisely, at the beginning of each period, lenders observe credit reports c_1, which display the outcome, either a success S or a default D, realized by borrower i in the preceding period. This information is used by the industry to update the prior p_0 on the borrower's type. I will focus on the choices of borrowers and lenders on the stationary path, along which their equilibrium strategies depend on the state c_1 and are otherwise history- and time-independent.

I reproduce here for convenience a type-θ borrower's problem, formulated in a time-invariant Bellman equation $V(c_1, \theta)$,

$$
\begin{aligned}
(4.1) \qquad V(c_1, \theta) = & \max_e U(e, c_1, \theta) \\
& + \beta \left[\Pr(c_1' = S \mid e, c_1) V(S, \theta) + \Pr(c_1' = D \mid e, c_1) V(D, \theta) \right]
\end{aligned}
$$

where $c_1' = S, D$ is the credit report obtained from c_1 by deleting the last reported outcome and replacing it with a repayment (default), and $\Pr(c_1' = S \mid e, c_1) = 1 - F(\bar{\omega}(e, r_{c_1}))$. The solution to the problem in (4.1) will be a stationary policy function $e_{\theta c_1} : C^1 \times \{H, L\} \to [0, 1]$.

To prove that the one-period equilibrium defined in section 2 exists and is unique, I will demonstrate that the borrowers' value function has a unique fixed point, which is the (unique) stationary policy function $e^*_{\theta c_1}$ that solves the dynamic program described above. The fact that borrowers' equilibrium strategies are stationary will allow me to claim that they induce a unique distribution of borrowers across credit scores. In turn, lenders in equilibrium will have a unique menu of posterior beliefs, one for each c_1, and thus offer a unique and time-invariant menu of interest rates.

The next proposition shows that Banach fixed point theorem applies to the value function in (4.1), granting that it is a contraction mapping with a unique fixed point which can be found by iteration.

Proposition 3. *There exists a unique stationary one-period equilibrium, with effort choices $e^*_{HS}, e^*_{HD}, e^*_{LS}, e^*_{LD}$, interest rates r^*_S, r^*_D and beliefs $p^*_{c_1}$.*

When $N = 1$, borrowers can be in one of two states and in the unique stationary equilibrium they exert two levels of effort, $e_{\theta c_1} = e(c_1, \theta)$ where $c_1 = S, D$. Optimal effort can be found by solving the borrower's dynamic optimization in (4.1), the first-order condition of which is given by,

$$Y'_e(e,\omega) - \beta E F' \bar{\omega}'_e(e_{\theta c_1}, r_{c_1}) \left(V(S,\theta) - V(D,\theta)\right) = c'_e(e,\theta) \tag{4.2}$$

According to equation (4.2), the unique optimal effort policy $e^*_{\theta c_1}$ is found by balancing its marginal benefit (LHS) and its marginal cost (RHS). The former is given by a direct increase in revenues today, $Y'_e(e,\omega)$, and by a marginal improvement in today's likelihood of repaying the loan, $F'\bar{\omega}'_e(e_{\theta c_1}, r_{c_1})$, which in turn leads to an increase in the future stream of utility, $V(S,\theta) - V(D,\theta)$. Without making additional assumptions on the functional forms, it is possible to derive (see Appendix C) the following expression for a type-θ borrower's optimal effort,

$$U'_e(e^*_{\theta S}, r^*_S) = \frac{F'\bar{\omega}'_{e^*_{\theta S}}}{F'\bar{\omega}'_{e^*_{\theta D}}} U'_e(e^*_{\theta D}, r^*_D) \tag{4.3}$$

which implicitly defines the effort policy $e^*_{\theta c_1}$ for a type θ borrower. Inspecting equation (4.3) it can be seen that in the one-period equilibrium borrowers exert more effort when a repayment, rather than a default, is going to be erased in the next period. Further, high-ability borrowers exert more effort than low-ability borrowers for any credit report. The next corollary formalizes these two considerations,

Corollary 1. *In the one-period equilibrium, $e^*_{\theta S} > e^*_{\theta D}$ for $\theta = H, L$ and $e^*_{Hc} > e^*_{Lc}$ for $c = \{S, D\}$.*

The fact that high-ability borrowers exert more effort than those of low-ability is also true for the other two informational arrangements discussed in section 3, and it follows directly from the assumption on the cost function. That the optimal effort differs between the two states is, however, clearly a novelty that belongs to the limited records arrangement. Effort for both types is higher when they are in the repayment state. This result is prompted by the fact that in the repayment state borrowers are faced with a more attractive interest rate, which increases the marginal value of effort by lowering the probability of default.

Lenders, as borrowers, base their equilibrium behavior on the state c_1. In particular, for each state the credit industry anticipates the borrowers' policy function $e^*_{\theta c_1}$ described above, and uses it to update the belief $p^*_{c_1} = p(H \mid p_0, e^*_{\theta c_1}, r^*_{c_1})$ that a borrower is of high-ability. The policy $N = 1$ allows for two credit reports to be observed by the lenders, thus in equilibrium there will be two interest rates, $r^*_{c_1} = \{r^*_S, r^*_D\}$, which can be found by solving for each $c_1 = S, D$ the following zero-profit condition together with the first order condition (4.2),

$$p^*_{c_1}\Pi(e^*_{Hc_1}, r^*_{c_1}) + (1 - p^*_{c_1})\Pi(e^*_{Lc_1}, r^*_{c_1}) = 1 + \rho \tag{4.4}$$

where $\Pi(e^*_{\theta c_1}, r^*_{c_1}, \theta)$ is defined in equation (3.2).

In any period, the borrowers' strategies and the interest rates induce a probability mass function g_θ which gives the probability $g_\theta(c_N)$ that a borrower of type θ has credit report c_N. Let this function have a cumulative distribution G. Although individual borrowers will not have the same credit report over time, in the stationary equilibrium the distribution of reports at the population level will remain stable. Let this stationary distribution be G^*, and let g^* be the corresponding stationary density. When $N = 1$, it is the case that $g_\theta(D) = 1 - g_\theta(S)$, so that the steady state probability $g^*_\theta(c_N)$ can be rewritten (see the Appendix for the complete derivation) as,

$$g^*_\theta(S) = \frac{1 - F(\bar{\omega}(e^*_{\theta D}, r^*_D))}{1 - (F(\bar{\omega}(e^*_{\theta D}, r^*_D)) - F(\bar{\omega}(e^*_{\theta S}, r^*_S)))}$$

which is an explicit expression for the steady state densities. It follows that, according to Bayes' rules, the credit industry's posterior upon seeing report $c_1 = S, D$ is given by,

$$\begin{aligned} p^*_S &= \frac{p_0 g^*_H(S)}{p_0 g^*_H(S) + (1 - p_0) g^*_L(S)} \\ p^*_D &= \frac{p_0 g_H(D)}{p_0 g^*_H(D) + (1 - p_0) g^*_L(D)} \end{aligned} \tag{4.5}$$

where p_0 is the industry's prior on the borrower being of high ability. The numerator in equation (4.5) represents the stationary-state probability that the observed outcome c_1 can be ascribed to a high-ability borrower. The denominator can be explained in a similar vein, and acts as the normalization factor.

Importantly it can be demonstrated that, on the stationary path, the distribution of high-ability borrowers' credit reports first-order stochastically dominates that of low-ability borrowers. This feature guarantees that in equilibrium borrowers of the former type are more likely to have better reports than borrowers of the latter type, that is,

PROPOSITION 4. *G^*_H first-order stochastically dominates G^*_L.*

The result follows from the assumption on the types' marginal costs, which in equilibrium ensures that high-ability borrowers exert more effort and hence achieve better reports more frequently than low-ability ones.

Proposition 4 is important because it shows that having a better record has a tangible value for the borrower. Lenders, in fact, will correctly believe that a good record is a signal of high ability and will offer better contractual terms to applicants with better records. Anticipating this, borrowers of both types will strive to obtain a good report, in order to benefit from the lower cost of capital associated with it.

In other words, borrowers use their records to build themselves a *reputation* for being of high ability, which will ensure them access to capital at a lower interest rate. It is possible to find the counterpart of the concern for reputation in the term $V(S,\theta)-V(D,\theta)$ which appears in the borrowers' first order condition (equation (4.2)). This term encapsulates the borrower's concern for the effect that his current actions will have on his future payoffs. In particular, $V(S,\theta)-V(D,\theta)$ represents a part of the marginal benefits of today's effort, which, besides leading to a higher utility today, also helps to instill in the lenders the belief to be dealing with a high-ability borrower. Thus, under limited records, and differently from the other two informational arrangements (equation 3.1), all borrowers share a concern for reputation, which arises endogenously and that can be capitalized in terms of higher future payoff.

4.2. The equilibrium with multi-period records. Having characterized the equilibrium for $N = 1$, this section studies the general case for finite choices of N. The main difference with respect to the previous case is that now, for an arbitrary N, borrowers can find themselves in any of 2^N states, corresponding to the equally many possible credit reports. To simplify notation, I define the following. Given a credit report c_N, let c_{DN} be the same report with the most recent outcome erased and a default appended before the *last* outcome displayed in the report. Similarly, let c_{SN} be computed in the same way but with a repayment appended. This notation allows me to specify the possible reports which the borrowers could have had the period before the current one. Precisely, assuming we are in period t,

$$\begin{aligned} c_N &= \{y_{t-N}, y_{t-N+1} \ldots, y_{t-1}, y_t\} \\ c_{DN} &= \{D, y_{t-N}, y_{t-N+1} \ldots, y_{t-1}\} \\ c_{SN} &= \{S, y_{t-N}, y_{t-N+1} \ldots, y_{t-1}\} \end{aligned}$$

This notation will be useful in the following analysis because reports that only differ in the last outcome are associated with the same continuation payoff, since the last outcome will be erased by the passage of time. Due to the fact that there is no natural way to order the credit reports[18], most of the following analysis will focus on how effort choices, lenders' beliefs and interest rates differ for credit reports that differ in the last outcome only.

As argued for the case $N = 1$, to prove the existence and uniqueness of the stationary equilibrium just defined, it is enough to demonstrate that the value function defined in (2.2) has a unique fixed point. The next proposition makes sure that this is indeed the case, and hence that the borrowers' problem has a unique stationary solution.

PROPOSITION 5. *There exists a unique stationary N-period equilibrium.*

Borrowers solve a dynamic program similar to that in (4.1). Here however, a borrower that finds himself in any state c_N can move either to c_{NS} or to c_{ND}, and the first order condition now becomes,

$$Y'_e(e, \omega) - \beta F' \bar{\omega}'_e(e, r_{c_N}) \left(V(c_{NS}, \theta) - V(c_{ND}, \theta)\right) = c'_e(e, \theta) \tag{4.6}$$

According to equation (4.6), and differently from the previous case, the difference in future payoff streams associated with repaying or defaulting, $V(c_{NS}, \theta) - V(c_{ND}, \theta)$, changes[19] with the state the borrower finds himself in. By extending the analysis in the previous section, it is possible to demonstrate that the optimal policy function solves the following,

$$U'_{e^*_{c_{SN}}}(e^*_{c_{SN}}, r_{c_{SN}}, \theta) = \frac{F' \bar{\omega}'_{e^*_{c_{SN}}}}{F' \bar{\omega}'_{e^*_{c_{DN}}}} U'_{e^*_{c_{DN}}}(e^*_{c_{DN}}, r_{c_{DN}}, \theta) \tag{4.7}$$

for any two credit reports c_{SN}, c_{DN} that differ only in the last outcome. According to equation (4.7), borrowers exert more effort when a repayment, rather than a default, is going to be erased in the next period. That is, for any finite $N > 1$,

[18] It is not clear, for instance, if report $\{S, S, D\}$ is better than report $\{S, D, S\}$ or $\{D, S, S\}$, in the sense of being more likely to belong to high- rather than a low-ability borrower.

[19] Once decided on how to order credit reports, the optimal effort schedule for different c_N could be elicited by studying how $V(c_{NS}, \theta) - V(c_{ND}, \theta)$ varies with c_N.

COROLLARY 2. *In the stationary N-period equilibrium, $e^*_{\theta c_{SN}} > e^*_{\theta c_{DN}}$ for $\theta = H, L$ and $e^*_{Hc_N} > e^*_{Lc_N}$ for $c_N \in C^N$.*

Since there are 2^N possible states, lenders will have equally as many beliefs that a borrower is of high ability. These can be found by proceeding in the same way as before, leading to the expression for the equilibrium posterior belief $p^*_{c_N}$,

$$p^*_{c_N} = \frac{p_0 g^*_H(c_N)}{p_0 g^*_H(c_N) + (1 - p_0) g^*_L(c_N)} \tag{4.8}$$

which can be explained in the familiar way. Anticipating that borrowers will follow the policy $e^*_{\theta c_N}$ described above, lenders will use these beliefs to derive the equilibrium (zero profit) interest rate menu $\langle r^*_{c_N} \rangle$,

$$p^*_{c_N} \Pi(e^*_{Hc_N}, r^*_{c_N}) + (1 - p^*_{c_N}) \Pi_L(e^*_{Lc_N}, r^*_{c_N}) = 1 + \rho \tag{4.9}$$

where $\Pi(e^*_{\theta c_N}, r^*_{c_N})$ is defined in equation (3.2). A policy of $N > 1$, then, leads to an exponential increase in the number of interest rates, which, anticipating the results in the next section, are also closer to one another. This consideration will be important when gauging the effects of different choices of N on the welfare in the market.

The general case of N periods also preserves the feature that high-ability borrowers are more likely to have better reports than low-ability borrowers, that is,

PROPOSITION 6. *G^*_H first-order stochastically dominates G^*_L.*

The same considerations as mentioned in the previous section apply: the stochastic dominance demonstrated in proposition 6 is the key to understanding why borrowers care about having good reports. Given that lenders in equilibrium offer more favorable contractual terms to borrowers with better reports, these will exert more effort in order to build themselves a good reputation and enjoy the monetary benefits associated with it.

Concluding this section, it is worth pointing out yet again that the main novelty brought about by the limited records arrangement is that it allows borrowers to link their current effort choices to future gains, via building, maintaining and possibly restoring a reputation for being of the high type. For limited records to act as such a "reputational" device both adverse selection and moral hazard are necessary. Thanks to the former distortion, each borrower is in the position to influence his future contractual terms by taking certain actions in the current period, while thanks to the second, the borrower can actually affect the distribution of the returns.

5. Effort choices and the length of credit records

This section investigates the effects of the choice of different lengths N of the credit reports on lenders' beliefs, equilibrium interest rates and borrowers' behavior. I begin with the lenders' beliefs. As can be seen from equation (4.8), for any possible credit report c_N, lenders form a corresponding belief p_{c_N} that the report belongs to a high-ability borrower. Denote with $\bar{c}_N$ the ("clean") report of length N with only repayments, and similarly $\underline{c}_N$ the ("dirty") report with only defaults. Then, we have that the difference in posterior beliefs assigned to the clean and the dirty record is larger the larger is N. Further, for any two credit reports that differ only in the last outcome, the difference in posterior beliefs is smaller the larger is N.

Lemma 1. *$p^*_{\bar{c}_N} - p^*_{\underline{c}_N}$ increases in N and $p^*_{c_{NS}} - p^*_{c_{ND}}$ decreases in N.*

The result above has a direct repercussion on the structure of the interest rates. Any of the 2^N equilibrium beliefs, in fact, directly gives rise to an interest rate, $r^*_{c_N}$, defined according to equation (4.9). The menu of equilibrium interest rates $\langle r^*_{c_N} \rangle$ falls in the interval (r^*_H, r^*_L) bounded by the two full-disclosure interest rates. In fact, as $N \to \infty$ and $p_{\bar{c}_N} \to p$, the interest rates corresponding to the clean and dirty records approach the boundaries.

Remark 4 follows from the previous lemma and equation (4.9). It shows that the choice of a larger N induces a larger menu of more fine-grained interest rates, of which those corresponding to the clean and the dirty record are relatively closer to (r^*_H, r^*_L). Conversely, a smaller N results in a smaller and coarser menu of interest rates, the extremes of which are relatively further away from (r^*_H, r^*_L). That is,

Remark 4. *$\left| r^*_H - r^*_{\bar{c}_N} \right|$, $\left| r^*_L - r^*_{\underline{c}_N} \right|$ and $\left| r^*_{c_{NS}} - r^*_{c_{ND}} \right|$ decrease in N.*

This last remark and the previous lemma have implications for the difference in future payoff streams associated with a repayment and a default, $V_\theta(c_{NS}) - V_\theta(c_{ND})$, which appears in the borrowers' first order condition, in equation (4.6). Lemma 1, in fact, states that the difference in the posterior beliefs for any two credit reports that differ only in the last outcome increases as N decreases. Remark 4 extends this consideration for the interest rate schedule, that becomes coarser as N decreases. Then, we have that following an exogenous decrease in N, the difference between the future payoff stream associated with a repayment, and default, increases, while an increase in N acts symmetrically.

Remark 5. *$V(c_{NS}, \theta) - V(c_{ND}, \theta)$ decreases in N.*

Note that the remark above disregards the "size" of the future payoff streams, which depends on the borrower's credit score, and, for the clean (dirty) record becomes higher (lower) as N grows. Rather, it states that the difference between filing a repayment rather than a default is larger the smaller N is.

We are now ready to state the main result of this section. The next proposition makes use of the remarks above to characterize how different choices of N affect the supply of effort from part of the borrowers. It turns out that borrowers' effort choices are most responsive to limited records when $N = 1$, that is,

PROPOSITION 7. *For any N-period equilibrium,* $\frac{e_S^*}{e_D^*} > \frac{e_{c_{SN}}^*}{e_{c_{DN}}^*}$.

To understand the proposition, consider a type-θ borrower's first order condition (4.2), which governs his policy function $e_{\theta c_N}^*$. Optimal effort results from solving the trade-off between current and future benefit of marginal effort (LHS) and its cost (RHS). By looking at the LHS of the first order condition, it is possible to see that the benefit of marginal effort is increasing in the difference between the future payoff streams associated with a repayment and a default respectively. By setting $N = 1$, the register contributes to giving maximum value to today's marginal effort. In turn, this implies that the borrowers care a lot for not having a default in their records, and accordingly they exert their highest effort when a repayment is about to be erased. Note that the fact that effort is higher when a repayment rather than a default is about to be erased is true of any choice of N, because in the former state the interest rate is lower. A one-period memory policy further contributes to this effect, by maximizing the return of effort in terms of future payoff streams.

Note that proposition 7 does not make any reference to the default rate, or the welfare, that different choices of N achieve, but it just states that reporting only the last-period performance is the policy that is most likely to elicit the highest effort from the borrowers. A one-period limited-record policy, in fact, maximizes the difference in the future payoff stream that follows from realizing a repayment rather than a default, and thus it allows borrowers to internalize most of the value of their effort.

6. Comparative statics: default rates

This section addresses comparative statics exercises among the three informational arrangements. More precisely, I will focus on their impact on the expected default rate. By default rate I mean the ex ante probability of default under the different equilibrium interest rates and effort choices, while I do not look at the amount that is defaulted upon after the lenders appropriate whatever revenues that were generated. Due to the analytical complexity of the model, I will restrict the current analysis of limited records to the case $N = 1$. This special case is interesting per se and sheds some light on the general effects of censoring the register's memory, yet admittedly it restricts the generalizability of the results[20]. Furthermore, I will restrict my attention to the case $p \leq \frac{1}{2}$. This is because when high-ability borrowers outnumber those of low-ability it is difficult to find conditions under which full disclosure is outperformed by the other two arrangements. Yet, the case $p \leq \frac{1}{2}$ is probably more representative of real credit markets. Then, provided that high- and low-ability borrowers are either present in equal shares, or low-ability types are more frequent, the next propositions pin down the sufficient condition(s) for each informational arrangement to minimize the expected default rate.

I begin with investigating whether full withholding ever leads to a lower default rate than full disclosure. Normalizing the borrowers' population to one, this would be the case when,

$$pF(\bar{\omega}(e_H^*, r_H^*)) + (1-p)F(\bar{\omega}(e_L^*, r_L^*)) \geq pF(\bar{\omega}(e_{H0}^*, r_0^*)) + (1-p)F(\bar{\omega}(e_{L0}^*, r_0^*))$$

The LHS represents the proportion of loans that are defaulted upon under the full disclosure arrangement. More precisely, it is given by the sum of the (expected) defaults coming from the borrowers of high-ability, $pF(\bar{\omega}(e_H^*, r_H^*))$, and low-ability, $(1-p)F(\bar{\omega}(e_L^*, r_L^*))$. The RHS instead shows the default rate under full withholding and it has a similar interpretation.

Assuming a uniform distribution for the shock, $\omega \sim U[0,1]$, the inequality above can be rearranged to read,

$$\bar{\omega}(e_L^*, r_L^*) - \bar{\omega}(e_{L0}^*, r_0^*) \geq \frac{p}{1-p}\left(\bar{\omega}(e_{H0}^*, r_0^*) - \bar{\omega}(e_H^*, r_H^*)\right) \tag{6.1}$$

Equation (6.1) states that the rate of default is lower under full withholding if the improvement in the "default threshold" $\bar{\omega}(e_L^*, r_L^*) - \bar{\omega}(e_{L0}^*, r_0^*)$ that is brought about for the low-ability group by switching from the full disclosure to the full withholding regime (LHS) is larger than the corresponding decrease for the high-ability group (RHS), weighted by the proportion of high and low-ability borrowers in the market.

Given its importance in proving the following propositions, before proceeding I discuss the default threshold $\bar{\omega}(e, r)$ in further details. According to the definition given in section 2, the default threshold is the shock which, given e and r, delivers just enough revenues to repay the loan plus the interests, and it is implicitly defined by $Y(e, \bar{\omega}(e, r))$ $= 1+r$. It follows that $\bar{\omega}(e, r)$ corresponds to the level curves of $Y(e, \cdot)$ evaluated at the different r's of interest, that is, $\bar{\Omega}(r) = \{(e, \omega) \in [0,1] \times [0,1] : Y(e, \omega) = 1 + r\}$. Due to the concavity of $Y(e, \omega)$, it is the case[21] that $\bar{\omega}'_e(e, r) < 0$, $\bar{\omega}'_r(e, r) > 0$, $\bar{\omega}''_{ee}(e, r) > 0$

20 Memories longer than one period are better suited to be studied with numerical methods.

21 Where $\omega^*_{ee}(e, r) = \frac{d}{de}(\omega^*_e(e, r)) = \frac{d}{de}(-\frac{R_e}{R_\omega}) = -\left(\frac{R_{ee}R_\omega - R_{\omega e}R_e}{R_\omega^2}\right) > 0$ because $R_{ee} < 0$. While $\omega^*_{er}(e, r) > 0$ can be seen by drawing a vertical line in correspondence to any e and see how ω^*_e becomes steeper as r increases.

and $\bar{\omega}''_{er}(e,r) > 0$, that is, the (upper) level sets are convex sets. These properties can be seen in figure 2, which draws a few level curves for an arbitrary concave function.

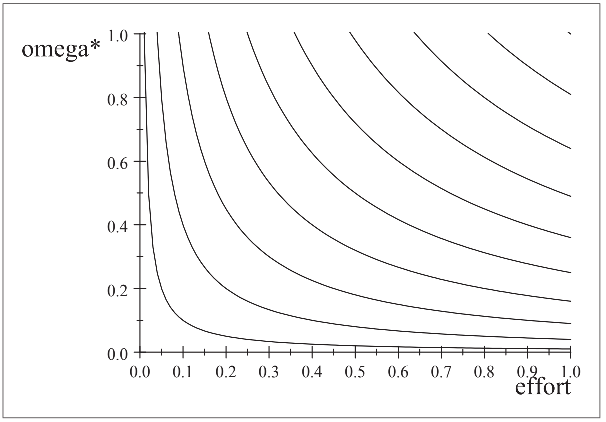


Figure 2. Level curves for $Y(e, \bar{\omega}) = 1 + r$

Each level curve (isoquant) is drawn holding fixed a value of r and it plots all the combinations of effort and the threshold $\bar{\omega}$ such that the realized revenues are just enough to repay the interest rate r to which the curve corresponds. Note that for any pair (e, r) there is only one threshold such that this is true. From the concavity of the revenue function it also follows that the distance between the level sets of fixed increments[22] in the value of Y (or $1+r$) increases with (e, ω). As mentioned, these considerations will be important in proving the next proposition, that outlines a sufficient condition for full withholding to induce a lower default rate than full disclosure.

Proposition 8. *Full withholding induces a lower default rate than full disclosure if $r_0^* < \frac{1}{2}(r_H^* + r_L^*)$ holds.*

The proof behind proposition 8 is straightforward. We know that full withholding leads to a lower default rate than full disclosure as long as the decrease in defaults from the low-ability group due to the shift from r_0^* to r_L^* is higher than the change (increase) in defaults for the high-ability group, accruing due to the shift from r_0^* to r_H^*. Then, the condition on r_0^*, together with the concavity of the revenue function, and thus the convexity of its contour sets, ensures that low-ability borrowers will increase their effort supply by more than what the high-ability borrowers will reduce their own, thus delivering the result.

The finding in proposition 8 paves the way to the question whether or not allowing a one-period memory, when condition $r_0^* < \frac{1}{2}(r_H^* + r_L^*)$ holds, would reduce the default

[22] The isoquants in figure 4 are drawn in correspondence of constant increases in r $(0.1, 0.2, ..., 1)$.

rates even further. Differently from the case of full withholding, the choice of a one-period memory leads to two equilibrium interest rates, r_D^* and r_S^*, which depend on the borrower's credit report. Proposition 9 shows that the policy $N = 1$ can indeed lead to a lower default rate than both full disclosure and full withholding, but only if an additional condition holds.

PROPOSITION 9. *If* $r_0^* < \frac{1}{2}(r_H^* + r_L^*)$ *holds, limited records induce a lower default rate than both full disclosure and full withholding if* $\bar{\omega}(e_{\theta D}^*, r_D^*) - \bar{\omega}(e_{\theta S}^*, r_S^*)$ *is small.*

In words, when the policy of full withholding leads to a lower default rate than full disclosure, so does the policy of limited records, provided that the additional condition on $\bar{\omega}(e_{\theta c_1}^*, r_{c_1}^*)$ holds. This second condition concerns the fact that the new equilibrium interest rates r_D^* and r_S^*, are such that $r_S^* < r_0^*$ but $r_D^* > r_0^*$. In proposition 8, I have shown that the condition $r_0^* < \frac{1}{2}(r_H^* + r_L^*)$ is enough to make sure that the shift from (r_H^*, r_L^*) to r_0^* leads to an overall lower default rate. The same condition would remain enough for proposition 9 to hold if the r_S^* was the only equilibrium interest rate, since we know that $r_S^* < r_0^*$. However, an additional condition is needed to make sure that the shift from r_0^* to r_D^* does not undermine the result. To this end, it is enough to require that the difference in default thresholds between the default state $\bar{\omega}(e_{\theta D}^*, r_D^*)$ and the repayment state $\bar{\omega}(e_{\theta S}^*, r_S^*)$ is small.

The next natural question, then, is whether limited records can outperform full disclosure when r_0 is *not* closer to r_H^* than to r_L^*, in which case we know from proposition 8 that full withholding may not be preferable to full disclosure.

In this case, we can directly compare the default rate under full disclosure to the one under limited records. The latter is lower when,

$$\text{(6.2)} \quad \bar{\omega}(e_L^*, r_L^*) - \frac{\bar{\omega}(e_{LS}^*, r_S^*)}{1-(\bar{\omega}(e_{LD}^*, r_D^*) - \bar{\omega}(e_{LS}^*, r_S^*))} \geq \frac{p}{1-p}\left[\frac{\bar{\omega}(e_{HS}^*, r_S^*)}{1-(\bar{\omega}(e_{HD}^*, r_D^*) - \bar{\omega}(e_{HS}^*, r_S^*))} - \bar{\omega}(e_H^*, r_H^*)\right]$$

where equation (6.2) has been derived by assuming a uniform distribution of the shock, and simplified using the equations in (4.5). Inequality (6.2) has the same interpretation as inequality (6.1) and it requires that the absolute difference in the thresholds $\bar{\omega}(e_{\theta c_1}^*, r_{c_1}^*)$ with complete information and limited records for the low-ability group is larger than that for the high-ability group, accounting for the proportion of high- and low-ability types in the economy. In other words, the fall in the defaults of low-ability borrowers, accruing from a decrease in the interest rates from r_L^* to (r_S^*, r_D^*), should more than compensate the increase in the high-ability defaults, that is brought about by the increase from r_H^* to (r_S^*, r_D^*).

It turns out that this result can be obtained under the same condition that was identified in proposition 9 for limited records to be preferable to full withholding, that is, Proposition 10. *If $r_0^* > \frac{1}{2}(r_H^* + r_L^*)$, limited records induce a lower default rate than full disclosure if $\bar{\omega}(e_{\theta D}^*, r_D^*) - \bar{\omega}(e_{\theta S}^*, r_S^*)$ is small.*

Again, the proof behind this result is simple. Since by the concavity of the revenue function the decrease in low-ability borrowers' default rate is larger than the corresponding increase in that of high-ability borrowers, to obtain the result it is only needed to make sure that the possibility of ending in a default state does not upset the original result.

As mentioned in the beginning of the section, the results in propositions 8-10 are only sufficient conditions, which moreover impinge on the assumptions of $p \leq \frac{1}{2}$ and uniform distribution of the shock, so they have to be taken with a grain of salt. However, they do provide some insights regarding how a credit register should set the length of its records in order to minimize the default rate.

First, it is possible to argue that among the three arrangements studied here, full disclosure is always preferable in a market where adverse selection is severe, or, at least, more severe than moral hazard. This is the case when $r_0^* > \frac{1}{2}(r_H^* + r_L^*)$ and when $\bar{\omega}(e_{\theta D}^*, r_D^*) - \bar{\omega}(e_{\theta S}^*, r_S^*)$ is large. Both statements, in fact, are true when the difference in cost structure between the two types is large and when moral hazard is not a particularly serious concern[23].

Instead, condition $r_0^* < \frac{1}{2}(r_H^* + r_L^*)$ is likely to hold in a market where adverse selection is not particularly grave, that is, when the two types face similar costs of effort. When this is the case, the full disclosure arrangement may excessively distort the effort choice of the low-ability borrowers[24], without succeeding in eliciting from the high-ability borrowers enough effort to compensate this distortion. Full withholding, on the other hand, may successfully lead to an equilibrium interest rate that alleviates the moral hazard problem for the low-ability group without distorting too much the effort supplied by the high-ability borrowers.

However, as moral hazard becomes more severe, limited records become the arrangement that delivers the lowest default rate. A situation of high moral hazard, and not too high adverse selection, is captured in this model by the statement that $\bar{\omega}(e_{\theta D}^*, r_D^*) - \bar{\omega}(e_{\theta S}^*, r_S^*)$ is small. The fact that the two thresholds are close, in fact, means that monitoring is noisy, and even when charged two different interest rates, the correspondingly different effort choices do not lead to different outcomes in a clear-cut

[23] This last statement can be seen from the fact that since the default thresholds in the two states are far apart monitoring works quite well.

[24] Because higher interest rates worsen the moral hazard problem.

way. When this is the case, limited records leads to the lowest default rate, by extracting the highest effort from borrowers of both types, thus counteracting the moral hazard problem in the most efficient way[25]. Table 3 summarizes the discussion,

Table 3. Informational arrangements by lowest default rate

Adv. Selection / Moral Hazard	High	Low
High	full disclosure	limited records ($N = 1$)
Low	full disclosure	full withholding

Again, the interpretation just given is subject to all the assumptions of the model and has to be subjected to further scrutiny. In particular, it has to be stressed that in this model adverse selection and moral hazard are to a certain extent intertwined (as can be seen in equation (2.1)), which calls for further care in extrapolating the results to the real world.

7. Comparative statics: welfare

While limited records give rise to an equilibrium that is different from those that arise under the other two informational arrangements, it is a policy that falls between the two. In fact, as N goes to infinity limited records approach full disclosure, and when N is zero they replicate full withholding. Denoting with W_D, W_W and W_{LR} the period-t welfare in the full disclosure, full withholding and limited records case, we have the following lemma,

LEMMA 2. *$W_{LR} \to W_D$ as $N \to \infty$, and $W_{LR} = W_W$ when $N = 0$.*

Lemma 2 states that aggregate welfare under limited records approaches aggregate welfare under full disclosure as $N \to \infty$, and aggregate welfare under full withholding when $N = 0$. Besides this taxonomic consideration, it is relevant to know how the three policies relate respect to each other in terms of the welfare that they induce (ex ante). To begin with, I compare the two extreme policies of full disclosure and

[25] The condition on $\bar{\omega}(e^*_{\theta c_1}, r_{c_1})$ has other implications that are worth mentioning. First, it holds when both types face similar moral hazard problems, and thus implicitly it makes sure that high ability types do not have to cross-subsidize low ability types too much. Secondly, it is possible to argue that when it holds, signalling (learning) is noisy, so it is more effective to enforce an arrangement that alleviates the moral hazard problem. Finally, due to the high noise, the probability of an unfair transition (the case in which a high ability borrowers defaults due to a high negative shock) to the low state is high, thus a choice of $N = 1$ may be preferable to prevent borrowers from being trapped in states with high interest rates.

full withholding, and demonstrate that the latter leads to a higher welfare under a well-defined condition.

The market's equilibrium under these two policies was characterized in Proposition 1 and 2. At a first glance, full disclosure looks appealing since it directly reduces the informational asymmetry in the market by eliminating adverse selection. However, such a policy does not deal with the problem of moral hazard, which results in an inefficiently low choice of effort for both types and distorts in particular the low-ability borrowers' effort supply, as discussed in section 3.

As first shown by Vercammen (2002) in a game-theoretic setup, the distortionary effects of moral hazard under full disclosure could be mitigated by charging a higher interest rate to the high-ability borrowers and simultaneously lowering the interest rate charged to the low-ability group[26]. This manipulation, in fact, can be shown to increase borrowers' *aggregate* welfare while leaving lenders' aggregate returns constant. As discussed in section 3, a cross-subsidization of this kind occurs naturally under full withholding where, thanks to adverse selection and the fixed-investment assumption, the equilibrium interest rate r_0^* falls in $r_H^* < r_0^* < r_L^*$. Thus in line of principle, full withholding *could* lead to a higher aggregate welfare, which may be interpreted as supporting strict privacy protection.

It turns out that under the assumptions of my model it is possible to write a sufficient, nonnecessary, condition under which aggregate welfare is higher with full withholding than with full disclosure. The proof is relegated to the Appendix, while here I only summarize its fundamental steps. I begin with considering the full disclosure interest rates and show that if r_H^* is increased marginally, while keeping constant the aggregate return for the lenders, r_L^* necessarily decreases. That is, $\left.\frac{dr_L^*}{dr_H^*}\right|_{Lenders} < 0$. Then I demonstrate that such a marginal increase in r_H^* produces an increases in the borrowers' aggregate welfare, or $\left.\frac{U^A}{dr_H^*}\right|_{Lenders} > 0$, provided that a condition, which I call CS for "cross-subsidization", holds. The CS condition requires that $\frac{1-F(\bar{\omega}(e_H^*, r_H^*))}{1-F(\bar{\omega}(e_L^*, r_L^*))}$ to be greater than one, but not too large, and it can be shown to hold if the two full disclosure interest rates are close to each other. When this holds, the interest rate r_0^* that arises under the full withholding equilibrium can be interpreted as the result of cross-subsidization with respect to the full disclosure equilibrium rates r_H^* and r_L^* and, according to the reasoning above, it leads to a higher aggregate welfare. It follows that,

PROPOSITION 11. *Full withholding is ex-ante welfare-improving with respect to full disclosure if the CS condition holds.*

[26] Who suffer the most from this distortion, due to the assumption on marginal costs.

By making sure that the full disclosure default thresholds are not too far away from each other, the CS condition allows cross-subsidization to work in a welfare-enhancing way. In particular, it implies that the decrease in effort from part of the high-ability borrowers can be compensated by the increase in effort from the low-ability borrowers, so that lenders' expected returns remain constant. Moreover, it implies that the respective decrease and increase in high- and low-ability borrowers' utility result in an higher aggregate welfare. Further, in order to interpret correctly Proposition 11, it is useful to recall that in this model moral hazard and adverse selection are intertwined, because by assumption the moral hazard is less severe for the high-ability borrowers, who face a lower marginal cost of effort. Thus, by requiring the interest rates r_H^* and r_L^* to be not too far from each another, the CS condition also implicitly constrains the two types not to be too unalike. Differently stated, the CS condition is satisfied in a credit market where adverse selection is low.

Informally, when the two types are not too unalike, the benefits from screening are not large and under CS they are outweighed by the costs of moral hazard. For this reason, the full withholding interest rate may provide the low-ability borrowers with the right incentives while not distorting "too much" the high type borrowers' effort choices, leading to a higher aggregate welfare[27]. This result is essentially an instance of the theory of second-best, here driven by the assumption of fixed investment-size that, by implying the uniqueness of the equilibrium interest rate under the full withholding arrangement, prevents adverse selection from causing net welfare losses.

In proving proposition 11, it has been shown that when the CS condition holds the borrowers' aggregate utility increases in r_H^*, and I have argued that it is possible to induce such an increase by withholding all the relevant information from the market, in order to force the credit industry to offer a unique interest rate $r_0^* \in (r_H^*, r_L^*)$. However, this is an extreme scenario. The next corollary shows that for the welfare result to hold, any policy that induce an increase in the interest rate faced by the high-ability types and a decrease in that faced by the low-ability types leads to higher welfare than the full disclosure case, provided that the original CS condition is satisfied. In particular, a policy mandating $N = 1$ leads to an increase in aggregate welfare. Thus, we have,

Corollary 3. *Under CS, limited records with $N = 1$ are ex-ante welfare-improving with respect to full disclosure.*

The corollary directly follows from proposition 7 and the proof is omitted. If we let the shock be uniformly distributed, it is possible to combine the results in the previous section and the corollary above, to have,

---

[27] Note that anyway in the full withholding equilibrium, however, high types will still be exerting a higher level of effort, and will be having a higher probability of success than low types.

Corollary 4. *Under CS, limited records with $N = 1$ is the arrangement that guarantees both the highest welfare ex ante and induces the lowest default rate.*

When CS holds in the full disclosure equilibrium, imposing a policy with limited records improves on the default rate of the low-ability group, while not distorting too much the effort supply, and hence the default rate, of the high-ability borrowers. This follows because $r^*_{c_1} \in (r^*_H, r^*_L)$, and r^*_H, r^*_L are already close to one another by CS. Then from the previous section, we know that, by the concavity of the revenue function, the positive effect on the performance of low-ability borrowers overcompensates the negative effect on the performance of the high-ability group.

Corollary 4 means that it is possible to outline sufficient conditions for one-period limited records to be the policy that performs best in both dimensions[28]. Although this result appears to speak in favor of limited records, I argued above that the CS condition may not be representative of all credit markets, and that $N = 1$ is in itself a very simple case. The question now becomes what are the effects of limited records, with $N > 1$, on defaults and welfare. In general, limited records elicit higher effort from part of the low-ability borrowers than full disclosure. In fact, if a lender knew the borrower's type, he would not offer him a lower interest rate because of his better records, since the borrower's future incentives would be unrelated to his credit history. Similarly, a borrower would not anymore be able to influence his future payoff by choosing a higher effort, and he would reduce it due to moral hazard. However, limiting memory also entails erasing past information and accordingly it prevents any screening, impeding borrowers from being charged the rate of interest that reflects the true value of their investment. This suggests that limited records deliver a higher welfare if the high-ability types have enough room to signal themselves out, and if simultaneously they provide low-ability types with the correct incentives to deal with the moral hazard problem. The findings in this section support these speculations, and suggest that censoring credit records is a viable alternative for credit registers to reduce the overall default rate.

8. Discussion

The following discussion further develops some of the points that were raised in the Introduction. First, I relate the model's findings to other results on limited records,

[28] Precisely, under CS, limited records score better than full disclosure in both dimensions, by corollary 3 and proposition 9 and 10. With respect to full withholding, instead, limited records certainly score better in terms of the default rate, proposition 9 and 10, while the effects on welfare are ambiguous.

demonstrated outside the context of credit markets. Secondly, I discuss how the framework presented in this paper can reconcile seemingly conflicting empirical evidence on the effects of data removal.

Data removal is relevant to several markets, because censoring public information often brings about reputation effects capable of disciplining informational asymmetries. In electronic marketplaces, for instance, opportunistic behavior from part of the sellers is disciplined by a reputation mechanisms that allows buyers to rate their experience with a particular seller. By compounding these ratings into a publicly available score, such mechanisms create scope for reputation concerns to induce the sellers to behave cooperatively. Contrary to what one may have expected, it turns out that mechanisms that aggregate the entire history of the sellers' past ratings into his score may not sustain cooperation[29] indefinitely. Cooperation, instead, has been proven to be sustainable over time by the *window aggregation mechanism* that aggregates a seller's ratings over a finite number of past periods only (Aperjis and Johari, 2010). Further, it can be shown that longer windows allow for more precise learning, and are effective in disclipining patient sellers[30]. Shorter windows, instead, provide stronger incentives against opportunistic behavior by punishing severely each single deviation, thus deterring sellers from accumulating a good reputation and regularly exploiting it.

The window aggregation mechanism is close in spirit to the limited records assumption in this paper, although here the records are presented as an ordered list of outcomes rather than a score. To fully investigate the optimal length of the register's memory goes beyond the scope of this work, however it is possible to speculate on how the results documented for the electronic marketplaces could inform the discussion on credit records. The results in propositions 8-10 seem to suggest that also in the credit market shorter records would be more effective when moral hazard is high. Longer records, instead, seem naturally preferable when adverse selection is high, because otherwise the performance of good borrowers would be hampered by a disadvantageous cross-subsidization. Future works should investigate these trade-offs more explicitly.

The present model is also related to the literature on behavior-based dynamic pricing (for an overview, see Fudenberg and Villas-Boas, 2005), which focuses on the recent

[29] For instance, good sellers would be tempted to free ride on their reputation every once in a while and then rebuilding it later on. In particular, Dellarocas (2005) shows that the choice of a one period memory allows for the greatest degree of cooperation to be achieved in an electronic marketplace where a single seller interacts with many buyers and faces the temptation to cheat on the quality of the good that he sells.

[30] Patient sellers, in fact,strongly dislike having a bad rating transiting in their report for a long time.

practice of using information on borrowers' previous purchasing decisions to target future prices[31]. A well-known result in this literature is that firms would like to commit not to collect information on customers' decisions, but they are induced to do so by competitive pressure, which potentially leads to lower welfare (Taylor, 2004a).

In the present model a similar concern may arise if it is the case that a market with full disclosure cannot settle on the full withholding equilibrium although cross-subsidization would lead to higher welfare. This concern is well founded, because if a lender was trying to charge the full-withholding interest rate to borrowers of both (known) types, he would lose his high-ability borrowers to other lenders who offer them a slightly lower interest rate. It follows that, due to the competitive pressure, the equilibrium that maximizes aggregate-welfare cannot be sustained. Interestingly, these considerations are aligned with a series of other papers that stress how disregarding privacy considerations can lead to socially undesirable outcomes (see for instance Hermalin and Katz, 2006; Hirshleifer, 1971).

Finally, the findings of my model can be related to the empirical literature that tries to assess the impact of existing data retention legislation on market's outcomes. This literature begins with Musto (2004) who documents that American borrowers are not able take advantage of the "bankruptcy flag" removal mandated by the Fair Credit Reporting Act. On the contrary, after experiencing a momentary increase in their assessed credit-worthiness, this quickly returns to the pre-removal level[32]. Musto interprets the evidence as suggesting that in the long-run data protection legislations only worsen the adverse selection problem, and that they only mandate "efficiency in reverse". This pioneering study has been followed by others that document a far less trenchant effect. In particular, Bos and Nakamura (2010) employ Musto's methodology to investigate the effects of the removal of credit remarks in Sweden[33]. Their analysis confirms the short-run boost in assessed credit-worthiness that was documented by

[31] In Krugman's words "[D]ynamic pricing is the new version of an old practice, price discrimination", The New York Times, October 4, 2000, p. A35.

[32] Musto's analysis makes use of a large panel dataset of borrowers in the unsecured bank-card credit market, which was appositely prepared for the study by the credit register Experia. In particular, the panel covers more than forty thousands borrowers that have a bankruptcy flag as of the starting date of the sample period, the "filers", together with a control group of twenty thousand "non-filers". Borrowers were followed for two years. By regressing the change in the filers' FICO credit scores between the last observation pre-removal and the first after removal, Musto is able to show a significant and positive impact of the removal of the flag. Immediately after the information is dropped, ex-filers gain substantial access to credit cards and enjoy an increase in credit limits up to a thousand dollars. Unfortunately, this boost in credit due to the removal of the flag is only short-lived. Tracking these borrowers for another year, in fact, shows that their assessed credit-worthiness declines back to the original level or below, and that delinquency rates go back to pre-removal levels.

[33] The authors employ an dataset comparable to Musto's, and follow borrowers up to two years after the removal.

Musto, however the authors do not find long-run reversal to the pre-removal levels. Rather, the improved credit scores are shown to remain stable over time.

These conflicting findings could be accommodated within the present model by arguing that the American and the Swedish markets are fundamentally different in their degree of moral hazard and adverse selection. More precisely, it is possible to argue that the former market, which is a market for small loans, is characterized by severe adverse selection. Thus removing past information is likely to be a detrimental policy, which effectively prevents lenders from screening out bad borrowers. In the case of Sweden, instead, the analysis concerns the removal of credit remarks rather than of bankruptcy events. This suggests that Bos and Nakamura are looking at a market in which average borrowers' quality is high. Consequently, data removal may actually succeed in providing the correct incentives for the borrowers, giving a fresh start to those that suffered from a negative shock realization.

9. Conclusions

With this paper, I put forward a model that tries to assess the impact of data removal legislation on credit markets, where it applies to the data on borrowers' past behavior stored by credit registers. I studied a stylized credit market wherein long-lived borrowers repeatedly seek funding from short-lived lenders and where a credit register regulates the disclosure of relevant information. By comparing different privacy protection arrangements, I demonstrate that a policy of limited records may lead to both a lower (ex ante) default rate and a higher (ex ante) aggregate welfare than either full disclosure or full withholding.

The model presented here is in an extremely stylized form and several extensions come to mind. While I have tried to justify the most stringent assumptions, more realistic modelling choices may improve our understanding of the reputational dynamics brought about by data retention policies. In particular, the next topics on the research agenda include a thorough study of the design of credit registers' optimal memory and the relaxation of the assumption that borrowers are of a certain type which is fixed over time. Further, it is important to note that the policy data retention is intimately intertwined with reputation formation, and thus it is not of unique interest to the credit market literature. This suggests that future works should take into account other related literatures that explore similar issues and try to benefit from them.

References

Akerlof, G. A. , 1970. The market for "lemons": quality uncertainty and the market mechanism. Quart. J. Econ. : 84, 488-500.

Aperjis, C. , Johari, R. , 2010. Optimal windows for aggregating ratings in electronic marketplaces. Manag. Sci. : 56, 864 -880.

Armour, J. , Cumming, D. , 2005. Bankruptcy law and entrepreneurship. Amer. Law Econ. Rev. :10, 303-350.

Bos, M. , Nakamura, L. , 2010. Loan applications and credit scores: evidence from legally mandated credit remark removal. *Manuscript*.

Bottero, M. , Spagnolo, G. , 2011. Reputation, privacy and limited records: a survey. *Manuscript*.

Brander, J. , Spencer, B. , 1989 . Moral hazard and limited liability: implications for the theory of the firm. Int. Econ. Rev., 30: 833-849.

Chatterjee, S. , Corbae, D. , Rios-Rull, V. , 2010. A theory of credit scoring and competitive pricing of default risk. *Mimeo.*

Cohen-Cole, E. , Duygan-Bump, D. , Montoriol-Garriga, J. , 2009. Forgive and Forget: Who Gets Credit after Bankruptcy and Why? Quantitative Analysis Unit W.P. QAU09-2, Federal Reserve Bank of Boston.

Cripps, M. W. , Mailath, G. J. , Samuelson, L. , 2004. Imperfect monitoring and impermanent reputation. Econometrica: 72, 407-432.

Diamond, D. W. , 1989. Reputation acquisition in debt markets. J. Polit. Economy : 97, 828-862.

Dellarocas, C. , 2005. Reputation mechanism design in online trading environments with pure moral hazard. Inform. Syst. Res. : 16, 209-230.

Elul, R. , Gottardi , P. , 2007. Bankruptcy: is it enough to forgive or must we also forget?W. P. 11-14, Federal Reserve Bank of Philadelphia.

Ekmekci, M. , 2010. Sustainable reputations with rating systems. J. Econ. Theory : 146, 479-503.

Galindo, A. , Miller, M. , 2001. Can credit registries reduce credit constraints? Empirical evidence on the role of credit registries in firm investment decisions. Annual Meetings of the Board of Governors, Inter-American Development Bank and Inter-American Investment Corporation.

Fudenberg, D. , Villas-Boas, J. M. , 2005. Behavior-based price discrimination and customer recognition. Handbook on Economics and Information Systems, Elsevier.

Hermalin, B. E. , Katz, M. L. , 2006. Privacy, property rights and efficiency: The economics of privacy as secrecy. Quant. Market. Econ. : 4, 209-239.

Hirshleifer, J. , 1971. The private and social value of information and the reward to inventive activity. Amer. Econ. Rev. : 61, 561-574.

Jacobson, T. , Lindé, J. , Roszbach, K. , 2006. Internal ratings system, implied credit risk and the consistency of banks' risk classification policies. J. Banking Finance : 30, 1899-1926.

Jappelli, T. , Pagano, M. , 2002. Information Sharing, Lending and Defaults: Cross-Country Evidence. J. Banking Finance : 26, 2017-2045.

Jentzsch, N. , 2001. The economics and regulation of financial privacy - A comparative analysis of the United States and Europe. W.P. 128, Freie Universität Berlin, John F. Kennedy Institute for North American Studies.

Liu, Q. , Skrzypacz, A. , 2011. Limited records and reputation. *Mimeo.*

Musto, D. K. , 2004. What happens when information leaves a market? Evidence from postbankruptcy consumers. J. Bus. : 77, 725-748.

Posner, R. A. , 1983. Privacy and related interests. In: The economics of Justice. Harvard University Press.

Rosen, J. , 2000. The unwanted gaze: The destruction of privacy in America. Random House, New York.

Report of Expert Group on Credit Histories. DG Internal Markets and Services. http://ec.europa.eu/ internal_market/ consultations/ docs/2009/ credit_histories/ egch_report_en .pdf

Sartor, G. , 2006. Privacy, reputation, and trust: some implications for data protection. Trust Management, Lecture Notes in Computer Science, 2006. 3986: 354-366.

Sharpe, S. A. , 1990. Asymmetric information, bank lending and implicit contracts: a stylized model of customer relationships. J. Finance : 45, 1069-1087.

Summary of replies to the public consultation about the future legal framework for protecting personal data.,2009, http://ec.europa.eu/justice/news/consulting_public/0003/summary _ replies_en.doc .

Taylor, C. , 2004a. Privacy and information acquisition in competitive markets. W.P. 13271, Olin Program in Law & Economics, UC Berkeley.

Taylor, C. , 2004b. Consumer privacy and the market for customer information. RAND: 35, 631-650.

United States House of Representatives. Committee on Banking and Currency. Subcommittee on Consumer Affairs (1970). Fair Credit Reporting: Hearings before the Subcommittee on Consumer Affairs of the Committee on Banking and Currency of the House of Representatives. Ninety-First Congress, Second Session on H.R. 16340, March 17, 19, 20, 23 and 24; April 8 and 8, 1970.

United States Senate. Committee on Banking and Currency. Subcommittee on Financial Institutions (1969). Fair Credit Reporting: Hearings before the Subcommittee on Financial Institutions of the Committee on Banking and Currency of the United States Senate. Ninety-First Congress, First Session on S. 823, May 19, 20, 21, 22 and 23, 1969.

Vercammen, J. , 2002. Welfare improving adverse selection in the credit markets. Int. Econ. Rev.: 43, 1017-1033.

Vercammen, J. , 1995. Credit register policy and sustainable reputation effects in credit markets. Economica: 62, 461-478.

Vogel, R. C. , Burkett, P. , 1986. Deposit mobilization in developing countries: the importance of reciprocity in lending. J. Develop Area: 20, 425-438.

Appendix

Here follows the proofs of the propositions, lemmas and corollaries in the text.

PROOF OF PROPOSITION 1. Equation (3.1) describes the borrowers' conditional effort function e_H and e_L. Equation (3.3) combines the lenders' beliefs with the conditional effort function of the borrowers to derive the lenders' reservation interest rate r_θ^*. The equilibrium interest rates are found by solving the zero-profit conditions,

$$\begin{cases} \Pi(e_H, r_H^*) = 1 + \rho \\ \Pi(e_L, r_L^*) = 1 + \rho \end{cases}$$

Then, in equilibrium, borrowers correctly anticipate r_H^*, r_L^* and exert e_H^* and e_L^*. Lenders cannot deviate from offering r_θ^*, because charging a lower interest rate would mean incurring losses, while no borrower would accept a higher interest rate due to perfect competition. At the interest rate r_θ^* it is optimal for all borrowers to accept the contract. In particular, this happens because both types' reservation utility is zero, meaning that a type θ borrower's participation constraint is,

$$\int_{\bar{\omega}(e_\theta, r)}^{1} \left[Y(e_\theta^*, \omega) - (1 + r)\right] dF(\omega) - c(e_\theta^*, \theta) \geq 0$$

which by construction is always satisfied, in expectation, by $r = r_\theta^*$.

The second part of the proposition follows directly from the assumption on marginal costs, $c_e'(e, H) < c_e'(e, L)$. □

PROOF OF PROPOSITION 3 AND 5. To prove these propositions, I will use Banach's fixed point theorem to show that the value function (4.1) has a unique fixed point. The proof consists of two steps. First I will describe the appropriate metric space and show that it is complete, and secondly I will show that the value function can be interpreted as a functional equation that maps a functional space in itself and is a contraction mapping.

Abusing notation somewhat, we can think of the value function as a functional equation,

$$V(c_N, \theta) = \max_e U(e, r_{c_N}, \theta) + \beta \sum_{c_N'} \Pr(c_N' \mid e, c_N) V_\theta(c_N', \theta)$$

$$TV(c_N, \theta) = \max_e U(e, r_{c_N}, \theta) + \beta \sum_{c_N'} \Pr(c_N' \mid e, c_N) V_\theta(c_N', \theta)$$

where T defines a mapping with domain equal to the space of functions V. We can restrict our attention to the space $\mathcal{V}$ of functions V which are continuous and bounded. Given that U is by assumption a continuous and bounded function, and so is V, TV is also bounded. Furthermore, by the theorem of the Maximum, TV is continuous as well. Then, the fact that the space of continuous and bounded real functions defined over some set X,

$$\mathcal{B} \equiv \{V : X \rightarrow \mathbb{R} | V \text{ is continuous and } m < V(x) < M, \forall x \text{ for some } m, M \in \mathbb{R}\}$$

and equipped with the sup metric

$$d_\infty(V_1, V_2) = \sup_x \mid V_1(x) - V_2(x) \mid$$

is a complete metric space, allows us to conclude that also $(\mathcal{V}, d_\infty)$ is a complete metric space. It follows that T is an operator that maps the metric space $(\mathcal{V}, d_\infty)$ to itself.

Thanks to the result above, to apply Banach's theorem we can use the Blackwell's (sufficient) conditions to establish that T is a contraction,

Blackwell's sufficient conditions: Let T be an operator on a metric space (S, d_∞) where S is a space of functions with domain X and d_∞ is the sup metric. Then T is a contraction mapping with modulus β if it satisfies the following:

(i) Monotonicity: For any pair of functions $(V_1, V_2) \in S$, $V_1(x) \geq V_2(x)$ for all $x \Rightarrow TV_1(x) \geq TV_2(x)$

(ii) Discounting: For any function $V(x) \in S(X)$, positive real numbers $q > 0$ and $\beta \in (0, 1)$ it is true that for any x on which V is defined $T(V + c) \leq T(V) + \beta c$

Monotonicity can be proven as follows. Take any two $(V_1, V_2) \in \mathcal{V}$ such that $V_1(x) \geq V_2(x)$. Then,

$$\begin{aligned} TV_1(c_N) &= \max_e U(e, c_N, \theta) + \beta \sum_{c'_N} \Pr(c'_N \mid e, c_N) V'_1(c'_N, \theta) \\ &\geq \max_e U(e, c_N, \theta) + \beta \sum_{c'_N} \Pr(c'_N \mid e, c_N) V'_2(c'_N, \theta) \\ &= TV_2(c_N, \omega_t) \end{aligned}$$

As for the Discounting condition, take any function $V(x) \in S(X)$, positive real numbers $q > 0$ and $\beta \in (0,1)$

$$\begin{aligned}
T(V+q) &= \max_e U(e, c_N, \theta) + \beta \sum_{c'_N} \Pr(c'_N \mid e, c_N)\,(V'_1(c'_N, \theta) + q) \\
&= \max_e U(e, c_N, \theta) + \beta \sum_{c'_N} \Pr(c'_N \mid e, c_N) V'_1(c'_N, \theta) + \beta E \sum_{c'_N} \Pr(c'_N \mid e, c_N) q \\
&= \max_e U(e, c_N, \theta) + \beta \sum_{c'_N} \Pr(c'_N \mid e, c_N) V'_1(c'_N, \theta) + \beta q \\
&= T(V) + \beta q
\end{aligned}$$

where $\sum_{c'_N} \Pr(c'_N \mid e, c_N) = 1$ by definition. Thus T is a contraction mapping on a complete metric space and, thanks to Banach's Theorem, we know it has a unique fixed point. $\square$

Proof of proposition 4. I momentarily suppress the asterisk notation to indicate equilibrium variables, with the understanding that $\langle e_{\theta c_1} \rangle$ and $\langle r_{c_1} \rangle$ indicate equilibrium quantities. Take any two G_{Ht}, G_{Lt}. Next period distribution is given by,

$$G_{Ht+1} = TG_{Ht}$$
$$G_{Lt+1} = TG_{Lt}$$

and

$$\begin{aligned}
G_{\theta t+1}(S) &= G_{\theta t}(S)(1 - F(\bar{\omega}(e_{\theta S}, r_S)) + G_{\theta t}(D)(1 - F(\bar{\omega}(e_{\theta D}, r_D)) \\
G_{\theta t+1}(D) &= G_{\theta t}(S)(F(\bar{\omega}(e_{\theta S}, r_S)) + G_{\theta t}(D)(F(\bar{\omega}(e_{\theta D}, r_D))
\end{aligned}$$

Case (a) Let us assume that $G_{Lt}(D) = G_{Ht}(D)$ so at time t high- and low-ability borrowers are distributed in the same way over the credit reports S and D. Then in the next period,

$$\begin{aligned}
G_{Lt+1}(D) - G_{Ht+1}(D) &= G_{Lt}(S)(F(\bar{\omega}(e_{LS}, r_S)) + G_{Lt}(D)(F(\bar{\omega}(e_{LD}, r_D)) \\
&\quad - G_{Ht}(S)(F(\bar{\omega}(e_{HS}, r_S)) - G_{Ht}(D)(F(\bar{\omega}(e_{HD}, r_D)) \\
&= G_t(S)\,[F(\bar{\omega}(e_{LS}, r_S) - F(\bar{\omega}(e_{HS}, r_S))] \\
&\quad + G_t(D)\,[F(\bar{\omega}(e_{LD}, r_D) - F(\bar{\omega}(e_{HD}, r_D))] \\
&> 0
\end{aligned}$$

where the last step follows from the assumption on marginal costs that makes sure that,

$$F(\bar{\omega}(e_{Lc_1}, r_{c_1})) \geq F(\bar{\omega}(e_{Hc_1}, r_{c_1}))$$

in any equilibrium, and for any $c_1 = \{S, D\}$.

Case (b) Let us now assume that $G_{Lt}(D) \geq G_{Ht}(D)$ so that at time t the distribution of L borrowers is already dominated by that of H borrowers. Then,

$$\begin{aligned}
G_{Lt+1}(D) - G_{Ht+1}(D) &= G_{Lt}(S)(F(\bar{\omega}(e_{LS}, r_S)) + G_{Lt}(D)(F(\bar{\omega}(e_{LD}, r_D)) \\
&- G_{Ht}(S)(F(\bar{\omega}(e_{HS}, r_S)) - G_{Ht}(D)(F(\bar{\omega}(e_{HD}, r_D)) \\
&= G_{Lt}(D)(F(\bar{\omega}(e_{LD}, r_D)) - G_{Ht}(D)(F(\bar{\omega}(e_{HD}, r_D)) \\
&+ (1 - G_{Lt}(D))(F(\bar{\omega}(e_{LS}, r_S)) - (1 - G_{Ht}(D))(F(\bar{\omega}(e_{HS}, r_S)) \\
&= G_{Lt}(D)(F(\bar{\omega}(e_{LD}, r_D)) - G_{Ht}(D)(F(\bar{\omega}(e_{HD}, r_D)) \\
&+ (F(\bar{\omega}(e_{LS}, r_S)) - F(\bar{\omega}(e_{HS}, r_S)) \\
&+ G_{Ht}(D))(F(\bar{\omega}(e_{HS}, r_S) - G_{Lt}(D))(F(\bar{\omega}(e_{LS}, r_S)) \\
&\geq 0
\end{aligned}$$

where the last step follows from acknowledging that,

$$\begin{aligned}
&G_{Lt}(D)(F(\bar{\omega}(e_{LD}, r_D)) - G_{Ht}(D)(F(\bar{\omega}(e_{HD}, r_D)) + (F(\bar{\omega}(e_{LS}, r_S)) - F(\bar{\omega}(e_{HS}, r_S)) \\
&\geq G_{Lt}(D)(F(\bar{\omega}(e_{LS}, r_S)) - G_{Ht}(D)(F(\bar{\omega}(e_{HS}, r_S)
\end{aligned}$$

Case (c) Finally, consider the case $G_{Lt}(D) \leq G_{Ht}(D)$. Using the same algebra as in case (b), we have once again that $G_{Lt+1}(D) - G_{Ht+1}(D) \geq 0$. □

Solving for the value function, optimal effort and posterior beliefs:

Case N = 1

To compute explicitly the value function $V(c_1, \theta)$ and the optimal effort $e^*_{\theta c_1}$ in each state c_1, one has to solve the following system,

$$\begin{cases}
V(S, \theta) = \max_e U(e^*_{\theta S}, r^*_S, \theta) + \beta E \sum_{c_1'=S,D} \Pr(c_1' \mid e^*_{\theta S}, r^*_S) V(c_1', \theta) \\
V(D, \theta) = \max_e U(e^*_{\theta D}, r^*_D, \theta) + \beta E \sum_{c_1'=S,D} \Pr(c_1' \mid e^*_{\theta D}, r^*_D) V(c_1', \theta) \\
0 = \frac{dU(e^*_{\theta S}, r^*_S, \theta))}{de^*_{\theta S}} + \beta E F' \frac{d\bar{\omega}(e^*_{\theta S}, r^*_S)}{de^*_{\theta S}} (V(S, \theta) - V(D, \theta)) \\
0 = \frac{dU(e^*_{\theta D}, r^*_D, \theta)}{de^*_{\theta D}} + \beta E F' \frac{d\bar{\omega}(e^*_{\theta D}, r^*_D)}{de^*_{\theta D}} (V(S, \theta) - V(D, \theta))
\end{cases}$$

Assuming that borrowers have perfect foresight of the steady state interest rates $\langle r_S^*, r_D^* \rangle$, the system above consists of four equations in four unknowns, $\{V(S,\theta), V(D,\theta), e_S^*, e_D^*\}$.

CaseN > 1

By extending the analysis above, it is possible to set up a system of equations to determine, for any $c_N \in C_N$, which are the value of $V(c_N,\theta)$ and $e^*_{\theta c_N}$. To facilitate reading, I substitute momentarily the notation $e^*_{c_N}$ with $e^*(c_N)$, and I use c_{NS}, c_{ND} to indicate the report c_N with farthest outcome erased and S, D added in front of the most recent position,

(9.1)

$$\begin{cases} V(c_{SN},\theta) = \max_e U(e^*(c_{SN}), r^*(c_{SN}),\theta) + \beta \sum_{c'_N = c_{NS}, c_{ND}} \Pr(c'_N \mid e^*(c_{SN}), r^*(c_{SN}))V(c'_N,\theta) \\ V(c_{DN},\theta) = \max_e U(e^*(c_{DN}), r^*(c_{DN}),\theta) + \beta \sum_{c'_N = c_{NS}, c_{ND}} \Pr(c'_N \mid e^*(c_{DN}), r^*(c_{DN}))V(c'_N,\theta) \\ 0 = \frac{dU(e^*(c_{SN}), r^*(c_{SN}),\theta)}{de^*(c_{SN})} + \beta E F' \frac{d\bar{\omega}(e^*(c_{SN}), r^*(c_{SN}))}{de^*(c_{SN})} \left(V(c_{NS},\theta) - V(c_{ND},\theta)\right) \\ 0 = \frac{dU(e^*(c_{DN}), r^*(c_{DN}),\theta)}{de^*(c_{DN})} + \beta E F' \frac{d\bar{\omega}(e^*(c_{DN}), r^*(c_{DN}))}{de^*(c_{DN})} \left(V(c_{NS},\theta) - V(c_{ND},\theta)\right) \end{cases}$$

Computations of the stationary density functions:

Case N $= 1$

For the case $N = 1$, if in period t the densities are g_{Lt} and g_{Ht}, in the next period it will be the case that,

$$g_{Ht+1}(S) = \Pr(S \mid e^*_{HS}, H)g_{Ht}(S) + \Pr(S \mid e^*_{HD}, H)g_{Ht}(D)$$
$$g_{Lt+1}(S) = \Pr(S \mid e^*_{LS}, L)g_{Lt}(S) + \Pr(S \mid e^*_{LD}, L)g_{Lt}(D)$$

and similarly for $g_{Ht+1}(D)$, $g_{Lt+1}(D)$. In steady state, the densities $\langle g_\theta \rangle$ solve,

$$g_\theta(S) = \Pr(S \mid e^*_{\theta S}, \theta)g_\theta(S) + \Pr(S \mid e^*_{\theta D}, \theta)g_\theta(D)$$
$$g_\theta(D) = \Pr(D \mid e^*_{\theta S}, \theta)g_\theta(S) + \Pr(D \mid e^*_{\theta D}, \theta)g_\theta(D)$$

Since $g_\theta(D) = 1 - g_\theta(S)$, the steady state distribution can be rewritten as

$$\begin{aligned} g_\theta(S) &= \frac{\Pr(S \mid e^*_{\theta D}, \theta)}{1 + \Pr(S \mid e^*_{\theta D}, \theta) - \Pr(S \mid e^*_{\theta S}, \theta)} \\ &= \frac{1 - F(\bar{\omega}(e^*_{\theta D}, r_D))}{1 - (F(\bar{\omega}(e^*_{\theta D}, r_D)) - F(\bar{\omega}(e^*_{\theta S}, r_S)))} \end{aligned}$$

Case N > 1

Proceeding as before, for any c_N, the borrowers' strategies and the interest rates induce probability density functions g_θ, with distribution G_θ. For any period t densities

g_{Lt} and g_{Ht}, in the next period we will have that,

$$g_{Ht+1}(c_N) = \Pr(c_N \mid e^*_{Hc^+_N}, H)g_{Ht}(c_{SN}) + \Pr(S \mid e^*_{Hc^-_N}, H)g_{Ht}(c_{DN})$$
$$g_{Lt+1}(c_N) = \Pr(c_N \mid e^*_{Lc^+_N}, L)g_{Lt}(c_{SN}) + \Pr(S \mid e^*_{Lc^-_N}, L)g_{Lt}(c_{DN})$$

and similarly for $g_{Ht+1}(D)$, $g_{Lt+1}(D)$. In steady state, the densities $\langle g_\theta \rangle$ solve,

$$g_\theta(c_N) = \Pr(c_N \mid e^*_{\theta c^+_N}, \theta)g_\theta(c_{SN}) + \Pr(S \mid e^*_{\theta c^-_N}, \theta)g_\theta(c_{DN})$$

which is an explicit expression for the steady state densities for any given effort function. It follows that the corresponding belief function should be given by lenders' updating of their prior p_0 using Bayes' rule,

$$p(c_N) = \frac{p_0 g_H(c_N)}{p_0 g_H(c_N) + (1 - p_0) g_L(c_N)} \tag{9.2}$$

which can be explained in the familiar way.

Proof of corollary 1. To prove the first part of the proposition consider again the system in (9.1). Rearranging the four equations, it is possible to find out that the policy function solves the following,

$$\frac{dU(e^*(c_{SN}), r^*(c_{SN}), \theta)}{de^*(c_{SN})} = \frac{F' \frac{d\bar{\omega}}{de^*_\theta(c_{SN})}}{F' \frac{d\bar{\omega}}{de^*_\theta(c_{DN})}} \frac{dU(e^*(c_{DN}), r^*(c_{DN}), \theta)}{de^*(c_{DN})} \tag{9.3}$$

for any two c_{SN}, c_{DN}. The expression $F' \frac{d\bar{\omega}}{de^*}$ stands for the change in the default likelihood that follows from an increase in effort. In the case $N = 1$, knowing that $r_S < r_D$ and, since $\bar{\omega}'_e < 0$ and $\bar{\omega}''_{er} < 0$ (see Section 6), it is also true that $\left| F' \frac{d\bar{\omega}(e^*, r^*_S)}{de^*} \right| > \left| F' \frac{d\bar{\omega}(e^*, r^*_D)}{de^*} \right|$, and we can conclude that for both types $e^*_S > e^*_D$.

We now only have to show that $e^*_{Hc} < e^*_{Lc}$ for $c = \{S, D\}$. Take equation (4.3) and rewrite,

$$\frac{dU(e_c, r_c, \theta)}{de_c} = \frac{dY(e_c, \omega)}{de_c} - \frac{dc(e_c, \theta)}{de_c}$$

Since $\frac{dc(e,L)}{de} > \frac{dc(e,H)}{de}$ by assumption, it follows that although the two types of borrowers are charged the same interest rate, it the case that,

$$\frac{dU(e^*_{Hc}, r_c)}{de^*_{Hc}} > \frac{dU(e^*_{Lc}, r_c)}{de^*_{Lc}}$$

meaning that the marginal value of effort is higher for high-ability borrowers rather than for low-ability ones. It follows that $e^*_{Hc} > e^*_{Lc}$. □

Proof of proposition 6. In proving stochastic dominance for the case $N = 1$ (proposition 4), I demonstrated that regardless of whether $G_{Lt}(D) \lesseqgtr G_{Ht}(D)$ in period t, the equilibrium strategies are such that in the following period $t+1$ it is going to be the case that $G_{Lt+1}(D) \geq G_{Ht+1}(D)$. To prove the current proposition it is enough to apply to each $c_N \in C^N$ the same computations, with the difference that now, the expressions

$$G_{\theta t+1}(S) = G_{\theta t}(S)(1 - F(\bar{\omega}(e_{HS}, r_S)) + G_{\theta t}(D)(1 - F(\bar{\omega}(e_{HD}, r_D))$$
$$G_{\theta t+1}(D) = G_{\theta t}(S)(F(\bar{\omega}(e_{HS}, r_S)) + G_{\theta t}(D)(F(\bar{\omega}(e_{HD}, r_D))$$

become (using a slightly different notation),

$$G_{\theta t+1}(c_{NS}) = G_{\theta t}(c_{SN})(1 - F(\bar{\omega}(e_\theta(c_{SN}), r(c_{SN}))) + G_{\theta t}(c_{DN})(1 - F(\bar{\omega}(e_\theta(c_{DN}), r(c_{DN})))$$
$$G_{\theta t+1}(c_{ND}) = G_{\theta t}(c_{SN})(F(\bar{\omega}(e_\theta(c_{SN}), r(c_{SN})) + G_{\theta t}(c_{DN})(F(\bar{\omega}(e_\theta(c_{DN}), r(c_{DN})))$$

Using the same algebra it is possible to show that regardless of whether $G_{Lt}(c_N) \lesseqgtr G_{Ht}(c_N)$ in period t, in the next period G_{Ht+1} will first-order stochastically dominate G_{Lt+1}. □

Proof of corollary 2. The lemma follows by inspecting (4.7) and applying the same reasoning used in proving corollary 1. □

Proof of lemma 1. The first part of the remark follows from the assumption on marginal costs. Denote with $\bar{c}_N$ the "clean" report of length N with only repayments. Consider the statement,

$$p_{\bar{c}_N} - p_{\bar{c}_{N-1}} > 0 \text{ for any finite } N$$

This is true as long as,

$$\frac{g_H(\bar{c}_N)}{g_H(\bar{c}_{N-1})} > \frac{p_0(g_H(\bar{c}_N) - g_L(\bar{c}_N)) + g_L(\bar{c}_N)}{p_0(g_H(\bar{c}_{N-1}) - g_L(\bar{c}_{N-1})) + g_L(\bar{c}_{N-1})}$$
$$= \frac{p_0 g_H(\bar{c}_N) + (1-p_0)\ g_L(\bar{c}_N))}{p_0 g_H(\bar{c}_{N-1}) + (1-p_0)\ g_L(\bar{c}_{N-1}))}$$

To prove that the inequality above holds for any (finite) N, consider first that, thanks to the assumption on costs and the stochastic dominance result, we have,

$$g_H(\bar{c}_N) = 1 - \sum_{c_N \neq \bar{c}_N} g_H(c_N) > 1 - \sum_{c_N \neq \bar{c}_N} g_L(c_N) = g_L(\bar{c}_N)$$

Further, for any c_N,

$$g(c_{N-1}) = g(c_{SN}) + g(c_{DN})$$

because when the register discloses one period less (moving from N to $N-1$). effectively it collapses into the same density, $g(c_{N-1})$, the densities of the two credit reports of length N equal in everything but in the farthermost outcome, $g(c_{SN})$ and $g(c_{DN})$. So we have that,

$$g_\theta(\bar{c}_N) \leq g_\theta(\bar{c}_{N-1})$$

Together, these results demonstrate the initial inequality holds, and that $p_{\bar{c}_N} - p_{\bar{c}_{N-1}} > 0$ for any finite N (because $p_0 \in (0,1)$).

Building on these results, we have that,

$$p_{\bar{c}_1} < p_{\bar{c}_2} < p_{\bar{c}_3} < ...$$

implying that the longer the history of repayment, the higher the belief in the high ability of the borrower. A similar reasoning can be applied to the report with only defaults, $\underline{c}_N$, to conclude that,

$$p_{\underline{c}_1} > p_{\underline{c}_2} > p_{\underline{c}_3} > ...$$

Thus, putting these two results together,

$$p_{\bar{c}_1} - p_{\underline{c}_1} < p_{\bar{c}_2} - p_{\underline{c}_2} < ... < p_{\bar{c}_{N-1}} - p_{\underline{c}_{N-1}} < p_{\bar{c}_N} - p_{\underline{c}_N}$$

which demonstrates the first part of the lemma.

As for the second part of the proof, we have to show that $p_{c_N^S} - p_{c_N^D}$ decreases in N. This is equivalent to showing that,

$$\frac{p_0 g_H(c_{NS})}{p_0 g_H(c_{NS}) + (1-p_0) g_L(c_{NS})} - \frac{p_0 g_H(c_{ND})}{p_0 g_H(c_{ND}) + (1-p_0) g_L(c_{ND})} \to 0 \text{ as } N \to \infty$$

or, that,

$$g_\theta(c_{NS}) - g_\theta(c_{ND}) \to 0 \text{ as } N \to \infty$$

Note that,

$$g(c_{N-1}) = g(c_{SN}) + g(c_{DN})$$

for any c_N. That is, when the register chooses to disclose one period less (moving from N to $N-1$) effectively it collapses into the same density, $g(c_{N-1})$ the densities of the two credit reports of length N equal in everything but in the farthermost outcome, $g(c_{SN})$ and $g(c_{DN})$. Then, we have that,

$$\begin{aligned} g_\theta(S) - g_\theta(D) &= (g_\theta(SS) + g_\theta(DS)) - (g_\theta(DD) + g_\theta(SD)) \\ &= (g_\theta(SS) - g_\theta(SD)) + (g_\theta(DS) - g_\theta(DD)) \end{aligned}$$

Thus,

$$g_\theta(S) - g_\theta(D) \geq g_\theta(SS) - g_\theta(SD)$$
$$g_\theta(S) - g_\theta(D) \geq g_\theta(DS) - g_\theta(DD)$$

Applying recursively this decomposition, it is easy to prove that $g_\theta(c_{NS}) - g_\theta(c_{ND}) \to 0$ as $N \to \infty$ □

PROOF OF PROPOSITION 7. To prove the proposition consider the results in corollaries (1) and (2). For $N = 1$,

$$\frac{dU(e_S^*, r_S^*, \theta)}{de_S^*} = \frac{F' \frac{d\bar{\omega}(e_S^*, r_S^*)}{de_S^*}}{F' \frac{d\bar{\omega}(e_D^*, r_D^*)}{de_D^*}} \frac{dU(e_D^*, r_D^*, \theta)}{de_D^*}$$

while for any arbitrary choice of N,

$$\frac{dU(e_{c_{SN}}^*, r_{c_{SN}}^*, \theta)}{de_{c_{SN}}^*} = \frac{F' \frac{d\bar{\omega}(e_{c_{SN}}^*, r_{c_{SN}}^*)}{de_{c_{SN}}^*}}{F' \frac{d\bar{\omega}(e_{c_{DN}}^*, r_{c_{DN}}^*)}{de_{c_{DN}}^*}} \frac{dU(e_{c_{DN}}^*, r_{c_{DN}}^*, \theta)}{de_{c_{DN}}^*}$$

Since

$$\frac{F' \frac{d\bar{\omega}(e_S^*, r_S^*)}{de_S^*}}{F' \frac{d\bar{\omega}(e_D^*, r_D^*)}{de_D^*}} > \frac{F' \frac{d\bar{\omega}(e_{c_{SN}}^*, r_{c_{SN}}^*)}{de_{c_{SN}}^*}}{F' \frac{d\bar{\omega}(e_{c_{DN}}^*, r_{c_{DN}}^*)}{de_{c_{DN}}^*}}$$

because $|r_S - r_D| > |r_{c_{SN}} - r_{c_{DN}}|$ and $\bar{\omega}_e < 0$ and $\bar{\omega}_{er} < 0$. it follows also that

$$\frac{\frac{dU(e_S^*, r_S^*, \theta)}{de_S^*}}{\frac{dU(e_D^*, r_D^*, \theta)}{de_D^*}} > \frac{\frac{dU(e_{c_{SN}}^*, r_{c_{SN}}^*, \theta)}{de_{c_{SN}}^*}}{\frac{dU(e_{c_{DN}}^*, r_{c_{DN}}^*, \theta)}{de_{c_{DN}}^*}}$$

meaning that the ratio of the marginal value of effort between a state where a success is about to be erased (numerator) and on where a default is about to be erased (denominator) decreases with N. Then it follows that,

$$\frac{e_S^*}{e_D^*} > \frac{e_{c_{SN}}^*}{e_{c_{DN}}^*}$$

demonstrating that borrowers are most responsive to limited records when $N = 1$. □

PROOF OF PROPOSITION 8. Take any two separating interest rates, r_L^* and r_H^*. We know from Proposition 1 that $r_H^* < r_L^*$ and $e_H^* > e_L^*$. Further, by (3.3) we have that $r_0^* \in (r_L^*, r_H^*)$. Under the assumption $r_0^* < \frac{1}{2}(r_H^* + r_L^*)$, we can write,

$$|r_H^* - r_0^*| < |r_L^* - r_0^*|$$

from which we also have that, for any $e \in [0,1]$,

$$|\bar{\omega}(e, r_0^*) - \bar{\omega}(e, r_H^*)| < |\bar{\omega}(e, r_L^*) - \bar{\omega}(e, r_0^*)|$$

by the concavity of the revenue function. Together, the inequalities above lead to,

$$|e_H^* - e_{0H}^*| < |e_L^* - e_{0L}^*|$$

because due to $\bar{\omega}''_{er}(e,r) > 0$ marginal effort responds more to decreases than increases in the interest rate.

Then, using the convexity of $\bar{\omega}(e_\theta, r_\theta)$ and knowing that, by remark 3, $r_H^* < r_0^* < r_L^*$ implies $e_H^* < e_{H0}^* < e_{L0}^* < e_L^*$ (where the intermediate step follows from the assumption on marginal cost), we also have that,

$$\bar{\omega}(e_{H0}^*, r_0^*) - \bar{\omega}(e_H^*, r_H^*) < \bar{\omega}(e_L^*, r_L^*) - \bar{\omega}(e_{L0}^*, r_0^*)$$

which proves the proposition whenever $p \leq \frac{1}{2}$. □

Proof of proposition 9. The proof begins with showing the following lemma, that lays down the condition under which limited records lead to a lower default rate than full disclosure.

Lemma 3. *When $p \leq \frac{1}{2}$, the policy $N = 1$ induces a lower default rate than full disclosure if $r_0^* < \frac{1}{2}(r_H^* + r_L^*)$ and if the difference $\bar{\omega}(e_{\theta D}^*, r_D^*) - \bar{\omega}(e_{\theta S}^*, r_S^*)$ is small.*

Proof. From the previous proposition we know that,

$$\bar{\omega}(e_L^*, r_L^*) - \bar{\omega}(e_{L0}^*, r_0^*) \geq \frac{p}{1-p}\left(\bar{\omega}(e_{H0}^*, r_0^*) - \bar{\omega}(e_H^*, r_H^*)\right)$$

By the same computations as for (6.1), limited records lead to lower defaults than full disclosure when,

$$\bar{\omega}(e_L^*, r_L^*) - \frac{\bar{\omega}(e_{LS}^*, r_S^*)}{1-(\bar{\omega}(e_{LD}^*, r_D^*) - \bar{\omega}(e_{LS}^*, r_S^*))} \geq \frac{p}{1-p}\left[\frac{\bar{\omega}(e_{HS}^*, r_S^*)}{1-(\bar{\omega}(e_{HD}^*, r_D^*) - \bar{\omega}(e_{HS}^*, r_S^*))} - \bar{\omega}(e_H^*, r_H^*)\right]$$

Thus as long as we can show that,

(9.4)
$$\bar{\omega}(e_{L0}^*, r_0^*) > \frac{\bar{\omega}(e_{LS}^*, r_S^*)}{1-(\bar{\omega}(e_{LD}^*, r_D^*) - \bar{\omega}(e_{LS}^*, r_S^*))}$$
$$\bar{\omega}(e_{H0}^*, r_0^*) > \frac{\bar{\omega}(e_{HS}^*, r_S^*)}{1-(\bar{\omega}(e_{HD}^*, r_D^*) - \bar{\omega}(e_{HS}^*, r_S^*))}$$

the proposition will be demonstrated. By (4.5), we know that,

$$p_D < p_0 < p_S$$

due to the lenders' updating process and the stochastic dominance of G_H over G_L. Then it follows that,

$$r_S^* < r_0^* < r_D^*$$

which implies that $\bar{\omega}(e, r_S^*) < \bar{\omega}(e, r_0^*)$ by the concavity of the revenue function. By remark 4, it also implies that $e_{\theta 0}^* < e_{\theta R}^*$. Then, it follows that,

$$\bar{\omega}(e_{L0}^*, r_0^*) > \bar{\omega}(e_{LS}^*, r_S^*)$$
$$\bar{\omega}(e_{H0}^*, r_0^*) > \bar{\omega}(e_{HS}^*, r_S^*)$$

which together with the assumption that $\bar{\omega}(e_{\theta D}^*, r_D^*) - \bar{\omega}(e_{\theta R}^*, r_S^*)$ is small proves the inequalities (9.4) and hence the lemma.

Lemma 1 demonstrates that, when $N = 0$ is preferable to full disclosure, for $N = 1$ to also induce a lower default rate than full disclosure we need inequalities (9.4) to hold. However, when they are satisfied, $N = 1$ also leads to lower default rate than $N = 0$. In fact, this is the case when the following holds,

$$\bar{\omega}(e_{L0}^*, r_0^*) - \frac{\bar{\omega}(e_{LS}^*, r_S^*)}{1-(\bar{\omega}(e_{LD}^*, r_D^*) - \bar{\omega}(e_{LS}^*, r_S^*))} \geq \frac{p}{1-p}\left[\frac{\bar{\omega}(e_{HS}^*, r_S^*)}{1-(\bar{\omega}(e_{HD}^*, r_D^*) - \bar{\omega}(e_{HS}^*, r_S^*))} - \bar{\omega}(e_{H0}^*, r_0^*)\right]$$

which follows directly form (9.4). □

□

Proof of proposition 10. Having assumed that $\bar{\omega}(e_{\theta D}^*, r_D^*) - \bar{\omega}(e_{\theta S}^*, r_S^*)$ is small, to prove inequality (6.2) we only need to show that $\bar{\omega}(e_L^*, r_L^*) - \bar{\omega}(e_{LS}^*, r_S^*)$ is greater than $\bar{\omega}(e_{HS}^*, r_S^*) - \bar{\omega}(e_H^*, r_H^*)$. When $r_0^* > \frac{1}{2}(r_H^* + r_L^*)$, we also have that, $|r_0^* - r_S^*| > |r_0^* - r_L^*|$ which means that

$$|\bar{\omega}(e, r_0) - \bar{\omega}(e, r_S)| > |\bar{\omega}(e, r_0) - \bar{\omega}(e, r_L)|$$

Further, the assumption on $\bar{\omega}(e_{\theta D}^*, r_D^*) - \bar{\omega}(e_{\theta R}^*, r_S^*)$ implies that $\bar{\omega}(e, r_D) - \bar{\omega}(e, r_S)$ are close, which mean that,

$$|\bar{\omega}(e_L^*, r_L^*) - \bar{\omega}(e_{LR}^*, r_S^*)| > |\bar{\omega}(e_{HS}^*, r_S^*) - \bar{\omega}(e_H^*, r_H^*)|$$

Then if $\bar{\omega}(e_{\theta D}^*, r_D) - \bar{\omega}(e_{\theta S}^*, r_S)$ is small enough, inequality (6.2) holds. □

Proof of lemma 2. Let us look first at the case $N \to \infty$. According to Bayes' law the lenders will eventually learn each borrower's type correctly. Once this happens, $p(c_{iN}) = 1$ iff $\theta_i = H$. Using this correct belief, the minimum interest rate $\hat{r}$ that a lender will offer in equilibrium to a borrower of high type can be derived from (4.9)

$$
\begin{aligned}
&1 \cdot \int_0^{\bar{\omega}(e_H^*,\hat{r})} Y(e_H,\omega)dF(\omega) + (1 - F(\bar{\omega}(e_H,\hat{r}))(1+\hat{r}) \\
&+ 0 \cdot \int_0^{\bar{\omega}(e_L^*,\hat{r})} Y(e_L,\omega)dF(\omega) + (1 - F(\bar{\omega}(e_L,\hat{r}))(1+\hat{r}) \\
&= 1 + \rho
\end{aligned}
$$

This equation is identical to equation (3.2) which describes the lenders' optimal strategy in the case of full disclosure, $\hat{r} = r_H^*$. Similarly, once a borrower is discovered to be of the low type, the interest rate charged to him is going to be the same as in the full disclosure case, r_L^*. Since the borrowers anticipate these interest rates, their effort is again identical to the one exerted in the case of full disclosure. Then, as $N \to \infty$, $W_G = W_C$.

We can now look at $N = 0$. In this case, $p(c_{i0}) = p_0$. Thus, lenders will offer the interest rate that solves,

$$
\begin{aligned}
&p_0 \cdot \int_0^{\bar{\omega}(e_H^*,r_0)} Y(e_H,\omega)dF(\omega) + (1 - F(\bar{\omega}(e_H,r_0))(1+r_0) \\
&+ (1-p_0) \cdot \int_0^{\bar{\omega}(e_L^*,r_0)} Y(e_L,\omega)dF(\omega) + (1 - F(\bar{\omega}(e_L,r_0))(1+r_0) \\
&= 1 + \rho
\end{aligned}
$$

which delivers the same pooling interest rate as in the full withholding case, $r_0 = r_0^*$. It follows that $W_G = W_0$ when $N = 0$. □

PROOF OF PROPOSITION 11. The proof unfolds in three steps. In order to simplify the notation, I suppress the asterisks that denote equilibrium variables (such as e_θ^*, r_θ^*), with the convention that variables are understood in equilibrium unless otherwise stated.

Step (i) I begin with considering the full disclosure equilibrium, to show that a marginal increase in r_H leads to a marginal decrease in r_L, holding the lenders' aggregate return constant. I suppress the time index, with the understanding that we are referring to the same period t. The lenders' *aggregate* break even (BE) condition reads,

$$
p\Pi(\bar{\omega}(e_H,r_H), e_H, r_H) + (1-p)\Pi(\bar{\omega}(e_L,r_L), e_L, r_L) = 1 + \rho
$$

and its total derivative is given by,

$$dBE = p\left(\frac{dBE}{d\Pi}\frac{d\Pi}{d\bar{\omega}}\frac{d\bar{\omega}}{de_H}\frac{de_H}{dr_H} + \frac{dBE}{d\bar{\omega}}\frac{d\bar{\omega}}{de_H}\frac{de_H}{dr_H} + \frac{dBE}{de_H}\frac{de_H}{dr_H} + \frac{dBE}{dr_H}\right)dr_H + (1-p)\left(\frac{dBE}{d\Pi_L}\frac{d\Pi_L}{d\bar{\omega}}\frac{d\bar{\omega}}{de_L}\frac{de_L}{dr_L} + \frac{dBE}{d\bar{\omega}}\frac{d\bar{\omega}}{de_L}\frac{de_L}{dr_L} + \frac{dBE}{de_L}\frac{de_L}{dr_L} + \frac{dBE}{dr_L}\right)dr_L \tag{9.5}$$

Setting $dBE = 0$, in order to hold the return for the lenders constant, the above can be rearranged into,

$$\left.\frac{dr_L}{dr_H}\right|_{Lenders} = -\frac{p}{1-p}\frac{\left(\frac{dBE}{d\Pi}\frac{d\Pi}{d\bar{\omega}}\frac{d\bar{\omega}}{de_H}\frac{de_H}{dr_H} + \frac{dBE}{d\bar{\omega}}\frac{d\bar{\omega}}{de_H}\frac{de_H}{dr_H} + \frac{dBE}{de_H}\frac{de_H}{dr_H} + \frac{dBE}{dr_H}\right)}{\left(\frac{dBE}{d\Pi_L}\frac{d\Pi_L}{d\bar{\omega}}\frac{d\bar{\omega}}{de_L}\frac{de_L}{dr_L} + \frac{dBE}{d\bar{\omega}}\frac{d\bar{\omega}}{de_L}\frac{de_L}{dr_L} + \frac{dBE}{de_L}\frac{de_L}{dr_L} + \frac{dBE}{dr_L}\right)} \tag{9.6}$$

which describes how r_L changes following an increase in r_H. Next, I proceed with computing separately the terms within brackets. In doing so, I will suppress the type index because each term is referring to the same type. For the time being, I will also suppress the p, to reintroduce them later. Beginning with $\frac{dBE}{d\Pi}\frac{d\Pi}{d\bar{\omega}}\frac{d\bar{\omega}}{de}\frac{de}{dr}$, it turns out that this term disappears in both numerator and denominator, since,

$$\frac{d\Pi}{d\bar{\omega}} = Y(e,\bar{\omega}(e,r))f(\bar{\omega}(e,r)) - (1+r)f(\bar{\omega}(e,r)) = 0$$

where I made use of the fact that $dF(\omega) = f(\omega)d(\omega)$, and because by assumption $Y(e,\bar{\omega}(e,r)) = (1+r)$. I now consider $\frac{dBE}{d\bar{\omega}}\frac{d\bar{\omega}}{de}\frac{de}{dr}$. First note that the shock $\bar{\omega}$ affects the BE condition only through Π. Then, the derivative can be rewritten as $\frac{d\Pi}{d\bar{\omega}}\frac{d\bar{\omega}}{de}\frac{de}{dr}$ and by the same reasoning as before, this is equivalent to zero. I now turn my attention now to $\frac{dBE}{de}\frac{de}{dr}$. The effort e of type θ affects the BE only via Π_θ, so that,

$$\frac{dBE}{de} = \frac{d}{de}\left(\int_0^{\bar{\omega}(e(r),r)} Y(e,\omega)dF(\omega) + \int_{\bar{\omega}(e(r),r)}^{1} (1+r)dF(\omega)\right)$$

$$Y(e,\bar{\omega})f(\bar{\omega})\frac{d\bar{\omega}(e,r)}{de} + \int_0^{\bar{\omega}(e,r)} Y_e'(e,\omega)df(\omega)d\omega - (1+r)f(\bar{\omega})\frac{d\omega^*(e,r)}{de}$$

$$= \int_0^{\bar{\omega}(e,r)} Y_e'(e,\omega)df(\omega)d\omega$$

by Leibniz rule. Since $\frac{de}{dr} = e_r(r)$, I can write,

$$\frac{dBE}{de}\frac{de}{dr} = \int_0^{\bar{\omega}(e,r)} Y_e'(e,\omega)dF(\omega)e_r(r)$$

Similarly I can compute $\frac{dBE}{dr}$ as,

$$\frac{dBE}{dr} = \frac{d}{dr}\left(\int_0^{\bar{\omega}(e(r),r)} Y(e,\omega)dF(\omega) + \int_{\bar{\omega}(e(r),r)}^{1} (1+r)dF(\omega) \right)$$

The first derivative can be found by using the chain rule,

$$\begin{aligned}\frac{d}{dr}\int_0^{\bar{\omega}(e(r),r)} Y(e,\omega)f(\omega)d\omega &= Y(e,\bar{\omega}(e(r),r))f(\bar{\omega}(e(r),r))\bar{\omega}'_r(e(r),r)) \\ &= (1+r)\bar{\omega}'_r(e(r),r))f(\bar{\omega}(e(r),r)))\end{aligned}$$

The second derivative can instead be computed using the Leibniz' rule,

$$\begin{aligned}\frac{d}{dr}\int_{\bar{\omega}(e(r),r)}^{1} (1+r)f(\omega)d\omega &= -f(\bar{\omega}(e(r),r))\bar{\omega}'_r(e(r),r)(1+r) + \int_{\bar{\omega}(e(r),r)}^{1} f(\omega)d\omega \\ &= -(1+r)\bar{\omega}'_r(e(r),r)f(\bar{\omega}(e(r),r))) + 1 - F(\bar{\omega}(e(r),r))\end{aligned}$$

Thus,

$$\frac{dBE}{dr} = 1 - F(\bar{\omega}^*(e(r),r))$$

Merging the results above we get that for each θ,

$$\begin{aligned}&\left(\frac{dBE}{d\Pi}\frac{d\Pi}{d\bar{\omega}}\frac{d\bar{\omega}}{de}\frac{de}{dr} + \frac{dBE}{d\bar{\omega}}\frac{d\bar{\omega}}{de}\frac{de}{dr} + \frac{dBE}{de}\frac{de}{dr} + \frac{dBE}{dr} \right) \\ &= \hat{Y} + (1 - F(\omega^*(e(r),r)))\end{aligned}$$

where $\hat{Y} = \int_0^{\bar{\omega}(e,r)} Y'_e(e,\omega)dF(\omega)e_r(r)$. This expression describes lenders' marginal return with respect to r_θ and is positive in equilibrium. Reintroducing the types' indexes, equation (9.6) now reads,

$$\left.\frac{dr_L}{dr_H}\right|_{Lenders} = -\frac{p_0}{1-p_0}\frac{\hat{Y}_H \; + (1 - F(\bar{\omega}(e_H(r_H),r_H)))}{\hat{Y}_L \; + (1 - F(\bar{\omega}(e_L(r_L),r_L)))} < 0 \tag{9.7}$$

demonstrating that in the full disclosure equilibrium a marginally higher value of r_H necessarily implies a marginal reduction in r_L.

Step (ii) Next, I derive the conditions under which a marginal increase in r_H has a positive marginal impact on the borrowers' aggregate welfare. Aggregate borrowers' welfare $\mathcal{W}_b$ in a group of $(1-p_0)N$ low-ability borrowers and $p_0 N$ high-ability borrowers can be written as,

$$\mathcal{W}_b = p_0 NU(\bar{\omega}(e_H,r_H),e_H,r_H) + (1-p_0)NU(\bar{\omega}\,(e_L,r_L)\,,e_L,r_L)$$

where,

$$U(e,r,\theta)=\max_{e\in[0,1]}\int\limits_{\bar{\omega}(e_\theta,r)}^{1} Y(e_\theta,\omega)-(1+r_\theta)dF(\omega)-c(e,\theta)$$

Writing down the total derivative and rearranging, we get,

$$\begin{aligned}\frac{U^A}{dr_H}&=p_0N\left(\frac{dU}{d\bar{\omega}}\frac{d\bar{\omega}}{de_H}\frac{de_H}{dr_H}+\frac{dU}{de_H}\frac{de_H}{dr_H}+\frac{dU}{dr_H}\right)\\&+(1-p_0)N\left(\frac{dU}{d\bar{\omega}}\frac{d\bar{\omega}}{de_L}\frac{de_L}{dr_L}+\frac{dU}{de_L}\frac{de_L}{dr_L}+\frac{dU}{dr_L}\right)\frac{dr_L}{dr_H}\end{aligned}$$

We can now make use of equation (9.7) and (3.1) to rearrange and simplify, yielding,

(9.8)

$$\begin{aligned}\frac{U^A}{dr_H}&=\frac{p_0N}{\hat{Y}_L+(1-F(\bar{\omega}(e_L(r_L),r_L)))}\\&\left(\hat{Y}_H\left[(1-F(\bar{\omega}(e_L,r_L)))+c_e'(e_L)e_r(r_L)\right]-\hat{Y}_L\left[(1-F(\bar{\omega}(e_H,r_H)))+c_e'(e_H)e_r(r_H)\right]\right.\\&\left.+(1-F(\bar{\omega}(e_L,r_L)))\frac{dU}{de_H}\frac{de_H}{dr_H}-(1-F(\bar{\omega}(e_H,r_H)))\frac{dU}{de_L}\frac{de_L}{dr_L}\right)\end{aligned}$$

where as before where $\hat{Y}_\theta$ has been defined above.

When equation (9.8) is positive, raising the interest rate for the high-ability group (and thus lowering it for the low-ability ones) results in higher aggregate utility for borrowers, while leaving lenders' expected returns constant.

I have already argued in step (i) that the denominator is positive. It remains to establish the conditions under which the term within brackets is non-negative. This is the case when,

$$\begin{aligned}&\hat{Y}_H\left[(1-F(\bar{\omega}(e_L,r_L)))+c_e'(e_L)e_r(r_L)\right]+(1-F(\bar{\omega}(e_L,r_L)))\frac{dU}{de_H}\frac{de_H}{dr_H}\\&\geq\hat{Y}_L\left[(1-F(\bar{\omega}(eH,r_H)))+c_e'(e_H)e_r(r_H)\right]+(1-F(\bar{\omega}(e_H,r_H)))\frac{dU}{de_L}\frac{de_L}{dr_L}\end{aligned}$$

or, rearranging, when,

$$\begin{aligned}&(1-F(\omega^*(e(r_L),r_L)))\left(\hat{Y}_H+\frac{dU}{de_H}\frac{de_H}{dr_H}\right)+\hat{Y}_Hc_e'(e_L)e_r(r_L)\\&\geq(1-F(\omega^*(e(r_H),r_H)))\left(\hat{Y}_L+\frac{dU}{de_L}\frac{de_L}{dr_L}\right)+\hat{Y}_Lc_e'(e_H)e_r(r_H)\end{aligned}$$

Consider first $\hat{Y}_Hc_e'(e_L)e_r(r_L)$ and $\hat{Y}_Lc_e'(e_H)e_r(r_H)$, and rearrange and expand the first term to $\int\limits_{\bar{\omega}(e_H,r_H)}^{1} Y_e(e(r_H),\omega_t)dF(\omega)e_r(r_H)c_e'(e_L)e_r(r_L)$ and proceed in the same way with the second. Since in equilibrium $r_H < r_L$, $e_t(r_H) > e_t(r_L)$, and $c_e'(e, L) > c_e'(e, H)$ for all $e \in [0, 1]$ it follows that,

$$\hat{Y}_H c_e'(e_L) e_r(r_L) > \hat{Y}_L c_e'(e_H) e_r(r_H)$$

Since we are after a sufficient, not necessary, condition, we can focus on when the following holds,

$$\hat{Y}_H + \frac{dU}{de_H}\frac{de_H}{dr_H} \geq \mathcal{F}(\bar{\omega}_H, \bar{\omega}_L)\left(\hat{Y}_L + \frac{dU}{de_L}\frac{de_L}{dr_L}\right) \tag{9.9}$$

where $\mathcal{F}(\bar{\omega}_H, \bar{\omega}_L) = \frac{1-F(\bar{\omega}(eH, r_H))}{1-F(\bar{\omega}(e_L, r_L))} > 1$. Since all the terms in the RHS are larger than their counterparts in the LHS, and since$\frac{1-F(\bar{\omega}(eH, r_H))}{1-F(\bar{\omega}(e_L, r_L))}$ satisfies the model's hypotheses, a sufficient condition for $\frac{U^A}{dr_H} > 0$ is given by constraining $\frac{1-F(\bar{\omega}(e_H, r_H))}{1-F(\bar{\omega}(e_L, r_L))}$ to be greater than one, but not too large.

Step (iii) When there is incomplete information in the market, lenders in equilibrium offer a single contract r_0^*, which falls in $r_H^* < r_0^* < r_L^*$. When the full information interest rates $\langle r_\theta^* \rangle$ satisfy the CScondition, any r_0^* such that $(r_0^* - r_H^*)$ and $(r_L^* - r_0^*)$ is compatible with equation (9.7), incomplete information increases ex ante aggregate welfare. □

Proof of corollary 4. The first part of this corollary follows from corollary 1. The second follows from proposition 10, since,

$$\frac{1 - F(\bar{\omega}(e_H^*, r_H^*))}{1 - F(\bar{\omega}(e_L^*, r_L^*))} > 1 \text{ but not too large}$$

may be taken to imply that, under a uniform distribution of the shock,

$$\bar{\omega}(e_L^*, r_L^*) - \bar{\omega}(e_H^*, r_H^*) > 1 \text{ but close}$$

which also implies,

$$\bar{\omega}(e_{\theta D}^*, r_D) - \bar{\omega}(e_{\theta S}^*, r_S) \text{ small}$$

Then, under CS, proposition 10 holds, and the result obtains. □

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 verksamhetsutveckling.*
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 Euro-pean 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.*

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.*