Transaction Costs and Institutions: Investments in Exchange[*]

Charles Nolan[†] Alex Trew[‡]

March 20, 2015

Abstract

This paper proposes a simple model for understanding transaction costs – their composition, size and policy implications. We distinguish between investments in institutions that facilitate exchange and the cost of conducting exchange itself. Institutional quality and market size are determined by the decisions of risk averse agents and conditions are discussed under which the efficient allocation may be decentralized. We highlight a number of differences with models where transaction costs are exogenous, including the implications for taxation and measurement issues.

JEL Classification: D02; D51; H20; L14.
Keywords: Exchange costs, transaction costs, general equilibrium, institutions.

[*]Forthcoming in *B.E. Journal of Theoretical Economics (Contributions)*. We are grateful to the editor and three referees for comments that have improved this paper. We thank colleagues at the Universities of Edinburgh, Exeter, Koç and St. Andrews, and also at the RES Conference at Warwick University and the ISNIE Conference in Toronto. In particular, we thank, without implicating, Sumru Altug, Vladislav Damjanovic, John Driffill, Oliver Kirchkamp, John Moore, Helmut Rainer, József Sákovics, David Ulph and especially Tatiana Damjanovic, Daniel Danau and Gary Shea.

[†]Department of Economics, University of Glasgow. Tel: +44(0) 141 330 8693. Email: charles.nolan@glasgow.ac.uk.

[‡]Corresponding author: School of Economics and Finance, University of St. Andrews: Castlecliffe, The Scores, St. Andrews, Fife KY16 9AL, Scotland, UK. Tel: +44(0) 1334 461950. Email: alex.trew@st-andrews.ac.uk.

1 Introduction

Models that incorporate transaction costs generally treat them as a 'useful formalism' (Townsend, 1983). They are meant to capture the costs of collecting information, of bargaining, organization, decision making, writing and enforcing contracts between individuals, firms and the state (Coase, 1960). The perception of such costs as exogenously impeding trade or inhibiting the formation of complete contracts suggests that reducing, eliminating or avoiding those costs is generally welfare enhancing.[1] As the quality of institutions is thought to be a part of what explains those transaction costs,[2] the implication is that better institutions always improve economic outcomes.

Many of these transaction costs are not directly impeding trade, however, but are resources allocated to technologies (or institutions) that *facilitate* exchange, even though those resources could otherwise have been allocated directly to the production of a consumption good. For example, the organization of the firm, the formation and nature of contracts, the emergence and use of a legal system are all themselves technologies employed to ease the conduct of exchange. Investments in such technologies – in the form of legal or judiciary arrangements, management consultants, and so on – are investments in an 'institutional capital'. The consequence of such investment is that the cost of an individual exchange can be lower. For example, the expected loss from a trade may be reduced, or it may be less costly to assess the quality of a traded good, if we have established standardized reporting practices; an economy with a stronger contracting environment can limit the losses from opportunistic behavior in trading; and so on. Moreover, the investments in exchange technologies can be private or public. Private investment in such capital could involve the formulation of trading standards within a coalition of traders, for example: There is an upfront cost to establishing and enforcing those standards but these may lower intermediation costs because trading risk is reduced. Public investment may be improve-

[1]See, for example, Greenwood and Jovanovic (1990) and Townsend and Ueda (2006) in relation to finance, growth and inequality; Levchenko (2007) on international trade; and Dixit (1996) on political economy.

[2]See Levchenko (2007) and Acemoglu et al. (2007) for a similar perspective.

ments in property rights legislation that make the transfer of assets more secure; again, such improvements are costly but they may reduce the costs of individual trades if it leads to fewer losses from disputed exchanges. Each of these costs may be categorized as 'transaction costs' but they can serve quite different purposes: Some are investments that facilitate exchange; some, such as trading risk or legal fees, are the subsequent costs of exchange.

We develop a model in which risk-averse agents who do not know their own production technology may, in advance of the productivity realization, form a coalition to share consumption risk. Agents face a cost to exchange output, however, and that cost of exchange is determined by investment in exchange cost-reducing institutions. Total transaction costs are then the sum of two components: There is a cost to forming the public and private institutions that govern, ex ante, the terms of exchange, and there are costs to *conducting* exchange once the state of nature is resolved.[3] Agents choose the resources allocated to reducing exchange costs *and* the extent of diversification (how extensively they will trade with others).

A number of results follow from modelling transaction costs as an endogenous component of a general equilibrium set-up. We first characterize the optimal allocation. Naturally, while the costs of exchange can be can be too high, they can also be *too low.* A high exchange cost reflects fewer resources directed towards facilitating transactions but may be associated with greater expected utility if those free resources are put to productive use and if agents choose to make fewer costly exchanges. Understanding these issues is directly important for policy design since many public institutions, such as the legal system, are bound up in the costs of trade. Real-world policies are generally based around simple objectives such as maximizing the size of markets or minimizing the cost of an individual exchange. Given the absence of a framework in which to account for the general equilibrium consequences of transaction costs, we cannot understand the welfare-ranking of different simple policies. Having established the optimal allocation, and since our

[3]Throughout, an 'exchange cost' is the cost of conducting a particular exchange and what we refer to as the 'transaction cost' is the sum of investments in institutions and the subsequent costs of exchanges that occur.

model can be considered quantitatively, we can conduct such an analysis.

By far the most damaging type of simple policy, in welfare terms, is one that focuses on minimizing the costs of individual exchange. The intuition is as follows: The optimal allocation represents a trade-off between the portion of the endowment that goes to production (i.e., net of transaction costs) and the amount of consumption variability. The minimum-exchange cost policy is so damaging because it ignores both consumption variability and the overall costs of transactions. A policy that targets market size is less damaging since it minimizes the overall consumption variability and allows agents to respond in their private investment decisions. The least damaging of the simple policies we consider is a policy which minimizes the overall size of the transaction sector. Under this policy, agents can respond to deviations from an optimal tax policy by varying both the amount they trade and their private investments in exchange.

In addition to policies that distort allocations to the institutional capital, we also consider the effect of a transactions tax. Agents invest more in institutions to ameliorate the effect of the tax on the costs of exchange, thereby making diversification decisions less sensitive to increases in the transactions tax. However, while apparently robust to the imposition of a transaction tax, agents opt for autarky at a lower transaction tax than might be anticipated using a model of exogenous transaction costs.

We can also use our quantitative model to put the empirical evidence on transaction costs into some context. Coase (1992, p.716) argues that "a large part" of economic activity is directed at alleviating transaction costs. In a first attempt to quantify the aggregate extent of such resources, Wallis and North (1986) estimate that what they term the transaction sector comprised half of US GNP in 1970, a proportion which had *grown* significantly over the preceding century. Moreover, Klaes (2008) concludes that economies with less sophisticated transaction sectors appear, at an aggregate level, to exhibit lower levels of transaction costs. In our model, a more wealthy economy is characterized by a smaller transaction sector, but one based on greater investments in institutions, larger markets and lower exchange costs. This is consistent with the evidence in Klaes (2008), that, at a micro-level, the costs of exchange are high when the aggregate transaction sector appears to be small.

Our framework relates to a number of other papers. The idea that investments in better transaction technologies can be part of an efficient economic system has been put forward by De Alessi (1983), Barzel (1985) and Williamson (1998). In making transaction costs endogenous, we are also blurring the distinction between institutional and technological efficiency, as Antràs and Rossi-Hansberg (2009) have noted in a survey on organizations and trade. In that literature, decisions on organizational arrangments and on trade are interrelated. In our approach, agents make joint-decisions about their productive capacity, risk sharing and the investments in institutions that govern the costliness of trade. Finally, the gains from allocating resources to institutions relates to the idea of a 'state capacity' in Besley and Persson (2010). State capacity is partly "legal infrastructure investments such as building court systems, educating and employing judges and registering property or credit" (p.6). In that model, equilibrium investments in legal capacity are increasing in national income. Our approach also considers the possibility of a 'private capacity' that might substitute for or complement investments in public institutions.

The model is set out in Section 2. Section 3 characterizes efficient equilibria, that is the optimal investment in institutional capital, the optimal extent of transaction costs and the optimal market size. The conditions are discussed in Section 4 under which the efficient equilibrium may be decentralized. Section 5 reports the implications of our model in the light of extant empirical evidence. Section 6 examines the impact of (simple) non-optimal institutions and also looks at the impact of an exogenous transaction tax. Having established the efficient allocation in Section 3, we can compute the welfare cost of each of these policies. Section 7 summarizes and concludes. Appendices contain some proofs, further numerical analysis of the model and a detailed description of the decentralized equilibrium.

2 The model: overview

We briefly outline the model (which is motivated by Townsend, 1978) before presenting it in detail. A large number of risk averse agents are each endowed with the same amount of capital. That capital can be combined with a production technology to produce a consumption good. Agents can differ in the technology with which they can produce the consumption good but, initially, they know only the distribution of possible technologies. Consumption risk can be reduced by forming markets with other agents but due to exchange costs it is not feasible to replicate a complete Arrow-Debreu allocation. The cost of each bilateral exchange is determined in a simple way by the quality of contracting institutions in the economy. As such, *before* they realise their productivity, agents decide whether or not to form a market with other agents and, if they do, how large that market should be and how much to invest in private (i.e., excludable) and public institutional quality.[4]

One agent per market becomes an intermediary, buying outputs and selling consumption bundles. Intermediaries are here the productive unit of the transaction sector, using the institutional capital as input to a common 'exchange cost technology' (ECT), the output of which is the exchange cost incurred by agents in its market. Ex post, agents honor their obligations even if it would be preferable to renege. If agents do not join a market then no institutional investment takes place. The primary aim of Sections 3-4 is to characterize the efficient level of institutional capital, exchange costs and market size (consumption risk-sharing) for such a model economy.

2.1 Preferences and production technology

The economy is populated by a countable infinity of agents, $i \in I$. All agents have the same utility function, $u(c)$, with a constant degree of relative risk aversion. Each is endowed with the same amount of capital, $0 < k < \infty$. Agent i produces amount $\lambda^i y^i$ of the non-storable consumption good, where

[4] Intuitively, consider the adoption, ex ante, of standard accounting practices by a set of firms. It is doubtless costly to establish such a framework. However, ex post, there are still costs to running the system such as inputting data or prudential auditing.

$y^i \leq k$ is the amount of capital used in production and λ^i the idiosyncratic production technology. The set of possible technologies, Λ, is finite and bounded away from infinity. λ^i is distributed i.i.d. across agents with $p\left(\lambda^i\right)$ denoting the probability of any agent drawing λ^i, and $\sum_{\lambda \in \Lambda} p(\lambda) = 1$.

Let ω represent the state of nature, i.e., a list of λ^i for all $i \in I$. Let Ω be the set of all possible states of nature, and $p(\omega)$ the probability of some ω occurring and $\int_{\omega \in \Omega} p(\omega) = 1$.[5]

2.2 Diversification and intermediation

To diversify against risk, agents can form markets in a star network around a single intermediary, as in Townsend (1978). A *market* is a set of agents $M \subset I$ with cardinality $\#M < \aleph_0$; so a market is finite-sized. The set of agents with whom agent i exchanges directly is denoted N^i. Agent $h \in M$ is an *intermediary* if $N^i = h$ for $i \in M \backslash h$ and $N^h = M \backslash h$. So defined, markets are disjoint and agents only exchange with one intermediary. M^h denotes the market intermediated by agent $h \in H$ where $H \subseteq I$ is the set of all intermediaries in the economy.

Agents can exchange some of their endowment of capital on a one-for-one basis for shares in the consumption portfolio compiled by the intermediary, less their contributions to the total costs of transactions. Each agent in a market exchanges twice with the intermediary and pays an equal share of market-wide exchange costs. So for a given exchange cost, α, the per-agent cost of exchange in a market of size $\#M$ is $2\alpha\left(\frac{\#M-1}{\#M}\right)$.

2.3 Institutional capital and exchange costs

Intermediaries form markets using institutional capital as an input to an exchange cost technology (ECT) which determines the cost of exchange in that market. There is free-entry to intermediation (i.e., the ECT is accessible to all agents). Institutional capital takes two forms: There is a market-specific institutional capital (S-capital) and a general economy-wide

[5] Precisely, $p(\omega)$ is the probability of the state of nature in a small interval $d\omega$ occurring.

institutional capital (G-capital). We assume that the specific capital is excludable to a market and agent i allocates a proportion τ_s^i of the endowment as a private investment in this local capital. The general capital is a public good, each an agent i making a public investment τ_g^i of the endowment. S-capital thus reflects activities specific to the transaction being made – such as organizational choices, standardization of quality, private infrastructures or learning about property rights – and is excludable to the market. On the other hand, G-capital reflects the general legal enforcement of contracts, fiduciary duties, public infrastructures or competition policy and is, by contrast, a public good.

In a market intermediated by agent h, the exchange cost, α^h, is determined as follows:

$$\alpha^h = \left[1 - F\left(S^h, G\right)\right] k. \tag{1}$$

where $F()$ is the ECT, which is concave, continuous and increasing in each of its arguments and satisfies some Inada-type conditions.[6] The range of F is the unit interval, so $0 \leq \alpha^h \leq k$. Since the exchange cost is proportional to k it is akin to an 'iceberg cost'. We assume that S-capital is the average contribution of agents in a market and G-capital is the average across the whole economy.[7] Given that all agents are identical ex ante, all will make the same allocations and so we generally omit superscripts. Clearly, then, $S \leq \tau_s k$ and $G \leq \tau_g k$. The ex ante allocations, $(\tau_s + \tau_g)\, k$, combined with the ex post costs, $2\alpha\left(\frac{\#M-1}{\#M}\right)$, make up total transaction costs.

Since the intention of this paper is to explore the general equilibrium implications of agent investments in reducing exchange costs, the equation for the exchange cost, α, is left as a reduced-form expression. The expression for α imposes: i) That zero ex ante spending on institutions means no trade; ii) that greater ex ante investments reduce the costs of individual ex post

[6]Namely: $F(0,0) = 0$ and $F(S,G) \to 1$ as $\{S,G\} \to \{\infty,\infty\}$. Next, $F_S(S,G) \to \infty$ as $S \to 0$, and $F_S(S,G) \to 0$ as $S \to \infty$; analogous conditions obtain with respect to G. So defined, F ensures that the 'null' arrangement of zero investment in S-capital and G-capital is always equivalent to there being no gains from trade.

[7]In other words, there is no *institutional* scale effect: The introduction of an agent into a market requires an equal additional contribution to keep the exchange cost for that market constant.

trades; iii) there are decreasing returns to investing in the institutions, and iv) zero exchange costs require infinite investment. Nevertheless, there are a number of specific mechanisms that may generate transaction costs in a way captured by equation (1).

First, we may consider the transaction cost to the be product of the "transfer of property rights" (Allen, 2000, p.901). That is, there are frictions associated with a transaction because of the "costs of bookkeeping, the cost of enforcement, the cost of monitoring..." (Townsend, 1983; p.259). That 'cost' is related, however, to the institutional capital from which the agents in the exchange can draw; the form of (1) imposes a natural relationship between investments in that institutional capital and the consequent cost of conducting exchange. The costs of enforcing a contract is a product of the quality of the legal system, for example. Consider, for example, the problem of trading an item of unknown quality. This mechanism can be motivated using an example in Langlois (2006). Prior to the coming of the railroad to the American Midwest in the mid-nineteenth century, the trade of grain could be conducted on a personal basis with quality levels maintained through the observed reputation of individual farmers. Once the railroads vastly increased the scale of trade, individual farmer grains became mixed and the grains were traded as commodities in a way detached from their original producer, thus breaking the prior system of quality control. In response, the Chicago Board of Trade (CBOT) was formed to standardize the nature of the grain trade. Setting up such standards was costly but there were also continuing costs of inspecting each transaction for conformance. The CBOT reduced the risks involved in making an individual exchange of grain, reducing the margin required by traders engaged in buying and selling grain.

A second perspective is one of incomplete contracts, i.e., that transaction costs are "the costs of establishing and maintaining property rights" (Allen, 2000; p.898). Hart and Moore (2008) introduce a model in which the broad outlines of trade may be defined ex ante. Ex post, there are costs to fill in the detail of the finer points. A contract embodies a trade-off between flexibility and rigidity in which an optimal arrangement may not be fully specified over all possible states of nature. For Hart and Moore (2008), the costs from such partial incompleteness can take the form of a deadweight loss. We may also think of this in the context of complexity (cf. Anderlini and Felli, 1999). In the context of our model, goods may be produced with different technologies that require specific contracts that are complex (costly) to write. To write ex ante contracts for each possible technology would be exorbitantly expensive, but to write no contracts ex ante would mean that, ex post, there would be no trade. Ex ante investments in legal and accounting systems could mean that agents are committed to trade under the rough outlines of an agreement, leaving the costs of writing specific contracts for individual trades to be incurred after the state of the world is known.

The distinction between the specific and general determinants of the costs of exchange also emerges from this literature. Williamson (1979) refers to the 'governance of contractual relations'. Anderlini and Felli (1999) consider that some of what determines the extent of complexity costs is environmental, since a contract is "'embedded' in a larger legal system" (*ibid.* p.25). A second determinant is specific to the market in which the contract is formed. The distinction between S- and G-capital can also be motivated using the example of the Chicago Board of Trade. The CBOT was established by local businesses, the standards were private to the CBOT and benefited the farmers and traders that used it. However, there were also public goods that made the CBOT effective. In 1859, for example, it obtained the authority of the State legislature to appoint grain inspectors with powers of enforcement. Without the public institutional input, the sector-specific arrangement may have been less effective.

Each type of institutional capital may impact upon the other in determination of exchange costs. In our baseline case, we assume that S and G are complementary inputs, so $F_{SG} = F_{GS} > 0$. This is the form of relationship described in connection to the CBOT: Without a public institution, the S-capital would be less effective; without S-capital, there may be little point in an economy-wide institution. In our robustness analysis below we also consider the impact of their being substitutes. There is an additional form of complementarity that we will not consider: Private investments in the specific capital of one market may not be excludable, instead benefiting other markets. For example, it may have been that the establishment of the CBOT demonstrated the feasibilty of such an arrangement, serving as a blueprint for additional Boards elsewhere. A version of this model with ex ante heterogeneity may naturally incorporate such a form of investment if investments in markets composed of one type of agent affected exchange costs in other markets also composed of that agent type. In this model, all agents are ex ante identical and so we leave this consideration for future work.

3 Efficient equilibria

The purpose of this Section is to characterize efficient allocations in a cooperative setting; i.e., in an economy where intermediaries arise exogenously and where there is no difficulty providing the public good, G. We subsequently show that the efficient cooperative allocations coincide with those that are in the core (Proposition 5), and then that competitive equilibria are Pareto optimal under some conditions on the provision of G (Section 4). The intention is to use the efficient allocation in order to quantify the welfare implications of distortive institutional arrangements in Section 6. We can draw a number of contrasts with Townsend (1978) before proceeding. In that paper, the optimal cost of bilateral exchange is always zero and Townsend shows that, in the presence of such a cost, core allocations may include finite-sized coalitions of agents organized around a single intermediary. We use the exogenous exchange cost model in Townsend (1978) as a starting point, but in making exchange costs endogenous, we need then to characterise the existence and uniqueness of equilibria with optimally non-zero exchange cost. That also entails establishing the existance and uniqueness of an optimal market size given the optimal exchange costs.

An intermediary h makes an allocation for a market which is an n-tuple $\{\#M, \{\tau_s, \tau_g\}, \alpha, y, c^i(\omega)\}$. The intermediary allocation is that which maximises his own expected utility,

$$\max_{\#M,\tau_s,\tau_g,\alpha} \left\{ \int_{\omega\in\Omega} p\left(\omega\right) u\left[c^h(\omega)\right] \right\} \tag{2}$$

subject to,

$$\begin{aligned}
\sum_{i\in M} c^i(\omega) &\leq \sum_{i\in M} \lambda^i(\omega) y && (3)\\
2\alpha(\#M-1) &\leq \#M\left[(1-\tau_s-\tau_g)k-y\right]; && (4)\\
G &\leq \tau_g k; && (5)\\
S &\leq \tau_s k; && (6)\\
\tau_s+\tau_g &\leq 1; \tau_s \geq 0; \tau_g \geq 0; && (7)\\
\alpha &= \left[1-F\left(S,G\right)\right]k; && (8)\\
E\left[u\left(c^i\right)|h\right] &\geq E\left[u\left(c^i|h'\right)\right] \quad \forall i, \forall h', \text{ for } h' \text{ feasible.} && (9)
\end{aligned}$$

Equation (3) limits consumption to be less than or equal to output in each state of nature, on a market-wide basis. Equation (4) says that, in any market, the sum of endowments net of production and institutional investment must be sufficient to cover the total of market exchange costs. Equations (5)-(6) describe institutional capital, (7) restricts the range of feasible $\{\tau_s, \tau_g\}$, and (8) describes the exchange cost. Finally, equation (9) is a participation constraint requiring that the utility of all agents in market h be at least as high as participating in another feasible market. Implicit in the maximisation of (2) is a participation constraint for the intermediary; if,

$$\sum_{\lambda\in\Lambda} p\left(\lambda\right) u\left[\lambda k\right] > \int_{\omega\in\Omega} p\left(\omega\right) u\left[c^h(\omega)|h\right], \text{ for all } h \text{ feasible,}$$

then no intermediated market with exchange is formed (i.e., $\#M=1$, $\tau_s=\tau_g=0$).

3.1 Optimality

All agents share an aliquot consumption payout in each market as determined by the observed average technology for that market, $\bar{\lambda}(\omega,h)=(\#M)^{-1}\sum_{i\in M^h}\lambda^i(\omega)$. One more assumption is required before stating Proposition 1 which establishes that well-defined solutions to (2) exist. Recall that total per capita costs are $2\alpha\left(\frac{\#M-1}{\#M}\right)$. If for all k it is the case that $\left(1-\tau_s-\tau_g\right)k \leq 2\alpha$, then the optimal market size will be bounded since perfect diversification would imply non-positive consumption. Hence, let an arbitrary maximum value of k be $\bar{k}$. Let $K \equiv [\underline{k}, \bar{k}]$ where $\underline{k} < \bar{k}$, be the set of potential endowments. When $k = \bar{k}$, as shown below, investments in ECT will be at their maximum level: $\bar{\tau}_s, \bar{\tau}_g$.[8] A sufficient condition for finite-sized markets to be part of the optimal plan is then,

$$2(1 - \max F(\cdot)) \geq (1 - \bar{\tau}_s - \bar{\tau}_g). \tag{10}$$

(10) ensures that there exists a $\bar{M} \geq \#M(\bar{k})$, where $\bar{M}$ is a finite integer which may serve as an upper bound on optimal market size. Let $\mathbb{M} \equiv \left[1, ..., \bar{M}\right]$ be the set of feasible market sizes. In short, (10) imposes that exchange costs do not fall 'too quickly' as institutional capital rises.

Proposition 1 *For each* (k, ω) , *(i) the maximum of (2) is attained and (ii) the value function,* $V(k)$, *is well-defined and continuous. (iii) The optimal policy correspondence may not be unique.*

Proof. The relevant measure space is given by the triple (Ω, ω, p), where ω is a σ-algebra, the collection of all the subsets of Ω, and p is the measure defined on ω. An agent's expected utility is, therefore, $Eu(c) \equiv \int_{\omega \in \Omega} p(\omega) u\left(c\left(\omega\right)\right)$, where the utility function is strictly concave. Also, note that:

$$k \in K \subset R_{++}; \tag{11}$$

$$\{\tau_s, \tau_g\} \in T \subset R^2_+; \tag{12}$$

$$\#M \in \mathbb{M} \subset \mathbb{N}. \tag{13}$$

Decisions on τ_s, τ_g and $\#M$ are taken after k is known but before agents' productivities are revealed. Using (12) and (13), define the feasible policy choices as follows: $\Gamma : K \to T \times \mathbb{M}$. That is, $T \times \mathbb{M}$ is the product space,

$$\left\{\left\langle [0,1]^2, 1\right\rangle, \left\langle [0,1]^2, 2\right\rangle, \left\langle [0,1]^2, 3\right\rangle, ..., \left\langle [0,1]^2, \bar{M}\right\rangle\right\}.$$

[8] S-capital and G-capital are essentially normal goods, 'purchases' of both rising in k.

For each $\omega \in \Omega$, Γ is clearly a non-empty, compact-valued, continuous correspondence. A typical element of that mapping is denoted $z \in \Gamma(k, \omega)$. The optimization problem, therefore, involves a strictly concave criterion function and a non-empty, compact constraint set, so that the maximum is attained. Since the maximum is attained, the value function, $V(k)$, is well-defined. It follows from the Theorem of the Maximum that $V(k)$ is continuous. As the feasible policy set is not strictly convex, the optimal policy

$$G(k, \omega) = \{z \in \Gamma(k, \omega) : Eu(c) = V(k)\}$$

need not be unique. Finally, the equivalence between the cooperative case and core allocations is established in Proposition 5. ■

Some properties of efficient equilibria are immediate. Since the ECT is freely accessible, rents from intermediation must be zero so (5)–(6) hold with equality. The optimal allocations to the ECT is the pair $\{\tau_s^*, \tau_g^*\}$ which satisfies,

$$\theta k F_S = 1; \tag{14}$$

$$\theta k F_G = 1, \tag{15}$$

where $\theta \equiv 2\left(\frac{\#M-1}{\#M}\right)$, and F_X $(X \equiv S, G)$ denotes a partial derivative of $F\left(S^h, G\right)$. So, for any market size, the optimal institutional investments equate the marginal cost with the marginal gain from reducing the exchange cost. When there are no transactions, agents would not undertake such investment and the optimal choice of market size satisfies

$$\arg\max_{\theta} Eu((1-\theta)\, k\lambda). \tag{16}$$

Clearly, the optimal market size is unity, $\theta = 0$. The choices $\tau_s = 0$, $\tau_g = 0$ and $\#M = 1$ also define the unique reservation utility for any agent to be a member of any $\#M > 1$ coalition, $\overline{V} = \sum_{\lambda \in \Lambda} p(\lambda)\, u[\lambda k]$, $\forall k$. That furthermore ensure that $0 \leq (\tau_s + \tau_g) < 1$.[9] A direct implication of the

[9] That (16) generates a unique reservation value of utility is not quite so trivial as it may at first appear. In particular, it is different to Townsend (1978). In that paper, agents take the per capita exchange cost as given. Even so, the optimal market size may well be

reservation value of utility is that any equilibrium of the model in which $\tau_s > 0$ and $\tau_g > 0$ is one in which $\#M > 1$, necessarily.

Remark 1 follows from analysis of (14) and (15):

Remark 1 *For $\#M > 1$ and endowment k, Equations (14) and (15) define a unique pair, $\left\{\tau_s^{\#M}, \tau_g^{\#M}\right\}$ such that $\tau_s^{\#M} > 0$ and $\tau_g^{\#M} > 0$. For $\#M = 1$, equations (14) and (15) imply $\tau_s^{\#M} = 0$ and $\tau_g^{\#M} = 0$. For a given endowment, optimal investments in the ECT rise and bilateral exchange costs fall as the market size increases.*

The choice of any market size greater than one is determined by the ECT and agents' attitude to risk. (14) and (15) determine efficient investment in the ECT and hence total resources diverted from goods production. Agents' risk aversion provides an upper bound on how much they are willing to pay for consumption insurance, given an alternative not to diversify; that bound is independent of the ECT. Agents will optimally form markets with $\#M > 1$ when *efficient* investment in the ECT delivers transaction costs lower than that bound.[10]

The rest of the analysis of equilibrium is contained in the following four Propositions. One can show that in general there is a critical level of k below which transaction costs dominate and agents do not diversify:

Proposition 2 *There is a $k > 0$ small enough such that optimal market size is one.*

Proof. See Appendix A. ■

The key to understanding that Proposition is to recall that $S = \tau_s k$. So, a given τ has a larger proportionate impact on α as k rises, even though α itself is proportional to k. Hence, higher levels of k permit lower ex post

greater than one. In the present set-up, when agents take the exchange cost as given, the optimal market size is necessarily one. This is due to our scaling the exchange cost by capital.

[10] That explains why agents may move from 'autarky' ($\#M = 1$) to a market size $\#M > 2$ for a small change in k. This is apparent in the numerical simulations – diverisficiation at market sizes in between is too costly given the ECT and the degree of risk aversion. Proposition 3 shows that subsequent increases in market size will be in steps of one.

transaction costs and help sustain larger market sizes and the benefits from consumption risk sharing.

Proposition 2 also indicates the possibility of multiple optimal plans. Proposition 3 now shows that no more than two such alternative plans can exist.

Proposition 3 *The optimal policy correspondence contains at most two plans.*

Proof. See Appendix A. ■

As a corollary, it follows that for any $\#M > 1$ which is optimal, subsequent market size increases are single steps for (sufficiently large) increases in k. The final Proposition and following remarks establish that multiple equilibria can arise although they are, in a sense, special cases.

Proposition 4 *Let $\mathcal{K}$ denote the Borel sets of $K \subset R_{++}$. There exists a level of $k \in \mathcal{K}$, call it k^*, such that*

$$Eu\left\{\left[\left(1-\tau_1\right)k^* - \theta_1\alpha_1\right]\bar{\lambda}\left(\omega, M_1^h\right)\right\} = Eu\left\{\left[\left(1-\tau_2\right)k^* - \theta_2\alpha_2\right]\bar{\lambda}\left(\omega, M_2^h\right)\right\}$$

where $\tau_1 \neq \tau_2, \alpha_1 \neq \alpha_2$, $M_1^h \neq M_2^h$ and where maximized utility is identical under both programs.

Proof. See Appendix A. ■

The set of values of k which result in multiple equilibria is of measure zero. That is, such values of k^* correspond to pairs of θ's; these $\theta's$ are a proper subset of the rationals and hence themselves drawn from a set of measure zero.

To summarize: Low endowment economies may resort to autarky (Proposition 2). However, for economies with larger endowments ($k > \underline{k}$) it is optimal to invest in the ECT and to form markets (Proposition 1 and Remark 1). Equilibrium plans need not be unique (Proposition 4) but those equilibria are, in a sense, of limited interest (Proposition 3 and the brief discussion following Proposition 4). Finally:

Proposition 5 *The allocations of the cooperative economy coincide with core allocations.*

Proof. See Appendix A. ■

4 Equilibrium with competitive intermediation

The question now is: Can the efficient outcome be decentralized? First, note that given the public good nature of the G-capital, there is a free-riding problem associated with investment in G-capital:

Proposition 6 *In the core, voluntary allocations to general capital are zero.*

Proof. See Appendix B. ■

There are a number of ways in which the public good problem inherent in G may be addressed. The simplest is to suppose that a state institution exists and is capable of enforcing a tax rate. In the absence of any rent-seeking, that institution can deliver the optimal tax rate. However, given that our economy has free-entry to private, coalition-level provision of specific capital, we may think of a public analogue to that in terms of competitive provision of a public good. In particular, suppose that there is free-entry to the public, economy wide provision of general institutional capital. While there is no ex ante political conflict here, there may be a need to establish a mandate for a public institution to tax all agents equally even though individual agents may not wish to follow through in that way. This is reminiscent of the arguments of Schumpeter (1942, p. 269), who asserted that political markets work in much the same way as do competitive economic markets: "(T)he democratic method is that institutional arrangement for arriving at political decisions in which individuals acquire the power to decide by means of a competitive struggle for the people's vote." These arguments were later echoed forcefully by Wittman (1989).[11]

We formalize this perspective and consider the outcome to be the result of a market with free-entry: Any agent can costlessly seek a mandate that

[11] Wittman (1989, pp. 1395–6) argues "...that democratic political markets are organized to promote wealth-maximizing outcomes, that these markets are highly competitive and that political entrepreneurs are rewarded for efficient behavior."

specifies taxes for all agents and the level of G-capital to be provided. The successful mandate is that with the largest share of votes and is, we assume, then enforced ex post.[12] Proposition 7 shows that such competition then yields an efficient outcome; the competitive G provision will be $G^* = \tau_g^* k$.

Proposition 7 *If any agent may propose an enforceable tax plan then all agents will be taxed according to (15). It follows that $G = G^*$.*

Proof. See Appendix B. ■

There is no such problem in the provision of S-capital as free-entry to intermediation (i.e., any agent may become an intermediary) ensures that all rents are competed away. In Appendix C the decentralized economy is studied. It is established that there exists a unique, Pareto-optimal allocation in each market with a unique intermediary in each market.

Proposition 8 *If the public good is provided according to Proposition 7, the provision of S-capital and G-capital is Pareto optimal in the competitive equilibrium.*

Proof. See Appendix C. ■

5 The size of the transaction sector

The empirical literature on transaction costs has focused on the costs of individual transactions, or individual organizational arrangements. Insofar as they can be measured, the size of such costs are considered a measure of the working of the market. Examples in a variety of different contexts, from property rights to finance, are noted in Allen (2000). In contrast to the study of individual transactions, there has been limited work to establish the size of transaction costs on aggregate. Wallis and North (1986) were the first to quantify the size of the US transaction sector. By dividing occupations into those classified as providing 'transaction services' to firms and those

[12]So we invoke ex ante perfect competition and ex post monopoly in the process of allocating a mandate. Relaxation of either assumption might provide a focus for understanding the existence and power of elites.

that provide primarly 'transformative' services, Wallis and North calculate the total remuneration to transaction occupations. This sum constitutes the size of the transaction sector. They found that in 1870, the transaction sector accounted for 25% of GNP, rising to 50% in 1970.[13] The Wallis and North methodology has been applied to Australia (Dollery and Leong, 2002), Bulgaria (Chobanov and Egbert, 2007) and New Zealand (Hazledine, 2001) with similar magnitudes and trends found in each.

We thus have two observations from different empirical literatures on transaction costs: First, the size of the transaction sector is large and positively related to wealth; and, second, low transaction costs are one of the spurs of development.[14] The model presented here offers a natural way to consider both observations and to suggest directions for future empirical work. As we have defined them, transaction costs are composed of the costs of making individual exchanges and the investments in exchange institutions. Our model, then, hinges on a distinction between ex ante and ex post costs which is not a focus of empirical study. However, we take the empirical evidence on individual arrangements to reflect an individual ex post exchange cost (α) in the model. While the aggregate evidence will likewise capture the ex post exchange cost, it will also include resources allocated to exchange cost-reducing activities. We can draw some tentative comparisons between the model implications and the data, therefore. However, one of the implications of this framework is that if empirical work accounted for both components of total transaction costs, the evidence presented by Wallis and North (1986) may be more fully understood.

We let κ measure the size of the transaction sector (the proportion of the endowment not allocated to the production of the consumption god):

$$\kappa \equiv (k-y)/k = \tau_s + \tau_g + \theta\left[1 - F\left(S, G\right)\right] \tag{17}$$

[13]The transaction sector appears to be important in other advanced economies. See Wang (2003) and Klaes (2008) for surveys.

[14]Consider the conclusions in Klaes (2008): " ...economies with less well-developed transaction sectors appear to exhibit lower levels of transaction costs if those costs are measured in terms of sector size, whereas micro-structurally those economies in fact suffer from higher levels of transaction costs due to significant barriers to smooth exchange and coordination of economic activity."

Some comparative static results are intuitive. Suppose that at some k, the optimal market size is $\#M' > 1$; κ is necessarily higher at any $\#M'' > \#M'$ since otherwise an equilibrium with lower transaction costs and better risk-sharing properties is available. The higher is k, the lower is κ for any given market size since agents invest more to reduce the costs of exchange (see Appendix D). These results may be compared with those of Besley and Persson (2010) who find equilibrium investments in 'legal capacity' increase increase with national income. As k increases, then, the cost of further diversification approaches the additional consumption-smoothing gain and for a sufficiently large increase in k optimal market size increases.[15] Figure 1 summarizes results from a numerical version of the model (see Appendix E).[16] One finds that optimal transaction costs can be substantial; that large part of market activity referred to by Coase (1992) is reflected in our simple general equilibrium framework.

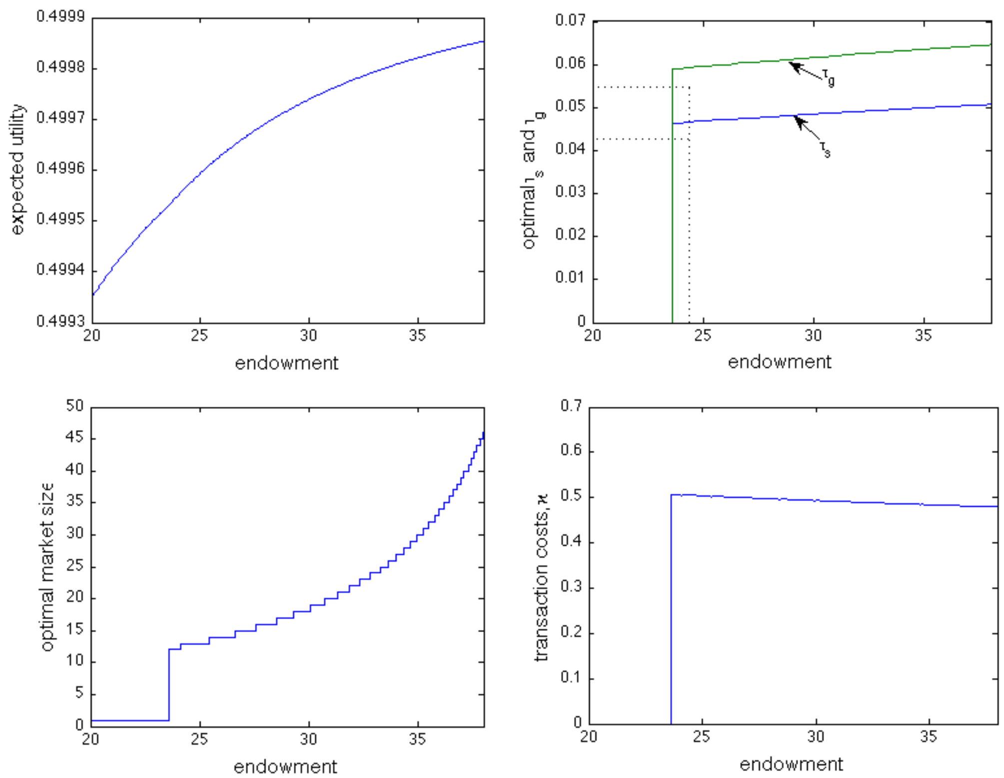
Proportionate investments in institutions can increase in k, as Figure 1 demonstrates;[17] while the number of exchanges increases and institutional investment grows, total transaction costs ultimately fall (although the relation is non-monotonic). In other words, a wealthier economy has a larger and more *sophisticated* transaction sector, one based on greater investments in contracting institutions, larger markets and lower exchange costs. Empirically, the *payoff* from the greater investments – lower costs of individual exchanges – are difficult to identify in sectoral analyses which then mistake the higher investments in institutions as higher transaction costs overall. What appears to be obscured in the Wallis and North analysis is the distinction between ex ante investments in institutional capital and the ex post cost of conducting exchange; empirical analyses which make that distinction would appear worthwhile.

[15] Proposition 3 shows that market size must increase by one from any $\#M > 1$, and the discussion preceding Proposition 2 explains how market size can jump by more than one when moving from $\#M = 1$.

[16] We adopt an ECT with constant elasticity of substitution of 2 whilst agents have a coefficient of relative risk aversion of 3. For Figure 1 we allow the effectiveness of S-capital and G-capital to differ slightly, although none of our results are sensitive to this assumption. See Appendix Table 2 for details.

[17] Appendix D identifies a sufficient condition on the ECT for that to be the case.

Figure 1: Equilibrium over a range of the endowment



These numerical results are robust to varying a number of different parameters, as shown in Appendix E. The coefficient or relative risk aversion drives the extent of diversification but, so long as it does not shut down all exchange, its impact on $\{\tau_s, \tau_g\}$ and so κ, is limited. The baseline relative risk aversion is $\gamma = 3$ and at $k = 30$, optimal $\#M = 18$ while $\tau_s = 0.0484$, $\tau_g = 0.0616$ and $\kappa = 0.4930$. Increasing the risk aversion parameter to 3.5 results in $\#M = 75$ and $\tau_s = 0.0507$, $\tau_g = 0.0684$ and $\kappa = 0.5100$. Reducing the risk aversion parameter to 2.5 causes all exchange to stop (i.e., optimal $\#M = 1$); with $\gamma = 2.5$, endogenous exchange occurs at $k = 32.3$. A further robustness is to consider the efficacy of the allocations to the institutional capital. In the baseline numerical simulations, the contributions $\tau_s k$ and $\tau_g k$ are both weighted by a parameter $\beta = 0.03$. Increasing the parameter makes the exchange cost lower for a given institutional capital. At $\beta = 0.035$, the optimal $\#M = 30$ while $\tau_s = 0.0499$, $\tau_s = 0.0666$ and $\kappa = 0.4845$; the more effective is institutional capital, the higher are the optimal allocations, the larger are markets and the lower is the size of the total transaction cost. A final robustness is to consider the assumption of complementarity between S and G capital. In the baseline we set the CES coefficient to $s = 2$. If we modify this to $s = -1$,then optimal $\#M = 19$ while $\tau_s = 0.0503$, $\tau_g = 0.0595$ and $\kappa = 0.4935$.

6 The impact of non-optimal public institutions

So far the analysis has focused on the nature of efficient equilibria; the efficient tax level, τ_g^*, provides an optimal trade-off between the expected level of consumption and its variability. This Section considers the impact of distortive (i.e., non-optimal) levels of the tax, τ_g.[18] First, we look at constrained optimal decisions about τ_s and $\#M$ over a range of imposed tax rates. Sec-

[18] In this section S-capital and G-capital are equally efficient in reducing transaction costs so any difference between τ_s and τ_g reflects the institutional distortion. Specifically, $\gamma_s = \gamma_g = 0.075$ and $k = 30$. All other parameters are as in Appendix Table 2 unless otherwise stated.

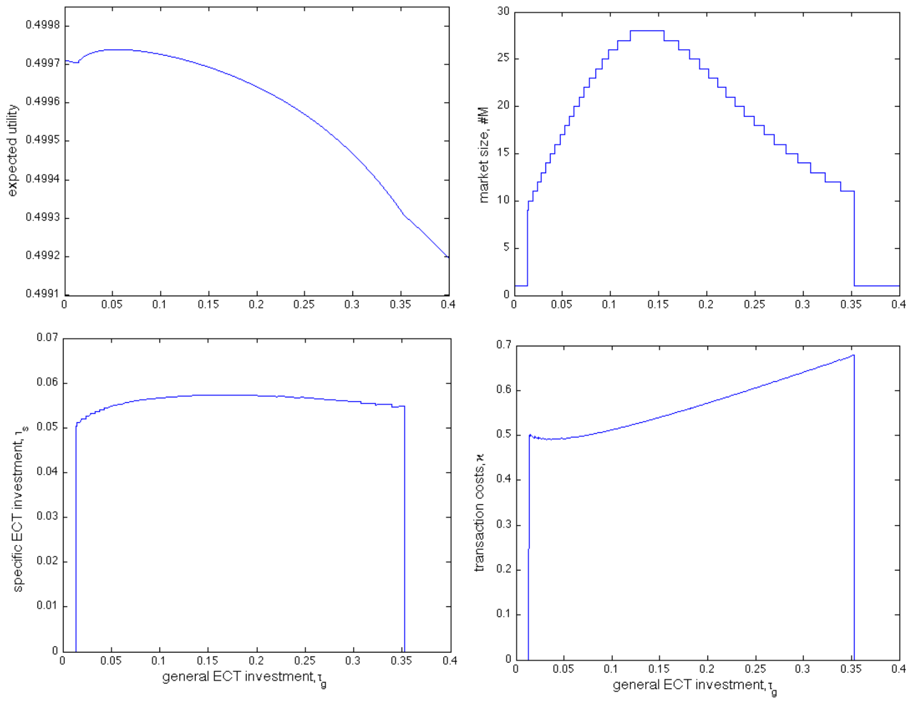
ond, we consider the welfare implications of various transaction cost policies. Third, we model the impact of a tax on transactions.

6.1 Equilibrium with exogenous τ_g

The distortive behavior centres on the resources turned over to G-capital via the tax τ_g. Such distortions might reflect deeper political tensions, perhaps between the electorate and a 'political class'. In an environment where the definition and measurement of the costs of exchange are not well understood, it may also be simply that the objective of tax policy is not straightforward to design. Alternatively, deviations from optimality may reflect 'irrational' voters, as in Caplan (2008). First, we look at the consequence of varying τ_g over some interval. In Subsection 6.2, we look at particular policy objectives that may drive specific deviations.

Figure 2 reports the consequences of varying τ_g over the interval $[0, 0.4]$. Responses in τ_s and $\#M$ are unconstrained but can be, of course, affected by deviations in τ_g from its efficient level. This can result in significant compensating changes in agent behavior, as Figure 2 shows. When τ_g deviates from τ_g^* agents can respond by changing τ_s and/or market size. If market size does not change, then τ_s varies positively with τ_g. This follows from (14) and the assumption that S and G are complementary inputs; a higher G increases the marginal return to S-capital investment. For a sufficiently large increase in τ_g, the constrained optimal choice of market size may also change by increasing or decreasing. Increasing τ_g holding $\#M$ fixed clearly reduces net endowment since the optimal τ_s also increases. One response is to recover some lost productive capacity by reducing the number of exchanges (lower $\#M$). Another response is to take advantage of the lower exchange cost created by the higher τ_g and diversify further (higher $\#M$). For this parameterization, market size is hump-shaped in the level of τ_g. At low levels of τ_g, there remains a substantial residual endowment and so the gain from additional risk-sharing dominates: A sufficiently large increase in τ_g lowers the exchange cost such that increasing market size (further risk-sharing) is attractive enough to incur the cost of a higher κ. However, for

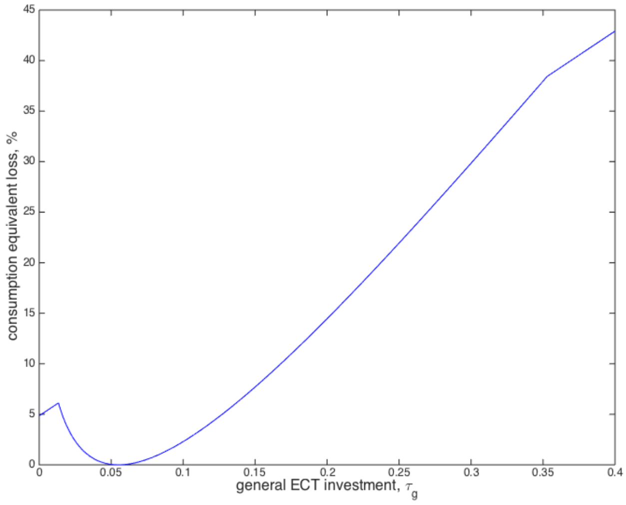
Figure 2: Equilibrium with distortive institutions



$\tau_g \gg \tau_g^*$ the residual endowment is severely diminished. Diverting additional resources away from production is more costly than the gain from further diversification; as such, market size and τ_s both fall because that is the only way to reduce κ.

At very low levels of τ_g, exchange costs are so high that no diversification is worthwhile (agents set $\tau_s = 0$ and $\#M = 1$). When τ_g is very high, the loss to productive capital is such that no diversification with $\tau_s = 0$ and $\#M = 1$ is better than some diversification with the small residual endowment, even though the endowment are still being taxed. In short, market size is hump-shaped in τ_g and, for some low and high values of τ_g, agents optimally choose not to diversify. The general pattern of robustness to parameters discussed in the previous section holds with regard to the effect of distortive institutions; in particular, the non-monotonic responses of τ_s and $\#M$ to changes in τ_g holds across a range of endowment levels and risk

Figure 3: Welfare Costs of Distortionary Institutions



aversion parameters. The non-monotonicity of $\#M$ is also robust to making S and G capital input-substitutes in the exchange cost technology, although the relationship between τ_s and τ_g is, of course, negative when we assume they are substitutes (aside from increases in τ_s that result from increasing $\#M$).

Given that Section 11 has established the efficient outcome, we can calculate the consumption equivalent loss for each level of τ_g relative to that efficient frontier. In particular, if $u^{\tau'_g}$ is the utility obtained under policy τ'_g, then the percentage loss in consumption that is equivalent to moving from τ_g^* to τ_g^d can be calculated as $\left[1 - u^{-1}\left(u^{\tau_g^d}\right)/u^{-1}\left(u^{\tau_g^*}\right)\right] \times 100$. As Figure 3 shows, deviation in τ_g from its optimal level can have significant welfare costs. A τ_g that is 0.05 higher than optimal is equivalent to a 2.7% drop in consumption; a τ_g that is 0.15 higher than optimal is equivalent to a 15.1% drop in consumption.

6.2 Specific transaction cost policies

The discussion of the empirical evidence above points to the complexity of classifying and measuring transaction costs in reality. In the light of that, it is reasonable to consider an environment where the optimal tax policy is unclear. The literature on transaction cost economics suggests that optimality is where "transactions... are aligned with governance structures... so as to effect a (mainly) transaction cost economizing outcome" (Williamson, 2010: p.681). The interpretation of that optimality condition in terms of policy may be complicated since there are multiple observable and potentially non-observable components: The number of exchanges (market size), the investments in reducing the cost of exchange, and the cost of exchange itself. As such, a policy maker may have as its objective some minimum or maximum of each of these and think it a reasonable interpretation of what is optimal: First, institutions could deliver zero exchange costs; second, institutions should maximise risk-sharing; third, the cost of individual exchanges should be minimized[19]; and, fourth, institutions should minimize the size of the transaction sector as a whole.

In our model, zero exchange costs are not feasible; institutions which deliver an infinite number of trades at zero cost are themselves infinitely costly. However, the analysis above shows that we can consider the second, third and fourth type of policy prescription. The second is a rule that maximizes the (constrained optimal) choice of market size; the third minimizes the cost of individual exchanges; the fourth minimizes size of the transaction sector. The market size policy is given by,

$$\tau_g^M := \left\{\min\left\{\tau_g | \#M \geq \#M'\right\} \text{ and } \tau_s \text{ optimal}\right\}. \tag{18}$$

That is, τ_g^M is the lowest tax required to induce the maximum market size, given that agents optimally choose market size and τ_s in response to τ_g. The tax that minimizes the exchange cost is given by,

[19] We are grateful to a referee for suggesting this policy.

$$\tau_g^{\alpha} := \min\{\alpha | \#M > 1 \text{ and } \tau_s \text{ optimal}\}. \tag{19}$$

Finally, the tax policy which minimizes the transaction sector is given by,

$$\tau_g^{\kappa} := \min\{\kappa | \#M > 1 \text{ and } \tau_s \text{ optimal}\}. \tag{20}$$

For each policy the percentage change in certainty equivalent consumption is calculated using the efficient frontier established in Section 3. Table 1 describes various features of equilibrium under the different rules.

Table 1: The Effects of Distortive Institutions.

rule	$\#M$	κ	τ_g	τ_s	α	$\%C$
τ_g^*	18	0.4940	0.0550	0.0550	6.0795	-
τ_g^M	28	0.5234	0.1210	0.0571	5.3711	4.3368
τ_g^{α}	11	0.6781	0.3520	0.0548	4.4776	38.2895
τ_g^{κ}	14	0.4906	0.0376	0.0537	6.4502	0.6400

None of the rules are optimal in a framework with endogenous transaction costs; the τ_g^M rule delivers too little *directly* productive capital and τ_g^{κ} too much consumption variability. The τ_g^{α} rule delivers a worse outcome in both regards. Under the τ_g^{α} rule, the policy to minimize individual exchange costs involves setting the tax to the highest level without shutting down all exchange, regardless of the effect on total transaction costs or on the amount of consumption variability. The economy is pushed toward low-market size (high consumption variability) *and* low productive capital. This results in a consumption equivalent loss far in excess of the other rules. Nonetheless, it is useful to analyze which of τ_g^M and τ_g^{κ} is the more costly, and why. Consider first the τ_g^M rule. Larger markets means, in short, higher taxation to lower the costs of bilateral exchange. The transaction sector is larger as a whole and its *composition* has shifted toward ex ante investments. τ_g cannot increase by too much, however, since agents always have the option of shifting resources to goods production by reducing market size and τ_s. For the τ_g^{κ} rule, minimizing transaction costs requires exchange costs to be too *high*. A low τ_g means that agents optimally form smaller markets and make smaller investment in S-capital, both actions reducing κ.

The loss from τ_g^M is greater than that from τ_g^κ, but both are eclipsed by the τ_g^α rule.[20] Expected utility is a product of the portion of the endowment that is productive $(1-\kappa)$ and the amount of consumption variability ($\#M$). The extreme costliness of the τ_g^α rule results from neglecting both of these. In this sense, our distinction between the costs of exchange and the total costs of the transaction sector is particularly important: An apparently sensible policy to minimize the costs of exchange, while neglecting the resources required to obtain that minimization, is highly distortive. The comparison between τ_g^M and τ_g^κ is also informative: Institutions fostering smaller exchange costs and bigger markets appear to be the more damaging in welfare terms. By targeting market size, the government 'distorts' choices of both $\#M$ *and* τ_g, leaving agents, in effect, with only one instrument to respond, τ_s. The alternative τ_g^κ, which minimizes the sum of all resources allocated to transactions, leaves private agents with the choice of market size and τ_s. The option for agents simply to consume their own endowment constrains policy such that the outcome is not too far from that which maximizes expected utility. That additional flexibility appears to reflect some of the empirical findings of Acemoglu and Johnson (2005). That paper finds that bad property rights institutions are more damaging that contracting institutions since individuals can respond to contractual distortions by using a variety of formal and informal mechanisms. In the case of this model, agents can respond via choice of τ_s and market size in the face of the τ_g^κ rule, but have only a limited range of response in the case of the τ_g^M rule.

6.3 Transaction taxes

Policies are sometimes designed to address the negative externalities from trade (such as trades of a pollutive item), or to raise revenue from a sector characterized by high-frequency trading (as in a Tobin-type tax). In such

[20] This welfare ranking is robust to a wide range of different endowment levels (numerical comparisons over $k \in (0, 35]$ all satisfy this ranking). It is also robust to assuming that S and G are input substitutes.

policies, individual exchanges are taxed. The European Commission analysis of its proposed Financial Transaction Tax (FTT), for example, finds "in a nutshell... very positive impacts on the functioning of the single market for financial instruments" (European Commission, 2013; p.16). The consequences of such a tax are not necessarily obvious, however, since agents may respond by increasing investment in private exchange cost reduction or may dramatically change the number of trades. Moreover, the robustness of a market to the imposition of a tax, i.e., its revenue-generating ability, may not be fully understood since those markets may shut down or relocate to avoid the tax.

In order to understand the impact of a transaction tax, we need a framework in which the nature of trade (i.e., the number of exchanges and investments in the cost of exchange) is endogenous. As such, our model can form a basis for a preliminary analysis of transaction taxes. We can consider a tax, $t > 0$, that simply makes individual exchanges more costly,

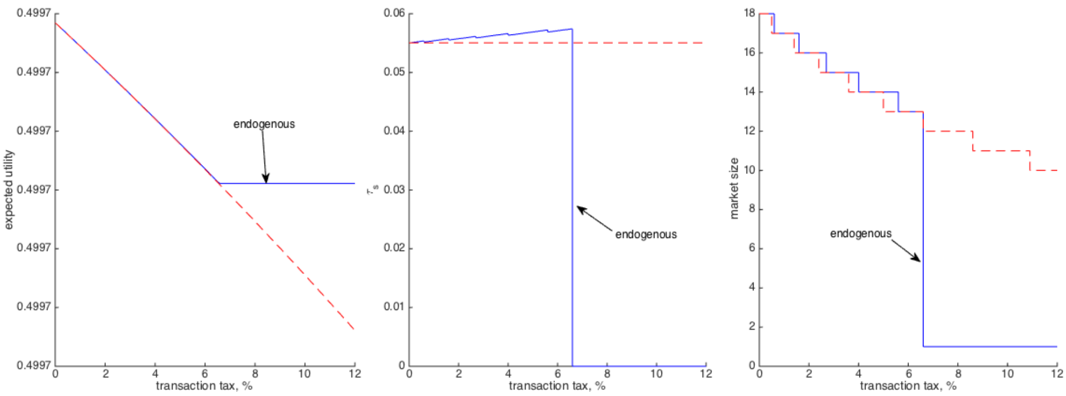
$$\alpha = (1 + t) [1 - F (S, G)] k. \tag{21}$$

Figure 4 demonstrates the effect of a transaction tax where transaction costs are endogenous (solid line). Agents respond first by increasing investments in institutions in order to dampen the effect of the tax on the cost of exchange. Relative to an environment where transaction costs are exogenous (dashed line),[21] market size is more robust to the introduction of a transaction tax. In each environment, agents can avoid the tax by simply resorting to autarky, recovering the investment in exchange and avoiding all tax, whenever the participation constraint is no longer satisfied. The important difference when transaction costs are endogenous is that agents have individually allocated a portion of their endowment to support institutions; in the exogenous case this investment has not occurred and so cannot be retrieved by agents. That means that although agents intially appear less affected by the tax, they will opt for autarky at a level of the tax far lower

[21]The exogenous transaction cost set-up fixes α and the net endowment such that the equilibrium when $t = 0$ is the same as that in the endogenous transaction cost environment (i.e., τ_g and τ_s are fixed at the values which are optimal for $t = 0$).

than that suggested when transaction costs are considered exogenous.[22] The implication for transaction tax policy is that, while markets respond as would be expected to the imposition of a tax by reducing market size, agents will resort to autarky earlier than may expected if the individual agents' own investments in that exchange are not taken into account. In case of the FTT, for example, this may mean that a transaction tax leads to the shifting of financial activities to geographies outwith the FTT's jurisdiction at much lower levels of the tax than may be anticipated. Revenues from such a tax may fall short of projections.

Figure 4: The effects of a transaction tax

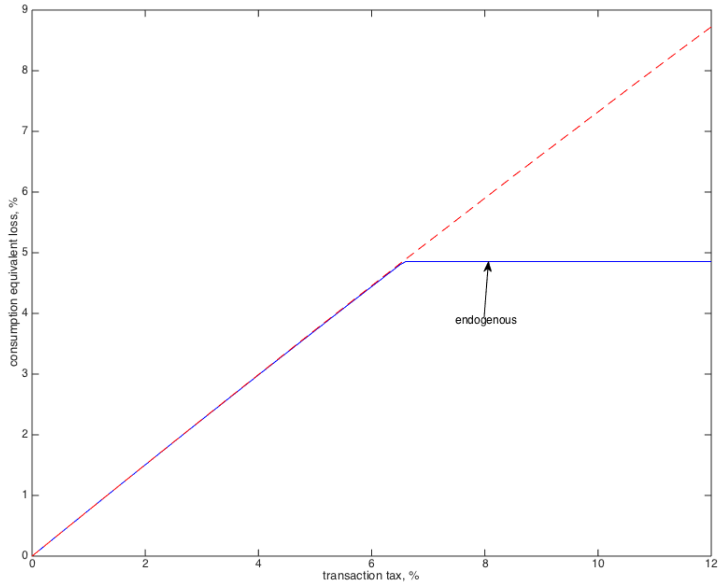


We can again calculate the losses from imposing a tax on transactions. Figure 5 depicts losses associated with Figure 4. A 5% transaction tax leads to 3.7% consumption equivalent loss in both the exogenous and endogenous cases, which is somewhat higher than an equal deviation in the τ_g. An increase in τ_g above optimum is at least reducing exchange costs; imposing a tax on exchange means that, even if agents can respond with higher τ_s and lower market size, the cost of exchange is increasing. In the case where agents can respond by removing their own investments in reducing exchange costs, the consumption equivalent loss is capped at 4.9% once the transaction

[22]With endogenous transaction costs, autarky occurs at 6.6%; if we take them to be exogenous, the autarky equilibrium is induced at 22.2%. Although these thresholds vary, this differential impact is robust across the different parameterizations discussed in the previous subsections.

tax induces no exchange.

Figure 5: Welfare Costs of Transaction Taxes



7 Discussion and concluding remarks

Economists are increasingly focusing on the role of good institutions in promoting growth, trade and other desiderata. The intention of this paper is to link in a simple way impediments to transactions, institutional quality and market size. The efficient equilibrium of the model is consistent with significant transaction costs and investments in institutions. A distinction between transaction costs and exchange costs was made. The impact of distortive institutions was also considered, although in a tentative and ultimately ad hoc way. We argued that a number of what might be thought of as 'good' institutions are actually sub-optimal when transaction costs are endogenous.

It would seem important to extend the analysis in a number of directions. First, although agents had different productive capabilities, this had a limited impact as decisions over τ_s and τ_g were made *before* types were revealed. If decisions over τ_s and τ_g were made *after* agents' productive capabilities were known (and capabilities are private information) then the analysis will be somewhat more complicated. Related to this, incentive compatibility issues were not to the fore because it was assumed that agents remained in markets even if, ex post, they might have been better off under autarky. Nevertheless, the framework developed above may prove useful in the analysis of optimal tax and the role of government.

References

Acemoglu, D., Antràs, P., and Helpman, E. (2007). 'Contracts and Technology Adoption'. *American Economic Review*, 97(3):916–943.

Acemoglu, D. and Johnson, S. (2005). 'Unbundling Institutions'. *Journal of Political Economy*, 113(5):949–95.

Allen, D. W. (2000). 'Transaction Costs'. In Bouckaert, B. and De Geest, G., editors, *Encylopedia of Law and Economics*. Edward Elgar.

Anderlini, L. and Felli, L. (1999). 'Incomplete Contracts and Complexity Costs'. *Theory and Decision*, 46(1):23–50.

Antràs, P. and Rossi-Hansberg, E. (2009). 'Organizations and Trade'. *Annual Review of Economics*, 1:43–64.

Barzel, Y. (1985). 'Transaction Costs: Are They Just Costs?'. *Journal of Institutional and Theoretical Economics*, 141:4–16.

Besley, T. and Persson, T. (2010). 'State Capacity, Conflict and Development'. *Econometrica*, 78(1):1–34.

Caplan, B. (2008). *The Myth of the Rational Voter: Why Democracies Choose Bad Policies*. Princeton University Press.

Chobanov, G. and Egbert, H. (2007). 'The Rise of the Transaction Sector in the Bulgarian Economy'. *Comparative Economic Studies*, 49:683–98.

Coase, R. H. (1960). 'The Problem of Social Cost'. *Journal of Law and Economics*, 3:1–44.

Coase, R. H. (1992). 'The Institutional Structure of Production'. *American Economic Review*, 82(4):713–9.

Commission, E. (2013). 'Implementing enhanced cooperation in the area of financial transaction tax Analysis of policy options and impacts'. *Commission Staff Working Document*.

De Alessi, L. (1983). 'Property Rights, Transaction Costs, and X-Efficiency: An Essay in Economic Theory'. *American Economic Review*, 73(1):64–81.

Dixit, A. (1996). *The Making of Economic Policy: A Transaction-Cost Perspective*. MIT Press.

Dollery, B. and Leong, W. H. (2002). 'Measuring the Transaction Sector in the Australian Economy, 1911–1991". *Australian Economic History Review*, 38(3):207–31.

Greenwood, J. and Jovanovic, B. (1990). 'Financial Development, Growth and the Distribution of Income'. *Journal of Political Economy*, 98(5):1076–1107.

Hart, O. and Moore, J. H. (2008). 'Contracts as Reference Points'. *Quarterly Journal of Economics*, 123(1):Forthcoming.

Hazledine, T. (2001). 'Measuring the New Zealand Transaction Sector, 1956-98, with an Australian Comparison'. *New Zealand Economic Papers*, 35(1):77–100.

Klaes, M. (2008). 'The History of Transaction Costs'. *The New Palgrave Dictionary of Economics, 2nd Edition*. Durlauf, S. N. and Blume. L. E. (eds.), Palgrave Macmillan.

Langlois, R. N. (2006). 'The Secret Life of Mundane Transaction Costs'. *Organization Studies*, 27:1389–1410.

Levchenko, A. A. (2007). 'Institutional Quality and International Trade'. *Review of Economic Studies*, 74(3):791–819.

Schumpeter, J. A. (1942). *Capitalism, Socialism and Democracy*. Harper & Row, New York.

Townsend, R. M. (1978). 'Intermediation with Costly Bilateral Exchange'. *Review of Economic Studies*, 55(3):417–25.

Townsend, R. M. (1983). 'Theories of Intermediated Structures'. *Carnegie Rochester Conference Series on Public Policy*, 18(Spring):221–72.

Townsend, R. M. and Ueda, K. (2006). 'Financial Deepening, Inequality, and Growth: A Model-Based Quantitative Evaluation'. *Review of Economic Studies*, 73(1):251–293.

Wallis, J. J. and North, D. C. (1986). 'Measuring the Transaction Sector in the American Economy, 1870–1970'. Chapter 3 in Engerman, S. L. and Gallman, R. E. (eds.). *Long-Term Factors in American Economic Growth*. University of Chicago Press.

Wang, N. (2003). 'Measuring Transaction Costs: An Incomplete Survey'. Ronald Coase Institute, Working Paper Number 2.

Williamson, O. E. (1979). ‘Transaction-Cost Economics: The Governance of Contractual Relations’. *Journal of Law and Economics*, 22(2):233–61.

Williamson, O. E. (1998). ‘Transaction Cost Economics: How it works; where it is headed’. *De Economist*, 146(1):23–58.

Williamson, O. E. (2010). ‘Transaction Cost Economics: The Natural Progression’. *American Economic Review*, 100:673–90.

Wittman, D. (1989). ‘Why Democracies Produce Efficient Results’. *Journal of Political Economy*, 97(6):1395–1424.

A Proofs of Section 3 propositions

Proposition 2 *There is a $k > 0$ small enough such that optimal market size is one.*
Proof. To the contrary, assume that

$$Eu\left(\left[(1-\tau_s-\tau_g)\,k-\theta\alpha\right]\bar{\lambda}\left(\omega, M^h\right)\right) \geq Eu\left(\lambda^i k\right), \tag{22}$$

for all k. All variables on the left-hand side reflect optimizing decisions. Note, in particular, that $\theta > 1, \forall k$. Observe that as $k \to 0$, $(1-\tau_s-\tau_g) - \theta(1 - F(\cdot)) \to (1-\theta) < 0$, and that $\bar{V}$ is positive for all $k > 0$. Thus, there exists a level of k, call it $\underline{k}$, such that

$$\lim_{k \to \underline{k}} Eu\left(\left[(1-\tau_s-\tau_g)\,k-\theta\alpha\right]\bar{\lambda}\left(\omega, M^h\right)\right) \to 0,$$

As $k \to \underline{k}$, the inequality in (22) is reversed. ■

Proposition 3 *The optimal policy correspondence contains at most two plans.*
Proof. Let k^* denote a value of k such that $Eu(P_1|k^*) = Eu(P_3|k^*) = V(k^*)$. Assume that there also exists a P_2 such that $Eu(P_1|k^*) = Eu(P_2|k^*) = Eu(P_3|k^*)$. Let

$$\tau_3 > \tau_2 > \tau_1;\ \theta_3 > \theta_2 > \theta_1.$$

Now, denote

$$\theta_3 = \frac{x}{y}; \theta_2 = \frac{p}{q}; \theta_1 = \frac{m}{n}, \tag{23}$$

where x, y, p, q, m and n are all positive integers. Define χ as follows:

$$\chi = \frac{\frac{p}{q} - \frac{x}{y}}{\frac{m}{n} - \frac{x}{y}}.$$

By assumption $\frac{p}{q} - \frac{x}{y} < 0$, $\frac{m}{n} - \frac{x}{y} < 0$, and $\frac{p}{q} - \frac{m}{n} < 0$, so $1 \geq \chi \geq 0$. Finally, note that $q = p + 1$, $n = m + 1$ and $y = x + 1$. Thus

$$\frac{p}{q} = \left[\frac{p(p+1)^{-1} - x(x+1)^{-1}}{m(m+1)^{-1} - x(x+1)^{-1}}\right]\frac{m}{n} + \left[\frac{m(m+1)^{-1} - p(p+1)^{-1}}{m(m+1)^{-1} - x(x+1)^{-1}}\right]\frac{x}{y}; \tag{24}$$

that is, market size associated with P_2 is a weighted average of the other two optimal market sizes. Hence it follows, by the strict concavity of the utility function, that either: *(i)* P_2 is indeed an optimal plan and P_1 and P_3 are not;

or, *(ii)* P_2 is optimal and identical to either P_1 or P_3; or, *(iii)* $P_2 = P_1 = P_3$. $\blacksquare$

Proposition 4 *Let K denote the Borel sets of $K \subset R_{++}$. There exists a level of $k \in K$, call it k^*, such that*

$$Eu\left\{\left[(1-\tau_1)\,k^* - \theta_1\alpha_1\right]\bar{\lambda}\left(\omega, M_1^h\right)\right\} = Eu\left\{\left[(1-\tau_2)\,k^* - \theta_2\alpha_2\right]\bar{\lambda}\left(\omega, M_2^h\right)\right\}$$

where $\tau_1 \neq \tau_2, \alpha_1 \neq \alpha_2$, $M_1^h \neq M_2^h$ and where maximized utility is identical under both programs.

Proof. Let $k \in [\underline{k}, \bar{k}]$. Partition that set into into $[\underline{k}, \hat{k})$ and $(\hat{k}, \bar{k}]$ such that the $\max Eu(\cdot \mid \forall k < \hat{k}) < \max Eu(\cdot \mid \forall k > \hat{k})$, and where $\tau(k > \hat{k}) > \tau(k < \hat{k})$, $\#M(k > \hat{k}) > \#M(k < \hat{k})$. By Proposition 2, such a partition is possible; it is also implied by (14)–(15). Let $\{k_1\}$ denote any sequence in $[\underline{k}, \hat{k})$ converging to $\hat{k}$ and let $\{k_2\}$ be any sequence in $(\hat{k}, \bar{k}]$ converging to $\hat{k}$. Let $V^1(k_1)$ denote $\sup Eu(\cdot|k \in [\underline{k}, \hat{k}))$, $V^2(k_2)$ denote $\sup Eu(\cdot|k \in (\hat{k}, \bar{k}])$ and $\hat{V}(\hat{k})$ denote $\sup Eu(\cdot|k_1 = \hat{k})$. By Proposition 1 these value functions are well defined and, by the Theorem of the Maximum, continuous. Hence, there exists a δ such that for $\left|k_1 - \hat{k}\right| < \delta/2$ and $\left|k_2 - \hat{k}\right| < \delta/2$ one has that,

$$\begin{aligned}\left|V^1(k_1) - \hat{V}(\hat{k})\right| &< \varepsilon/2;\\ \left|V^2(k_2) - \hat{V}(\hat{k})\right| &< \varepsilon/2.\end{aligned}$$

Hence

$$\begin{aligned}\left|V^1(k_1) - V^2(k_2)\right| &\leq \left|V^1(k_1) - \hat{V}(\hat{k})\right| + \left|V^2(k_2) - \hat{V}(\hat{k})\right|\\ &\leq \varepsilon/2 + \varepsilon/2\\ &= \varepsilon.\end{aligned}$$

for k_1, k_2 close to $\hat{k}$ market size will not be changing. Hence, market size and taxes are higher for all $k \in (\hat{k}, \bar{k}]$ compared with $k \in [\underline{k}, \hat{k})$. Expected utility is identical with different optimal plans at $k^* = \hat{k}$. $\blacksquare$

Following Townsend (1978) and Boyd and Prescott (1985) one may characterize core allocations directly. In the discussion of the efficient equilibrium in the main text we studied the equilibrium decision rules of an intermediary. Townsend (1978) labelled that analysis the "cooperative" solution. Hence, the equivalence of the core and cooperative solutions is now established, Proposition 5 in the text.

Proposition 5 *The allocations of the cooperative economy coincide with core* *allocations.*

Proof. Consider the unique equilibrium: $x^h = \{c^*, y^*, \tau^*, \#M^*; i \in M^h\}$ for all $h \in H$. This n-tuple determines $F(S^*, G^*)$ and hence α^*. Suppose a strict subset of agents in the market intermediated by h, $B \subset M^h$, can form a blocking coalition (i.e., market). The cooperative equilibrium necessitates that the blocking coalition cannot deviate from contributing τ_g^* on average. The blocking intermediary chooses τ_s^B given the market size $\#B < \#M^h$. The agents in the blocking coalitions are better off with the following program:

$$\begin{aligned} c^i &= \bar{c}, \qquad \tau_g^i = \tau_g^*, \qquad \forall i \in B; \\ 1 &= 2\left(\frac{\#B-1}{\#B}\right) kF_S\left(S^B, G^*\right). \end{aligned}$$

The consumption profile follows from optimizing over agents with identical, strictly concave utility functions and the second condition was derived in the text. By Remark 1, S^B-capital is strictly lower and exchange costs strictly higher. There are fewer transactions in this proposed market but each is more costly. In addition, the investment portfolio is less diversified. Given that market size and investment in S-capital deviate from the optimum, it must be that the higher exchange cost and less diversification are not compensated by fewer transactions; expected utility is necessarily lower. Now consider the case $B \supset M$. The same argument applies: In this proposed market, exchange costs are strictly lower and investments higher. The deviation from first-best means that expected utility must be lower than in market M. Hence, the cooperative allocation is in the core. Further, since the core is non-empty and $\#M$ is the unique optimal market size, it follows that the core allocations and the allocations of the cooperative economy coincide. ■

B Proofs of Section 4 propositions

Proposition 6 *In the core, voluntary allocations to general capital are zero.*

Proof. Consider an equilibrium in which $\tau_g^i = 0 \ \forall \ i \in I$, so that $G = 0$. By Proposition 2, for k big enough, there is a positive level of S-capital that is optimal, $\tau_s^i = \tau_s^{**}$, $\forall \ i \in I$; $EU^i = EU_0$ denotes expected utility under this plan. Suppose a blocking coalition B exists such that an agent $b \in B$ proposes $\tau_g^b > 0$ for $i \in B$. If $\#B < \aleph_0$, then $G = 0$ obtains and it follows that $EU^i < EU_0$ for all $i \in B$. Suppose, however that $\#B = \aleph_0$ and that an agent $b \in B$ proposes $\tau_g^b > 0$ for $i \in B$. In that case $G = G^b > 0$. Since $\#B > 0$, some positive level of G is optimal, and so $EU^i > EU_0$ for each

$i \in B$. But now there exists a blocking coalition $B' \subset B$ in which some agent $b' \in B'$ proposes $\tau_g^{b'} = 0$ for each $i \in B'$. Since $\#B' < \aleph_0$ it remains the case that $G = G^b$. Therefore, EU^i (for $i \in B'$) must be greater than EU^i (for $i \in B$). So while there can exist blocking coalitions which propose $\tau_g^b > 0$ for some $i \in B$, they are not in the core; 'voluntary' contributions to G-capital are zero. ■

Proposition 7 *If any agent may propose an enforceable tax plan then all agents will be taxed according to (15). It follows that $G = G^*$.*

Proof. Suppose that a group of agents $\mathcal{V} \subseteq I$ each seek a mandate for their proposals. Let $\mathcal{G} = \tau_g k$. The proposal $\mathcal{M}^g = \{\{\tau^{ig}\}_{i \in I} | G \leq \mathcal{G}\}$ of each agent $g \in \mathcal{V}$ includes taxation levels for each agent $i \in I$ as well as a proposed level of G-capital, $G \leq \mathcal{G}$. Agent i votes for the tax plan that will provide her with the highest expected utility; $EU^i|_{g'}$ is the expected utility to agent i if agent g' holds the mandate to tax. $V^{g'}$ denotes the set of agents who vote for the tac plan of agent g'. So, if $\#V^{g'} > \#V^{g''}$ for every $g'' \in \mathcal{V} \backslash g'$, then agent g' holds the mandate to tax, imposes taxation levels and delivers the level of G-capital in return. In the core, all agents are taxed equally and no rents accrue: $G = \mathcal{G}$. Consider the alternative to this. If an agent $g' \in \mathcal{V}$ offers a tax plan $\mathcal{M}^{g'} = \{\{\tau^{ig'} = \tau'_g\}_{i \in I} | \mathcal{G}'\}$ in which $\tau'_g k > \mathcal{G}'$ there is some other agent $g'' \in \mathcal{V}$ who offers a plan $\mathcal{M}^{g''} = \{\{\tau^{ig''} = \tau'_g\}_{i \in I} | \mathcal{G}''\}$ in which $\tau'_g k > \mathcal{G}'' > \mathcal{G}'$ which delivers $\#V^{g''} > \#V^{g'}$. Given free entry to proposing a tax plan, the rent from holding the mandate is driven to zero, so that agent g^*, who sets τ_g^* to satisfy (15), ensures that $G = \mathcal{G} = G^*$. ■

C Proof of proposition 8

The equivalence of core and competitive equilibria is established by extending the arguments of Townsend (1978). First, some notation is developed. Let any agent $h \in I$ propose strategy P^h for intermediating in a market. This strategy has eight components: M^h is the market proposed by agent h; P_1^h is the yield in terms of the consumption good of one share in the portfolio of agent h; P_2^h is the price in terms of the capital good at which agent h is willing to buy an unlimited number of shares in the goods production of any agent in M^h; P_3^h is a fixed fee in terms of the capital good for the purchase of shares in the portfolio of agent h by $i \in M^h$; P_4^h is the price in terms of the capital good at which agent h is willing to sell an unlimited number of shares in her portfolio to agents in M^h; τ_s^h is the proportion of the capital endowment that agent h proposes to invest in S-capital for that market; τ_g^h is similarly defined. Recall that with free political entry agent g^* delivers on the manifesto promise and ensures that $G = \mathcal{G}$. Finally, $\alpha^h(F)$ is the ex post exchange cost. It is not strictly necessary to include $\alpha^h(F)$ in the definition of the strategy space but it aids intuition to do so. In what follows, Q_D^{ih} is the quantity of shares purchased by i in h's portfolio, whilst Q_S^{hi} is the quantity of shares sold by i to h. A^{ih} is a switching function, where $A^{ih} = 1$ if agent i buys shares in the portfolio of intermediary h, and $A^{ih} = 0$ otherwise. One may characterize the optimal strategies for intermediaries and non-intermediaries for a given market size. The following unconstrained optimization delivers the supporting price vector. Finally, we reduce on notation by writing x when we really mean $x(\omega)$.

Definition 1 *A competitive equilibrium is a set of actions* $\left\{Q_{S*}^{ij}, Q_{D*}^{ji}, A_*^{ij}\right\}$ *and a strategy* $P_*^i = \left\{M_*^i, P_{1*}^i, P_{2*}^i, P_{3*}^i, P_{4*}^i, \tau_{s*}^i, \tau_{g*}^i, \alpha_*^i(F)\right\}$ *for each agent* $i \in I$ *(where for any variable* x*, the convention* $x^{ii} = 0$ *is adopted), an allocation* $\{c_*^i, y_*^i; i \in I\}$ *and a set of markets which satisfy:*

1. If agent i is not an intermediary ($Q_{D*}^{ji} = 0, \forall j \in I$) then $\left\{y_*^i, Q_{S*}^{ij}, Q_{D*}^{ji}, A_*^{ih}, \tau_{s*}^{hi}, \tau_{g*}^{hi}, \alpha_*^{hi}(F)\right\}$ maximizes (25) subject to (26), (27) and (28) and $P^h = P_*^h$ $\forall h \neq i$. c_*^i is given by (27), and given (30) – (33).

2. Agents participate in, at most, one market, M. In each market there is one intermediary such that $h \in M$, $M_*^h = M$, and $\left\{\tau_s^{ih}, \tau_g^{ih}, \alpha^{ih}(F)\right\} = \left\{\tau_{s*}^{ih}, \tau_{g*}^{ih}, \alpha_*^{ih}(F)\right\}$. For each $i \in M \backslash h$, $A^{ih} = 1$. For every such h, P^h is feasible, with y^h chosen to maximize (29) subject to (30) and (31). c_*^h is given by (31).

3. There exist no blocking strategies for any agent of I.

Hence, a non-intermediary faces the following problem:

$$\max_{Q_D^{ih}, Q_S^{hi}, y^i, A^{ih}} \sum_{h:i\in M^h} \left[A^{ih}\right] \left[EU\left(Q_D^{ih}P_1^h + y^i\lambda^i - Q_S^{hi}\lambda^i\right)\right], \tag{25}$$

subject to the following constraints regarding feasibility and participation,

$$\left[\sum_{h:i\in M^h} \left(A^{ih}\right)\left((1-\tau_s^h-\tau_g^h)k^i + Q_S^{hi}P_2^h - Q_D^{ih}P_4^h - P_3^h - \alpha(F^h) - y^i\right)\right] \geq 0; \tag{26}$$

$$c^i(\omega) = \left[\sum_{h:i\in M^h} \left(A^{ih}\right)\left(Q_D^{ih}P_1^h(\omega) + y^i\lambda^i(\omega) - Q_S^{hi}\lambda^i(\omega)\right)\right] \geq 0, \forall\omega \in \Omega; \tag{27}$$

$$A^{ih} = 1 \Rightarrow A^{ij} = 0, \forall i \text{ such that } i \in M^j, \forall j \neq h. \tag{28}$$

An intermediary chooses a strategy to maximize,

$$\max EU\left(y^h\lambda^h+\sum_{i\in M^h}\left[Q_S^{hi}\lambda^i-Q_D^{ih}P_1^h\right]\right), \tag{29}$$

subject to,

$$\left[(1-\tau_s^h-\tau_g^h)k^h-\alpha(F^h)\left(\#M-1\right)-y^h+\sum_{i\in M^h}\left[Q_D^{ih}P_4^h+P_3^h-Q_S^{hi}P_2^h\right]\right]\geq 0; \tag{30}$$

$$c^h(\omega)=\left[\sum_{i\in M^h}\left[Q_S^{hi}\lambda^i(\omega)-Q_D^{ih}P_1^h(\omega)\right]+y^h\lambda^h(\omega)\right]\geq 0,\forall\omega\in\Omega; \tag{31}$$

and the following relations:

$$\begin{aligned}
G &\leq \tau_g k; && (32)\\
S^h &\leq \tau_s k; && (33)\\
\alpha^h &= \left[1-F\left(S^h,G\right)\right]k; && (34)\\
\sum_{i\in M^h}P_3^h-\alpha\left(\#M-1\right) &= -2\alpha\frac{(\#M-1)}{\#M}. && (35)
\end{aligned}$$

Equation (35) may be regarded as the free-entry constraint upon the equilibrium intermediary strategy.

Proposition 9 *If there exists a competitive equilibrium, with political free-entry as defined above, then the equilibrium is in the core.*

Proof. Consider an equilibrium, for a given endowment k, for which a blocking coalition exists. Further, assume that this blocking coalition, denoted $\#B$, is a market in which transaction costs, production and consumption are Pareto optimal. In such an equilibrium, consumption is equalized across all agents in the blocking coalition: $y=(1-\tau_s-\tau_g)\,k-\left(\frac{\#B^b-1}{\#B^b}\right)2\alpha^b$ and $c^b=\bar{c},\forall i\in B$; $EU(c^b)>EU(c^h)$, $h\neq b$. For some agent $b\in B$, the following are necessary conditions for an optimal strategy:

$$\begin{aligned}
U'(\cdot)\sum_{i\in B^b}\lambda^i+\phi^b\left(\#M\right)P_2^b+\mu^b\sum_{i\in B^b}\lambda^i &= 0; && (36)\\
-U'(\cdot)\left(\#B\right)P_1^b-\phi^b\left(\#B\right)P_4^b-\mu^b\left(\#B\right)P_1^b &= 0; && (37)\\
U'(\cdot)\lambda^b+\phi^b+\mu^b\lambda^b &= 0; && (38)
\end{aligned}$$

$$\phi^b k - \phi^b F'(S^b, \mathcal{G})(\#B-1)k^2 - \eta^b F'(S^b, \mathcal{G})(\#B-1)k^2 + 2\eta^b F'(S^b, \mathcal{G})k^2 \frac{(\#B-1)}{\#B} = 0, \tag{39}$$

where ϕ^b, μ^b, η^b, are unknown multipliers on constraints (30), (31) and (35), respectively. For non-intermediaries, necessary conditions for an optimum are,

$$\begin{aligned} U'(\cdot)\lambda^i + \phi + \mu\lambda^i &= 0; \quad (40) \\ -U'(\cdot)\lambda^i - P_2^b\phi - \mu\lambda^i &= 0; \quad (41) \\ U'(\cdot)P_1^b + \phi P_4^b + \mu P_1^b &= 0, \quad (42) \end{aligned}$$

where ϕ and μ are multipliers on constraints (26) and (27), respectively. Equations (40) and (41) together imply that $P_2^b = 1$, and so it follows that $P_4^b = 1$. Use these in the intermediary's first-order conditions:

$$P_1^b = -\frac{\phi^B}{U'(\cdot) + \mu^B}, \tag{43}$$

and,

$$U'(\cdot) = -\frac{\phi^h \, (\#B) \, P_2^b}{\sum_{i \in B^b} \lambda^i} - \mu^b. \tag{44}$$

Therefore,

$$P_1^b = \left(\frac{1}{\#B}\right) \sum_{i \in B^b} \lambda^i. \tag{45}$$

Finally, we have,

$$P_3^b = [1 - F(S^b, \mathcal{G})]k \frac{(\#B-2)}{\#B}. \tag{46}$$

Since they are the shadow prices of dual constraints, it follows $\phi^b = -\eta^b$, so then (39) implies,

$$1 = 2F'_S(S^b, \mathcal{G})k \frac{(\#B-1)}{\#B}. \tag{47}$$

Equation (47) determines the unique optimal, $\tau_s^{\#B}$, for a given market size, $\#B$. Finally, the agent who decides on general capital investment will choose

$$1 = 2F'_G(S^b, G)k \frac{(\#B-1)}{\#B}. \tag{48}$$

If this agent is a non-intermediary this follows from $0 = \phi k + \phi \frac{\partial P_3^b}{\partial \tau_g} + \phi \frac{\partial \alpha}{\partial \tau_g}$; if the agent is an intermediary, it follows from the analogue to equation (39).

Consider, a political manifesto that proposes a τ_g greater or less than that proposed by (46). Since, the ECT is strictly concave, as is the utility function, a tax rate exists such that the tax burden is no higher, but exchange costs are lower. That is, consumption is strictly higher. However, an equilibrium outcome has been constructed that is consistent with efficient, optimizing behavior (across intermediaries and non-intermediaries), that is nevertheless inconsistent with property 3 in the definition of equilibrium. ■

Proposition 10 *All core allocations can be supported as equilibria.*

Proof. Assume that the optimal market size is greater than unity and less than infinity. Then the above first-order conditions can be used directly to construct an equilibrium that is a competitive equilibrium (since core and competitive equilibria coincide). Since the optimal market size exists, one may construct equilibrium markets. Hence, Properties 1 and 2 of the definition of equilibrium are met. Finally, no set of agents will block this allocation since it would be unable to attain a higher level of utility than under the cooperative equilibrium. ■

D Comparative Static Results

Remark 1 shows that market size and ECT allocations increase in $\#M$. Propositions 3 and 4 demonstrates that there may exist a k^* at which two (and no more than two) plans are optimal. It remains to characterize the relationship between k and transaction costs in between the k^* at which the optimal market size changes.

Proposition 11 *For a given market size greater than one, (i) transaction costs are strictly decreasing in k; and (ii) if F_S and F_G are inelastic with respect to S and G respectively, optimal investments in ECT rise unambiguously in k.*

Proof. *Transaction costs:* Recall $\kappa = (k-y)/k = \tau_s + \tau_g + 2\left(\frac{\#M-1}{\#M}\right)(1-F)$.

Thus,

$$\begin{aligned}\frac{\partial \kappa}{\partial k} &= \frac{\partial \tau_s}{\partial k} + \frac{\partial \tau_g}{\partial k} - \theta\left[\frac{\partial F}{\partial S}\left(k\frac{\partial \tau_s}{\partial k} + \tau_s\right) + \frac{\partial F}{\partial G}\left(k\frac{\partial \tau_g}{\partial k} + \tau_g\right)\right]; \\ &= -\left(\theta k\frac{\partial F}{\partial S} - 1\right)\frac{\partial \tau_s}{\partial k} - \left(\theta k\frac{\partial F}{\partial G} - 1\right)\frac{\partial \tau_g}{\partial k} - \theta\left(\tau_s\frac{\partial F}{\partial S} + \tau_g\frac{\partial F}{\partial G}\right),\end{aligned}$$

where $\theta = 2\left(\frac{\#M-1}{\#M}\right)$. And using the optimality conditions,

$$\frac{\partial \kappa}{\partial k} = -(\tau_s + \tau_g)/k < 0. \tag{49}$$

ECT investment: Throughout, it is assumed that $F_{SG} = F_{GS} > 0$; this is consistent with the view that specific and general capital investments are positively correlated. (14) and (15) together imply,

$$\left\{F_{SS}(S,G)\left[\tau_s + \frac{\partial \tau_s}{\partial k}k\right] + F_{SG}(S,G)\left[\tau_g + \frac{\partial \tau_g}{\partial k}k\right]\right\}k + F_S(S,G) = 0; \tag{50}$$

$$\left\{F_{GG}(S,G)\left[\tau_g + \frac{\partial \tau_g}{\partial k}k\right] + F_{GS}(S,G)\left[\tau_s + \frac{\partial \tau_s}{\partial k}k\right]\right\}k + F_G(S,G) = 0. \tag{51}$$

It should be recalled that we are characterizing optimal outcomes for a given $\#M$. From equation (51) it follows that $\frac{\partial \tau_g}{\partial k}k + \tau_g = -\frac{F_G}{kF_{GG}} - \frac{F_{SG}}{kF_{GG}}\left(\tau_s + \frac{\partial \tau_s}{\partial k}k\right)$, and combining with (50), one recovers an expression for the elasticity of τ_s with respect to k:

$$\frac{\partial \tau_s}{\partial k}\frac{k}{\tau_s} = \frac{F_G F_{SG} - F_S F_{GG}}{\tau_s k\left[F_{SS}F_{GG} - (F_{SG})^2\right]} - 1. \tag{52}$$

Equation (52) allows one to check the sign of the elasticity; $\tau_s k\left[F_{SS}F_{GG} - (F_{SG})^2\right]$ is positive by the assumption of strict concavity and $F_G F_{SG} - F_S F_{GG}$ is positive since $F_G, F_{SG}, F_S > 0$ and $F_{GG} < 0$. As such, for $\frac{\partial \tau_s}{\partial k}\frac{k}{\tau_s} > 0$, one requires that $F_{SG}(F_G + \tau_s k F_{SG}) > F_{GG}(F_S + \tau_s k F_{SS})$. The left hand side is positive; a sufficient condition then is that the right hand side is negative. This requires that F_S is inelastic with respect to S, i.e., that$-\frac{SF_{SS}}{F_S} < 1$. Assuming the equivalent condition holds for general capital, optimal investments in general and specific capital are, for a given market size, strictly increasing in the level of the endowment. ■

E Numerical solution to the model

First, the choice of functional form for the ECT and utility function is discussed, as is the modelling of consumption-good productivity. Next, the solution of the model is explained and analyzed holding fixed the level of the endowment. The solution of the model is then studied when k varies. Finally, the robustness of the analysis to the risk aversion and ECT parameters is examined.

E.1 A numerically tractable model

Consider a general CES form for the ECT,

$$F(S,G) = \left[\delta F_s(S)^{\sigma} + (1-\delta) F_g(G)^{\sigma}\right]^{1/\sigma},$$

where $\delta \in (0,1)$ and $1/(1-\sigma)$ is the constant elasticity of substitution. Functions F_x ($x = s$, $x = g$) are continuous, decreasing and strictly concave in their only argument; in particular, $F_x' > 0$, $F_x'' < 0$ and $F_x(0) = 0$, $F_x(y) \to 1$ as $y \to \infty$, $F_x'(y) \to 0$ as $y \to \infty$ and $F_x'(0) = \infty$ for $x \in \{s, g\}$. F_g and F_s need not be identical functions. Thus,

$$\begin{aligned} F_s &= \left[1 - \exp(-\beta \tau_s k)\right]^{\gamma_s}; \\ F_g &= \left[1 - \exp(-\beta \tau_g k)\right]^{\gamma_g}, \end{aligned}$$

where γ_s need not be equal to γ_g and where $\beta, \gamma_s, \gamma_g \in (0,1)$. These functions and parametric restrictions have a number of useful properties that both satisfy the restrictions on the ECT and facilitate numerical analysis. Taking the first partial of F_s with respect to τ_s one obtains,

$$F_s' = \beta k \gamma_s \exp(-\beta \tau_s k)\left[1 - \exp(-\beta \tau_s k)\right]^{\gamma_s - 1}$$

The Inada-type requirements are readily confirmed. In addition, the parameters $\beta, \gamma_s, \gamma_g \in (0,1)$ allow one to vary the effectiveness of the ECT.[23] one may use the parameters γ_s, γ_g to make distinctions between the relative efficiency of specific and general investments to exchange cost alleviation.

Agents have identical CRRA utility, $U(x) = \left[x^{1-\gamma} - 1\right]/(1-\gamma)$ with $\gamma > 0$ being the coefficient of relative risk aversion. For numerical tractability, we specialize to a two-state case in which production technologies are restricted to $\Lambda = \{\lambda_1, \lambda_2\}$ where $\lambda_1 < \lambda_2$ and $p(\lambda_1) = \rho$ and $p(\lambda_2) = 1 - \rho$.

The expected average technology for a market of size $\#M$ is given by the following expression

$$E\left[\bar{\lambda}_{\#M}\right] = \sum_{i=0}^{\#M} \left\{ \binom{\#M}{i} \rho^{\#M-i} (1-\rho)^i \left[(\#M - i)\lambda_1 + i\lambda_2\right] \right\} / \#M, \quad (53)$$

which is invariant to $\#M$. One may now state the problem:

$$\Gamma = \arg\max_{\#M} \sum_{i=0}^{\#M} \binom{\#M}{i} \rho^{\#M-i} (1-\rho)^i U \left[\begin{array}{c} \left\{ (1 - \tau_s - \tau_g)\, k - \frac{(2\alpha)(\#M-1)}{\#M} \right\} \times \\ \times \left[(\#M - i)\lambda_1 + i\lambda_2\right] / \#M \end{array} \right]$$

[23]We use MATLAB to compute equilibria. MATLAB code and simulation output is available from the authors.

subject to

$$2\left(\frac{\#M-1}{\#M}\right)\left[\delta\left[1-\exp\left(-\beta\tau_s k\right)\right]^{\gamma_s\sigma}+(1-\delta)\left[1-\exp\left(-\beta\tau_g k\right)\right]^{\gamma_g\sigma}\right]^{\frac{1-\sigma}{\sigma}}\times$$
$$\times\,\delta\beta k\gamma_s\exp\left(-\beta\tau_s k\right)\left[1-\exp\left(-\beta\tau_s k\right)\right]^{\gamma_s\sigma-1}=1,$$

$$2\left(\frac{\#M-1}{\#M}\right)\left\{\delta\left[1-\exp\left(-\beta\tau_s k\right)\right]^{\gamma_s\sigma}+(1-\delta)\left[1-\exp\left(-\beta\tau_g k\right)\right]^{\gamma_g\sigma}\right\}^{\frac{1-\sigma}{\sigma}}\times$$
$$\times\,(1-\delta)\,\beta k\gamma_g\exp\left(-\beta\tau_g k\right)\left[1-\exp\left(-\beta\tau_g k\right)\right]^{\gamma_g\sigma-1}=1,$$

where $\alpha=\left\{1-\left[\delta\left[1-\exp\left(-\beta\tau_s k\right)\right]^{\gamma_s\sigma}+(1-\delta)\left[1-\exp\left(-\beta\tau_g k\right)\right]^{\gamma_g\sigma}\right]^{1/\sigma}\right\}k$ and $\max\{\Gamma\}$ is selected as the unique solution. The two constraints to solve simultaneously for the optimal $\tau^*=\left\{\tau_s^*,\tau_g^*\right\}$.

E.2 ECT, diversification and utility with a fixed endowment

Table 2 gives the parameter values in the baseline case.

Table 2: Baseline Calibration for Endogenous Exchange Costs

endowment	k	25
coefficient of relative risk aversion	γ	3
specific ECT curvature	γ_s	0.065
general ECT curvature	γ_g	0.085
ECT factor	β	0.03
weight on specific ECT	δ	0.5
CES coefficient s, $1-\frac{1}{s}=$	σ	0.5
low technology	λ_1	1
high technology	λ_2	5
probability of low technology	ρ	0.5

The baseline case supposes a difference in the effectiveness of S-capital and G-capital; $\gamma_g>\gamma_s$ and so G-capital is the less efficient.

First the effect on optimal market size and expected utility of different levels of ECT investment is examined. Agents choose how far (if at all) to diversify given their residual endowment and exchange costs. Figure 6 displays expected utility solving at optimal market size over a grid of $\{\tau_s, \tau_g\}$; figure 7 gives choices of market size.

Figures 6 and 7 make clear that over some combinations of the $\{\tau_s, \tau_g\}$ pairs, expected utility from diversification is higher than from not diversifying. There is a peak in expected utility at some unique combination of τ_s and τ_g. Clearly, market size does not peak at the peak of expected utility: Sub-optimal institutions can induce *larger* markets, as was explored in Section 6.

E.3 ECT, diversification and utility with a varying endowment

Using the parameter values given in Table 2, k is permitted to vary. The solution algorithm evaluates $\left\{\tau_s^*, \tau_g^*\right\}$ using the marginal conditions for market sizes up to some $\#\hat{M} > 1$ and selects the diversification level that maximizes expected utility (increasing $\#\hat{M}$ as required when $\#\hat{M}$ is the utility maximizing choice). Figure 1 gives results for these simulations.

E.4 Robustness to risk aversion and ECT parameters

There are computational limitations to numerical analysis of the model. MATLAB v.7, for example, will only calculate up to 170!. As can be seen in the bottom-left panel of Figure 1, this will quickly become a limiting problem. However, one can see the effect of changes in some other parameters. Figures 8 and 9 give results under different parameterizations of risk aversion and ECT efficiency. In each, the central case (i.e., the middle line) reflects the parameterization in Table 2. Variations on the coefficient of relative risk aversion, γ, and on the ECT, β, are Table 2 values multiplied by $1 \pm \frac{1}{6}$.

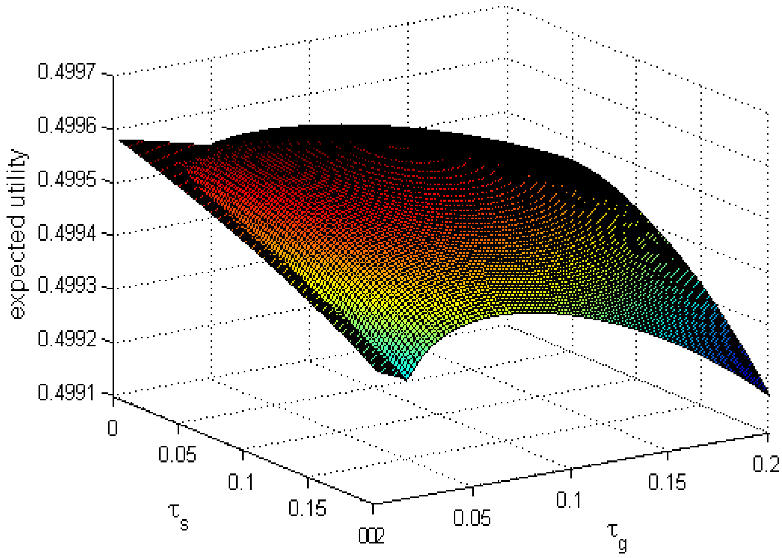
Figure 8 shows some variation in optimal behavior in regard to risk aversion. For a given endowment, the optimal market size is decreasing in risk aversion. Further, agents are willing to spend more on diversification (i.e., forming markets), as can be seen by the size of transaction costs in each case. Figure 9 demonstrates the effect of varying the coefficient on the ECT. Reducing β means that agents diversify at lower endowments. The size of transaction costs is lower for a given endowment, but relatively unchanged for a given market size. The effect of changing the exchange technology is primarily to make it feasible for agents to diversify with a lower endowment: The nature of that diversification is little affected.

E.5 Distortive institutions and the transaction tax

For distortive institutions, the algorithm is a simple extension of that described above: One uses the marginal condition to find τ_s^* for a grid of τ_g and identify the τ_g which obtains the two rules of thumb. For the transaction tax, the value for t is introduced to the expression for α and agents optimize over investments and market size taking t as given.

F Figures

Figure 6: Expected Utility and ECT Investment



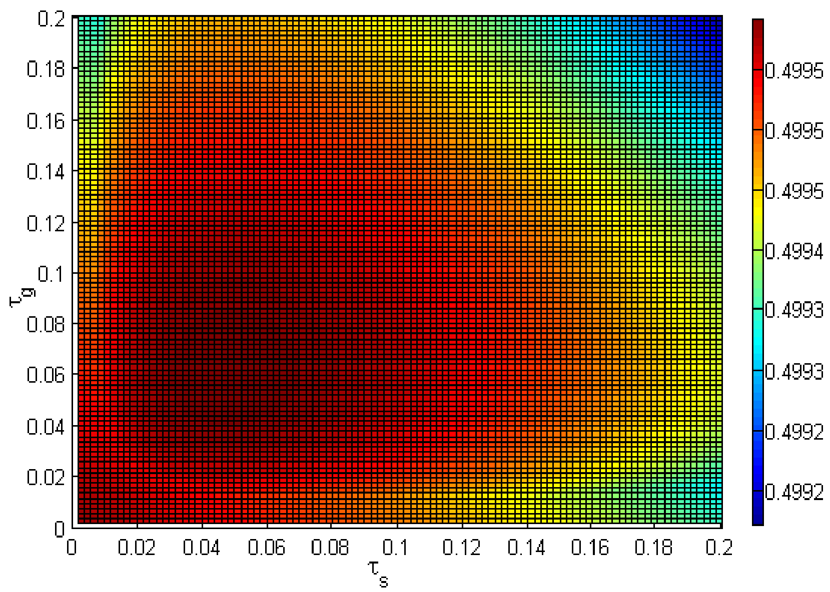


Figure 7: Optimal Market Size and ECT Investment

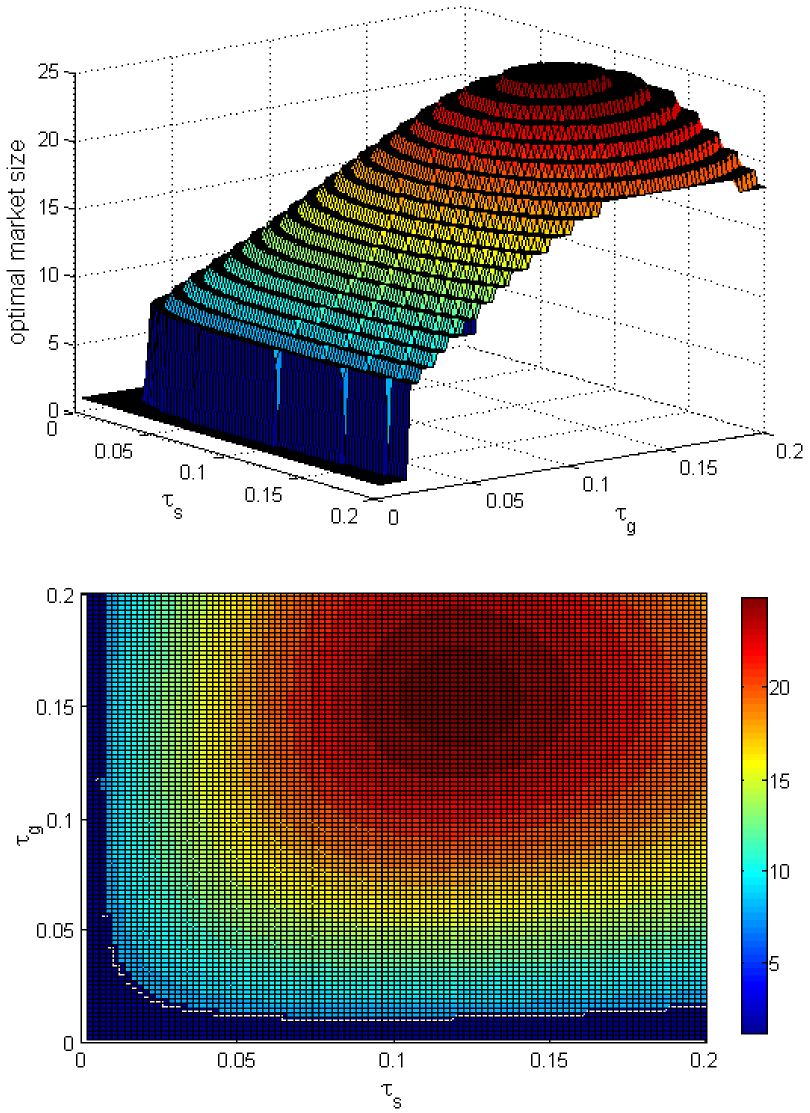


Figure 8: Risk Aversion and Endogenous Exchange

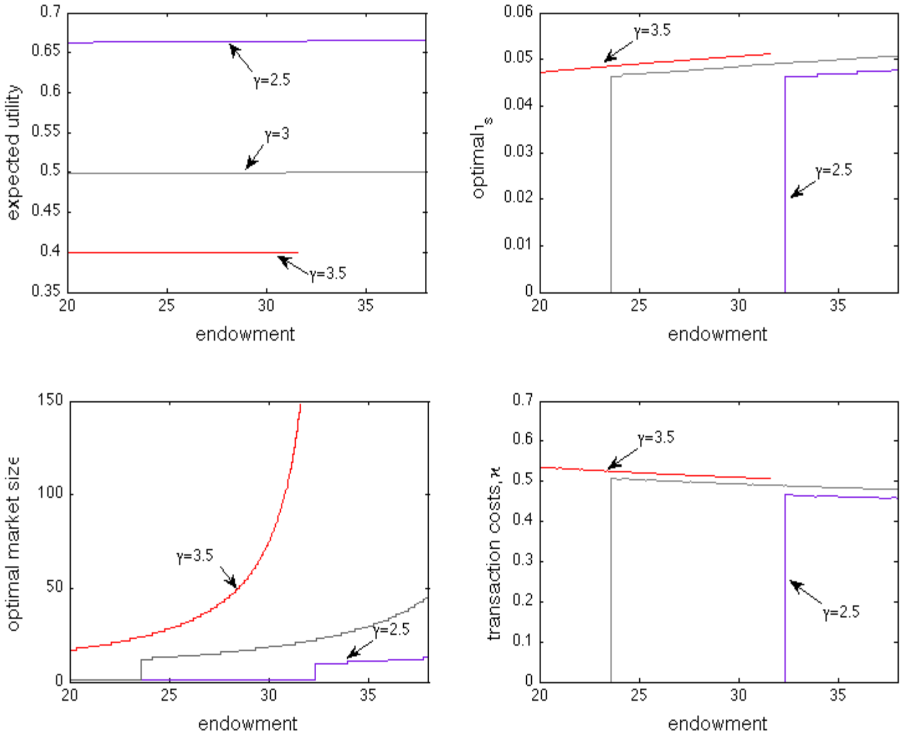


Figure 9: ECT Calibration and Endogenous Exchange