Can Contract Theory Explain Social Preferences?*

W. Bentley MacLeod

Department of Economics

Mail Code 3308

Columbia University

420 West 118th Street

New York, NY

Tel: 212-854-4212

Fax:212-280-6304

Email: wbm2103@columbia.edu

Session: Contracts and Fairness

Chair: Ulrike Malmendier, Stanford

Discussant: Robert Gibbons, MIT

January 10, 2007

Can Contract Theory Explain Social Preferences?

W. Bentley MacLeod

Columbia University and IZA

For several decades, a growing body of research has shown that humans do not always choose to maximize material payoffs. Economists following the lead of psychologists Daniel Kahneman and Amos Tversky (1979) and Matthew Rabin (1993) have built on such research to suppose that individuals are attentive to fair distribution rewards between themselves as well as personal payoffs. (Ernst Fehr and Klaus Schmidt (1999)) An alternative approach, suggested by Elizabeth Hoffman, Kevin McCabe and Vernon Smith (1996) argue that laboratory subjects that perceive a potential for future interaction as approximated by social distance act on a preference for reciprocity.

Both approaches capitalize on the power of psychology to enhance understanding of economic exchanges between particular people in specific laboratory conditions. Yet, psychology is not only concerned with the problem of explaining behavior, but it is also in the business of *modifying* behavior. On of the shelves of the psychology section of any bookstore one will find numerous works devoted to helping individuals modify their behavior so that they drink and smoke less, and learn to have more rewarding social interactions at work and in private life.

Economists suppose that individuals maximize rewards not because they believe that people do so in every case, but because the maximizing supposition provides a useful unified model of behavior in designing better economic institutions. The model not only rationalizes the profit motive, but explains how even successful institutions may be destroyed by the common human desire to maximize possessions–if that leads to malfeasance, theft, or general corruption.

In this note I suggest that economics may also learn more from psychology about positive behaviors – such as the complex dynamics of honesty, fairness, and trust. Rather than insist

that people are endowed with a stable set of unchanging preferences, we may ask instead how small modifications in preferences can lead to significant improvements in economic performance.

Specifically, I suggest that contract theory can be a starting point for a larger enquiry into exactly how preferences might be modified to enhance market performance. I consider two substantive themes. First I show how contract theory may help to anchor a theory of fairness. Second, I show that a small taste for trustworthiness can lead to a large increase in cooperation in a relational contract.

Fairness is a crucial theme because notions about it are decisive in determining whether a party has breached an agreement. For example, an authoritative and widely used treatise on contract law by Edward A. Farnsworth has twenty index entries under the heading "Good Faith and Fair Dealing". The Uniform Commercial Code of the United States requires all parties to a contract to act in "good faith," defined to mean "the observance of reasonable commercial standards of fair dealing. "

Yet "fair" is never precisely defined. The next section outlines a model of fairness developed with Lorne Carmichael (Carmichael and MacLeod (2003)). Following the insights of Williamson (1975) and Hart (1995), we suggest that one can develop a concept of fairness based upon the idea that it is optimal to reward sunk investment, and hence "fair" bargains should take this into account. This may help explain the observations of Kahneman, Knetsch and Thaler (1986) and many others who identify the *endowment effect* - individuals value more highly assets they own than those with similar attributes that they must purchase. Contract theory predicts that this effect arises from the investment, either psychic or pecuniary, that owner has made into the asset that increases the value from ownership.

My second point begins with the observation that we invest in teaching our children to abide by their obligations as a matter of principle, a behavior that makes little sense if we expect them to simply maximize their material payoff. Such training will never produce completely trustworthy offspring, but it may be expected to produce children who experience

some remorse or disutility from breaking promises. ("You're grounded.") In section II, I modify a behavioral model introduced by Hart and Holmström (1987) that supposes a few individuals are inherently honest but most not. We relax this assumption and suppose that all individuals experience *some* disutility from breaching a contract.

Section II shows that a small amount of remorse is enough to support a high level of cooperation in a long term relationship. Moreover, the theory makes some predictions on the form of the optimal contract that are consistent with available evidence, such as the work of Ernst Fehr, Simon Gachter and George Kirchsteiger (1997). They find, as our model predicts, that reciprocity plays an important role in supporting cooperative behavior in a long term relationship.

I. Fairness

Students in first year economics are taught that opportunity costs are important, but sunk costs – already paid, unrecoverable, and irrelevant to current or future choices – are not. Students often find this lesson counter-intuitive, even though rational choice theory predicts that they have been making optimal choices all their lives.

In fact we know that even costs that have been sunk by other people seem to affect behavior. In a well-known survey, Kahneman et al. (1986) asked people whether they thought it would be fair for a store to increase its prices under various circumstances. They found that individuals are unwilling to accept a higher price due to higher demand, but accept higher prices due to higher input costs. For a hardware store to raise prices for snow shovels after a serious snowstorm is considered gouging, but it seems to be perfectly acceptable to raise prices if the cost of shovels rose even if they were already bought and on the shelves.

In Carmichael and MacLeod (2003) we introduce a model in which risk neutral agents meet randomly in pairs. Each pairing involves a productive opportunity followed by a bargain over the division of the surplus created. In the *ex ante* stage, agents independently make a decision as to the character of an investment. In the *ex post* stage, the parties meet

and bargain over the resulting surplus. We assume the matching process is efficient, and concentrate on the events that occur within a match. In particular, we focus on the norms that might govern bargaining in the *ex post* period.

A bargaining strategy in the model is a mapping from the *ex post* state of the world to a "fair demand" for the agent. The *ex post* state is known to the parties, but need not be verifiable. "What's fair" can therefore depend on things that would be impossible to establish in a court of law. Nonetheless, an agent who deviates from the norm may be subject to costly sanctions.

More formally, the stages in the history of a match for a pair are outlined below.

1. Two agents $i \in \{1, 2\}$ make private *ex ante* decisions I_i that affect the distribution of realized costs, and the surplus from trade via a state ω. It is assumed that decision I_i affects her own costs, $C^i(\omega)$, but not her partner's costs $C^j(\omega)$, $j \neq i$. Both decisions affect surplus from future trade, denoted $S(\omega)$. Costs and the surplus are observable by both parties.

2. Each agent plays a Nash demand game and chooses his or her fair share d^i of S. Agent i's payoff is:

$$V^i(I_1, I_2, \omega) = \begin{cases} -C^i(\omega) \text{ if } S(\omega) - (d^i + d^j) < 0 \\ d^i + (S(\omega) - d^i + d^j)/2 - C^i(\omega), \text{ if not.} \end{cases} \tag{1}$$

If demands are less than the total surplus, then each agent gets their demand, an equal share of the remaining surplus, less their costs. If demands exceed total surplus, then each agent gets none of the surplus, but is responsible for their out-of-pocket costs. Carmichael and MacLeod (2003) show that the following "fair share" rule is, under appropriate conditions, the unique efficient equilibrium in this model.

Definition 1. *The* fair share *rule is defined by:*

$$d^i = \textit{ sunk costs paid by } i \ + \ \textit{an equal share of the net surplus,} \tag{2}$$

$$= C^i + \frac{(S - C^i - C^j)}{2}. \tag{3}$$

At the *ex post* stage, any division of the surplus is efficient, and as Ken Binmore (1994) highlights, different solutions can be interpreted as different possible social norms. Yet, most economic exchange entail some sunk costs. A plumber on a home visit will typically not set the price in advance, even though upon arrival there may be no alternative job available. The homeowner might try to offer the plumber a lower price, yet the likely consequence is that he will be upset and simply leave. Such behavior, while *ex post* inefficient, does ensures the plumber can earn a competitive return upon his investment in skills and trave time to the job.

The theory provides terms to the agreement that parties would have signed had they met before making a sunk investment. In addition to providing a normative benchmark, the results are consistent with the effects documented by Kahneman et al. (1986). They find that buyers may be unwilling to patronize sellers who increase prices due to a shift in market demand that generate a pure rent to the seller, yet they are willing to accept price increases arising from higher costs.

II. Reciprocity and a Taste for Trustworthiness

A contract is an instrument between two parties that specifies each party's obligations. In traditional societies a contract, or a *covenant,* often had an elaborate set accompanying rituals. Parties might be expected to take an oath, or place their seal upon a document.[2] All such rituals were imperfect mechanisms to ensure that parties would understand their agreement and feel a sense of *obligation* to perform. Yet the time and effort they took suggests their importance. Rituals are in fact conscious efforts to change each party's preferences to

instill a sense of obligation that a covenant will be executed as agreed.

Realistically, all enforcement is imperfect, and must rely upon some level of trustworthiness. The question is how much. I shall show that if upon entering an agreement a party feels or experiences some loss of utility from breaching an agreement, this greatly enhances the performance of the relationship. In particular, in a finitely repeated relationship, parties can achieve close to the first best if one party has some preference for performing the contract.

Consider a buyer and seller who meet f times (the frequency of the relationship) over a unit length of time (say a year). For simplicity, there is no discounting, and periods are denoted by $\tau = 0, 1, 2, .., f-1$ corresponding to time $\tau\delta$ after the start of the relationship, where $\delta = 1/f$ is the length of one trading period. The goal of the relationship is to trade a good of flow quality $q_\tau \geq 0$ in exchange for a flow price p_τ in period τ. The payoff of the buyer in period t is $U_t^B = \sum_{\tau=t}^{f-1} \delta \left(v\left(q_\tau\right) - p_\tau \right)$, while the seller has payoff $U_t^S = \sum_{\tau=t}^{f-1} \delta \left(p_\tau - q_\tau \right)$, where $v\left(q\right)$ is the value of quality q. This value function is assumed to satisfy $v\left(0\right) = 0, v' > 0, v'' < 0$. The efficient level of quality is $q^* > 0$, and it is given by the unique solution to $v'\left(q^*\right) = 1$.

Let us suppose that the best alternative for the seller is zero, while the buyer has all the bargaining power. This is modelled by supposing each period the buyer makes a take-it-or-leave-it offer to the seller for one period of trade. It is assumed that there are no enforceable contracts. Rather, the role of the contract is to specify the obligations of the parties in the subsequent period.

A. Contracts: Pay in Advance Contract

Two contracts are possible. The first is the pay-in-advance contract denoted by $\{p_t, q_t\}$. At the beginning of period t the buyer pays a flow price p_t, and the seller promises to deliver a flow of q_t at the end of the period.[3] Further suppose that this contract induces social preferences in the seller such that she or he will suffer a flow utility loss of u_c should the

seller breach the agreement. It is assumed that $u_c > 0$, and significantly smaller than the efficient level of quality q^*.

Given this, we can solve for the optimal contract by backwards induction. Let $\hat{U}_t^B$ and $\hat{U}_t^S$ be the equilibrium utility of the buyer and seller respectively in period t. The dynamic programming algorithm can be used to derive the optimal contract in period t :

$$\hat{U}_t^B = \max_{q_t, p_t} \delta\left(v\left(q_t\right) - p_t\right) + \hat{U}_{t+1}^B \text{ subject to:}$$

$$\text{IR: } \delta\left(p_t - q_t\right) + \hat{U}_{t+1}^S \geq 0,$$

$$\text{IC: } \delta\left(p_t - q_t\right) + \hat{U}_{t+1}^S \geq \delta\left(p_t - u_c\right).$$

The first constraint is the requirement that the seller get at least his or her alternative payoff. The second constraint incorporates the seller's preference to perform. If the seller does not perform, the buyer ceases all trade because they believe the seller is untrustworthy. At the same time, the seller suffers a utility lost δu_c. The solution is as follows. The relationship stops in the last period, $f-1$, and therefore $\hat{U}_f^S = \hat{U}_f^B = 0$. From the IC constraint this implies that $q_{f-1} \leq u_c$. The IR constraint implies that the buyer will not offer more than the seller's alternative payoff, and hence we conclude $\hat{U}_{f-1}^S = 0$ and $p_{f-1} = q_{f-1} = u_c$. The concavity of $v\left(q\right)$ ensures that it is always profitable, even when the quantity is very small. Working backwards, it follows that the unique solution is $p_t = q_t = u_c$ and $\hat{U}_t^S$ for all $t = 0, ..., f-1$. Thus, under this solution, quality is completely determined by the seller's preference to perform.

B. Contracts: Payment after Delivery

If instead of having the seller reciprocate with quality, suppose that the seller delivers the good, and then sends an invoice to the buyer, who then decides whether to pay or not. This contract is denoted by $\{q_t, p_t\}$. Should the seller deliver $\hat{q}_t < q_t$, then the buyer has no obligation to make any payment. If at any time during the relationship the buyer does not

reciprocate with a payment p_t, then the seller would choose to cease the relationship. In this case the dynamic programming problem becomes:

$$\hat{U}_t^B = \max_{q_t, p_t} \ \delta\left(v\left(q_t\right) - p_t\right) + \hat{U}_{t+1}^B \text{ subject to:}$$

$$\text{IR: } \delta\left(p_t - q_t\right) + \hat{U}_{t-1}^S \geq 0,$$

$$\text{ICS: } \delta\left(p_t - q_t\right) + \hat{U}_{t-1}^S \geq -\delta\hat{q}_t + \hat{U}_{t-1}^S, \text{ if } \hat{q}_t < q_t$$

$$\text{ICB: } \delta\left(v\left(q_t\right) - p_t\right) + \hat{U}_{t-1}^B \geq \delta\left(v\left(q_t\right) - u_c\right).$$

It will still be the case that the seller gets zero utility, but now ICS implies that she or he will agree to choose any quality satisfying $q_t \leq p_t$. It is also interesting to note that in this formulation the seller does not feel any obligation to perform if they are not paid. This formally captures the Uri Gneezy and Aldo Rustichini (2002) observation that explicit prices (or fines) result in individuals conforming behavior to the letter of the agreement.

Solving this backwards, the solution is characterized by a price and quantity that rise as we move backwards from the endpoint until we reach an amount close to the social optimum. Figure 1 illustrates the efficiency of the relationship as a function of the period. Notice that the simulated output is efficient early in the relationship, and then declines at the end.

On this graph I also plot the results reported in Armin Falk, David Huffman and W. Bentley MacLeod (2007), from an experimental labor market in which the buyer can offer bonus pay to the seller. In that case, the efficiency is relatively flat, then dips down near the end. Hence, the time structure is very similar to the one predicted by the simple model with limited trust. These theoretical and empirical results contrast the predictions from a reputation model with private types, as in Anat Admati and Motty Perry (1991) and Joel Watson (1999). In their models with adverse selection models the level of trade should rise with time, not fall.

III. Conclusions

As Henrich, Boyd, Bowles, Camerer, Fehr, Gintis and McElreach (2001) and Jean Ensminger (2004) document, there is large variation in social preferences across individuals within the same society and across societies. This work also finds that market integration results in more, rather than less importance attached to fairness in exchange. These results are consistent with the hypothesis that social preferences are endogenous, and may be explained by a need to provide individuals with social preferences that enhance trade.

In this note I have shown that in the context of asset ownership, an efficient fairness rule ensures that parties are compensated for their sunk investments. Such a theory may eventually help us understand surplus division norms, and phenomena such as wage rigidity.[4]

In the context of a relational contract, I have shown that when the party with the bargaining power has some taste for honesty, and reciprocates good behavior, parties can achieve close to the first best, with cooperation decreasing as we reach the end of the relationship. As Robert Gibbons (1997) emphasizes, individuals do respond to incentives, and hence these results suggest that in addition to rewarding measured performance, there may also be real efficiency gains from rewarding individuals who value sunk costs and have a taste for trustworthiness. This perspective is also consistent with work in the development literature, such as George Akerlof (1970), Marcel Fafchamps (1996) and Jean-Phillippe Platteau Platteau (2000), that highlights the central role that norms and social preferences play in explaining market performance and economic growth.[5]

REFERENCES

Admati, Anat R. and Motty Perry (1991). "Joint Projects Without Commitment." *Review of Economic Studies.* April **58** (2), 259–76.

Akerlof, George A. (2007). "Presidential Address: The Missing Motivation in Macroeconomics." in "American Economics Association Annual Meeting" Chicago, IL.

——— (1970). "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics.* **84** (3), 488–500.

Binmore, Ken (1994). *Game theory and the social contract. Volume 1. Playing fair*, Cambridge and London: MIT Press.

Carmichael, H. Lorne and W. Bentley MacLeod (2003). "Caring About Sunk Costs: A Behavioral Solution to Hold-Up Problems with Small Stakes." *Journal of Law, Economics and Organization.* Spring **19** (1), 106–118.

Ensminger, Jean (2004). "Market Integration and Fairness: Evidence from Ultimatum, Dictator, and Public Goods Experiments in East Africa." in J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, and H. Gintis, eds., *Cooperation, Reciprocity and Punishment: Experiments in 15 Small-Scale Societies*, London: Oxford University Press, pp. 356–381.

Fafchamps, Marcel (1996). "The enforcement of commercial contracts in Ghana." *World Development.* Mar **24** (3), 427–448.

Falk, Armin, David Huffman, and W. Bentley MacLeod (2007). "New Evidence on Employment Protection, Bonus Pay, and Labor Market Performance." Technical Report, IZA.

Farnsworth, E. Allan (1999). *Contracts, 3nd edition*, Boston: Little, Brown and Company.

Fehr, Ernst and Klaus M. Schmidt (1999). "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics.* **114** (3), 817–68.

——— , Simon Gächter, and Gerd Kirchsteiger (1997). "Reciprocity as a contract enforcement device: Experimental evidence." *Econometrica.* July **65** (4), 833–860.

Gibbons, Robert (1997). "Incentives and Careers in Organizations." in D. M. Kreps and K. F. Wallis, eds., *Advances in Economics and Econometrics: Theory and Applications*, Cambridge, UK: Cambridge University Press, pp. 1–37.

Gneezy, Uri and Aldo Rustichini (2000). "A fine is a price." *Journal of Legal Studies.* Jan **29** (1), 1–17.

Hart, Oliver D. (1995). *Firms, Contracts and Financial Structure*, Oxford, UK: Oxford University Press.

——— and Bengt Holmström (1987). "The Theory of Contracts." in T. Bewley, ed., *Advances in Economic Theory: Fifth World Congress.*, Cambridge, U.K.: Cambridge University Press, pp. 71–155.

Henrich, J., R. Boyd, S. Bowles, S. Camerer, E. Fehr, H. Gintis, and R. McElreach (2001). "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies." *American Economic Review.* **91** (2), 73–78.

Hoffman, Elizabeth, Kenneth McCabe, and Vernon L. Smith (1996). "Social Distance and Other Regarding Behavior in Dictator Games." *American Economic Review.* June **86** (3), 1–8.

Kahneman, Daneil and Amos Tversky (1979). "Prospect Theory: An Analysis of Decisions Under Risk." *Econometrica.* **47**, 262–91.

——— , Jack L. Knetsch, and Richard Thaler (1986). "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." *American Economic Review.* September **76** (4), 728–741.

Platteau, Jean-Phillipe (2000). *Institutions, social norms, and economic development* Fundamentals of development economics ; v. 1, Amsterdam, The Netherlands: Harwood Academic Publishers.

Rabin, Matthew (1993). "Incorporating Fairness Into Game Theory and Economics." *American Economics Review.* December **83** (5), 1281–1302.

Watson, Joel (1999). "Starting Small and Renegotiation." *Journal of Economic Theory.* March **85** (1), 52–90.

Williamson, Oliver E. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications*, New York: The Free Press.

Notes

*I would like to thank Ernst Fehr, Robert Gibbons, Oliver Hart and Abraham Wickelgren for very helpful comments. I am particularly indebted to Lorne Carmichael for our many discussions on these issues.

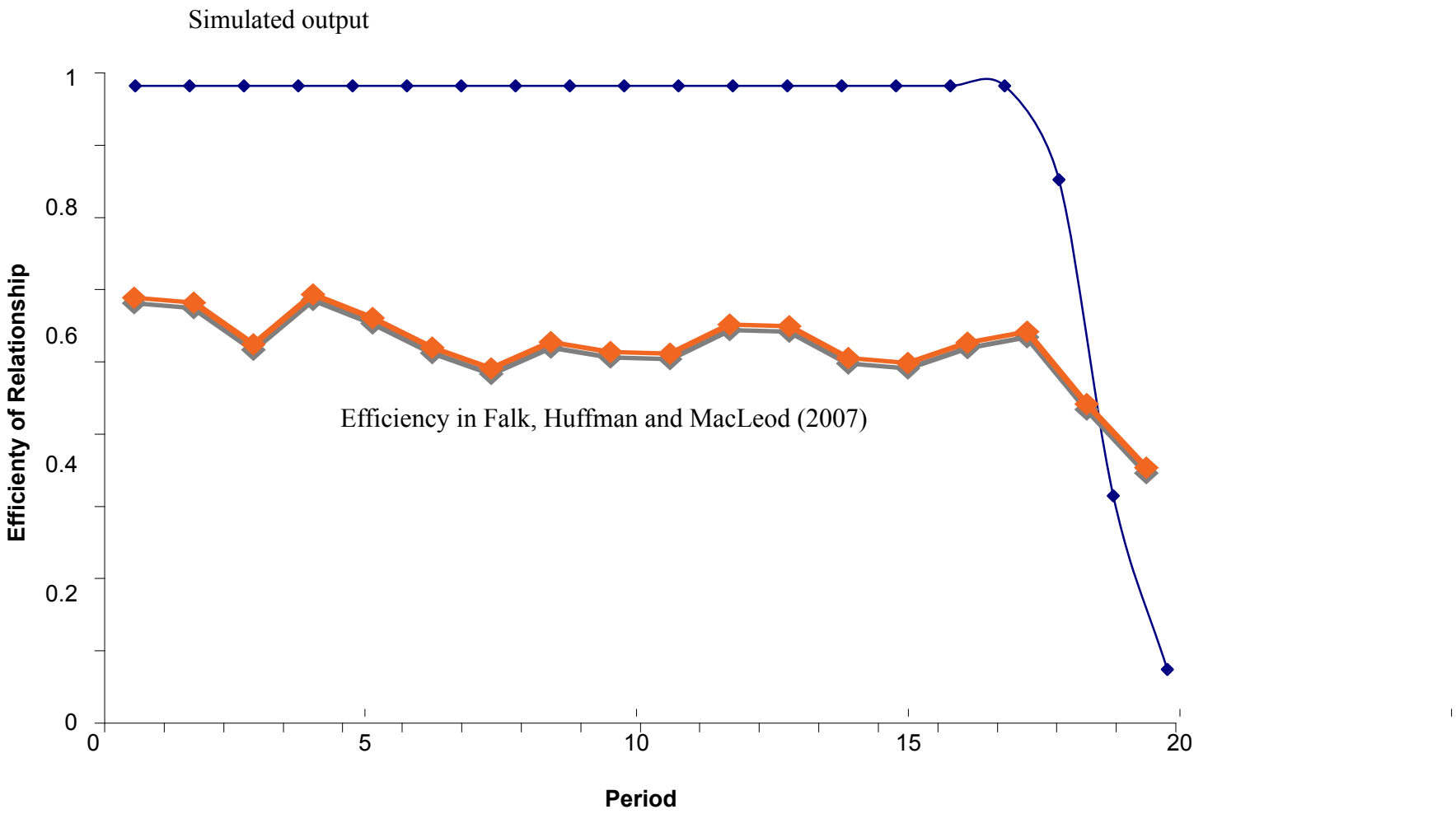
[2]See the Encyclopaedia Britanica entry on "convenant".

[3]Flow terms are used to simplify the analysis when varying the frequency. The total price in period t is therefore δp_t, total cost δq_t, resulting is a value of $\delta v\left(q_t\right)$.

[4]See George Akerlof (2007)'s discussion of the importance of norms for macroeconomics.

[5]Though Akerlof's famous "Market for Lemons" is typically cited for the introduction of the concept of *adverse selection,* the paper is motivated by the problems that arise when there is a lack of *trust* between buyer and seller.

Figure 1: Productive Efficiency as a Function of Time


Simulated output
Efficiency in Falk, Huffman and MacLeod (2007)
Efficienty of Relationship
Period
1
0.8
0.6
0.4
0.2
0
0
5
10
15
20