COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1247

DEFAULT IN A GENERAL EQUILIBRIUM MODEL WITH INCOMPLETE MARKETS

Pradeep Dubey, John Geanakoplos and Martin Shubik

January 2000

Default in a General Equilibrium Model with Incomplete Markets[1]

Pradeep Dubey, John Geanakoplos, and Martin Shubik

Abstract

We extend the standard model of general equilibrium with incomplete markets (GEI) to allow for default. The equilibrating variables include aggregate default levels, as well as prices of assets and commodities. Default can be either strategic, or due to ill-fortune. It can be caused by events directly affecting the borrower, or indirectly as part of a chain reaction in which a borrower cannot repay because he himself has not been repaid.

Each asset is defined by its promises A, the penalties λ for default, and the limitations Q on its sale. The model is thus named $GE(A, \lambda, Q)$. Each asset is regarded as a pool of promises. Different sellers will often exercise their default options differently, while each buyer of an asset receives the same pro rata share of all deliveries. This model of assets represents for example the securitized mortgage market and the securitized credit card market.

Given any collection of assets, we prove that equilibrium exists under conditions similar to those necessary to guarantee the existence of GEI equilibrium. We argue that default is thus reasonably modeled as an equilibrium phenomenon. Moreover, we show that more lenient λ which encourage default may be Pareto improving because they allow for better risk spreading.

Our definition of equilibrium includes a condition on expected deliveries for untraded assets that is similar to the trembling hand refinements used in game theory. Using this condition, we argue that the possibility of default is an important factor in explaining which assets are traded in equilibrium. Asset promises, default penalties, and quantity constraints can all be thought of as determined endogenously by the forces of supply and demand.

Our model encompasses a broad range of moral hazard, adverse selection, and signalling phenomena (including the Akerlof lemons model and Rothschild–Stiglitz insurance model) in a general equilibrium framework. Many authors (including Akerlof , Rothschild and Stiglitz) have suggested that equilibrium may not exist in the presence of adverse selection. But our existence theorem shows that it must. The problem is the inefficiency of the resulting equilibrium, not its nonexistence. The power of perfect competition simplifies many of the complications attending the finite player, game theoretic analyses of the same topics.

The Modigliani–Miller theorem typically fails to hold when there is the possibility that the firm or one of its investors might default.

Keywords: default, incomplete markets, adverse selection, moral hazard, equilibrium refinement, endogenous assets

[1] This work is supported in part by NSF Grant DMS 8705294 and SES–881205. The first version of this paper appeared as Cowles Foundation Discussion Paper #773, in 1990. That version, containing essentially the basic model and Theorems 1 and 2 from this paper, was never published, though our model has been frequently used and cited: see, for example, Zame [1993]. An expanded version, very similar to the present version, was circulated in 1994. The current version corrects some minor errors in the examples in the 1994 version.

Contents

I. Default in Equilibrium

1 Introduction

There is a substantial amount of default in the American economy. At first glance this would seem to be a sign of disequilibrium, and to call for economic models that radically depart from the orthodox paradigm of general equilibrium and market clearing.

Indeed, general equilibrium theory has for the most part not made room for default. In the Arrow–Debreu model of general equilibrium with complete contingent markets (GE), and likewise in the general equilibrium model with incomplete markets (GEI), agents keep all their promises by assumption. More specifically, in the GE model, agents never promise to deliver more goods than they personally own. In the GEI model, the definition of equilibrium (that has been developed in a rapidly growing literature) allows agents to promise more of some goods than they themselves have, provided they are sure to get the difference elsewhere. Agents there too must honor their commitments, though no longer exclusively out of their own endowments. Each agent can keep his promises because other agents keep their promises to him.

We build a model that explicitly allows for default, but is broad enough to incorporate conventional general equilibrium theory as a special case. We call the model $GE(A, \lambda, Q)$ because each asset j is defined by its promise A_j, the penalty λ_j for default on the promise, and the quantity restriction Q_j attendant on those who sell it.

Fixing exogenously the set $\mathcal{A}$ of tradeable assets,

$$\mathcal{A} = \{(A_j, \lambda_j, Q_j) : (A_j, \lambda_j, Q_j) \text{ is tradeable}\},$$

we solve for equilibrium $E(\mathcal{A})$. The levels of trade of every asset in $\mathcal{A}$, the rates of default of every asset in $\mathcal{A}$, and the prices of every asset in $\mathcal{A}$, emerge endogenously as part of the equilibrium $E(\mathcal{A})$. In contrast to the GE and GEI models, in which prices are the only equilibrating variables, in the $GE(A, \lambda, Q)$ model, prices and expected default rates are needed to clear markets.

One of the central features of our model is that assets must be thought of as pooling devices. Different sellers of the same asset will typically default in different events, and in different proportions. The buyers of the asset receives a pro rata share of all the different sellers' deliveries, just as an investor does today in the securitized mortgage market, or in the securitized credit card market. Our general equilibrium model thus also stands in contrast to models in which a single lender and a single borrower must negotiate with each other.

We have avoided a finite-player, game-theoretic treatment of default because, for the massive anonymous financial markets on which we focus attention, perfect competition is a better approximation to reality, and much more analytically tractable. We have also avoided a (perfectly competitive but) partial equilibrium treatment of our subject because we wanted to evaluate the system wide consequences of default. In a world in which promises can exceed physical endowments, each default can begin a chain reaction. A creditor in one market where payment does not occur is deprived of the means of delivery in another market where he is the debtor, thereby causing a further default in some other market, etc. The indirect effects of default might be as important as the direct effects, but they are missed in partial equilibrium models. We emphasize that these chain reactions occur exclusively in economies with intermediate levels of financial development, such as the system now in place in the United States. Once the asset markets become complete, the system of interlocking debts will be broken, as in the GE model, and no chain reactions will occur.

Another central feature of our model is that the subset $\mathcal{A}^* \subset \mathcal{A}$ of actively traded assets

$$\mathcal{A}^* = \{(A_j, \lambda_j, Q_j) \in \mathcal{A} : (A_j, \lambda_j, Q_j) \text{ is positively traded in } E(\mathcal{A})\}$$

also emerges in equilibrium. The promises, penalties, and sales limitations corresponding to actively traded assets can thus themselves be regarded as endogenous. Whereas in GEI the selection of assets is usually regarded as outside the model, here we can resolve the asset selection problem by taking the set $\mathcal{A}$ of traded assets to be inclusive and focusing on the endogenous determination of positively traded assets $\mathcal{A}^*$. In equilibrium there will typically be many assets in $\mathcal{A} \backslash \mathcal{A}^*$ which are priced by the market, but neither bought nor sold.[2]

Special cases in the literature on asset selection can be obtained by specializing the model. For example, we render the Rothschild–Stiglitz insurance model with quantity signals a special case of our model by holding promises and penalties fixed exogenously at a single level (say, $A_j = \bar{A}$, $\lambda_j = \bar{\lambda}$ for each traded asset (A_j, λ_j, Q_j) in $\mathcal{A}$), while allowing for assets corresponding to every possible level of the quantity dimension. If in equilibrium only one or two assets j are positively traded, then equilibrium has endogenously determined the quantity signals Q_j of positively traded assets. The forces of supply and demand in equilibrium also determine a different price for each possible quantity signal, not just for the two that are actively traded, giving what has come to be called a non-linear pricing schedule.

The equilibrating macrovariables in our model are commodity prices, asset prices, and asset default rates (or, equivalently, delivery rates). A crucial role in the endogenous determination of asset trade is played by the expectations agents have over the deliveries of assets that are not positively traded. (In game theoretic terms, this is analogous to beliefs off the equilibrium path.) We fix these expectations for non-traded assets at reasonable levels by a straightforward equilibrium refinement. The simplicity of the refinement is due to our hypothesis of perfect competition consistently applied.[3]

In Sections 2–4 we describe the model and explain who bears the loss from default and how the penalties are administered. We explain that endogenous default necessarily involves adverse selection and moral hazard. Indeed we note that the

[2] In some applications we might choose to limit $\mathcal{A}$ exogenously; the point is that even if $\mathcal{A}$ is inclusive, $\mathcal{A}^*$ will still be limited.

[3] To the best of our knowledge, this appears to be the first application of the standard "trembling hand" refinement to perfectly competitive equilibrium.

standard adverse selection and signalling models of Akerlof (1972) and Rothschild and Stiglitz (1976) are special cases of our model. Section 4 describes and justifies our equilibrium refinement.

Our first goal is to show that if agents have the mental powers to anticipate future rates of default (contingent on future events), just as they are presumed by conventional equilibrium theory to have the mental powers to anticipate future prices (contingent on future events), then default is consistent with the orderly function of markets. In Section 5 we prove the existence of equilibrium with default under exactly the same conditions necessary to prove the existence of equilibrium in the GEI model (where default is ruled out by assumption.) More precisely, we show that our refined equilibrium $E(\mathcal{A})$ exists for every collection $\mathcal{A}$ of assets (A, λ, Q) for which $Q < \infty$, or for which $Q = \infty$ but the promises A are all paid in the same numeraire.

This general existence of equilibrium is somewhat surprising, because default seems linked to disequilibrium, and because we know from the GEI literature that the existence of equilibrium can be compromised when the asset span is endogenous, and because Akerlof and Rothschild and Stiglitz all suggested that equilibrium might not exist. Our general existence proof is also surprising in that it seems to counter the suspicion that asymmetric information creates an obstacle to competitive equilibrium. (See for example, Helpmann and Laffont, 1975). The key to existence is that the asymmetry is one-sided. Each seller has the option to deliver whatever he wants, while all buyers get the same payoff (per unit purchased). Were the asymmetry two-sided, then indeed equilibrium would be more problematic.[4]

In Section 6 we describe how chain reactions could occur in the model. We also make the obvious but important point that in a very primitive financial world with one or no assets, there cannot be chain reactions of default. Furthermore, at the opposite extreme, in an Arrow–Debreu world, there will also be no chain reactions because no agent need ever promise to deliver more than he himself has on hand. One agent's default will therefore not compromise any other debtor's ability to repay. Chain reactions are thus characteristic of financial economies with intermediate levels of development.

Section 7 describes an example of equilibrium with default whose variations will be analyzed in subsequent sections.

[4] Consider an outdoor market at which different farmers can put apples up for sale in the same bin. Each farmer may know how many of his own apples are rotten, but suppose all the apples are mixed together in the bin. If buyers canot pick out their favorite apples, but must order by the number of randomly chosen apples, then a single price can clear the market for these (non-homogeneous) apples. But if buyers have asymmetric capacities for detecting rotten apples, and if the buyers were allowed to examine the fruit and to choose their apples, then a homogeneous price per apple might not be enough to clear the apple market.

There is also no problem in clearing the (heterogeneous) apple market with one price if, in addition, buyers can pay with real or counterfeit coins, provided that all sellers get the same distribution of coins. What is crucial is that each market can be separated into two sides α and β such that every trader on the α-side receives the same relative bundle of β-goods, though each β-trader may deliver a different bundle to the market, and similarly each β-trader receives the same proportions of α-goods, though each α-trader may deliver a distinct α-bundle to the market.

We elaborate this general situation in other work.

The consequences of default are potentially ruinous, yet many economic systems permit them, at least to a certain extent. (To be sure, some societies are more tolerant of default than others.) Since the imposition of default penalties causes a deadweight loss of utility which could be avoided altogether, either by abolishing the penalties or else by making them so harsh that nobody dares incur them, a rationale for intermediate levels of default penalties is called for. From a historical or legal perspective, many explanations suggest themselves: protection of creditors, deterrence of debtors, punishment commensurate with the crime, etc.

In Section 8 we give a purely economic explanation for intermediate default penalties by showing that when markets are incomplete, intermediate levels of penalties that encourage a limited amount of default can raise the level of overall economic efficiency, making both creditors and debtors better off, even when the whole chain of indirect effects is accounted for. Sometimes this range of appropriate intermediate levels of default penalties consists of no more than a single point, as in the example presented in Sections 7 and 8. In such cases we speak of the optimal default penalty. In the Arrow–Debreu world where all contingencies can be foreseen and written into the contract, the first welfare theorem demonstrates that contracts should be strictly enforced, so the optimal default penalty is infinitely harsh. But if some contingencies cannot be written into the contract, as will be the case when markets are incomplete, then it may be advisable not to punish severely those who default, even when the penalties cannot be varied with the reason for defaults.

We are careful to explain the two reasons why lenient default penalties are advantageous when markets are incomplete. First, default allows agents to tailor-make promises into deliveries that suit them best. In effect they can replace the given assets by more appropriate assets. Second, the span of the asset deliveries can be made much larger than the span of the asset promises, since a single given asset can be made into as many different assets as there are agents, if different agents default differently on the same promises. These reasons are seen more clearly in Section 9 where we embed our model of default in another model which includes factors tending to make markets incomplete.

In Sections 9–12 we study the set $\mathcal{A}^*$ of positively traded assets in equilibrium $E(\mathcal{A}^*)$, with an eye to making asset promises, default penalties, and sales restrictions endogenous. We note first that if all assets are available, so $\mathcal{A} = \overline{\mathcal{A}}$, the comprehensive collection of assets, then in equilibrium, essentially only the Arrow security promises, with infinite penalties and unlimited sales restrictions will be actively traded, $\overline{\mathcal{A}}^* = \{\text{Arrow securities}\}$. The Arrow–Debreu economy thus emerges endogenously as the equilibrium choice in a world without frictions.

In Section 9 we postulate transactions costs (that might decrease with increased liquidity), and we postulate that contingent securities cannot be traded until after a fixed "evaluation" cost is paid. Both of these market impediments decrease the social efficiency of trading in many asset markets and increase the social benefit of packaging heterogeneous promises into a single security. Under these circumstances we show that even if it were possible to write any conceivable contingent contract, and to set any degree of harshness for the default penalties, active equilibrium trade will involve fewer markets and lenient penalties.

In Sections 10–12 we drop the transactions and evaluation costs of Section 9, and we study how the set of actively traded assets $\mathcal{A}^*$ varies with $\mathcal{A}$. Each asset (A_j, λ_j, Q_j) is characterized by three dimensions. We show that exogenously restricting two dimensions of available assets in $\mathcal{A}$ leads to endogenous restrictions in the other dimension in $\mathcal{A}^*$. We show by example in Section 10 that when the set of potential promises in $\mathcal{A}$ is incomplete, the forces of supply and demand will select default penalties of intermediate harshness. We show by example in Section 11 that if $\mathcal{A}$ contains only assets whose default penalties are not too harsh, then active equilibrium trade will involve a narrow span of asset promises, even if all asset promises are available and priced by the market. The Arrow security promises are particularly vulnerable to adverse selection and may not be actively traded.

In Section 12 we fix the penalties and promises and see which sales restrictions Q_j emerge in active equilibrium trade. This enables us to show how the phenomenon of signalling can be treated in perfect competition, moreover without jeopardizing the existence of equilibrium. Furthermore, by suitable choices of default penalties, we subsume insurance contracts in our framework of default. In particular, the models of Akerlof (1972) and Rothschild and Stiglitz (1976) can be embedded as special cases in our model. Since both of those models were used by their authors to show that equilibrium may not exist in the presence of adverse selection, we carefully describe the difference between our definition of equilibrium and theirs.

The Rothschild–Stiglitz insurance model presumes that asset sales are exclusive, that is that each agent can sell at most one asset (i.e., obtain at most one insurance contract). Our model in Sections 1–11 does not make any exclusivity assumptions. In Section 12 we add the exclusivity hypothesis to our model, and we show that equilibrium must still exist. Specializing the model to the Rothschild–Stiglitz insurance setting, we show that their "separating" contracts always form a $GE(A, \lambda, Q)$ equilibrium, even when they say there is no equilibrium. The "pooling" contract is never a $GE(A, \lambda, Q)$ equilibrium when the exclusivity hypothesis is in play. The crucial difference between the Rothschild–Stiglitz definition of equilibrium and ours can be understood in terms of the assumption each makes about the reliability of untraded contracts $j \in \mathcal{A}/\mathcal{A}^*$. We argue in Section 2 and Section 12 that our assumption is natural when there are many buyers and sellers, and corresponds to cautious expectations. These expectations are determined by the standard "trembling hand" refinement to equilibrium that we give. By contrast, the expectations attributed to agents by Rothschild and Stiglitz are not compatible (to our way of thinking) with perfect competition.[5]

A second difference between our model of insurance and the Rothschild–Stiglitz model is that we also consider situations in which agents have access to a mixture of insurance policies. In the first part of Section 12 we consider insurance as a special case of the model of Sections 1–11, that is, without the exclusivity hypothesis.

[5] Indeed, Rothschild and Stiglitz seem to have in mind oligopolistic insurance companies designing contracts for a continuum of private consumers. We have recast the story into a perfectly competitive setting.

We permit an agent to take out fire insurance and theft insurance from different companies, and even multiple life insurance policies.

Equilibrium in the Rothschild–Stiglitz model, when it exists, is the "separating" equilibrium in which the least reliable agents can purchase as much insurance as they like, albeit at a very bad price because they reveal themselves to be unreliable, while the most reliable agents are constrained from buying much insurance, albeit at a very good rate, since by accepting the constraints they signal that they are reliable. By contrast, when we allow for multiple insurance policies, we find that it is the unreliable agents who are quantity constrained, albeit at rates that are better than actuarially fair for them. In our equilibrium one insurance contract (the "standard" insurance package) is sold to every agent, but in quantity only up to the level equal to what the most reliable agents want to purchase. The standard insurance price reflects the average reliability of the whole population. The least reliable agents purchase the standard insurance package, but since they know themselves to be more accident prone, they feel constrained. The least reliable agents therefore purchase additional insurance from a second carrier, at a much higher price since by doing so they reveal themselves to be unreliable. This equilibrium appears to conform better with common practice than the "separating" equilibrium identified by Rothschild–Stiglitz. The equilibrium we obtain was not considered by Rothschild and Stiglitz because they assumed that no agent could take out more than one insurance policy.

Finally, once we have a model with default in equilibrium, we can investigate the Modigliani–Miller assertion that the form of the promises is irrelevant to the outcome. Our model does not include equity, but we do include junior and senior debt. The question becomes whether it matters how a firm divides its debt between junior and senior issues. Our analysis of Modigliani–Miller uses the same apparatus of quantity constraints that we developed in Section 12 for the Rothschild–Stiglitz model. In Section 13 we find that in general the Modigliani–Miller claim is false if markets are incomplete, and for two reasons. The first reason is that if there is default on the senior debt, then the span of asset deliveries will typically change as junior debt is substituted for senior debt, provided there are more states of nature than assets.

Second, different agents have different propensities to repay. If a very reliable firm reduces its senior debt (deleverages), then according to the Modigliani–Miller theorem, its junior debt owners can compensate for that by issuing their own private debt, that is by privately leveraging. But those individuals may not get the same favorable borrowing terms the firm got, and so private leveraging may be impossible. Indeed, one crucial explanation for why firms leverage is that their shareholders cannot leverage themselves.

2 Adverse Selection and Moral Hazard in Perfect Competition

In keeping with the spirit of perfect competition, which is the hallmark of general equilibrium, we suppose that all trades are mediated by the market at market-given prices. This situation arises in practice when agents trade small quantities with each of many partners via a market. It differs from the standard framework of adverse selection and moral hazard found in so-called "principal–agent" models and "matching" models in which a single buyer confronts a single seller to negotiate a large transaction (see, e.g., Gale [1992]). But it still leaves plenty of room for adverse selection and moral hazard.

In the GEI model agents sell prespecified assets, i.e., promises to deliver commodities and money in the future, contingent on an observable state of nature. We extend that model by giving the seller of an asset the *option* of delivering whatever he chooses, i.e., of *defaulting*. However, an agent who sells an asset but does not completely pay what the asset promises in a state of nature incurs a penalty. As a result of the option, different agents may pay off differently on the same asset, so that the revenue from purchasing an asset depends on the asset's promises and on the identities of the sellers.

Adverse selection enters the picture because different sellers may have different proclivities to keep promises, for example because they have different disutilities for the penalties incurred by defaulting, or because they have different endowments out of which to pay their debts. Since there is potentially a (negative) correlation between an agent's proclivity to repay and the quantity of promises he is likely to try and sell, buyers must be aware that the default rates they face will be different from those they would get from the median seller. Moral hazard enters the picture twice, first because agents have a choice not to repay, and second because an agent who sells many assets will be less able to fully deliver on any one of them than he would if he had refrained from overextending himself.

In finite player, game theoretic analyses of the strategic role of asymmetric information, moral hazard and adverse selection play additional roles, that we do not allow here, stemming from the supposition that each agent has a large impact on traders he deals with. For example, those models posit that anyone who lends another agent more money must take into account the moral hazard that the borrower might as a result pursue a larger and riskier project, and hence the probability of repayment might be affected. Similarly, anyone who unilaterally offers a higher price for the same promise (equivalently, a lower interest rate for the same loan) must take into account the adverse (or favorable) selection effect on the kind of people who want that loan. We ignore these complexities and retain the hypothesis of perfect competition.

In our model, agents do not unilaterally set price; the market sets the price. If the market price changes, then indeed agents must rationally anticipate that the selection of sellers (borrowers) will change, and that even the same borrowers may repay differently due to the moral hazard. But in our model no agent has the power, or perhaps the visibility, to set a price different from the market price.[6] Equivalently, we might say that a buyer of an asset can set any price he wants, but in doing so he makes the cautious assumption that the selection of sellers he will find is no different from what the market price elicited. (Cautious expectations are defendable on their own merits, but it is probably worth pointing out that in markets with a large number

[6] We think of each buyer as a point in a continuum of buyers.

of traders, a buyer who credibly and visibly offers a price above the market price will be deluged with more sellers than he can accommodate. From which seller is he likely to buy? The sellers with the greatest incentive to get to him first are the ones who would have already been willing to sell at the low market price and now find an opportunity to make a surplus, not the sellers who just barely prefer to sell at all at the new high price.) With cautious expectations, there is no reason for a buyer to offer more than the market price, since he can already purchase whatever quantity he desires at the market price.

In our model, lenders provide money to a pool of heterogeneous borrowers. No lender can observe the personal characteristics of any particular borrower, but he can formulate a judgment about the (state-contingent) rate of repayment for the pool as a whole. Ultimately he will receive a share of the deliveries from the whole pool of borrowers. Furthermore, he supposes that he is so negligible compared to the size of the pool he lends to that nothing he does can affect the general terms of trade. Therefore he should figure that no matter how much money he lends, the rate of repayment in any state of nature will be unaffected. He obtains a prorated share of the aggregate default. Adverse selection and moral hazard are nevertheless incorporated into a framework of perfect competition by enlarging the traditional set of equilibrating price variables to include rates of repayment.

The large markets on Wall Street conform to our spirit of perfect competition and anonymous trade. For example, in the mortgage backed securities (MBS) market, investors buy shares of a pool of home mortgages. The homeowners have the option to default on their mortgage payments. The investing agent, however, is not matched with a particular homeowner. On the contrary, he gets a share of the payments of all the homeowners in the pool, so that his risk is diversified. He collects potentially different amounts from different homeowners selling the same asset. MBS payments also differ depending on the identity of the homeowners because the homeowners are given a second option, to prepay the mortgage. In the MBS market there are widely disseminated predictions of the future average rate of default and prepayment, conditional on the realized state of the world (typically specified by interest rates and perhaps one or two other parameters). Similarly we suppose in our model that the state contingent rates of repayment are known as part of the definition of equilibrium.

In the mortgage market, banks act as intermediaries, approving each mortgage after checking the homeowner's credentials, and then selling the mortgage to a Wall Street firm, or first to GNMA or FNMA, which in turn sell them to Wall Street. Many of the finite game theoretic models of default emphasize the bilateral negotiation between the bank and the homeowner. We take the opposite view in this paper, that the bank plays a mostly mechanical role not requiring any judgment about the quality of homeowners beyond checking customer assertions of objective facts that must be passed on to the mortgage market. We therefore concentrate on the decisions made by the homeowner–borrowers who sell the assets and the investor–lenders who purchase the assets, leaving the banks entirely out of the picture. Indeed we find in recent practice that banks are compensated in their mortgage efforts not for their judgment in choosing reliable homeowners, but primarily from servicing fees for collecting payments and other administration.

Mutual funds are another prominent example of securities that aggregate the payments from many parties. For that matter, virtually all companies whose stocks are traded over the New York Stock Exchange can be regarded as conglomerations of different businesses whose profits are summed and distributed to the shareholders. In practice, the purchase of a single asset often brings revenues from many different sources, which can be conveniently approximated by the limiting case of an infinite pool of sources.

In this paper we do not allow for options beyond default, but they could be handled in the same manner. One should note that in many options markets there is a central clearing house. Agents trade the options against the clearing house, not against each other in bilateral negotiations. Often the clearing house guarantees payment (here the option is held by the buyer of the asset instead of by the seller). This guarantee tends to make the payments independent of the identity of the sellers, but if the guarantee should fail then the system would revert to one akin to our model.

3 Failure Laws

Once we allow for default it is evident that society has much to gain from punishing those agents who fail to keep their promises. In a multiperiod world, market forces themselves might provide some incentive to keep promises, since agents who acquired a bad reputation for previous defaults might find it more difficult to obtain new loans. Collateral is also a very important device for guaranteeing at least partial payment (see Geanakoplos (1996)); but here we ignore it. In practice, market incentives are almost always supplemented by third party penalties such as prison terms or garnishing of future income. For reasons of simplicity and tractability, we confine attention to a two period model with exogenously specified default penalties which are increasing in the size of the default. These penalties might be interpreted as the sum of third party punishment, future (unmodeled) reputation losses, and pangs of conscience. If the reader prefers a more literal interpretation of the model, in which the world actually ends after the second period, then the default penalties must be understood as the utility loss imposed on agents who default.

The seasoned economic theorist might argue that it would be "nicer" or more logically satisfying to model all failure or bankruptcy penalties as strictly economic. We agree only in part. Historically, the punishment has rarely been strictly economic, and the technical details of specific schemes can quickly become overwhelming. For example, modeling penalties by some form of rule of recapture where part of the individual's possessions are confiscated or his income is garnished appears to be reasonable. But even here basically noneconomic legal and administrative factors must be considered. Can one capture the can of caviar before it is eaten? How is the search for assets carried out, and by whom, at what cost? If assets can be confiscated, at what price should they be evaluated? If future income is to be garnished, how much should be taken? It is easy to become swamped in institutional detail in considering a recapture scheme or the mechanics of restrictions on future borrowing. Our main conclusions do not depend on these important details. What they do depend on is a failure law which can be adjusted to various degrees of unpleasantness for those who fail to honor their commitments.[7]

Default in our model can either be strategic or due to ill fortune. Penalties are imposed on agents who fail to deliver, whatever the cause. Debtors choose whether to repay or to bear the penalty for defaulting; creditors cannot observe why default occurs. Agents who have no resources to repay will be punished as severely as they would if they had the resources but chose not to repay.[8] The consequences of default penalties are therefore two-fold: they tend to induce agents to keep promises when they are able, and they tend to discourage agents from making promises that they know in advance they will not always be able to keep.

Let d^h_{sj} be the nominal market value of default by agent h in state s on asset j. Although in practice the severity of the penalty (e.g., a felony vs. a misdemeanor) depends on the nominal amount, and that is only adjusted slowly in the face of inflation, we suppose the adjustment is instantaneous, so that the penalties depend on the "real" default. Accordingly, we divide d^h_{sj} by the market price in state s of a fixed basket of goods v_s.

We introduce the parameters λ^h_{sj} to represent the utility penalty on agent h for each real dollar of default in state s on asset j. The payoff to agent h in state s can thus be written

$$w^h_s\left(x_s, (d^h_{sj})_{j\in J}, p_s\right) = u^h_s(x_s) - \frac{\sum_{j\in J}\lambda^h_{sj}[d^h_{sj}]^+}{p_s\cdot v_s}$$

where x_s denotes consumption and p_s the prices in state s, and $[y]^+$ denotes the maximum of y and 0. The formula implies that the agent is punished for defaulting but not rewarded for overpaying his promises.

This simple parameterization of extra-economic default penalties was first introduced by Shubik and Wilson (1977). It is meant to capture the idea that as the default gets gradually higher, the penalty gets gradually higher, so utility is continuous and monotonically decreasing in the level of default. Furthermore, by raising the default penalty parameters λ^h_{sj}, we can increase the marginal disutility of default. All other properties implied by our formulation of the default penalties are irrelevant.

In our formulation, the default penalties do not affect the marginal rate of substitution between goods. If we had wished to allow for the possibility that time in jail

[7]We have chosen not to analyze the complications that arise when garnishing or confiscation is introduced, but there are two important lessons to be drawn for that possibility from the simpler case we do examine. First, were we to allow an outside agency (like the courts) to enforce delivery when debtors had resources to make good on their promises, then the argument we shall make later in favor of intermediate rather than harsh default penalties would be still more compelling, since the need to induce agents to keep promises when they are able would already be met. Second, even if the courts could observe perfectly the possessions of all agents, and had the power to transfer ownership without any transactions costs, it would not be optimal in a world with incomplete markets to always force agents to repay all debts, up to the point they were able. Bankruptcy law implicitly recognizes this, putting limits on how much can be confiscated.

[8]In our model default penalties do not distinguish fraud from ill fortune. In reality they are hard to separate, but ever since Las siete Partitas of Don Alfonso X "the wise," bankruptcy law has sought to distinguish them.

affects the relative utility for different kinds of goods, we could have easily dropped the separable form of the utilities we have assumed and inserted both λ_{sj}^h and d_{sj}^h into the utility function u_s^h. Our main results would remain intact. Similarly, the function w_s^h is concave in the level of default, which is convenient in deriving continuous demand functions. However, since we will be assuming a continuum of agents anyway, there is no difficulty in proving the existence of equilibrium with nonconcave utilities.

In specifying the penalties λ_{sj}^h in our model we must decide on the grounds of realism whether the individual default penalties can be set independently, and whether the penalties should be allowed to depend on the state or on the asset. Of course since utilities are not presumed to be interpersonally comparable, it makes no sense to require that $\lambda^h = \lambda^{h'}$. The question is whether policy makers have the freedom to increase penalties for one group of people and not for another. As a good first approximation the answer is no. In practice a distinction may be made in some societies between personal and corporate default. Furthermore, in practice it is difficult to make the penalties state dependent, although we sometimes see contracts with an escape clause for an "act of God," war or other occurrences which can be construed as "force majeure." At most a society will have two or three essentially anonymous, more or less state independent default conditions modified by a disaster clause and qualified by a host of special considerations decided on an ad hoc basis in the courts of law. Thus when we refer to the benefits of harsher or weaker default penalties, we mean the comparative statics thought experiment in which all the λ_{sj}^h rise, or all fall, in proportions policy makers cannot fine tune.

An interesting interpretation can be given to the condition

$$\lambda_{sj}^h > \lambda_{sj'}^h$$

namely, that asset j' is junior debt compared to asset j for agent h in state s. The rational agent h will pay off his j debt entirely in state s before redeeming a single dollar of j' debt. This distinction between junior and senior debt will not concern us until we reconsider the Modigliani–Miller principle.

In view of the foregoing, we shall be especially concerned with two special cases. In the first, $\lambda_{sj}^h = \lambda$ for all h, s, and j. This is done for simple analytical convenience to reflect the idea that policy makers cannot fine tune the default parameters between people, states, or assets. Since we can always rescale the utilities of different agents differently, this case does not require interpersonal utility comparisons, that is it does not imply that a day in jail is regarded with the same dread by every agent. (It does however suggest that the same agent regards a day in jail with the same horror no matter which state it occurs in.) The extreme version of this case occurs when every λ_{sj}^h is set to infinity, which reduces our model to the standard GEI model.

The second case which shall concern us occurs when we set

$$\lambda_{sj}^h = \begin{cases} \infty & \text{if} \quad s \in \bar{S}^h \subset S \\ 0 & \text{if} \quad s \notin \bar{S}^h \end{cases} .$$

Agents are forgiven completely in some states (perhaps when their endowments are zero) and compelled to repay otherwise. This case will allow us to make insurance a special case of our model. We replace a single insurance contract that says agent i will receive money if an accident occurs with two contracts, one which delivers x dollars to agent i in every state, and the other in which agent i promises to deliver x dollars in every state. Since agent i will deliver everything he promised in states with λ_{sj}^h infinite, and default completely without penalty when λ_{sj}^h is 0, on net agent i will receive the same money as he would if he bought insurance paying x dollars in those states $S \setminus \bar{S}^h$ where his λ_{sj}^h is 0.

4 Default in Equilibrium: The $GE(A, \lambda, Q)$ Model

4.1 The Economy

As in the canonical model of general equilibrium with incomplete markets (GEI), we consider a two-period economy, where agents know the present but face an uncertain future. In period 0 (the present) there is just one state of nature (called state 0), in which H agents trade in L commodities and J assets. Then chance moves and selects one of S states which occur in period 1 (the future). Commodity trades take place again, and assets pay off. The difference from GEI is that in our $GE(A, \lambda, Q)$ model, assets pay off in accordance with what agents opt to deliver. Our notation can be formalized as follows:

$\ell \in L = \{1, ..., L\}$ = set of commodities
$s \in S = \{1, ..., S\}$ = set of states in period 1
$S^* = \{0\} \cup S$ = set of all states
$h \in H = \{1, ..., H\}$ = set of agents
$e^h \in \mathbb{R}_+^{S^* \times L}$ = initial endowment of agent h
$j \in J = \{1, ..., J\}$ = set of assets
$A^j \in \mathbb{R}_+^{S \times L}$ = promises per unit of asset j of each commodity $\ell \in L$ in each state $s \in S$
$u^h : \mathbb{R}_+^{S^* \times L} \to \mathbb{R}$ = utility function of agent h
$\lambda_{sj}^h \in \mathbb{R}_+$ = real default penalty on agent h for asset j in state s
$Q_j^h \in \mathbb{R}_+$ = bound on sale of asset j by agent h

We assume that no agent has the null endowment, and that all named commodities are present in the aggregate, i.e.,

$$e_s^h = (e_{s1}^h, ..., e_{sL}^h) \neq 0$$

for all $h \in H$ and $s \in S^*$, and

$$e_{s\ell} = \sum_{h \in H} e_{s\ell}^h > 0$$

for all $s\ell \in S^* \times L$. Also each u^h is continuous, concave and strictly increasing in each of its $S^* \times L$ variables.

We can visualize the state space as a simple tree:

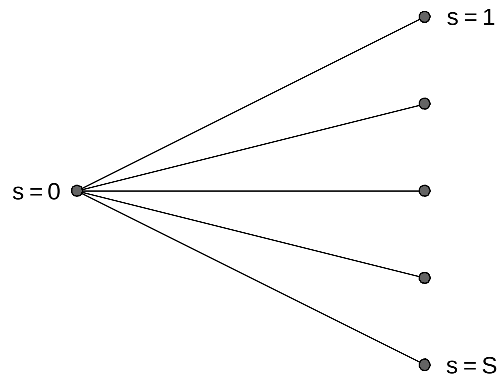

Figure 1

Agents h have state-dependent endowments $e^h_s \in \mathbb{R}^L_+$ and state-dependent and heterogeneous disutilities of default λ^h_{sj}, depending on which assets they default, and in what state, as we discussed earlier.

Adverse selection enters the picture because agents have different endowments out of which to keep their promises, and also different disutilities of default.

Promises must be of a limited kind $j \in J$ fixed a priori. A promise $j \in J$ specifies bundles of goods (or services) to be delivered in each state:

$$\text{Promise } A^j = \left(\begin{array}{c} \\ \\ \\ \\ \end{array} \right) \begin{array}{l} \} - \text{state 1 goods} \\ \} - \text{state 2 goods} \\ \\ \} - \textit{state } S \text{ goods.} \end{array}$$

Agents h make promises by selling various quantities φ^h_j of each asset j. An agent's ability to keep a promise depends on how many promises he sells, both of the same kind j, and of other kinds $j' \neq j$. Moral hazard enters the picture, since a buyer of an asset (i.e., lender) does not know which other promises the seller (i.e., borrower) has made, and because borrowers have the option to default.

Each kind of asset has a limit on the quantity of assets of that type that can be sold, $\varphi^h_j \leq Q^h_j$. Limits on sales of promises are necessary to any realistic model of credit.[9] If $Q^h_j = 0$, then agent h is essentially forbidden from selling asset j. If the limits Q^h_j are very large, they may be entirely irrelevant, as they mostly are in Sections 1-11.[10] But if they are small, then they may be used as a signal that the sellers are not making many promises, and hence that the promises are reliable. We shall turn to the question of signalling Section 12.

An economy is defined as a vector

$$\mathcal{E} = \left((u^h, e^h)_{h \in H}, \left(A^j, ((\lambda^h_{sj})_{s \in S}, Q^h_j)_{h \in H} \right)_{j \in J} \right),$$

[9] Evidence abounds that finite bounds are always imposed in the extension of credit. Even the best "name" among borrowers has a limited credit line.

[10] In Section 5 we are able to prove the existence of equilibrium even when $Q^h_j = \infty$, provided $\lambda \gg 0$ and the A^j all deliver in the same good. The reader who is worried about the meaning of the Q^h_j can ignore them entirely (by assuming they are infinite) until Section 12.

where our notation is described above. Note that an asset consists of a promise, penalties for default, and a limit on sales.

4.2 Equilibrium

As in conventional general equilibrium theory, we think of each agent as very small and unable to affect market prices. As usual, this point of view is embodied in our model by the assumption of price taking behavior. In the next section we make this interpretation more tangible by replacing each household h by a continuum of identical households $t \in (h-1, h]$.

The possibility of default, however, forces us to extend the definition of perfect competition. We do so by continuing to suppose that agents trade through "the market," that is, they are not able to observe the identity of the agents taking the other side of the trade. This is what is meant by the anonymity of markets. In accordance with this anonymity, we suppose that each buyer ends up via his purchase of a quantity θ_j^h of asset j with sales from *every* seller in proportion to how much they sell. Thus if households of type $h = 1$ and $h = 2$ are the only sellers of asset j, and households of type 2 sell twice as many units as households of type 1, then each buyer of asset j receives 2/3 of his purchases from households of type 2 and 1/3 from households of type 1. Similarly, when a borrower (i.e., a seller of an asset promise) defaults on some asset delivery, the loss is spread out proportionally to all owners of that type of asset.

This sharing of losses is implemented in the real world by financial intermediation. For example, many homeowners' mortgage promises are pooled together and shares of the pool are purchased. Even without the financial intermediary who organizes the pools, the same effect could be obtained if lenders lent as shareholders of banks that took on many separate mortgages on a one-by-one basis, or even by lenders who went directly to the homeowners and made many small loans to many homeowners, thereby achieving a perfect random sample.

If we wish to allow for the possibility that some of the characteristics of the sellers are observed, this is easily accommodated in our model by setting some of the sales limits $Q_j^h = 0$. For example, in the extreme case that the type of the sellers of asset j is perfectly observable, we can think of H different copies of asset j, namely j_1, ..., j_H such that $Q_{j_h}^{h'} = 0$ if $h \neq h'$. In that case a buyer of asset j_h knows that he must be buying from households of type h. Even in this case a buyer makes his purchases from a continuum of sellers, so that the buyer never has to worry about the strategic effect of his own loan on an individual seller's actions.

Lenders will naturally try to forecast what fraction of their investments actually deliver. They recognize that their own loans are spread among many borrowers, so that the rates of default will not be affected by how much they loan. Their forecasts of default will of course be conditional forecasts, depending on the state of nature that prevails in the future. In exactly this way the great Wall Street investment banks make forecasts of homeowner prepayment and default rates, conditional on the future level of interest rates and other parts of the "state of nature." In our model we will make the heroic (though standard) "rational expectations" assumption that these conditional forecasts are all correct: the realized rate of default (or delivery) in each state on an asset is exactly the rate anticipated, conditional on that state.

One might suppose that it is a simple matter to describe a $GE(A, \lambda, Q)$ equilibrium with default by respecifying the assets according to what is actually delivered as opposed to what is promised. But what is delivered is determined endogenously and cannot be predicted without solving for the equilibrium. Moreover different agents will make different deliveries on the same asset even though the lenders receive the same aggregated payoffs. Thus our model cannot be fitted into the standard GEI framework.

When an asset market is active, the informational requirements for the $GE(A, \lambda, Q)$ equilibrium in our model are roughly the same as in competitive equilibrium: agents must know the promised delivery of each asset in each state, and the average fraction of delivery of each asset in each state. No trader needs to bother about the identities of those he is trading with.

To define a $GE(A, \lambda, Q)$ equilibrium, first consider the "macrovariables" p, π, K that each agent takes as fixed. Here $p \in \mathbb{R}_{++}^{S^* \times L}$ is the vector of commodity prices; $\pi \in \mathbb{R}_+^J$ is the vector of asset prices; and K is an $S \times J$ matrix with entries K_{sj} between 0 and 1, representing the fraction expected to be delivered of payments promised by asset j in state s. The *budget set* $B^h(p, \pi, K)$ of agent h is given by:

$$
\begin{aligned}
B^h(p,\pi,K) \;=\; \Big\{ & (x,\theta,\varphi,D) \in \mathbb{R}_+^{S^*\times L} \times \mathbb{R}_+^J \times \mathbb{R}_+^J \times \mathbb{R}_+^{J\times S\times L} : \\
& p_0 \cdot (x_0 - e_0^h) + \pi \cdot (\theta - \varphi) \leq 0;\;\; \varphi_j \leq Q_j^h \text{ for } j \in J;\; \text{and, } \forall s \in S, \\
& p_s \cdot (x_s - e_s^h) + \sum_{j \in J} p_s \cdot D_{sj} \leq \sum_{j \in J} \theta_j K_{sj} p_s \cdot A_s^j \Big\}
\end{aligned}
$$

Here $x \in \mathbb{R}_+^{S^* \times L}$ is the final consumption of commodities, $\theta \in \mathbb{R}_+^J$ (respectively, $\varphi \in \mathbb{R}_+^J$) gives the purchases (respectively, sales) of the J assets, and $D_{sj} \in \mathbb{R}_+^L$ is the vector of goods delivered by agent h on asset j in state s.

The budget set allows agent h to deliver whatever he pleases. On the other hand, the agent expects to receive a fraction K_{sj} of the promises made to him on asset j in state s. The first constraint says that agent h cannot spend more on purchases of commodities x_0 and assets θ than the revenue he receives from the sale of commodities e_0^h and assets φ. Moreover he can never sell more than Q_j^h of any asset j. The second constraint applies separately in each state $s \in S$. It says that agent h cannot spend more on the purchase of commodities x_s and asset deliveries $\sum_j D_{sj}$ in state s than the revenue he gets in state s from commodity sales e_s^h and asset receipts $\sum_j \theta_j K_{sj} p_s A_s^j$.

The only reason that agents deliver anything on their promises is that they feel a disutility λ_{sj}^h from defaulting. The payoff of (x, θ, φ, D) given prices p, to agent h is

$$
w^h(x,\theta,\varphi,D,p) = u^h(x) - \sum_{j \in J} \sum_{s \in S} \frac{\lambda_{sj}^h [\varphi_j p_s \cdot A_s^j - p_s \cdot D_{sj}]^+}{p_s \cdot v_s}.
$$

where $v_s \in \mathbb{R}^L_+$ with $v_s \neq 0$. Note that $[\varphi_j p_s \cdot A^j_s - p_s \cdot D_{sj}]^+ \equiv \max\{0, \varphi_j p_s \cdot A^j_s - p_s \cdot D_{sj}\}$ is exactly the money value of the default of h on his promise to deliver on asset j in state s. For simplicity (and the facility of doing comparative statics) we have taken the default penalty to be linear and separable in the amount of default. But more general functions can be allowed for our existence theorems. Indeed, for Theorems 3 and 4, any continuous function w^h would do if $w^h(x, \theta, \varphi, D, p) \leq u^h(x)$ always, and $w^h(x, \theta, \varphi, D, p) = u^h(x)$ whenever there is no default. For Theorems 1 and 2, w^h must be concave; and for Theorem 2 we need to assume, in addition, that given any x, $w^h(x, \theta, \varphi, D, p) < u^h(e^h)$ if the default in any state, on any asset, is sufficiently large.

We are now in a position to define a $GE(A, \lambda, Q)$ equilibrium. It is a list $\langle p, \pi, K, (x^h, \theta^h, \varphi^h, D^h)_{h \in H}\rangle$ such that (1) to (4) below hold.

(1) For $h \in H$, $(x^h, \theta^h, \varphi^h, D^h) \in \arg\max w^h(x, \theta, \varphi, D, p)$ over $B^h(p, \pi, K)$

(2) $\sum_{h \in H}(x^h - e^h) = 0$

(3) $\sum_{h \in H}(\theta^h - \varphi^h) = 0$

(4) $K_{sj} = \begin{cases} \sum_{h \in H} p_s \cdot D^h_{sj} / \sum_{h \in H} p_s \cdot A^j_s \varphi^h_j, & \text{if } \sum_{h \in H} p_s \cdot A^j_s \varphi^h_j > 0 \\ \text{arbitrary}, & \text{if } \sum_{h \in H} p_s \cdot A^j_s \varphi^h_j = 0 \end{cases}$

Condition (1) says that all agents optimize; (2) and (3) require commodity and asset markets to clear. Condition (4), together with the definition of the budget set, says that each potential lender (i.e., buyer) of an asset is correct in his expectation about the fraction of promises that do in fact get delivered. Moreover, his expectation $K^h_{sj} = K_{sj}$ of the rate of delivery does not depend on anything he does himself; in particular, it does not depend on the amount θ^h_j he loans (i.e., purchases) of the asset. Every lender gets the same rate of delivery.

Since heterogeneous borrowers may be selling the same asset, the realized rate of delivery K_{sj} is an average of the rates of delivery of each of the borrowers, weighted by the quantity of their sales. It might well happen that those borrowers with the highest rates of default are selling most of the asset, and this is the adverse selection and moral hazard that rational lenders must forecast.

We believe that our definition of $GE(A, \lambda, Q)$ equilibrium embodies the spirit of perfect, anonymous competition, and represents a significant fraction of the mass asset markets of a modern enterprise economy.

In the next sections we investigate the properties of equilibrium.

4.3 An Equilibrium Refinement

It is a curious fact that many of the large asset markets that our model seeks to describe have been initiated not by entrepreneurs but by government intervention. The government, for example, began the GNMA mortgage program by guaranteeing delivery on the promises of all borrowers eligible for the program (but not the timing[11] of delivery). It is likely, however, that these mortgage markets would function smoothly even without government guarantees. Private companies indeed do sell insurance on non-GNMA mortgages. A reasonable question to ask is why the pass through mortgage market did not begin on its own?

One possible explanation is provided by our model. When assets are traded, expected deliveries K_{sj} must be equal to actual deliveries. Expectations cannot therefore be unduly pessimistic. But for assets that are not traded, our model makes no assumption about expectations of delivery (see (4)). In the real world, investors with no experience in observing default rates might tend to overestimate their probability. This can create serious problems, in practice as in our model. In the model there is nothing to stop the expectations from being absurdly pessimistic, which in turn will support trivial equilibria with no trade in the asset. The point is easily seen by a simple example. Consider an equilibrium of an economy in which certain assets are missing. Introduce these new assets j but choose their prices π_j close to zero. Then no agent will be willing to sell them, for he gets very little in exchange, but undertakes a relatively large obligation either to deliver commodities or to pay default penalties. Also choose the K_{sj} to be positive but even smaller. Then in spite of their low price, no agent will be willing to buy the assets since he expects them to deliver virtually nothing. Thus we have obtained trivial equilibria in which there is no trade of the new assets on account of arbitrarily pessimistic expectations regarding their deliveries. Our last condition (5), given below, rules out such arbitrary pessimism.

We believe that unreasonable pessimism prevents many real world markets from opening, and provides an important role for government intervention. But it is interesting to study equilibrium in which expectations are always reasonably optimistic. It is of central importance for us to understand which markets are open and which are not, and we do not want our answer to depend on the agents' whimsical pessimism. To this end we add a condition (5) to the definition of equilibrium. This requires that if a small change in the macro parameters (p, π) could induce some agents to start selling some of an asset j, where none was being sold before, then buyers should expect at least the rate of delivery they would get had the world indeed been so perturbed. (If there are many ways of perturbing (p, π) to induce sales, then we allow the buyers to focus their attention on one of these perturbations.) If prices π_j are so low that no small perturbation will induce any agents to sell asset j, then buyers are required to expect full delivery, $K_{sj} = 1$. One can (but need not) interpret these expectations as if the government guaranteed delivery on the first infinitesimal promises. We shall see in the next section that condition (5) can always be realized by adding an extra agent to the economy who sells ε of every promise and always keeps his promises, and then letting $\varepsilon \to 0$.

This expectational refinement completely eliminates the kind of trivial (but possibly realistic) equilibria discussed above. Mathematically, the idea is to insist that a candidate equilibrium E satisfying (1)–(4) also has the property: for all small $\varepsilon > 0$,

[11] A default induces the government to prepay the loan immediately, even if the lender would prefer the scheduled payments.

there should exist an "optimistic ε–equilibrium" which is "ε–close" to the candidate equilibrium. Of course all the words within quotes must be made precise.

Let $\| \|_\infty$ denote the supremum norm, and let $E \equiv \langle p, \pi, K, (x^h, \theta^h, \varphi^h, D^h)_{h\in H}\rangle$, i.e., E is the candidate equilibrium which satisfies conditions (1) to (4). For $s \in S$, let $J(s) = \{j \in J : \sum_{h\in H} p_s \cdot A_s^j \varphi_j^h = 0\}$. Thus $J(s)$ is the set of assets in state s for which K_{sj} is not determined by market activity in E. We are ready to state

(5) For any $\varepsilon > 0$, there exists $E(\varepsilon) \equiv \langle p(\varepsilon), \pi(\varepsilon), K(\varepsilon), (x^h(\varepsilon), \theta^h(\varepsilon), \varphi^h(\varepsilon), D^h(\varepsilon))_{h\in H}\rangle$ such that

 (i) $\|E - E(\varepsilon)\|_\infty < \varepsilon$

 (ii) $(x^h(\varepsilon), \theta^h(\varepsilon), \varphi^h(\varepsilon), D^h(\varepsilon)) \in \arg\max w^h(x, \theta, \varphi, D)$ over $B^h(p(\varepsilon), \pi(\varepsilon), K(\varepsilon))$

 (iii) $$K_{sj}(\varepsilon) \geq \begin{cases} \sum_{h\in H} D_{sj}^h(\varepsilon) / \sum_{h\in H} p_s(\varepsilon) \cdot A_s^j \varphi_j^h(\varepsilon) & \text{if } \sum_{h\in H} p_s(\varepsilon) \cdot A_s^j \varphi_j^h(\varepsilon) > 0 \\ 1 & \text{if } \sum_{h\in H} p_s(\varepsilon) \cdot A_s^j \varphi_j^h(\varepsilon) = 0 \end{cases}$$

 for all $s \in S$ and $j \in J(s)$.

Condition (ii) says that agents' choices are maximal on their budget-sets. The main point is (iii), which requires expectations to be optimistic in $E(\varepsilon)$ for all those assets whose promises are going to 0. Combined with (i), which in particular requires $||K - K(\varepsilon)||_\infty < \varepsilon$, these conditions guarantee reasonable optimism for K.

Our condition (5) will enable us to ascertain whether the introduction of a new asset is likely to disrupt a prevailing equilibrium. The definition we give seems simplicity itself, and almost suggests itself. But it is important to notice one of its consequences. In the past, authors such as Rothschild and Stiglitz have implicitly assumed that if a buyer offered a price for a new asset that was high enough to attract the entire population of sellers, then he should expect the distribution of sellers he finds to mirror exactly the population distribution. Our condition (5) implies something quite different, namely that the buyer should recognize that he alone cannot serve all the potential sellers, and that he is likely to be reached first by the sellers who are most anxious to sell, that is by the sellers who have the lowest reservation price for the asset.

Consider the following heuristic example, illustrated in Figure 2 below. Suppose an equilibrium in an economy with assets $j = 1, ..., J$ is given. A new asset $J + 1$, promising 1 in every state, is added to the economy. Suppose at prices $\pi_{J+1} \leq 10$, no agent would sell it, while at prices $10 < \pi_{J+1} \leq 20$ only unreliable agents (with low default penalties $\lambda_{s,J+1}^h$) would sell it, and at prices $\pi_{J+1} > 20$ reliable and unreliable agents would sell it. Without the equilibrium refinement, we could always include the new asset in the old equilibrium by assigning it a price $\pi_{J+1} = 0$ with no trade, and with $K_{sJ+1} = 0\ \forall s \in S$. But our equilibrium refinement requires that if $\pi_{J+1} < 10$, then $K_{sJ+1} = 1$ for all $s \in S$, since no perturbation would induce sales. The equilibrium refinement thus rules out equilibria with $\pi_{J+1} < 10$ unless demand is zero even with expectations of full delivery. For concreteness, let us suppose demand is zero unless expected delivery per dollar invested is at least 0.034, after which demand becomes positive. Clearly there is no equilibrium with $0 \leq \pi_{J+1} < 10$, since expected delivery per dollar invested $\frac{1}{S}\sum_s K_{sj} 1/\pi_{J+1} = (1)(1/\pi_{J+1}) \geq (1)(1/10) = 0.100 > 0.034$.

If there is any equilibrium in this example in which asset $J+1$ is not traded, then there must be such an equilibrium at which $\pi_{J+1} = 10$. Since there are no sales, when $\pi_{J+1} = 10$, the refinement allows for $K_{sJ+1} < 1$, provided that there would be sales at $\pi_{J+1} = 10 + \varepsilon$ and that the delivery rate on those sales is approximately K_{sJ+1}. By hypothesis, at $\pi_{J+1} = 10 + \varepsilon$, only unreliable agents would be selling, hence we might have, say, $K_{sJ+1} = 1/3 \ \forall s \in S$. At these low levels of delivery, and at a price of 10, there would indeed be no buyers (as well as no sellers), since expected delivery per dollar invested in $(1/3)(1/10) = 0.033 < 0.034$. Since the K_{sJ+1} are obtained from the perturbation, we would regard the expectations as reasonable and call this a genuine equilibrium.

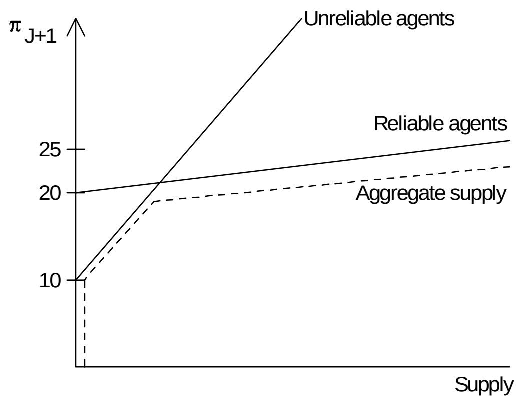

Figure 2

Consider also the situation where the unreliable and reliable supply curves are reversed, so that at $\pi_{J+1} = 10$ it is the reliable agents who begin to sell. Then according to our definition of equilibrium, in order for $\pi_{J+1} = 10$ to be an equilibrium, it must be that no demand would be forthcoming even with K_{sJ+1} set at the reliable rates of delivery (say $K_{sJ+1} = 9/10$, for concreteness). But $(9/10)(1/10) = 0.090 > 0.034$, so in this version of the example, there could be no equilibrium in which asset $J+1$ remains untraded. (Our existence theorem, proved in the next section then assures us that there must be some equilibrium in which asset $J+1$ is traded.)

The Rothschild and Stiglitz logic would say that there could be no equilibrium in which asset $J+1$ is untraded, even in the case, illustrated by the diagram, in which the unreliable agents begin selling at $\pi_j = 10$. Suppose that if the price were raised to $\pi_{J+1} = 25$, reliable agents would want to sell in such large quantities that, leaving aside the question of market clearing, the fraction of deliveries out of all desired sales on asset $J+1$ would be $K'_{sJ+1} = 6/7$. (If at price $\pi_{J+1} = 25$, reliable agents sell 110 units to every 9 units unreliable agents sell, then $K'_{sJ+1} = 6/7 = (110/119)(9/10) + (9/119)(1/3)$.)

If buyers took 6/7 as the rate of delivery, then at the price $\pi_{J+1} = 25$ they would be willing to buy, even though they had refused to buy at $\pi_{J+1} = 10$, because their returns per dollar invested would be better, $6/7 \cdot 1/25 > (0.34)(1/10) > 1/3 \cdot 1/10$. If this were the case, then according to the logic of Rothschild–Stiglitz (1972), equilibrium would not exist at $\pi_{J+1} = 10$, since buyers would have an incentive to raise the price to $\pi_{J+1} = 25$.

Our definition of equilibrium allows for $\pi_{J+1} = 10$, and $K_{sJ+1} = 1/3$, and we believe it does so for good reasons. We have in mind a competitive world with many small buyers. If a single buyer raised his offering price to 25, fully 15 points above the market price, he would be deluged with sellers. The people with the most to gain from selling to him would be those who already were willing to sell at 10, namely the unreliable agents. Why should he assume he would be equally likely to meet each seller? We feel justified in assigning him the cautious expectations of $K_{sJ+1} = 1/3$ no matter what price he offered, given that the market price is $\pi_{J+1} = 10$.

We can put the same point a little differently. We suppose that buyers are aware of the composition of sales at the market prices, and perhaps of the composition of sales at prices a penny off from market prices. But agents lack the knowledge or computing power to infer what the composition of demand would be at prices far from market prices.[12]

4.4 A Continuum of Traders

We have mentioned several times that our model is meant to embody the ideal of perfect competition, in which each agent is so small that by himself he cannot influence anyone else. We can make such an interpretation of our model more concrete by replacing each agent h by a continuum of identical agents parameterized by t lying in the interval $(h-1, h]$: each agent $t \in (h-1, h]$ has identical endowments e^h and utility u^h. The $GE(A, \lambda, Q)$ of the finite agent economy, whose existence we shall prove in Theorems 1 and 2, corresponds to a $GE(A, \lambda, Q)$ of the continuum model with the added feature that $(x^t, \theta^t, \varphi^t, D^t) = (x^{t'}, \theta^{t'}, \varphi^{t'}, D^{t'})$ whenever t and t' are both in $(h-1, h]$, i.e., all agents of the same type behave symmetrically. We shall call such equilibria *type-symmetric* when viewed in the continuum setting.[13]

[12] Putting the matter still differently, we regard an asset or contract as setting out the obligations of the seller, including the penalties if he fails to deliver, and the quantity limitations on his other sales. The price of the contract is set by competition between sellers and buyers, that is, by the market. Agents need only think about one prevailing price for each contract. In the Rothschild–Stiglitz view, the price is one of the terms of the contract. In this view, there is no such thing as a single contract; there are as many contracts as there are prices. Notice also that the Rothschild–Stiglitz view must regard market clearing as one of rationing. At most prices, the contract will not be traded, because *either* supply or demand is zero, and the other side of the market is rationed. This point of view has been admirably expressed by Gale. In our view competitive equilibrium should be defined by a single price at which both supply and demand are equal (possibly both zero), as long as expectations at that price are set at rational levels.

[13] We could pursue the ideal of perfect competition still further by defining a strategic market game in which prices were formed directly from the strategies of the agents. But in the present context, this would not add to the analysis. The continuum of agents is necessary when we consider nonconvex budget sets in Sections 6 and onwards.

More generally, if we wished, we could consider continuum models where the agent types were more than finite in number. Even if the characteristics divided the agents into a finite number H of types, the actions of agents of the same type might be different. Consider the interval $I \equiv [0,T]$ on the real line endowed with the Lebesgue measure μ. For each agent $t \in I$, let $e^t \in \mathbb{R}_+^{S^* \times L}$ be the endowment and $u^t : \mathbb{R}_+^{S^* \times L} \to \mathbb{R}$ the utility function of t. Assume that the map $e : I \to \mathbb{R}_+^{S^* \times L}$ (given by $e(t) \equiv e^t$) is integrable, and that the map $u : I \times \mathbb{R}^{S^* \times L} \to \mathbb{R}$ (given by $u(t,x) \equiv u^t(x)$) is measurable. Further let $\lambda : I \to \mathbb{R}_+^{SJ}$ (where $\lambda(t) \equiv (\lambda_{sj}^t)_{sj \in S \times J}$) and $Q : I \to \mathbb{R}_+^J$ (where $Q(t) \equiv (Q_j^t)_{j \in J}$) be measurable, and Q be furthermore bounded. Then for any $(p, \pi, K) \in \mathbb{R}_{++}^{S^* \times L} \times \mathbb{R}_+^J \times [0,1]^{S \times J}$ we can define $B^t(p, \pi, K)$ exactly as before, replacing h by t throughout. Also $GE(A, \lambda, Q)$ can be defined as before, replacing $\sum_{h \in H}$ by $\int_I d\mu$, "$\forall h \in H$" by "almost all $t \in I$," and the notion of convergence of x, θ, φ, D (which are now integrable functions on I) in condition (5) by almost everywhere pointwise convergence.

Although the definition of $GE(A, \lambda, Q)$ equilibrium for the continuum model does not require type symmetric behavior, when agents have convex budget sets and strictly concave payoffs, the equilibria are perforce type-symmetric.

But we shall shortly consider variants of our model in which the convexity of budget sets and/or concavity of payoffs fail to hold. Here the continuum model is necessary for establishing existence of $GE(A, \lambda, Q)$. Even if the economy is finite-type (a situation to which we adhere, for simplicity) its equilibria are no longer type-symmetric, and the consideration of a continuum becomes unavoidable.

5 The Orderly Function of Markets with Default

Our first goal in this paper is to establish that default is completely consistent with the orderly function of markets. To that end we prove that under fairly general conditions, equilibrium always exists in our model.

The universal existence of equilibrium is somewhat surprising because of the historical tendency to associate default with disequilibrium (or more accurately, to make full delivery part of the definition of equilibrium), as we have already remarked. Furthermore, endogeneity of the asset payoff structure is known to complicate the existence of equilibrium with incomplete markets. But we show that no new existence problems arise from the endogeneity of the asset payoffs due to default.

The universal existence of equilibrium with default is also surprising because the pioneering papers placing adverse selection in a model of competition, by Akerlof (1972) on the market for lemons, and Rothschild and Stiglitz (1976) on insurance markets, purportedly showed that adverse selection is quite commonly inconsistent with equilibrium. Since the Akerlof and Rothschild–Stiglitz models are special cases of our model, a word about them might be illuminating. We discuss Rothschild–Stiglitz in more detail later.

In Akerlof's lemons paper, each seller knows the value of his car, but the buyer only knows the average quality of the cars for sale. This is analogous to our model in which each seller knows his disutility of defaulting (and indeed his intentions to default) but the buyer knows only the overall average default rate. Akerlof's formal analysis consists of a special example with the property that at any price for used automobiles, the sellers will supply automobiles whose average quality is not worth the price. More precisely, suppose that the quality of cars is uniformly distributed between 0 and 1, and that if v is the quality of some car, then the owner (i.e., the potential seller) knows v and values it at v, whereas any potential buyer would value it at $1.1v$ once he got it, but unfortunately does not know what v is. Suppose that there is a large pool of potential buyers with the same preferences. At any price p, the cars with v less than p will be put up for sale, and the average quality to the buyer will be $(0.5)(1.1)p = .55p < p$. As the price falls, so does the average quality of the automobiles put up for sale. Under this extreme hypotheses, there cannot be any trade of automobiles at any price; even though at each price there are cars for sale that are worth more to the buyers than the price, a buyer must count on getting an average quality car, which is worth less than the price.

Needless to say, one could easily imagine less severe conditions under which there would still be some trade for automobiles, though to be sure the quantity would be less than would obtain under complete information about the quality of every car. For example, if the minimum quality level were $m > 0$ instead of 0, then there would be an equilibrium with $p = (11/9)m$. All the cars with $m < v < p$ would be sold, some buyers would be disappointed, others would be pleasantly surprised, but on average the buyers would get value $(1/2)(m + (11/9)m)(1.1) = (11/9)m = p$ equal to what they paid for. Indeed it is accurate to describe the extreme situation in the Akerlof model as one in which there is an equilibrium, with price set at 0. If the equilibrium with no trade seems bad, it is just that: bad in the welfare sense. But it still is an equilibrium. Interpreted properly, Akerlof's paper shows that adverse selection may in extreme situations result in an equilibrium with no trade; it does not provide any reason to suppose that equilibrium and adverse selection are incompatible.

Insurance contracts promise payments conditional on the state of nature, and so can be viewed as assets such as we describe in this paper, as we mentioned earlier. In particular, the Rothschild–Stiglitz model can be expressed as a special case of our general equilibrium model, as we show in Section 12. The reason Rothschild and Stiglitz found robust regions with no equilibrium is that they defined equilibrium expectations differently, as we have explained. If buyers had the perfectly competitive expectations that we invoke, namely that each thinks he cannot improve his selection of sellers by unilaterally offering a higher price, then the Rothschild–Stiglitz model would always have an equilibrium, as we show in Section 12, even using their "exclusionary" hypothesis.

We are now ready to state our main theorem, which is that $GE(A, \lambda, Q)$ equilibrium always exists, even if we insist on the equilibrium refinement discussed in Section 4.3.

Theorem 1 *For any $\lambda \in \mathbb{R}_+^{HSJ}$ and $Q \in \mathbb{R}_+^{RJ}$, a $GE(A, \lambda, Q)$ equilibrium satisfying (1)–(5) exists.*

Proof Define, for $\varepsilon > 0$,

$$\Delta_\varepsilon = \left\{ (p,\pi) \in \mathbb{R}_+^{S^* \times L} \times \mathbb{R}_+^J : \sum_{\ell=1}^{L} p_{s\ell} = 1 \ \forall s \in S^*, \right.$$
$$\left. \varepsilon \leq p_{s\ell} \quad \forall s\ell \in S^* \times L, \text{ and } 0 \leq \pi_j \leq \tfrac{1}{\varepsilon} \ \forall j \in J \right\};$$

define, for each $h \in H$, $\square^h = \{(x,\theta,\varphi,D) \in \mathbb{R}_+^{S^* \times L} \times \mathbb{R}_+^J \times \mathbb{R}_+^J \times \mathbb{R}_+^{SLJ} : \|x\|_\infty \leq 2\|e\|_\infty,\ \theta_j \leq 2\sum_{h' \in H} Q_j^{h'},\ \varphi_j^h \leq Q_j^h, \text{ and } ||D||_\infty \leq ||Q||_\infty ||A||_\infty\}$.

For each $h \in H$, define the correspondence $\psi_\varepsilon^h : \Delta_\varepsilon \times [0,1]^{S \times J} \Rightarrow \square^h$ by

$$\psi_\varepsilon^h(p,\pi,K) = \arg\max_{x,\theta,\varphi,D} \left\{ w^h(x,\theta,\varphi,D,p) : (x,\theta,\varphi,D) \in B^h(p,\pi,K) \cap \square^h \right\}.$$

Notice that ψ_ε^h is non-empty valued and convex-valued, thanks to the continuity and concavity of w^h, for all $h \in H$. To check that $B^h(p,\pi,K) \cap \square^h$ is LSC, let p^n, π^n, $K^n \xrightarrow{n} \bar{p}$, $\bar{\pi}$, $\bar{K}$ with $\bar{p} \gg 0$. Let $(\bar{x},\bar{\theta},\bar{\varphi},\bar{D}) \in B^h(\bar{p},\bar{\pi},\bar{K})$. Fix $0 < \alpha < 1$. Then $(\alpha\overline{x},\alpha\overline{\theta},\alpha\overline{\varphi},\alpha\bar{D}) \in B^h(p^n,\pi^n,K^n) \cap \square^h$ for sufficiently large n because $\overline{p}_s \cdot e_s^h > 0$ $\forall s \in S^*$. Since α was arbitrary, this shows that $B^h(p,\pi,K) \cap \square^h$ is LSC in (p,π,K) whenever $p \gg 0$. Hence ψ_ε^h is USC by the maximum principle.

Consider the correspondence $\psi_\varepsilon^0 : \square^H \equiv \bigtimes_{h \in H} \square^h \Rightarrow \Delta_\varepsilon$ defined by

$$\begin{aligned} &\psi_\varepsilon^0((x^h,\theta^h,\varphi^h,D^h)_{h\in H}) \\ &= \arg\max_{(p,\pi)\in\Delta_\varepsilon} \left\{ p_0 \cdot \sum_{h\in H}(x_0^h - e_0^h) + \pi \cdot \sum_{h \in H}(\theta^h - \varphi^h) + \sum_{s \in S}\sum_{h\in H} p_s \cdot (x_s^h - e_s^h) \right\}. \end{aligned}$$

Clearly ψ_ε^0 is non-empty-valued, convex-valued, and USC. Consider the map $\overline{\psi}_\varepsilon : \Delta_\varepsilon \times \square^H \to [0,1]^{S \times J}$ defined by

$$\overline{\psi}_{\varepsilon sj}((p,\pi,K),\ (x^h,\theta^h,\varphi^h,D^h)_{h\in H}) = \frac{\varepsilon + \sum_{h\in H} \min\{p_s \cdot D_{sj}^h, \varphi_j^h p_s \cdot A_{sj}\}}{\varepsilon + \sum_{h \in H} \varphi_j^h p_s \cdot A_{sj}}$$

for each $s \in S$, $j \in J$. Clearly $\overline{\psi}_{\varepsilon sj}$ is a continuous function. Let $\psi_\varepsilon : \Delta_\varepsilon \times [0,1]^{S\times J} \times \square^H \Rightarrow \Delta_\varepsilon \times [0,1]^{S \times J} \times \square^H$ be the correspondence defined by

$$\begin{aligned} \psi_\varepsilon((p,\pi,K),((x^h,\theta^h,\varphi^h,\ D^h)_{h\in H})) \ = \ &\psi_\varepsilon^0((x^h,\theta^h,\varphi^h,D^h)_{h\in H}) \times \overline{\psi}_\varepsilon\Big((p,\pi,K), \\ &(x^h,\theta^h,\varphi^h,D^h)_{h\in H}\Big) \times \bigtimes_{h\in H} \psi_\varepsilon^h(p,\pi,K). \end{aligned}$$

By Kakutani's theorem ψ_ε has a fixed point $(p^\varepsilon$, π^ε, K^ε, $(x^h(\varepsilon)$, $\theta^h(\varepsilon)$, $\varphi^h(\varepsilon)$, $D^h(\varepsilon))_{h\in H})$. To avoid notational clutter, we temporarily suppress the ε. We shall now show that this fixed point constitutes an "ε-*GE*(A,λ,Q)," i.e., all markets clear up to a factor of ε, for small enough ε.

By strict monotonicity of the utilities, for small enough $\varepsilon > 0$, if $p_{0\ell} > 1/L$ for some h and ℓ with $e^h_{0\ell} > 0$, then $x^h_{0i} > \sum_{h' \in H} e^{h'}_{0i}$ for all i with $p_{0i} = \varepsilon$. Hence if any $p_{0\ell} = \varepsilon$, $\sum_{h\in H}(x^h_{0\ell} - e^h_{0\ell}) > 0$.

Note that in state 0, $p_0 \cdot (\sum_h (x^h_0 - e^h_0)) + \pi \cdot (\sum_h (\theta^h - \varphi^h)) = 0$ (since this equality holds for each h individually in his budget-set). Then $\sum_h (x^h_0 - e^h_0) \le 0$; if some $\sum_{h\in H}(x^h_{0\ell} - e^h_{0\ell}) > 0$, then either for some $i \in L$ with $p_{0i} > \varepsilon$, $\sum_{h\in H}(x^h_{0i} - e^h_{0i}) < 0$, or else for some $j \in J$ with $\pi_j > 0$, $\sum_{h\in H}(\theta^h_j - \varphi^h_j) < 0$. By raising $p_{0\ell}$ a little and lowering p_{0i} by the same amount, or else simply by lowering π_j a little, we contradict the optimization in ψ^0_ε. Hence $\sum_{h\in H} x^h_0 \le \sum_{h\in H} e^h_0$.

If $0 < \pi_j < 1/\varepsilon$, then we must have that $\sum_{h\in H}(\theta^h_j - \varphi^h_j) = 0$, for otherwise, by raising or lowering π_j, we contradict the optimization in ψ^0_ε. If $\pi_j = 0$, then no agent h can be selling the asset unless $\lambda^h_{sj} = 0$, $\forall s$, in which case he is always defaulting, and so WLOG we can take $\theta^h_j = \varphi^h_j = 0$, $\forall h \in H$. But if $\pi_j = 1/\varepsilon$, then $0 \le \sum_{h\in H}(\theta^h_j - \varphi^h_j) \le \varepsilon \|e_0\|_\infty$. The first inequality must hold, for otherwise lowering π_j would contradict the optimality of ψ^0_ε. The second inequality then follows from the period 0 budget equality.

Observe that from the fact that ψ_ε fixes K, we can conclude that $\sum_h p_s \cdot D^h_{sj} \ge K_{sj} \sum_h p_s \cdot A_{sj} \varphi^h_j - \varepsilon$, for all sj . From the budget set for each h we have $p_s \cdot (x^h_s - e^h_s) \le \sum_j K_{sj} p_s \cdot A_{sj} \theta^h_j - \sum_j p_s \cdot D^h_{sj}$. Adding over H and combining inequalities yields

$$\begin{aligned}\sum_h p_s \cdot (x^h_s - e^h_s) &\le \sum_j K_{sj} p_s \cdot A_{sj} \sum_h (\theta^h_j - \varphi^h_j) + J\varepsilon \\ &\le \|A\|_\infty \|e_0\|_\infty J\varepsilon + J\varepsilon.\end{aligned}$$

It follows by the same argument used for state 0 that for each state $s \in S$,

$$\sum_h (x^h_s - e^h_s) \le ((\|A\|_\infty \|e_0\|_\infty J + J)\varepsilon.$$

Finally, π_j must remain bounded as $\varepsilon \to 0$, otherwise by increasing φ^h_j by Δ an agent can consume $\sim \pi_j \Delta$ units of some good in state 0, losing at most $\sim \Delta L \|A\|_\infty \sum_{s\in S} \lambda^h_{sj}$ utility, contradicting that his choice of φ^h_j is optimal.

Since all choices and all macrovariables are uniformly bounded for small ε, we can pass to convergent subsequences, obtaining $\bar{E} \equiv \langle \bar{p}, \bar{\pi}, \bar{K}, (\bar{x}^h, \bar{\theta}^h, \bar{\varphi}^h, \bar{D}^h)_{h\in H} \rangle$ as a limit point. When $\varepsilon = 0$, all the $GE(A, \lambda, Q)$ equilibrium conditions hold as inequalities. Since agents spend all their money, they must all be equalities. Moreover, since $w^h(x, \theta, \varphi, D, p)$ is concave in (x, θ, φ, D) and the constraints $2\|e\|_\infty$ and $\sum_{h'\in H} Q^{h'}_j$ are not binding on agents optimal choices, it follows that they have in fact optimized on their untruncated budget sets for small enough ε. Thus $\bar{E}$ satisfies conditions (1) to (4) of $GE(A, \lambda, Q)$. Finally the fixed point of ψ_ε yields $E(\varepsilon)$ which is easily seen to satisfy condition (5). ■

Our proof has used the fact that $\varphi^h_j \le Q^h_j$ by assumption. Later the Q^h_j will play an important role as signals, but now the reader may wonder what would happen if they were eliminated, or taken to be enormously large. Recall that there is a pathology that occasionally occurs even when there is no default, for example in the GEI model. Sometimes two assets j and j' that promise different commodities nevertheless become nearly equivalent at some spot prices $(p_s)_{s\in S}$ because they then promise nearly the same money. At these prices the number of independent assets suddenly drops, and demand blows up as agents try to go infinitely long in asset j' and infinitely short in asset j (or vice versa). This destroys the existence of equilibrium. The bounds Q_j^h prevent this, as Radner (1972) long ago pointed out for the GEI model.

In the GEI model without short sale constraints like the Q_j^h, equilibrium can only be guaranteed if all the assets promise payoffs exclusively in the same good (say L) in each state $s \in S$. (See Geanakoplos–Polemarchakis (1986).) The asset matrix A then is effectively reduced to $S \times J$ dimensions.

Default provides another reason why two assets that make different promises might, given certain macro variables (p, π, K), actually deliver the same money in every state. One should therefore wonder if default introduces additional difficulties in proving the existence of equilibrium. We have just seen that in the presence of the bounds Q_j^h it does not. Now we shall show that default also does not complicate the existence picture without the bounds Q_j^h.

Theorem 2 *Let all the assets A^j promise delivery exclusively in good L for all $s \in S$. Define $GE(A,\lambda) = GE(A,\lambda,Q)$ with $Q_j^h = \infty$, $\forall h \in H$, $j \in J$. Then $GE(A,\lambda)$ exists for any vector $\lambda \gg 0$.*

Proof Theorem 2 specializes the conditions of Theorem 1. Hence we have a $GE(A,\lambda,Q)$ equilibrium for all finite Q. Consider a sequence of equilibria, $\eta(Q) = (p(Q), \pi(Q), K(Q), (x^h(Q), \theta^h(Q), \varphi^h(Q), D^h(Q))_{h\in H})$, where $Q_j^h = Q \in \mathbb{N}$, for all $h \in H$, $j \in J$.

Define
$$v(Q) \equiv \sum_{j=1}^{J} \sum_{h=1}^{H} \theta_j^h(Q) = \sum_{j=1}^{J} \sum_{h=1}^{H} \varphi_j^h(Q).$$

If there is a single Q with $v(Q) < Q$, then by the concavity of each u^h, $\eta(Q)$ is a $GE(A,\lambda)$. Since for any fixed economy, the set of equilibria, $E(Q)$, is a closed set, we may assume that $\eta(Q)$ minimizes v over all equilibria in $E(Q)$. We shall suppose that $v(Q) \geq Q$ for all Q, and derive a contradiction.

Passing to a convergent subsequence if necessary, we may suppose that for all $h \in H$ and $j \in J$,
$$\frac{\theta_j^h(Q)}{v(Q)} \to \bar{\theta}_j^h, \ \frac{\varphi_j^h(Q)}{v(Q)} \to \bar{\varphi}_j^h.$$

Moreover, for at least one j and some h, h', $\bar{\theta}_j^h \neq 0$ and $\bar{\varphi}_j^{h'} \neq 0$.

For notational convenience, we shall write A_{sj} and D_{sj}, instead of the more accurate A_{sLj} and D_{sLj}, and we shall suppose that real default in each state $s \in S$

is measured in terms of the commodity bundle $v_s = 1_L$, which is one in the Lth coordinate, and zero elsewhere. Since all assets are exclusively delivering in the Lth good, no harm results from these simplifications. Finally, w.l.o.g. take $p_{sL} = 1$ for all $s \in S$.

Observe that for any $h \in H$, $s \in S$, $j \in J$, the level of default

$$\delta^h_{sj}(Q) \equiv [A_{sj}\varphi^h_j(Q) - D^h_{sj}(Q)]^+ \leq \frac{1}{\lambda^h_{sj}}[u^h(e) - u^h(e^h)],$$

for otherwise agent h would have done better not trading at all. Hence if $\varphi^h_j(Q) \to \infty$ and $A_{sj} > 0$,

$$\frac{[A_{sj}\varphi^h_j(Q) - D^h_{sj}(Q)]}{\varphi^h_j(Q)} = \frac{[A_{sj}\varphi^h_j(Q) - D^h_{sj}(Q)]^+}{\varphi^h_j(Q)} = \frac{\delta^h_{sj}(Q)}{\varphi^h_j(Q)} \to 0.$$

In particular, $K_{sj}(Q) \to 1$ for all $s \in S$ with $A_{sj} > 0$, provided that $\sum_{h\in H} \varphi^h_{sj}(Q) = \sum_{h\in H} \theta^h_{sj}(Q) \to \infty$.

Furthermore, since relative prices $p_{s\ell}(Q)/p_{sk}(Q)$ stay bounded,

$$\sum_{j\in J} K_{sj}(Q)A_{sj}\theta^h_j(Q) - \sum_{j\in J} D^h_{sj}(Q)$$

must stay bounded. Otherwise agent h would eventually be consuming a negative quantity in state s, or a quantity exceeding the aggregate endowment e_s, contradicting commodity market clearing.

Putting these last statements together, and recalling $v(Q) \to \infty$, we must have that

$$\lim_{Q\to\infty} \frac{\sum_{j\in J} K_{sj}(Q)A_{sj}\theta^h_j(Q) - \sum_{j\in J} D^h_{sj}(Q)}{v(Q)} = A_s(\bar{\theta}^h - \bar{\varphi}^h) = 0,$$

for all $h \in H$, $s \in S$.

Consider any h with $\bar{\varphi}^h \neq 0$, and hence $\bar{\theta}^h \neq 0$. At any large Q, the agent could feasibly have chosen

$$\begin{aligned}\hat{\theta}^h &= \theta^h(Q) - \bar{\theta}^h \\ \hat{\varphi}^h &= \varphi^h(Q) - \bar{\varphi}^h \\ \hat{D}^h_{sj} &= D_{sj}(Q) - A_{sj}\bar{\varphi}^h_j \text{ for all } j \in J.\end{aligned}$$

With these choices he would pay exactly the same penalty as in the equilibrium $n(Q)$. He would receive exactly the same consumption at time 1 if $K_{sj}(Q) = 1$ for all j with $\bar{\theta}^h_j > 0$, and strictly more consumption otherwise. In order for him not to prefer this deviation, we must therefore have

$$\pi(Q)[\bar{\theta}^h - \bar{\varphi}^h] \leq 0 \text{ for all } h \in H.$$

But since $\bar{\theta}^h$ and $\bar{\varphi}^h$ are limits of $GE(A, \lambda, Q)$ equilibrium portfolios,

$$\sum_{h \in H} \bar{\theta}^h = \sum_{h \in H} \bar{\varphi}^h, \text{ hence we must have}$$

$$\pi(Q)[\bar{\theta}^h - \bar{\varphi}^h] = 0 \text{ for all } h \in H.$$

It now follows that household h would still prefer this deviation unless

$$K_{sj}(Q) = 1 \ \forall s \in S \text{ with } A_{sj} > 0, \text{ and } \bar{\theta}^h_j > 0 \text{ for any } h \in H \text{ and } j \in J.$$

Note finally that if $\bar{\varphi}^h_j > 0$, there must be some agent i with $\bar{\theta}^i_j > 0$, hence $K_{sj}(Q) = 1$ for all $s \in S$ with $A_{sj} > 0$ and either $\bar{\theta}^h_j > 0$ or $\bar{\varphi}^h_j > 0$.

Replacing $(p(Q), \pi(Q), K(Q), (x^h(Q), \theta^h(Q), \varphi^h(Q), D^h(Q))_{h \in H})$ with $(p(Q), \pi(Q), K(Q), (x^h(Q), \hat{\theta}^h, \hat{\varphi}^h, \hat{D}^h)_{h \in H})$ we get another $GE(A, \lambda, Q)$ with a lower v. (Notice that we are reducing sales and purchases only for assets with $K_{sj} = 1$, which therefore leaves the K unchanged.) This is the contradiction we have been seeking and this concludes the proof. ■

6 Chain Reactions, Netting, and Supernetting

6.1 Chain Reactions

In modern financial economies, agents often are long and short in many different assets. They rely on revenues from their loans to keep their own promises. But these revenues are only as reliable as the loans other agents have made to yet different parties, thus opening the possibility of a chain reaction of defaults. If α defaults against β, forcing β to default against γ, forcing γ to default against δ, then in our definition of equilibrium, α, β, and γ will pay default penalties, and the total utility loss from defaults will be large. Curiously this phenomenon is at its most dangerous when the financial system is at an intermediate level of development, with smoothly functioning markets that permit agents to go short, but with missing asset markets which force agents to hold complicated portfolios of assets to achieve the risk spreading they desire.

Consider a world with four agents and three possible future events, each consisting of many different states of the world. Suppose β wants to consume in the first event, γ in the second event, and δ in the third event. Suppose agents β, γ, and δ have no endowment in the future states. Suppose α wants to consume in the present, but has a considerable endowment of goods in the future, except in one unlikely state ω in the third event.

If there were an advanced financial system of Arrow securities, agent α would in effect sell directly to each of the other three agents. For example, with just three Arrow securities, each one paying off exclusively in a different one of the three events, agent α would sell the first security to β, the second to γ, and the third to δ. Agent α by himself would default in state ω, and he alone would pay a default penalty.

Suppose, however, that in a less advanced financial system there are again three securities available. A^{123} promises 1 dollar in every state, A^{23} promises 1 dollar in (every state in) events 2 and 3, and A^3 promises 1 dollar in (every state in) event 3. Then in equilibrium we could expect α to sell A^{123}, β to buy A^{123} and to sell A^{23}, γ to buy A^{23} and to sell A^3, and δ to buy A^3. In the bad state ω in event three, the chain of defaults indicated above will take place. The penalty that α pays for starting the chain reaction may be very small compared to the total penalty incurred by the rest of the defaulters.

A diagram may make the situation clearer.

Figure 3

Notice that the asset span is exactly the same as with the three Arrow securities. What makes the chain of defaults possible is the interlocking asset trade, with some investors holding assets that other investors short, in a long chain. With Arrow securities this chain would never reach more than two links and one default.

One way around these chain reactions is to encourage market intermediation that nets payouts.

6.2 Netting

Consider the variation of our $GE(A, \lambda, Q)$ model in which an agent's purchases and sales of any given asset j are netted, so that he is deemed to have purchased $(\theta_j - \varphi_j)^+$ and sold $(\varphi_j - \theta_j)^+$. In this case, the budget-set and payoff function of an agent $t \in (h-1, h]$ of type h are modified as follows:

$$B^h(p, \pi, K) = \left\{ (x, \theta, \varphi, D) : p_0 \cdot (x_0 - e_h^h) + \pi \cdot (\theta - \varphi) \leq 0;\ \varphi_j \leq Q_j^h \text{ for } j \in J; \right.$$

$$\left. p_s \cdot (x_s - e_s^h) + \sum_{j \in J} p_s \cdot D_{sj} \leq \sum_{j \in J} K_{sj} p_s \cdot A_s^j (\theta_j - \varphi_j)^+ \text{ for all } s \in S \right\}$$

$$w^h(x,\theta,\varphi,D,p) = u^h(x) - \sum_{s\in S}\sum_{j\in J} \frac{\lambda^h_{sj}\left[p_s \cdot A^j_s(\varphi_j - \theta_j)^+ - p_s \cdot D_{sj}\right]^+}{p_s \cdot v_s}$$

Moreover, if $(x^t,\, \theta^t,\, \varphi^t,\, D^t)_{t\in(0,H]}$ are the equilibrium choices of the agents, then condition (4) on $GE(A,\lambda,Q)$ becomes:

$$K_{sj} = \int_0^H p_s \cdot D^t_{sj} d\mu(t) / \int_0^H p_s \cdot A^j_s(\varphi^t_j - \theta^t_j)^+ d\mu(t)$$

whenever the denominator is positive.

Notice that the budget set is no longer convex, hence an equilibrium may not exist in the finite agents model. However we have

Theorem 3 *A $GE(A,\lambda,Q)$ exists in the finite-type continuum model with netting.*

This equilibrium will, in general, no longer be type symmetric and will entail different choices by agents of the same type. We defer the proof until the next section.

6.3 Supernetting

Here we go a step further and consider netting across different assets that an agent has traded in.[14] Now deliveries are no longer made separately on each asset, but there is one combined payment in every state $s \in S$. This extends the idea in Theorem 1 that agents may deliver differently on the same asset because of default or some other option. Here they may even make different promises. Supernetting combines this extension with netting both sides of a trade. Thus the delivery choice of an agent is a vector $D \in \mathbb{R}^{SL}_+$, and the constraint in t's budget set must be rewritten (where t is of type h)

$$p_s \cdot (x_s - e^h_s) + p_s \cdot D_s \le K_s[\sum_{j\in J} p_s \cdot A^j_s(\theta_j - \varphi_j)]^+.$$

His payoff is modified to

$$u^h(x) - \sum_{s\in S} \lambda^h_s \left[\left(\sum_{j\in J} p_s \cdot A^j_s(\varphi_j - \theta_j)\right)^+ - p_s \cdot D_s\right]^+ .$$

Notice that the K_{sj}, λ^h_{sj} are reduced in this setting to K_s, λ^h_s. Finally, the condition on K_s is

$$K_s = \int_0^H p_s \cdot D^t d\mu(t) / \int_0^H \left(\sum_{j\in J} p_s \cdot A^j_s(\varphi^t_j - \theta^t_j)\right)^+ d\mu(t)$$

whenever the denominator is positive.

Along the same lines as Theorem 3, we have

[14] Institutionally this may be regarded as a clearinghouse.

Theorem 4 *A $GE(A, \lambda, Q)$ exists in the finite-type continuum model with supernetting.*

Remark (1) If the constraint Q is removed, then Theorems 3 and 4 still hold provided asset payoffs are designated in a single commodity.

(2) These theorems also hold when the privilege of netting (supernetting) varies with the type of a player. Indeed in this case $GE(A, \lambda, Q)$ will need to have non-type-symmetric behavior only for those types which are allowed to net (supernet).

Proofs of Theorems 3 and 4 We prove Theorem 3. The proof of Theorem 4 is similar. We observe first that since $[\alpha x]^+ = \alpha[x]^+$, the budget sets are lower semi continuous exactly as in the proof of Theorem 1: let $(\bar{x}, \bar{\theta}, \bar{\varphi}, \bar{D}) \in B^h(\bar{p}, \bar{\pi}, \bar{K})$, and let $(p(\varepsilon), \pi(\varepsilon), K(\varepsilon)) \xrightarrow{\varepsilon} (\bar{p}, \bar{\pi}, \bar{K})$, where $\bar{p} \gg 0$. Take $\alpha < 1$ and $(x(\varepsilon), \theta(\varepsilon), \varphi(\varepsilon), D(\varepsilon)) = (\alpha\bar{x}, \alpha\bar{\theta}, \varphi, \alpha\bar{D})$. For ε near 0, these points are clearly budget feasible, because $\bar{p}_s \cdot e_s^h > 0$, $\forall s \in S^*$. Since α was taken to be arbitrary, the budget is LSC.

But now we can repeat the rest of the proof of Theorem 1, replacing ψ_ε^h by $\text{conv}(\psi_\varepsilon^h)$. By Kakutani's Fixed Point Theorem there is a fixed point $(p(\varepsilon)$, $\pi(\varepsilon)$, $K(\varepsilon)$, $(x^h(\varepsilon)$, $\theta^h(\varepsilon)$, $\varphi^h(\varepsilon)$, $D^h(\varepsilon))_{h \in H})$ of ψ_ε. By Caratheodory's theorem $(x^h(\varepsilon)$, $\theta^h(\varepsilon)$, $\varphi^h(\varepsilon)$, $D^h(\varepsilon)) \in \text{conv}(\psi_\varepsilon^h(p(\varepsilon), \pi(\varepsilon), K(\varepsilon)))$ is in the convex hull of at most $n = S^*L + J + J + SL + 1$ points $(x^{hi}(\varepsilon)$, $\theta^{hi}(\varepsilon)$, $\varphi^{hi}(\varepsilon)$, $D^{hi}(\varepsilon))_{i=1}^n$ in $\psi_\varepsilon^h(p(\varepsilon), \pi(\varepsilon), K(\varepsilon))$. Passing to convergent subsequences as $\varepsilon \to 0$ gives a ${}^*GE(A, \lambda, Q)$ equilibrium for the continuum economy, in which each type h displays at most n different (but indifferent!) behaviors.

Remark Notice that our proof shows the existence of a $GE(A, \lambda, Q)$ in which the strategy-selection is a simple function (i.e., takes on a finite number of values). This feature would be lost if we considered a general continuum model which is not of finite type.

7 Examples

We present a basic example in this section. Variations to appear in later sections will illustrate most of the themes of this paper.

Let there be six potential agents and $S = 3$ states of nature, let there be one good in each state, $L = 1$, and suppose agents have no utility for consumption at $t = 0$. Each agent has the same utility

$$u(x_1, x_2, x_3) = \sum_{s=1}^{3} \log(x_s).$$

The endowments of the agents are

$$e^1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix};\ e^2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix};\ e^3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix};$$

$$e^4 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix};\ e^5 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix};\ e^6 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Notice that the first three agents have positive endowments in two states, and will therefore turn out to be more reliable borrowers than the last three agents, who have positive endowments in only one state, and will therefore be more unreliable borrowers.

We take the collection of asset promises to be

$$A^0 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix};\ A^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \equiv 1^1;\ A^2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \equiv 1^2;\ A^3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \equiv 1^3.$$

We take default penalties to be one of three types:

$$\lambda^h_{sj} = \infty,\ \forall h, s, j;\ \lambda^h_{sj} = \lambda > 0,\ \forall h, s, j;\ \text{or}$$

$$\lambda^h_{sj} = \begin{cases} \infty & \text{if } e^h_s = 1 \\ 0 & \text{if } e^h_s = 0 \end{cases}$$

Notice that the first two penalties are completely anonymous, since they are the same whatever the name of the defaulter, and whatever his circumstances. The last penalty type is infinite when agents have the resources to pay, and 0 otherwise. They do not depend on the name of the defaulter, but they do depend on his circumstances; they require more information to carry out. The information required is identical to the sort of information an insurance company must obtain to verify that an accident has occurred. Indeed we shall use these penalties precisely in order to render insurance a special case of default.

In all the examples in Sections 7–11, we shall take

$$Q^h_j = \infty.$$

In Section 12 we study sales limitations closely, and allow for many $Q^h_j < \infty$.[15]

The reader should recall that by combining a long position with a short position on which there may be default, loan (asset) markets may be interpreted as insurance markets. Since we shall take every e^h_s to be either 0 or 1, our entire series of examples may be interpreted as a study in accident insurance markets. In Version B adverse selection and the Rothschild–Stiglitz model of insurance will play the central roles.

Version A0: Complete Markets and Infinite Penalties

We suppose that $H = \{1, 2, 3\}$, so that endowments are $e^1 = (0, 1, 1)$, $e^2 = (1, 0, 1)$, and $e^3 = (1, 1, 0)$. The Arrow–Debreu equilibrium can easily be calculated

[15] In each example we shall compute the equilibrium, so no question of existence arises. It is worth noting that Theorem 1 also guarantees the existence of equilibrium in every example except for the case when $\lambda^h_{sj} = 0$ and $Q^h_j = \infty$. But had we set $\lambda^h_{sj} \geq \varepsilon > 0$ the Theorem would also have guaranteed the existence of equilibrium in every example.

as $p = (1,1,1)$ and $x^h = \left(\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right)$, $\forall h \in H$. The Arrow–Debreu equilibrium can be implemented as a $GE(A, \lambda, Q)$ if there are three assets $A^j = 1^j$, $j = 1, 2, 3$, where 1^j is the jth unit vector, and if default penalties and sales constraints are set at infinity, $\lambda^h_{sj} = Q^h_j = \infty$, $\forall h \in H$, $s \in S$, $j \in J$. The equilibrium is given by $(p, \pi, K, (x^h, \theta^h, \varphi^h, D^h)_{h \in H})$ where $p = (1,1,1)$, $\pi = (1,1,1)$, $K_{sj} = 1$, $\forall sj$, $x^h = \left(\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right)$, $\theta^h = \frac{2}{3}1^h$, $\varphi^h = \frac{1}{3}e^h$, and $D^h_{sj} = \frac{1}{3}$ if $h \neq s = j$, and 0 otherwise.

In the $GE(A, \lambda, Q)$ equilibrium just described, the volume of trade is 2/3 in each of the three asset markets. Notice that there is some trivial multiplicity in the equilibria, since agents could engage in wash sales and buy and sell the same asset. We could instead have taken $\theta^h = \left(\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right)$, $\varphi^h = e^h$, which has volume of trade equal to 2 in each of three asset markets. In this alternative equilibrium, asset prices remain the same. We return to this when we discuss the Modigliani–Miller principle.

8 The Economic Advantages of Intermediate Default Penalties with Incomplete Markets

In Example A0 we found that setting $\lambda^h_{sj} = \infty$ gave a Pareto efficient outcome, because it eliminated default. Setting $\lambda^h_{sj} = \lambda < \infty$ would have led to a Pareto worse outcome. Nevertheless, we shall argue in this section that when markets are incomplete, it is often better to set intermediate default penalties. In Example A0, markets for risk sharing were effectively complete.

There are four fundamental drawbacks to reducing the default penalties λ so far that some agents choose to default in at least some states in equilibrium: (1) creditors, rationally anticipating (on account of direct and indirect reasons) that they might not be repaid, are less likely to lend; (2) borrowers may not repay even in contingencies that have been foreseen, and even though they are able; (3) imposing penalties is a deadweight loss; (4) the default of unreliable agents imposes an externality on reliable agents who, because they cannot distinguish themselves from the unreliable agents, are forced to borrow on less favorable terms.

Akerlof regarded the fourth (externality) cost of default as so important that for this reason alone he suggested it would always be worthwhile to reduce default by imposing penalties on defaulters. By analogy one could ask manufacturers of products to issue guarantees to replace any defective parts, and in addition to pay for all damages caused by defective parts.

Our second goal in this paper is to show that despite myriad reasons why default is socially costly, the benefits from permitting some default often outweigh all of these costs. The benefits from allowing default are basically twofold, and both stem from the fact that markets are incomplete to begin with. First, an agent who defaults on a promise is in effect tailoring the given security and substituting a new security that is closer to his own needs, at a cost of the default penalty. With incomplete markets one set of assets may lead to a socially more desirable outcome than another set. Second, since *each* agent may be tailoring the same given security to his special needs, one asset is in effect replaced by as many assets as there are agents, and so the dimension of the asset span is greatly enlarged. A larger asset span is likely to improve social welfare (although this gain must be weighed against the deadweight loss of the default penalties that are thereby incurred). In short, permitting default allows for a plethora of additional assets that do not have to be specified in advance. Each agent can tailor the simple standard contract to fit his idiosyncratic situation.

A third benefit from allowing default, which is closely related to the first two, is that when there is no netting, agents can go long and short in the same security, thereby doubling their asset span. We make use of this in our examples. (The examples could be presented with netting, or supernetting, but then we would need more assets and a more cumbersome analysis to make the same points.)

Version A1: The Optimal Default Penalty

Consider the economy with $H = \{1, 2, 3\}$, as in Example A0, but with only one asset $A^0 = (1, 1, 1)$. Suppose that the reason for default cannot be observed, so $\lambda^h_{sj} = \lambda$, $\forall h$, s, j. Agents who promise delivery but do not have the good will default and suffer the penalty. Anticipating this they will make fewer promises, and risk-sharing will be reduced.

We can calculate the equilibrium and agent utilities for any value of $\lambda \in [0, \infty]$. When $\lambda = 0$ buyers realize that sellers will not deliver anything, so demand will be zero and equilibrium will involve no trade. When $\lambda \to \infty$ buyers will anticipate full delivery, but sellers will realize that with probability 1/3 they will not be able to avoid a crushing penalty, and so again equilibrium trade goes to 0. By setting an intermediate level of default penalties we can make everybody better off. We graph the situation schematically in welfare space:

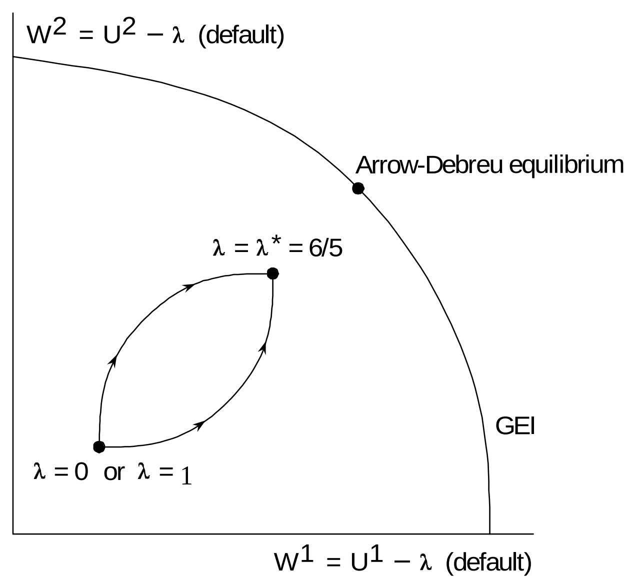

Figure 4

Note first that since the only object traded in period 0 is the asset A^0, we can always take its price $\pi_0 = 1$. Final consumption for agent $h = 1$ will be

$$x^1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} D^1_{10} \\ D^1_{20} \\ D^1_{30} \end{pmatrix} + \begin{pmatrix} K_{10}1 \\ K_{20}1 \\ K_{30}1 \end{pmatrix} \theta^1_0.$$

By symmetry we can guess $K_{s0} = K$, and $D^h_{h0} = 0$, $D^h_{s0} = D$, if $h \neq s$, and $\theta^h_0 = \varphi^h_0 = \theta = \varphi$. In any state, two agents will be delivering D, and since all three will be promising $\varphi = \theta$, we must have $K = 2D/3\varphi$. Hence

$$x^1 = \begin{pmatrix} \frac{2}{3}D \\ 1 - \frac{1}{3}D \\ 1 - \frac{1}{3}D \end{pmatrix}.$$

When $0 \leq \lambda < 1$, $K_{s0} = 0$, $\forall s$, and $\theta^h = \varphi^h = 0$. For $\lambda \geq 1$, let us guess that each agent delivers precisely up to the point where the marginal utility of consumption equals λ, defaulting on the rest of his promises. Then $x^1_2 = x^1_3 = 1/\lambda$, and so $D = 3 - 3/\lambda$. Consumption for agent $h = 1$ (the other h are handled symmetrically) is then

$$x^1 = \begin{pmatrix} 2(1 - 1/\lambda) \\ 1/\lambda \\ 1/\lambda \end{pmatrix}.$$

The marginal utility to buying a unit more of the asset is then

$$MU(\text{buyer}) = K \left[\frac{1}{2(1 - 1/\lambda)} + \lambda + \lambda \right]$$

and, since the agent is defaulting on the margin in all three states, the marginal disutility to selling a unit is

$$MDU(\text{selling}) = 3\lambda.$$

These two must be equal in equilibrium, hence

$$K = \frac{3\lambda}{\frac{\lambda}{2(\lambda-1)} + \lambda + \lambda} = \frac{6\lambda - 6}{4\lambda - 3}.$$

Moreover

$$\varphi = \frac{2}{3}\frac{D}{K} = \frac{(2 - \frac{2}{\lambda})\left(\frac{1}{2-\frac{2}{\lambda}} + \lambda + \lambda\right)}{3\lambda} = \frac{4}{3} - \frac{1}{\lambda}.$$

Notice that φ, D, and K are monotonically increasing in λ. For $1 \leq \lambda < 6/5$, $D < \varphi$, confirming that we have guessed a genuine equilibrium. Note that at $\lambda = 1$, $D = 0$, and the only equilibrium involves no trade. Because marginal utility is infinite at the no trade point, trade jumps immediately for $\lambda > 1$ to $\varphi = 4/3 - 1/\lambda$. As λ rises to $\lambda = 6/5$, φ rises to $1/2$ and D rises to $1/2$, and K rises to $2/3$.

At $\lambda = 6/5$, $x^1 = \left(\frac{1}{3}, \frac{5}{6}, \frac{5}{6}\right)$, $x^2 = \left(\frac{5}{6}, \frac{1}{3}, \frac{5}{6}\right)$ and $x^3 = \left(\frac{5}{6}, \frac{5}{6}, \frac{1}{3}\right)$. By buying and selling $1/2$ unit of the asset A^0, agent h gains $1/3$ when $s = h$ and loses $1/6$ in the two states $s \neq h$. Agent h delivers fully when $s \neq h$ because his marginal utility of consumption after delivery is $\frac{1}{(5/6)} = \frac{6}{5} = \lambda^*$. When $s = h$, agent h defaults completely since his marginal utility of consumption $\frac{1}{(1/3)} = 3 > \lambda^*$. Since for any $s \in S$ we have 2 agents with $h \neq s$, $K_{s0} = \frac{2}{3}$. Thus the asset promise $A^0 = (1,1,1)$ actually delivers $\left(\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right)$ per unit promise. Agent $h = 1$ delivers $\frac{1}{2}(0,1,1)$, agent $h = 2$ delivers $\frac{1}{2}(1,0,1)$, and agent $h = 3$ delivers $\frac{1}{2}(1,1,0)$. The reason each agent buys and sells only 1/2 a unit of asset A^0 instead of a full unit to get to the Arrow–Debreu allocation is that the sale of φ units of the asset is accompanied by the loss of $\varphi\lambda$ utiles for the inevitable default in state $s = h$. The marginal utility from buying the asset is $\left(\frac{2}{3}\right)\left(\frac{6}{5}\right) + \left(\frac{2}{3}\right)\left(\frac{6}{5}\right) + \left(\frac{2}{3}\right)(3) = 18/5$; the marginal disutility from selling is also $\left(\frac{6}{5}\right) + \left(\frac{6}{5}\right) + \left(\frac{6}{5}\right) = \frac{18}{5}$. (It is therefore more convenient to take $\pi_0 = \frac{18}{5}$.)

For $\lambda \geq 6/5$, the agents always deliver if they have the goods on hand. Thus for $\lambda > 6/5$ we can no longer maintain our guess that agents default until marginal utility equals λ. When λ is increased beyond $\lambda^* = 6/5$, marginal utility is less than λ in the good state, and K_{s0} is maintained at 2/3, but asset trade again begins to drop because the inevitable punishment makes selling less attractive. The formulas are messy and we do not bother to present them here. An increase in the penalty rate beyond $\lambda = 6/5$ does not improve risk bearing (since φ begins to drop), and it also increases the deadweight loss from punishing agents who cannot deliver anyway. It thus strictly lowers welfare.

Furthermore, observe that as λ rises from 1 to $\lambda = 6/5$, the deadweight utility loss from default

$$\lambda\varphi + 2\lambda(\varphi - D) = \frac{4}{3}\lambda - 1 - \frac{10}{3}\lambda + 4 = 3 - 2\lambda$$

actually falls, to 3/5. Since the allocation is improving, and the default penalty is falling, we deduce that $\lambda^* = 6/5$ leads to the Pareto best outcome among all economies with $\lambda^h_{sj} = \lambda$.

Example A1 illustrates that the optimal default penalty might be low enough to encourage some real default, and the attendant deadweight loss, when markets are incomplete. It also illustrates that the possibility of default makes the asset payoffs endogenous, since we do not know before an equilibrium is calculated what the default rates will turn out to be. If we change the utilities or endowments of the agents, or the default penalties, the equilibrium will change, the default rates will change, and the asset payoffs will be different. In the next sections, we turn to a more radical kind of asset endogeneity: the endogeneity of the whole asset structure.

II. Endogenous Asset Structures

In some contexts it has become customary to think of endogenizing the asset structure by allowing atomic agents to invent new assets (often one at a time) to upset a prevailing equilibrium. These asset-creating agents are hypothesized to be motivated by payoffs that might depend on the perceived volume of trade which would take place in their new asset if no other prices changed (or in the new trading equilibrium after all prices equilibrated), or in some other way on their perceived profits from introducing the new asset. When the status quo assets are chosen so that none of these agents has an incentive to introduce a new asset, the asset structure is said to have been endogenously determined. This approach to endogenizing the asset structure almost inevitably involves a combination of price taking behavior and oligopolistic-Nash thinking on the part of the asset creating agents. (An example will be treated in detail when we discuss the Rothschild–Stiglitz model of insurance in Section 12.)

By contrast we follow a relentlessly competitive approach to the problem of endogenous assets. Every agent is a price taker. An asset is endogenously missing in our approach only if there is a price at which no agent wants to sell or buy it. Our approach is to include all potential assets in the original asset structure J and to see which assets are priced in such a way that they are not traded in equilibrium (i.e., at their equilibrium prices both demand and supply are zero). When this occurs we say that the actively traded assets have been determined endogenously.[16] Our main conclusion is that a bit of exogenous incompleteness in one asset dimension can lead to substantial endogenous incompleteness in other dimensions.

It is important to note that an asset is specified not just by its vector A^j of promises across states, but also by the associated default penalties λ^h_{sj}, and quantity constraints Q^h_j. When we say that the asset structure is determined endogenously, we mean the promises, the default penalties, and the quantity constraints. In Sections 9–12 we shall elaborate our meaning.

As an illustration of our approach, let N be a very large nonnegative integer. Let the set of potential promises be given by

$$\mathbb{P} = \left\{ \frac{n_{11}}{N^3}, ..., \frac{n_{SL}}{N^3} : 0 \leq n_{s\ell} \leq N^3,\ n_{s\ell} \in \mathbb{Z},\ \sum_{s=1}^{S} \sum_{\ell=1}^{L} n_{s\ell} = N^3 \right\}.$$

Let the set of potential default penalties be given by

$$\Lambda = \left\{ \left(\frac{n_1}{N}, ..., \frac{n_S}{N} \right) : 0 \leq n_s \leq N^2,\ n_s \in \mathbb{Z} \right\}$$

and let the set of potential quantity constraints be given by

$$\mathbb{Q} = \left\{ \frac{n}{N} : 0 \leq n \leq N^2,\ n \in \mathbb{Z} \right\}.$$

The comprehensive set of potential assets is given by $\overline{\mathcal{A}} \equiv \mathbb{P} \times \Lambda^H \times \mathbb{Q}$, which can be indexed by $\bar{J} = \{1, ..., \bar{J}\}$. Sometimes we shall take $\mathcal{A} = \overline{\mathcal{A}}$, and see which assets are traded in positive quantities in equilibrium. This will usually be a small subset. Often we shall take $\mathcal{A} \subsetneqq \overline{\mathcal{A}}$; typically we shall then find still fewer assets actively traded in equilibrium. Let us begin with the comprehensive set of assets $\overline{\mathcal{A}}$, and no transactions costs.

[16] Our approach in this section, including our existence theorems, is similar in spirit to the pioneering efforts of Allen and Gale (1988, 1991) and Pesendorfer (1995). None of these papers considered default or signalling via quantity constraints.

If the government could simultaneously and without limitations choose promises, default penalties, and quantity constraints, it would select promises equivalent to a full set of Arrow securities, infinite penalties, and arbitrarily high quantity constraints. Now we show the market would do the same.

Version A2: Arrow Securities Emerge When Default Penalties Are Infinite

Consider our standard example with $H = \{1, 2, 3\}$, but with four assets $A^j = 1^j$, $j = 1, 2, 3$, and $A^0 = (1, 1, 1)$. Let the penalties be $\lambda^h_{sj} = \infty$ if $j = 1, 2, 3$, and $\lambda^h_{s0} = \lambda^* = \frac{6}{5}$ for all h and s. Despite the fact that the default penalty for asset 0 has been chosen "optimally," the unique equilibrium (ignoring redundant trades) is the Arrow–Debreu equilibrium of Version V0, so that asset 0 is not traded at all. The forces of supply and demand determine that the Arrow securities are traded and other assets are not.

We elevate this example to a theorem:

Theorem 5 *Let $\mathcal{E} = ((u^h, e^h)_{h\in H}, (\mathcal{A}_j, ((\lambda^h_{sj})_{s\in S}, Q^h_j)_{j\in J}))$ be an economy which includes all the Arrow securities: for each $s \in S$, there is an asset $j = j(s)$ such that $A^j_{sL} = 1$ and $A^j_{s'\ell} = 0$ otherwise, with $Q^h_j = \infty$ $\forall h$ and $\lambda^h_{sj} = \infty$ $\forall h$ and $\forall s$. Then for any $GE(A, \lambda, Q)$ equilibrium $\eta = ((p, \pi, K), (x^h, \theta^h, \varphi^h, D^h)_{h\in H})$, we can find prices $q \in \mathbb{R}^{(1+S)L}_{++}$ such that $(q, (x^h)_{h\in H})$ is an Arrow–Debreu equilibrium. Moreover, if $\lambda \gg 0$, no agent defaults on any actively traded asset in η, even if there are assets $j \in J$ with low λ^h_{sj}. Finally, there is an equilibrium η', possibly η itself,with the same $((p, \pi, K), (x^h)_{h\in H})$ such that the only actively traded assets in η' are Arrow securities.*

Proof Let η be given. Let $q_0 = p_0$ and let $q_s = \pi_{j(s)}(p_s/p_{sL})$, $\forall s \in S$. Let

$$v^h(q) \equiv \max\{u^h(x) : q \cdot x \leq q \cdot e^h,\ x \in \mathbb{R}^{(1+S)L}_{+}\}.$$

Observe that $K_{sj} = 1$ for each asset j with $\lambda^h_{sj} = \infty$ $\forall h, s$, if $A_s \neq 0$, since no agent will default in the refinement. It follows that by never defaulting, each agent h could, by selling and buying the Arrow securities, achieve at least $v^h(q)$, that is,

$$u^h(x^h) \geq u^h(x^h) - \text{default penalty} \geq v^h(q).$$

It follows that $q \cdot x^h \geq q \cdot e^h$ $\forall h \in H$. Since η is an equilibrium $\sum_{h\in H} x^h = \sum_{h\in H} e^h$. Hence $q \cdot x^h = q \cdot e^h$ $\forall h \in H$, and $(q, (x^h)_{h\in H})$ is an Arrow–Debreu equilibrium, and the default penalty actually borne by each agent $h \in H$ is zero.

Clearly each agent is indifferent to achieving x^h via the actively traded assets in η, or via Arrow securities. If every agent trades exclusively via Arrow securities, then supply will equal demand, and η' is a genuine equilibrium. ■

The theorem tells us that if $\overline{\mathcal{A}}$ is the comprehensive set of asets, then $\overline{\mathcal{A}}^*$ consists of Arrow securities, or at least of a group of assets whose (unbroken) promises span all the state contingent transfers of wealth necessary to support the Arrow–Debreu equilibrium.

The rest of this paper is devoted to describing (realistic) situations to which Theorem 5 cannot be applied, and the actively traded assets $\mathcal{A}^*$ are not Arrow securities.

In Section 9 we introduce transactions costs and evaluation costs. We find that then $\overline{\mathcal{A}}^*$ typically has lower cardinality, and no intersection, with the collection of Arrow securities.

In Sections 10–12 we drop transactions costs, but we show that if the set $\mathcal{A} \subset \overline{\mathcal{A}} \equiv \mathbb{P} \times \Lambda^H \times \mathbb{Q}$ of available assets is limited in some dimension, then the set $\mathcal{A}^*$ of actively traded assets will be limited in another dimension, and in a different way than it would be in $\overline{\mathcal{A}}^*$. In Section 10 we limit the promises (and selling constraints), $\mathcal{A} = \{A_0\} \times \Lambda \times \{\infty\}$, and we find that only assets with intermediate default penalties are traded, $\mathcal{A}^* = \{A_0\} \times \{\lambda^*\} \times \{\infty\}$ where $\lambda^* < \infty$, even though all other penalties are available. In Section 11 we allow only low default penalties, $\mathcal{A} = \mathcal{P} \times \{\lambda\} \times \{\infty\}$, and we find in equlibrium that the span of actively traded asset promises is low-dimensional, $\mathcal{A}^* = \{A_0\} \times \{\lambda\} \times \{\infty\}$, even though all possible promises are available. Finally in Section 12 we limit the available promises and penalties $\mathcal{A} = \{A_0\} \times \{\bar{\lambda}\} \times \mathbb{Q}$, and we show that the endogenous choice of actively traded quantity constraints in $\mathcal{A}^*$ is essentially equivalent to the Rothschild–Stiglitz problem of equilibrium insurance design.

Before moving to Sections 9–12, let us reflect for a moment on why endogenous asset trade is a suitable problem in the $GE(A, \lambda, Q)$ model. In the GEI model, where there is no default by assumption, every asset in J will almost surely be traded itself, or be spanned by the actively traded assets. The problem of active trade in GEI equilibrium is thus almost never of interest.[17] In $GE(A, \lambda, Q)$ equilibrium, it will turn out that many assets will not be actively traded, even though their promises are not spanned by the actively traded assets. The reason is that in the $GE(A, \lambda, Q)$ model, there may be a wedge between the marginal utility of buying an asset and the marginal disutility of selling the same asset. The possibility of default means that a buyer does not expect the benefit of full payoffs, but if he sells the asset he incurs the obligation of making full payoffs (although he can choose to default and suffer a penalty instead).

Thus in $GE(A, \lambda, Q)$ equilibrium we should expect many assets not to be traded. Which of the assets turns out to be traded cannot be anticipated in advance without knowledge of the utilities, endowments, and so on of the agents. The assets that do get traded are therefore endogenously determined by the forces of supply and demand.

[17] Of course for special GEI models with special utilities, like the capital asset pricing model, it is very interesting to see which assets are actively traded. But these are non-generic economies.

9 Endogenous Assets, Transactions Costs, and Incomplete Markets

Even if assets and penalties could be chosen simultaneously, there is good reason to suppose that not every Arrow security would be actively traded. In practice, that would be much too costly. In the next two subsections we formalize two kinds of costs.

We begin by noting that in general, the Arrow–Debreu allocation can be achieved without using all the Arrow securities. For example, suppose $2H < S$. Let $\tilde{z}^h = \tilde{x}^h - \tilde{e}^h$ be the net trade vector of time 1 commodities at a Walrasian equilibrium. Consider the collection of the set of $2H$ nonnegative vectors $\bigcup_{h\in H}\{\tilde{z}^h_+, \tilde{z}^h_-\}$, where $\tilde{z}^h_+ - \tilde{z}^h_- = \tilde{z}^h$. We can always find J^* linearly independent, nonnegative vectors $B^1, ..., B^{J^*}$, with $J^* \le 2H < S$, such that $\bigcup_{h\in H}\{\tilde{z}^h_+, \tilde{z}^h_-\} \subset \text{cone}\{B^1, ..., B^{J^*}\}$. Attaching infinite default penalties, each agent could sell and buy precisely what he wanted by trading the B assets, and given the independence of the assets, if any agent did so, then supply would equal demand. If $2H \ll S$, one might expect this system of assets to be traded rather than the Arrow securities.

Far greater savings can often be obtained from the observation that agents often differ more in their idiosyncratic selling than in their buying. All risk averse agents, for example, prefer to buy riskless consumption over risky consumption with the same expected payoff. By the same token, agents have idiosyncratic endowments and income streams, so they each have different objects to sell. In our equilibrium, each agent can deliver different amounts on the same asset; buyers can make the same purchases while sellers make idiosyncratic deliveries.

We now present an example which shows that in some cases the Arrow–Debreu equilibrium can be achieved with just one asset.

Version A3: The Advantages of Standardized Assets

Consider the economy described in Section 7 with $H = \{1, 2, 3\}$ and with just one asset $A^0 = (1, 1, 1)$. Suppose that the default penalties are

$$\lambda^h_{sj} = \begin{cases} \infty & \text{if} \quad s \neq h \text{ (i.e., if } e^h_s = 1) \\ 0 & \text{if} \quad s = h \text{ (i.e., if } e^h_s = 0) \end{cases}$$

that is, default penalties are infinite when agents have the resources to pay, and 0 otherwise. We might interpret state s as the state in which a bad accident happens to agent $h = s$. Suppose for now that there are no transactions costs.

The interesting thing is that the Pareto efficient Arrow–Debreu equilibrium can now be implemented with only one asset. Let $\pi_0 = 3$, $p = (1, 1, 1)$. Agent h buys *and* sells 1 unit of the asset, delivering fully when his endowment is 1, and defaulting completely when his endowment is 0. Since in every state two agent types deliver and the other type defaults, $K_{s0} = \frac{2}{3}$, $\forall s \in S$. Consumption by h in the state $s = h$ where he has no endowment is thus $K_{s0}\theta^h_{s0}A_{s0} = \left(\frac{2}{3}\right)(1)(1) = \frac{2}{3}$. Consumption in the other states where he delivers is $e^h_s + K_{s0}\theta^h_{s0}A_{s0} - D^h_{s0} = 1 + \frac{2}{3}(1)(1) - 1 = \frac{2}{3}$. We verify that this is a $GE(A, \lambda, \infty)$ equilibrium by noting that the marginal utility of owning an extra unit of the asset is $\sum_{s=1}^{S} \frac{\partial u}{\partial x_s} K_{s0} A_{s0} = \frac{3}{2}\left(\frac{2}{3}\right) + \frac{3}{2}\left(\frac{2}{3}\right) + \frac{3}{2}\left(\frac{2}{3}\right)$, which is equal to the marginal disutility of selling the asset $\sum_{s=1}^{S} A_{s0} \min\left[\frac{\partial u}{\partial x_s}, \lambda_{s0}^h\right] = \frac{3}{2}(1) + \frac{3}{2}(1) + 0$, where $\frac{3}{2} = \frac{d \log(2/3)}{dx}$ is the marginal utility of consumption in each state.

Version A3 seems at first glance like an artificial example,because the penalties themselves are partly idiosyncratic. But, as we said earlier, they are no more personalized than insurance contracts. Indeed one could regard example A3 as a single insurance contract, as we will in Section 12. The same point can be made anyway (and will be in Example A6) with default penalties λ that are the ssame for all people, and all states.

9.1 Transactions Costs, Liquidity Costs and Endogenous Assets

Trade in any market usually involves some sort of transactions cost. The "competitive market" itself is an abstraction of a complicated set of interrelationships between brokers, middlemen, buyers and sellers, and it should come as no surprise that final buyers do not receive the full value of what sellers give up. As a first approximation, we can represent this wedge by a simple utility loss to transacting, proportional to the quantity of the transaction. We can, however, be a little bit more specific about which asset markets are likely to have higher transactions costs.

When an asset is very finely defined, so as to pay off in exactly those states that a particular small group of people is interested in, then it is not likely to be heavily traded. A seller may have to wait a long time to find a suitable buyer and vice versa. And when such a buyer is found, he will exercise some temporary monopsony power. We say that the market lacks liquidity.

As a first approximation we can incorporate liquidity costs simply by assuming that more utility is lost in transactions in less liquid markets. For each asset $j \in J$, denote the total volume of purchases and sales by $\theta_j = \sum_{h \in H} \theta_j^h$, and $\varphi_j = \sum_{h \in H} \varphi_j^h$. Then we can define agent h's utility at an equilibrium $E = (p, \pi, K, (x^h, \theta^h, \varphi^h, D^h)_{h \in H})$ by

$$^*W^h(E) = W^h(E) - \sum_{j=1}^{J} \theta_j^h a_j^h(\varphi_j) - \sum_{j=1}^{J} \varphi_j^h b_j^h(\theta_j)$$

where a_j^h and b_j^h are continuous, nonnegative, decreasing functions. When a_j^h and b_j^h are constant functions, we get a simple transactions costs economy. When they are strictly decreasing in the quantity of trade on the other side of the market, they indicate that part of the transactions costs are due to the difficulty of finding an agent with whom to trade.

$^*GE(A, \lambda, Q)$ equilibrium is defined exactly the same way as $GE(A, \lambda, Q)$ equilibrium, except that $^*W^h$ replaces W^h for each agent h. Each agent h regards θ and φ as fixed when he ponders changing his θ^h and φ^h. Under these circumstances, equilibrium always exists.

Theorem 6.1 *Under the conditions of Theorem* 1*, a* $^*GE(A, \lambda, Q)$ *equilibrium exists.*

Proof The theorem needs no additional proof, since under the hypothesis that each agent regards himself as so small that he cannot affect either θ or φ, his payoff $^*W^h$ is still a concave function of his choice variables. ■

The advantages of conducting trade through large, standardized, liquid markets as opposed to many specialized markets becomes quite clear when we consider transactions costs. It is possible to standardize contracts to some degree because idiosyncratic default on the same standardized contract can sometimes offer almost the same flexibility as completely separate contracts, as we saw in example Version A3.

If we introduced high liquidity costs into the example, efficiency gains of the allocation with default described in Version A3 would be quite striking compared to the complete markets allocation described in Version A0. The same allocation is achieved via one asset, instead of via three assets. And the liquidity of the single asset is 3, instead of 2/3 for each of the three assets in Version A0.[18] Evidently it is socially preferable to have many agents sell the same promise to deliver a dollar unconditionally, and then default in the idiosyncratic states where they cannot pay, rather than to define a separate, idiosyncratic asset for each agent, which only he will sell.

We turn now to an equally important source of trading costs.

9.2 Information and Evaluation Costs

When an agent considers buying a contingent asset he must think carefully about its implications. This computation cost is usually highly nontrivial in practice, and causes most people to shy away from most securities. As a first approximation we can formalize this cost by subtracting a fixed entry cost for buying or selling an asset:

$$^{**}W^h(E) = ^*W^h(E) - \sum_{j=1}^{J} \bar{a}_j^h(A^j, (K_{sj})_{s\in S}, \varphi_j)\chi(\theta_j^h) - \sum_{j=1}^{J} \bar{b}_j^h(A^j, (\lambda_{sj}^h)_{s\in S}, \theta_j)\chi(\varphi_j^h)$$

where $\chi(x) = 1$ if $x > 0$, and 0 otherwise, and $\bar{a}_j^h$ and $\bar{b}_j^h$ are continuous, nonnegative functions which are decreasing in θ_j and φ_j, respectively.

The "evaluation" costs $\bar{a}_j^h$ and $\bar{b}_j^h$ of studying a security may depend on the sources of its riskiness. For example, it may be harder to think through contingent defaults

[18] We could formally modify our example to incorporate transaction costs $a_j^h(\theta_j)$ and $b_j^h(\varphi_j)$ which are very high for small volumes of trade, but decline (continuously) to near 0 as θ_j and φ_j approach 3. Then the equilibrium described in Version A3 would indeed be very close to a genuine equilibrium with transactions costs (which we do not bother to compute). This equilibrium easily Pareto dominates what could also be accomplished with Arrow securities. Observe that with wash sales, the Arrow securities could be traded in large enough volume to nearly elimiante transactions costs. But as long as the transactions cost is positive, no matter how small, no agent will engage in these wash sales, rendering the Arrow securities prohibitively expensive.

than contingent promises. To allow for these possibilities we made the cost of evaluation depend on the rates of payment, as well as on the promises. Furthermore, a large volume of trade in a market may make it very easy to learn about the security, whereas a small volume of trade means an esoteric security for which it is hard to find an expert who can explain it. Thus we also allowed the fixed cost to depend on the liquidity of the market.

We define $^{**}GE(A, \lambda, Q)$ exactly as $GE(A, \lambda, Q)$ was defined, except that $^{**}W^h$ replaces W^h for all $h \in H$. Again equilibrium must exist.

Theorem 6.2 *In the finite-type, continuum model, $^{**}GE(A, \lambda, Q)$ equilibrium always exists.*

Sketch of Proof The fixed cost of buying an asset destroys both the continuity of $^{**}W^h$ and its concavity, so at first glance it seems to compromise our existence proof. But in fact on closer inspection one sees that demand is still upper semi continuous, because utility jumps up at zero demand, and zero demand is always feasible. To be slightly more precise, let us abuse notation and use a transparent shorthand for the budget set, demand, and the macro variables. Let x^n be optimal in the budget sets $B^h(p^n)$, and let p^n approach p and x^n approach x. If x is not optimal in $B^h(p)$, then there must be some y in $B^h(p)$ with $^{**}W^h(y) > {}^{**}W^h(x)$. But from the proof of lower semicontinuity of the budget set in Theorems 1 and 4, we know that we can approximate y by points y^n in $B^h(p^n)$ in such a way that the asset purchases and sales in y^n are less than or equal respectively to the asset purchases and sales in y. Hence if y involves no purchases or no sales of some asset, then so does y^n, and the utilities $^{**}W^h(y^n)$ must approach $^{**}W^h(y)$. But for the same reason the utilities $^{**}W^h(x^n)$ approach a number at most equal to $^{**}W^h(x)$. Hence we deduce that x^n is not an optimal demand in $B^h(p^n)$ after all, a contradiction showing that demand is USC.

Since the lack of continuity of $^{**}W^h$ has no effect on the upper semi continuity of demand, the problem is reduced to the lack of concavity of $^{**}W^h$. But that is exactly analogous to the lack of convexity of the budget set, and we dealt with that in Theorems 3 and 4. ■

We now have two reasons (liquidity costs and information-evaluation costs) why equilibrium cannot support a full set of traded assets. Both reasons give advantages to the standardized contract in example A3 over the Arrow securities described in A0. In general if we begin with a full complement in J of asset promises, default penalties, and quantity constraints, in equilibrium only a very few of them will be traded on account of the liquidity costs and information costs. These will be the endogenously determined assets.

The effects of liquidity costs and information-evaluation costs on endogenous asset choice is the subject of further planned work. For the rest of the paper, we drop these costs. Instead, we illustrate how limitations in the availability of assets with some asset dimensions can limit the endogenous choices of the rest of the asset dimensions.

We shall show in an example that if only a limited range of (intermediate) default penalties is available, then only a few promises will be traded, even if every conceivable promise is available. Similarly, we shall see how penalties might be chosen by the market, fixing a limited set of promises, and how quantity constraints might be chosen, fixing both promises and penalties. After all, in practice the choice is never to introduce all potential assets immediately, but rather incrementally to add some new promises, or to modify default penalties.

10 Endogenous Default Penalties When Promises Are Incomplete

We have already seen that when all asset promises are available, the market should and will exclusively trade promises with infinite penalties. Let us suppose that on account of the transactions costs described in Section 9, or perhaps for other reasons like simplicity, the set of assets J contains only a limited variety of promises, far short of a complete set of Arrow promises. Given these limitations on promises, in Section 8 we were able to ask how severe the default penalties *should* be to promote economic efficiency. Since our model allows for the possibility that different punishment regimes coexist at the same time, we can also ask how harsh the punishment scheme *will* be that endogenously emerges in equilibrium. For example, an agent could indicate his intention to perform a service, he could orally commit to performing the service, he could put in writing that he promised to perform a service, or he could draw up a contract with a lawyer announcing his promise to perform a service. If all four of these promises are treated equally by the courts, then there is no issue of selecting a punishment. But if the punishment in case of default is different for these different manners of making the same promise, then in effect the parties to the agreement are choosing the severity of default penalties attached to the promise. We shall now show that in our example, the forces of supply and demand select the optimal default penalties.

Version A4: Endogenous Default Penalties

Consider the model of version A1 with only one asset promise $A^0 = (1,1,1)$ and $\lambda^h_{s0} = \lambda^* = 6/5$, $\forall h \in H$ and $\forall s \in S$. It is natural to regard the penalty λ^* as imposed by a beneficent and knowledgeable government. But we may also regard λ^* as emerging from the equilibrium forces of supply and demand.

Now let there be a finite number of additional assets A^j, all making the same promises $A^j = (1,1,1)$, but with default penalties $\lambda_j = \lambda^h_{sj}$ for all $h \in H$, $s \in S$, ranging at intervals of $\lambda^*/100$ from 0 to $100\lambda^*$. We shall now show that despite the myraid of available assets, in every (symmetric) equilibrium, all trade will be conducted in the assets j for which $\lambda^h_{sj} = \lambda^*$. We begin by describing an equilibrium of this type, and then we show it is essentially the only (symmetric) equilibrium.

The equilibrium will involve exactly the same prices, delivery rates, trades, and consumption as described in example A1 for the case $\lambda = \lambda^* = \frac{6}{5}$. We must now extend that equilibrium to define prices π_j and delivery rates K_{sj} for all the new assets. Set $\pi^* = \frac{18}{5}$, and set $\pi_j = \frac{6}{5} + \frac{6}{5} + \min\{\lambda_j, 3\}$ for $\lambda_j \geq \lambda^* = \frac{6}{5}$, which is the marginal disutility of selling asset j, when $\lambda_j \geq \lambda^*$. At these prices optimal supply is zero. Recall in example A1, $\pi_0 = \frac{18}{5}$ = the marginal utility of buying or selling asset A^0, hence $\pi_j \geq \pi_0$. For $\lambda^* \leq \lambda_j < 3$, set $K_{sj} = 2/3$ for all $s \in S$. The marginal utility of buying such assets j is then 18/5, and this is never higher than the price π_j, hence optimal demand $\theta_j^h = 0$. For $\lambda_j \geq 3$, we let $K_{sj} = 1$, $\forall s \in S$. Here $\pi_j = \frac{12}{5} + 3 = \frac{27}{5}$. But the marginal utility of buying such an asset is also $\frac{6}{5} + \frac{6}{5} + 3 = \frac{27}{5}$, again optimal demand and supply are zero.

For $\lambda_j < \lambda^*$, define $\pi_j = 3\lambda_j$ and $K_{sj} = 0$, $\forall s \in S$. Since utilities are concave, to check that no agent would sell asset j, we only need to look at the marginal utilities at the allocation achieved in example A1. Clearly on the margin, the disutility of selling asset j is $3\lambda_j$. Notice that every seller would in fact completely default, hence buyers are rational to anticipate this, and to demand zero.

In every case we set the price equal to the marginal utility of the sellers, and the K_{sj} equal to the rates of payments that would be made with infinitesimally higher π_j. Hence in every case buyer expectations are rational, i.e., they satisfy condition (5) of equilibrium. We have thus displayed an equilibrium in which (almost) any default penalty is available, yet only a single one (namely the Pareto efficient penalty) is used in equilibrium.

We now argue that there can be no other (symmetric) equilibrium. In any (symmetric) equilibrium we have consumption $x^1 = (2x, 1-x, 1-x)$, and similarly $x^2 = (1-x, 2x, 1-x)$, and $x^3 = (1-x, 1-x, 2x)$, with $x \leq \frac{1}{3}$. If $x = 1/6$, we must be in our original equilibrium. If $x > 1/6$, then agent 1 has delivered up to a point in states 2 and 3 where his marginal utility of consumption $\frac{1}{1-x} > \frac{6}{5}$. He would not have done that unless he was selling an asset with default penalty $\lambda_j \geq \frac{1}{1-x} > \frac{6}{5}$. If asset j delivers fully in every state, then it is irrelevant, since by symmetry each agent is buying and selling an equal amount of it. But from the argument in the proof of Theorem 2, if the asset did not deliver everywhere, then any agent buying and selling it would default completely in at least one state. Since by symmetry every agent buys and sells it, $K_{sj} \leq 2/3$, $\forall s \in S$. The marginal utility to purchasing asset j is at most $\frac{2}{3}\left(\frac{1}{2x} + \frac{1}{1-x} + \frac{1}{1-x}\right) = \frac{2}{3}\frac{3x+1}{(1-x)2x} = \frac{1}{1-x} + \frac{1}{(1-x)3x} < \frac{3}{1-x}$ (if $x > \frac{1}{6}$) of utility in period 1. The marginal disutility of selling asset j is at least $\frac{1}{1-x} + \frac{1}{1-x} + \frac{1}{1-x} = \frac{3}{1-x}$, a contradiction.

If $x < 1/6$, we shall show there can be no equilibrium price π^* for asset $j = j^*$. The marginal disutility of selling asset j^* is $\frac{1}{1-x} + \frac{1}{1-x} + \frac{6}{5}$. Since $\frac{1}{1-x} < \frac{6}{5}$, the marginal disutility of selling is less than 18/5. It also follows that every agent would deliver in each of his two good states if he were selling asset j^*. Hence $K_{s0} = 2/3$, $\forall s \in S$. The marginal utility of buying asset j^* is then $\frac{2}{3}\frac{1}{1-x} + \frac{2}{3}\frac{1}{1-x} + \frac{2}{3}\frac{1}{2x}$. For $x < \frac{1}{6}$, the marginal utility of buying is always larger than 18/5, hence larger than the marginal utility of selling, a contradiction.

11 Endogenous Promises when Default Penalties Are Low Enough to Permit Default

Consider a situation in which default penalties are not allowed to be too severe, either because of the reasons suggested in Section 8, or because politics do not permit harsh penalties that don't "fit the crime." We shall show that then the "Arrow promises" will often not be actively traded, even if they are available without transactions costs.

Version A5: Adverse Selection with Differential Penalties

Let us introduce assets $A^1 = (1,0,0)$, $A^2 = (0,1,0)$ and $A^3 = (0,0,1)$ into the economy of version A3 (either in place of asset $A^0 = (1,1,1)$, or in addition to A^0), where the penalties are $\lambda^h_{sj} = \begin{cases} \infty \text{ if } e^h_s = 1 \\ 0 \text{ if } e^h_s = 0 \end{cases}$. In equilibrium none of the Arrow securities would be traded, since agents of type $h = j$ would be tempted to sell $j \in \{1,2,3\}$ like crazy, thereby reducing K_{sj} to zero, so that π_j would be zero, so that no agents besides those of type $h = j$ would sell asset j even if π_j went up a little.

When default penalties are not uniform, asset promises which pay off in very specific states are not likely to be traded even if they can be written, because there is bound to be some agent who can take advantage of the specificity of the conditions to escape punishment. These agents will debase the value of the asset and prevent others from selling it. In short, when there is a variety of penalties λ^h_{sj}, buyers must beware of an adverse selection of sellers with $\lambda^h_{sj} < \partial u^h / \partial x_s$.

Version A6: Span of Active Assets Shrinks as Default Penalties Fall

Consider the same three agents' utilities and endowments as in examples A0–A5. Suppose now, however, that there are asset promises $A^0 = (1,1,1)$, $A^1 = 1^1$, $A^2 = 1^2$, $A^3 = 1^3$. Fix all the penalties $\lambda^h_{sj} = \lambda$, for all $h \in H$, $s \in S$, $j = 0,1,...,J$, as in Version A1. We wish to illustrate two points. First, we shall see that in equilibrium not all available assets are traded, even though there are no transactions costs. Second, we shall see that as the default penalties λ decline, the span of actively traded asset promises shrinks.

When $\frac{3}{2} \leq \lambda \leq \infty$, there is (essentially) a unique equilibrium reproducing the Arrow–Debreu equilibrium of Version A0. Each agent h consumes $\left(\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right)$, obtained by selling $1/3$ units of assets $j \in \{1,2,3\} \setminus h$, and buying $2/3$ units of asset h. (Thus agent 1 puts $\varphi^1_2 = \varphi^1_3 = \frac{1}{3}$ and $\theta^1_1 = \frac{2}{3}$.) All assets deliver completely, $K_{sj} = 1\ \forall s \in S$, $j \in J$, and $1 = \pi_1 = \pi_2 = \pi_3 = \frac{1}{3}\pi_0$. There is no trade in asset 0.

When $\frac{3}{2} > \lambda > \frac{6}{5}$, default emerges, but traded asset promises are still "complete." In the unique GEI(A, λ, ∞) equilibrium, $x^h_s = \begin{cases} \frac{1}{\lambda} & \text{if } h \neq s \\ 2\left(1 - \frac{1}{\lambda}\right) & \text{if } h = s \end{cases}$. We can guess that each agent h sells φ units of each of the two "Arrow promises" $j \neq h$, and buys 2φ units of asset $j = h$. (The prices of all three assets is the same.) The marginal utility of buying asset $j = h$ is $MU_B = \frac{1}{2(1-1/\lambda)} K$ and the marginal utility of selling either "Arrow promise" $j \neq h$ is $MU_S = \lambda$. Equating the two marginal utilities gives $K = 2(\lambda - 1)$. Final consumption, say for agent $h = 1$, is

$$x' = \begin{pmatrix} 2\,(1 - 1/\lambda) \\ 1/\lambda \\ 1/\lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ K\varphi \\ K\varphi \end{pmatrix} + \begin{pmatrix} 2K\varphi \\ 0 \\ 0 \end{pmatrix}.$$

This gives $1/\lambda = 1 - K\varphi$. Replacing K with $2(\lambda - 1)$ gives $\varphi = 1/2\lambda$. We can now describe the rest of the equilibrium: $\varphi_j^h = \begin{cases} \frac{1}{2\lambda} & \text{if } j \in \{1,2,3\} \setminus h \\ 0 & \text{if } j \in \{0,h\} \end{cases}$; $\theta_j^h = \begin{cases} 0 & \text{if } j \in \{0,1,2,3\} \setminus h \\ \frac{1}{\lambda} & \text{if } j = h \end{cases}$; $K_{sj} = 2(\lambda - 1)\ \forall s \in S,\ j \in \{1,2,3\}$; $K_{s0} = \frac{2}{3}\ \forall s \in S$; $\lambda = \pi_1 = \pi_2 = \pi_3 = \frac{1}{3}\pi_0$. For example, if $\lambda = \frac{4}{3}$, then $x^1 = \left(\frac{1}{2}, \frac{3}{4}, \frac{3}{4}\right)$; $\varphi_0^1 = 0$, $\varphi_1^1 = 0$, $\varphi_2^1 = \frac{3}{8}$, $\varphi_3^1 = \frac{3}{8}$, $\theta_0^1 = 0$, $\theta_1^1 = \frac{3}{4}$, $\theta_2^1 = 0$, $\theta_3^1 = 0$; $K_{sj} = \frac{2}{3}$, $j = 1,2,3$, $\forall s \in S$; $K_{s0} = \frac{2}{3}$, $\forall s \in S$; $\pi_1 = \pi_2 = \pi_3 = \frac{4}{3}$; $\pi_0 = 4$. Note that agent 1 sells 3/8 units of asset 2, delivers $\frac{2}{3}\left(\frac{3}{8}\right) = \frac{2}{8}$, and thus consumes $x_2^1 = 1 - \frac{2}{8} = \frac{3}{4}$. Notice that the marginal disutility of selling one unit of asset 0 is to default in each state, which costs $3\left(\frac{4}{3}\right)$ utiles, or 4, which is equal to the price π_0. The marginal utility of buying one unit of asset 0 is $\frac{2}{3}\left(\frac{4}{3}\right) + \frac{2}{3}\left(\frac{4}{3}\right) + \frac{2}{3}(2) = \frac{28}{9} < 4$, so no agent wants to buy asset 0. For $6/5 \geq \lambda > 1$, the same equilibrium persists. But there is another.

When $\frac{6}{5} \geq \lambda > 1$, there is an equilibrium in which asset trades shrink to one dimension. We can guess from example Version A1 and from the above calculations that $x_s^h = \begin{cases} \frac{1}{\lambda} & \text{if } h \neq s \\ 2\left(1 - \frac{1}{\lambda}\right) & \text{if } h = s \end{cases}$; $\varphi_0^h = \theta_0^h = \frac{4}{3} - \frac{1}{\lambda}$, $\varphi_j^h = \theta_j^h = 0$, $\forall j \in \{1,2,3\}$, $K_{s0} = \frac{6\lambda - 6}{4\lambda - 3}$, $\forall s \in S$, $K_{sj} = 2(\lambda - 1)$, $\forall s \in S$, $j \in \{1,2,3\}$; $\lambda = \pi_1 = \pi_2 = \pi_3 = \frac{1}{3}\pi_0$. To verify equilibrium condition (5) for the given K_{sj}, for asset $j = 1,2,3$, note that each agent h could equally well have acheived exactly the same consumption and default penalties by trading via the assets $j \in \{1,2,3\}$, exactly as described in the last paragraph. Had all three agents done so, delivery rates on these assets really would have been $K_{sj} = 2(\lambda - 1)$. For example, if $\lambda = \frac{8}{7}$, then $x^1 = \left(\frac{1}{4}, \frac{7}{8}, \frac{7}{8}\right)$, $\varphi_0^1 = \theta_0^1 = \frac{11}{24}$, $K_{s0} = \frac{6}{11}$; $\pi_1 = \pi_2 = \pi_3 = \frac{8}{7}$, $\pi_0 = \frac{24}{7}$. Observe that by buying $\theta_0^1 = \frac{11}{24}$ units of asset 0, agent 1 consumes $x_1^1 = K_{10}\theta_0^1 1 = \frac{6}{11}\left(\frac{11}{24}\right) = \frac{1}{4}$, as claimed. Note that agent 1 delivers $D_0^1 = \left(0, \frac{9}{24}, \frac{9}{24}\right)$. Note also that the marginal disutility of selling asset 0 is 24/7, while the marginal utility of buying another unit of asset 0 is $\frac{6}{11}(4) + \frac{6}{11}\left(\frac{8}{7}\right) + \frac{6}{11}\left(\frac{8}{7}\right) = \frac{264}{77} = \frac{24}{7}$. Similarly the marginal utility to agent 1 from buying or selling asset 1 is $\frac{2}{7}(4) = \frac{8}{7}$. The marginal disutility to selling assets 2 or 3 is also 8/7, while the marginal utility of buying them is 0.

When $1 > \lambda \geq 0$, the actively traded asset span shrinks to zero dimensions, since there is no trade in equilibrium.

12 Screening and Signalling: Endogenous Quantity Constraints

Screening and signalling play prominent roles in many loan markets. They are embodied in our model via the quantity constraints Q introduced at the outset, and by additional restrictions that we shall describe below on the sales of assets. We note immediately that quantity constraints are compatible with perfect competition and with partial, but not total, anonymity. Some agency (like a credit card company or GNMA) must monitor its agents' sales to make sure they comply with the quantity constraints, but the buyers themselves may still be totally unaware of who the sellers are, beyond the fact that they have indeed complied with those constraints.

The propensity of agents to keep promises is often correlated with observable characteristics, and with their past behavior. Agents who lack certain "qualifications" or who have a bad "credit rating" are barred from taking out various kinds of loans and insurance. We incorporate screening practices into our model by constraining agents' participation in asset markets, that is by letting Q_j^h depend on h. The simplest kind of constraint $Q_j^h = 0$ completely bars some agents from selling some assets. When screening is occurring, agents who have good characteristics will be able to sell the same promises at higher prices than agents with riskier qualifications. Thus if $A_j = A_{j'}$, and $Q_j^{h'} = 0$, it may be that $\pi_j > \pi_{j'}$ because asset j excludes agents with the unfavorable characteristics $e^{h'}$ and $\lambda^{h'}$.

A moral hazard problem arises because agents take out too many loans. It can be ameliorated, if selling is monitored by a third party, by injunctions limiting the quantity of promises. This second role for the quantity restrictions Q_j^h does not require the lenders to observe borrower behavior, although they will usually be the ones doing the monitoring. For example, a credit card company will usually set a limit on the total amount of debt that can be outstanding; a university may set an arbitrary limit on the total amount of loans it will give any student. If the credit card holder or the student takes out fewer loans than allowed, this will typically have no bearing on the terms of the loans.

Sometimes lenders will give different rates depending on the size of the loan and the commitment the seller makes about taking out other loans. This can be embodied via the dependence of Q_j^h on j. Suppose that two assets j and j' make the same promises $A_j = A_{j'}$. But suppose that $Q_j^h < \infty$ for all h, while $Q_{j'}^h = \infty$ for all h. Then it may well be that $\pi_j > \pi_{j'}$ in equilibrium. Every borrower will sell asset j before asset j'. But only the borrowers who need a lot of money will go in addition to the j' market. Rational buyers realize that sellers of asset j include both large and small borrowers, whereas the sellers of asset j' are all large borrowers. If large borrowers are regarded as riskier than small borrowers, buyers will pay less per unit promise of asset j'. This is precisely what happens in insurance markets. A primary carrier for life insurance or accident insurance will set a standardized maximum contract size. Agents who want to increase the size of their insurance will go to secondary carriers who offer less favorable premiums per dollar of insurance.

Mortgage loans of (very) different sizes are also allocated to different pools. A buyer of shares in a pool cannot observe the names of any of the homeowners in the pool, but he is aware of the size of the loans in the pool. It is well-known that homeowners with very large mortgages, sometimes called jumbo mortgages, exercise their prepayment and default options differently from homeowners with small "conforming" mortgages, and this behavior is reflected in different prices of the respective pool shares. A homeowner contemplating how big a mortgage to take out should consider that the size of his loan will give a signal to the market and affect the rate he pays.

The most effective quantity signals are those which commit the seller not to sell any other kind of asset, or at least no other asset of the same type. Agents who sell assets that limit their ability to sell other assets do so with the expectation that they will receive a better price because they are signalling to the market that their moral hazard risks are limited. In these cases it is crucial to the seller that the lenders be aware of his commitment. If in the above example agents could commit to sell only one asset, then the buyers of asset j could be sure they were loaning exclusively to small borrowers, and the price would be even higher.

The extent of the commitment or constraint to forbear from other asset sales can vary widely. Thus a firm may issue bonds, and commit itself not to issue any more except as junior debt. On the other hand, a credit card allows an agent to issue asset promises to deliver money in the future up to some prespecified credit limit. The agent can have many credit cards, and also take out a mortgage loan etc., without limitation from the original credit card. An insurance contract might limit the total amount of insurance to be taken out of a particular type, but it usually does not restrict the agent to one carrier, and it does not prohibit him from taking out other kinds of insurance. In the MBS securities market described earlier, a homeowner can only take out one first mortgage on his house. But he can certainly take out other kinds of loans.

Often the relevant constraint is on the total amount of promises issued, whether they are issued to one party or to many parties. This occurs when there is a further moral hazard problem we have not considered in our formal analysis: an agent may be able to affect the probability of the states of nature, hence limiting the size of promises that are more burdensome in some states than in others is necessary. A homeowner might be limited to fire insurance up to \$100,000, for fear that otherwise he might burn his own house down to collect the insurance. Notice, however, that the temptation for arson is affected by the total insurance, and is completely independent of the number of insurance policies (provided the total value of insurance is the same).

The simple quantity constraints we have already introduced do not allow us to formalize a wide enough variety of signals. In particular, they cannot handle cases where the sales constraints interact across assets j. We can handle both screening and a broad range of endogenous signals simply by restricting the asset sale choices of an agent of type h to lie in some compact subset C^h of $\mathbb{R}^J_+$ (provided that C^h includes 0). The quantity constraints our model has used up to now define C^h as a product of intervals, which is certainly a special case. More generally, C^h might restrict an agent from selling more than a certain quantity of assets across types j and k, or across all asset types in some class $J' \subset J$. Under these restrictions the sets C^h are still convex, and our proof of the existence of equilibrium can be carried over with almost no change.

Exclusive, or partly exclusive, quantity constraints, however, not only limit the quantity of assets that can be sold, but also limit the number of different kinds of assets that can be sold. If an attractive new contract becomes available, an agent might be forced by the constraint to drop his old contract altogether in order to sell the new one. Under the extreme constraint (emphasized by Rothschild and Stiglitz) that an agent can sell only one asset, the choice set of the agent is not convex, but becomes star shaped, with each tentacle representing the quantities sellable in a different security. One could imagine assets that had to be sold, if at all, in quantities between some minimum and some maximum. These constraints also give rise to nonconvex C^h.

We show that equilibrium exists even when the C^h are not convex. For this result we must take seriously the idea that each agent is of infinitesimal size, and that there is a continuum of agents of each type. In these "nonatomic" economies, it may turn out that in equilibrium two agents of the same type will choose different contracts to trade, although of course they must be indifferent to the two choices.

12.1 A More General Model of Signalling

Let $C^h \subset \mathbb{R}^J_+$ be compact, and let $0 \in C^h$, for all $h \in H$. Define the budget set $\underline{B}^h(p, \pi, K) = B^h(p, \pi, K) \cap \{(x, \theta, \varphi, D) : \varphi \in C^h\}$. Define $GE(A, \lambda, \underline{Q})$ equilibrium by the same conditions as $GE(A, \lambda, Q)$, except with $\underline{B}^h$ replacing B^h.

The following theorem can be proved in the same manner as Theorems 3 and 4.

Theorem 7 *For any $\lambda \in \mathbb{R}^{HSJ}_+$, and any $Q \in \mathbb{R}^{HJ}_+ \cup \{\infty\}$, a $GE(A, \lambda, \underline{Q})$ satisfying* (1)–(5) *exists for the continuum economy.*

12.2 Endogenous Quantity Constraints

The restriction to choices in C^h embodies the idea of screening, as long as C^h depends on h. Since for simplicity we present only a two period model in this paper, in which asset trade occurs in period one and delivery (or default) occurs in period two, there is no credit history before the asset markets meet. If the characteristics of agents are unobservable, or if the law prohibits distinctions based on observable characteristics (race, gender, religion), then the C^h would not depend on h and screening would play no role.

Signalling is embodied in our model by the dependence of C^h on the assets j. To see for example how our model includes Rothschild–Stiglitz as a special case, fix one underlying asset promise A^0. Take a set of J distinct positive numbers Q_j, and now define each of J distinct assets by the promise A^0 and one of the quantity constraints Q_j.[19] Define C^h by restricting agents to sell at most one asset. By choosing to sell an asset j that has a particular quantity constraint attached to it, agent h is signalling something to the market place. We show exactly how our model includes Rothschild–Stiglitz in example B3 below.

Rothschild and Stiglitz permit agents to introduce new assets that might upset a "candidate" equilibrium. A new asset is said to upset a candidate equilibrium if there

[19] To keep attention focused on the quantity constraints, define $\lambda^h_{sj} = \lambda^h_s$ independent of j.

is a price at which both demand and supply are nonzero. If a candidate equilibrium cannot be so upset, Rothschild and Stiglitz call it a genuine equilibrium and they assert that they have found the endogenously determined set of assets (or at least of quantity constraints on the assets). By contrast the $GE(A, \lambda, Q)$ model begins with a set of specified assets that cannot be enlarged, and hence appears not to come to grips with the endogenous determination of assets.

This difference is more apparent than real, however, since the whole range of relevant potential assets in Rothschild–Stiglitz can be included in the initial set of $GE(A, \lambda, Q)$ assets. In the resulting $GE(A, \lambda, Q)$ equilibrium, many of these assets may not be traded, which corresponds to the idea in Rothschild–Stiglitz that they would not upset the active equilibrium markets. In $GE(A, \lambda, Q)$ equilibrium, each inactive asset is assigned a price at which no agent wants to buy it or sell it. With cautious expectations, any higher price would elicit zero demand, and any lower price would elicit zero supply.

The assets are endogenously determined in the sense that only in the equilibrium does it become clear which assets will be actively traded and which will not.

Since the Rothschild–Stiglitz model of insurance is a special case of our model, it might seem puzzling that Theorem 7 guarantees that equilibrium always exists, while they prove that equilibrium may robustly fail to exist. The reason of course is that the definitions of equilibrium differ in a subtle but crucial way. In the Rothschild–Stiglitz model, if an agent offers to buy a new asset that is not currently traded, at a price at which all agents would want to sell it, then the agent is assumed to expect that the distribution of sellers is equal to the general population distribution. We, on the other hand, argue that some sellers may have a much greater incentive to sell the asset than others, and since the potential buyer cannot possibly satisfy all the sellers who want to sell it, he should expect that he will get a selection of sellers drawn from those who are most anxious to sell it, i.e., those who would sell the asset at the lowest price. Under our definition of delivery expectations for nontraded assets, equilibrium always exists. And indeed the unique equilibrium under the exclusivity hypotheses that no agent can sell more than one asset is the so-called separating equilibrium described by Rothschild and Stiglitz.

We also differ substantively from Rothschild–Stiglitz in our understanding of insurance markets. In their analysis agents can only obtain one insurance contract, whereas we firmly believe that agents can obtain many insurance contracts. Of course if we think of insurance for different events like fire and theft and automobile collision and death and worker disability, it is obvious that people take out many insurance policies. In our general equilibrium formulation all these contracts can simultaneously be incorporated with little trouble, whereas in the partial equilibrium framework of Rothschild and Stiglitz this is problematic. But even if we confine our attention to one kind of accident, we maintain that many people still take out more than one policy. Professional athletes and musicians often buy additional insurance, and some people take out additional life insurance.

We begin our analysis in this section by introducing two types of agents and adverse selection. In example B1, we allow agents to sell without quantity constraints.

The unreliable agents choose to sell and buy more (i.e., take out more insurance), and thus the equilibrium aggregate delivery rates K_s are skewed toward unreliable rates.

Next, in example B2, we allow firms to limit the amount of insurance a customer obtained from them (while allowing customers to take out more insurance elsewhere). This leads to a new equilibrium with primary carriers and secondary carriers. All agents obtain the maximum allowable insurance from the primary carrier. This maximum is set at a level that does not constrain the most reliable type. Less reliable agents feel that they need more insurance, and go to secondary carriers who will then charge them a higher premium. In contrast to the separating equilibrium obtained with exclusive insurance contracts, it is the unreliable types who are constrained.

Finally, in example B3 we allow for exclusive contracts, and we show that equilibrium must always reproduce the separating equilibrium discovered by Rothschild–Stiglitz.

Version B0: Arrow–Debreu Equilibrium for an Economy with 2 Types

Suppose now we add to our example three additional agent types $h = 4, 5, 6$, with the same utilities, but with endowments $e^4 = (1,0,0)$, $e^5 = (0,1,0)$, and $e^6 = (0,0,1)$. Recall that $e^1 = (0,1,1)$, $e^2 = (1,0,1)$, and $e^3 = (1,1,0)$, and that all six agents have utility $u^h(x_1,x_2,x_3) = \sum_{s=1}^{S} \log x_s$.

The Arrow–Debreu equilibrium is $p = (1,1,1)$ and $x^h = \left(\frac{2}{3},\frac{2}{3},\frac{2}{3}\right)$ for $h \in \{1,2,3\}$ and $x^h = \left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$ for $h \in \{4,5,6\}$.

Version B1

We continue with the same 6 agents from version B0. Suppose now that the default penalties are

$$\lambda^h_{sj} = \begin{cases} \infty & \text{if} \quad e^h_s = 1 \\ 0 & \text{if} \quad e^h_s = 0 \end{cases} \quad \forall h \in H,\ s \in S,\ j \in J\ .$$

Suppose we have a single asset promise $A^0 = (1,1,1)$. We can think of this model as one big insurance contract, as mentioned in version A3. The difference is that this time we have adverse selection. The sellers $h \in \{1,2,3\}$ default 1/3 of the time, while the sellers $h \in \{4,5,6\}$ default 2/3 of the time. A buyer must anticipate that he may get more sellers of the bad type than of the good type.

We suppose that $Q^h_0 = \infty$, $\forall h \in H$. In the unique equilibrium, the unreliable agents $h \in \{4,5,6\}$ sell and buy twice as much of the asset as the reliable agents $h \in \{1,2,3\}$. Hence $K_{s0} = \frac{2}{3}\frac{1}{3} + \frac{1}{3}\frac{2}{3} = \frac{4}{9}$, $\forall s \in S$. It can be shown that $\theta^h_0 = \varphi^h_0 = \frac{3}{5}$ for $h \in \{1,2,3\}$ and $\theta^h_0 = \varphi^h_0 = \frac{6}{5}$ for $h \in \{4,5,6\}$, and $x^1 = \left(\frac{4}{15},\ \frac{2}{3},\ \frac{2}{3}\right)$, $x^2 = \left(\frac{2}{3},\ \frac{4}{15},\ \frac{2}{3}\right)$, $x^3 = \left(\frac{2}{3},\ \frac{2}{3},\ \frac{4}{15}\right)$, and $x^4 = \left(\frac{1}{3},\ \frac{8}{15},\ \frac{8}{15}\right)$, $x^5 = \left(\frac{8}{15},\ \frac{1}{3},\ \frac{8}{15}\right)$, $x^6 = \left(\frac{8}{15},\ \frac{8}{15},\ \frac{1}{3}\right)$. Compared to the Arrow–Debreu equilibrium, reliable agents are doing much worse since their insurance rates are debased by the unreliable agents. The unreliable agents are much better off than in the Arrow–Debreu equilibrium because they benefit from being pooled with the reliable agents.

The reader can verify that the equilibrium is correct by calculating that the marginal utilities to reliable and unreliable agents of buying and selling a unit of the asset is 3.

A useful way to conceptualize the problem is to think of each agent h having a good state, in which his endowment is 1, and a bad state, with endowment 0. (The only difference between agents $h \in \{1, 2, 3\}$ and agents $h' \in \{4, 5, 6\}$ is that the probability of the good state for agents h is 2/3, while the probability of the good state for agent h' is 1/3.) By selling and buying one unit of the single asset (with fixed delivery rates $K_{s0} = K$), every agent gets K in his bad state, and gives up $1 - K$ in his good state. Since the agents perceive K as fixed, this implicitly defines a price $q = (1 - K)/K$ of consumption in the bad state in terms of the good state. All agents are effectively maximizing $\text{prob}(G) \log x_G + (1 - \text{prob}(G)) \log x_B$ subject to the constraint $x_G + qx_B \leq 1$. With these Cobb–Douglas utilities, agents will always choose $x_G = \text{prob}(G)$, as illustrated in Figure 5. This demonstrates that reliable agents h trade half as much as unreliable agents h', and hence that $K = 4/9$, and hence that $1/q = 4/5$, and hence that $x_B^h = \frac{1}{3}\frac{1}{q} = \frac{4}{15}$, $x_B^{h'} = \frac{2}{3}\frac{1}{q} = \frac{8}{15}$, and so on. In this analysis the price π of the asset played no role, since all agents are buying and selling the same asset. In the next example, with multiple traded assets, the prices π_j are important

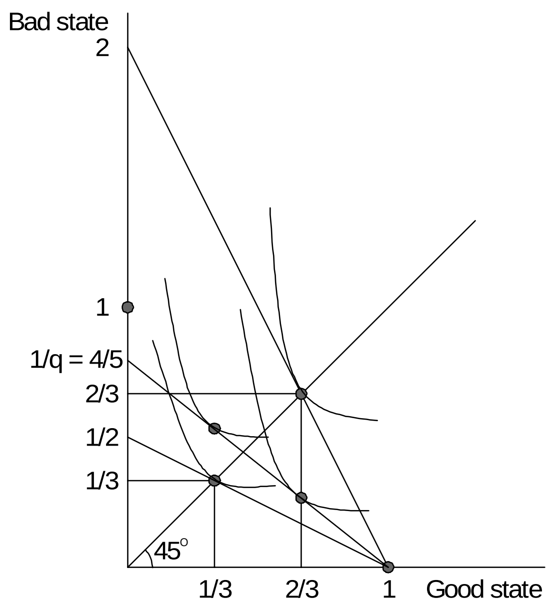

Figure 5

Figure 5 tells the same story. Every agent begins with an endowment of 1 in his "good" state(s) and 0 in his bad state(s). The top budget line represents the actuarially fair odds for the reliable agents, for whom the good state is twice as likely as the bad state. At those fair odds (in the Arrow–Debreu equilibrium) they completely insure by moving to the 45° line and consuming 2/3 in every state. Similarly the unreliable agents have a fair odds budget set that reflects the fact that for them the good state is only half as likely as the bad state. In the Arrow–Debreu equilibrium they completely insure by consuming 1/3 in every state.

When the odds are 4:5 of good to bad, the unreliable agents take advantage of the actuarially favorable odds to overinsure, while the reliable agents underinsure because for them the odds are unfair. The odds of 4:5 are closer to 1:2 than to 2:1, reflecting the fact that the unreliable agents take out twice as much insurance. Note finally that the assumption of Cobb–Douglas utilities fixes the same consumption for each agent in his good state, no matter what the budget line.

Version B2: Primary and Secondary Insurance

So far in our examples we have not paid any attention to the quantity constraints Q_j^h (perhaps because we regarded them as so high). Quantity constraints require monitoring to enforce. We must now think of some intermediary who is checking that agent h does not try to sell more than Q_j^h of asset j, for example by placing two sell orders under different names.

Let us now reconsider Version B1 with additional quantity constrained assets. We now take the set of assets $J = \{ j = 0, 1, ..., 100\}$ which each promise $A^j = (1,1,1)$, but with quantity constraints $Q_0^h = \infty$ and $Q_j^h = \frac{j}{30}$, for $j = 1, ..., 100$. (We have chosen the mesh of size 1/30 arbitrarily). The budget set of each agent h is constrained by the condition $\varphi_j^h \leq Q_j^h$ for all $j \in J$. The pooling equilibrium of Version B1 is still an equilibrium, provided we extend the asset prices to be $\pi_0 = \pi_1 = ... = \pi_{100} = \frac{4}{9}$ and $K_{sj} = \frac{4}{9}$ for all $s \in S$ and $j \in J$. Consumption and trades of asset 0 are as above, and none of the other assets is traded.

A new equilibrium now emerges with the quantity constraints. In this new "primary/secondary" equilibrium there is a primary or standard contract, which in our example turns out to be contract 20, which sells for the highest price $\pi_{20} = \frac{1}{2}$. Every agent sells this contract, in our example all the way up to the limit $Q_{20}^h = \frac{20}{30} = \frac{2}{3}$. Then there is a secondary contract, in our example contract 0, in which agents who want more insurance must sell at a lower price, in our example $\pi_0 = \frac{1}{3}$, because by trading in the secondary contract they reveal themselves to be risky individuals. All other assets $j \neq 20$ also sell at the price $\pi_j = \frac{1}{3}$. This primary/secondary equilibrium conforms to much general practice. Many institutions sell life and disability insurance to their employees, at some rate, up to a specified maximum. The employees are then free to purchase additional insurance on the open market, at what will often be less advantageous terms.

The rest of the equilibrium is specified precisely as follows: $K_{s20} = \frac{1}{2}$, $\forall s \in S$, $K_{sj} = \frac{1}{3}$, $\forall s \in S$, $j \neq 20$, $(\theta_0^h, \theta_{20}^h) = (\varphi_0^h, \varphi_{20}^h) = \left(0, \frac{2}{3}\right)$, $\forall h \in \{1,2,3\}$, and $(\theta_0^h, \theta_1^h) = (\varphi_0^h, \varphi_1^h) = \left(\frac{1}{3}, \frac{2}{3}\right)$, $\forall h \in \{4,5,6\}$, and $\theta_j^h = \varphi_j^h = 0$, $\forall j \in J \setminus \{0, 20\}$, $\forall h \in H$, and $x^1 = \left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right)$, $x^2 = \left(\frac{2}{3}, \frac{1}{3}, \frac{2}{3}\right)$, $x^3 = \left(\frac{2}{3}, \frac{2}{3}, \frac{1}{3}\right)$, and $x^h = \left(\frac{4}{9}, \frac{4}{9}, \frac{4}{9}\right)$, $\forall h \in \{4,5,6\}$. Since asset 0 delivers exactly 2/3 of asset 20 in every state, and both assets are being purchased, we know $3\pi_0 = 2\pi_{20}$. Since asset 20 is sold in equal amounts by reliable and unreliable agents, $K_{s20} = \frac{1}{2} = \frac{1}{2}\left(\frac{2}{3}\right) + \frac{1}{2}\left(\frac{1}{3}\right)$. Since asset 0 is sold only by unreliable agents, $K_{s0} = \frac{1}{3}$. Clearly agents $h \in \{4,5,6\}$ would prefer to sell A^{20} for a price $\pi_{20} = \frac{1}{2}$ to selling the same promise A^0 for a price $\frac{1}{3} < \frac{1}{2}$. But they are constrained by $Q^h_{20} = \frac{2}{3}$ from selling more. The reliable agents on the other hand are not affected by the constraint Q^h_{20}, since they would like to sell $\varphi^h_{20} = \frac{2}{3}$ anyway.

A diagram helps the analysis.

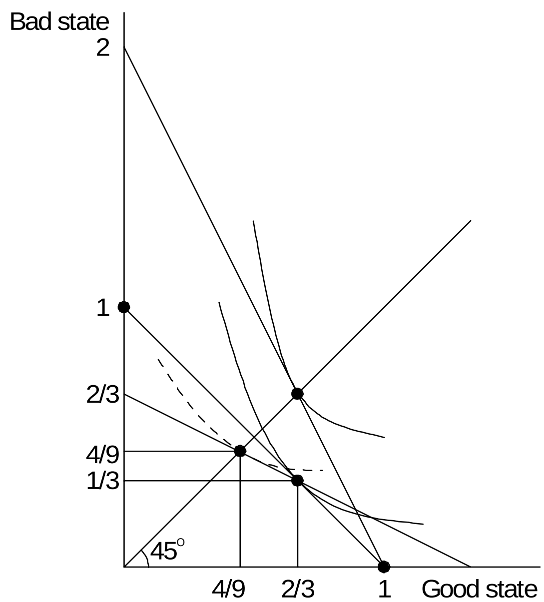

Figure 6

The unreliable agents trade along the 1:1 budget line as far as permitted, until reaching $(2/3, 1/3)$. Then they trade from there along the 1:2 budget line until they fully insure at 4/9 in every state.

In our example B there are two equilibria, the pooling equilibrium of Version B1 and the primary/secondary markets of Version B2.[20] Note incidentally that Rothschild and Stiglitz described a different "symmetric pooling equilibrium" in which all agents sold the same quantity $\varphi^h_{j^*} = \varphi^{h'}_{j^*} = Q^h_{j^*}$, $\forall h, h'$, of one contract A^{j^*}. That cannot be an equilibrium in our formulation, if there are other contracts available, since the unreliable agents will feel constrained. The final allocation will also be Pareto dominated by our primary/secondary equilibrium.

Note that the final utilities of the agents achieved in the primary/secondary equilibrium, $\log(4/27)$ for the reliable agents and $\log(64/729)$ for the unreliable agents are better for the reliable agents and worse for the unreliable agents than the utilities achieved by the pooling equilibrium, $\log(16/135)$ for the reliable agents, and $\log(64/675)$ for the unreliable agents. Which kind of equilibrium occurs in practice

[20] There are actually a continuum of primary/secondary equilibria, where the quantity limit on the primary carrier is arbitrary. All but the one we described will disappear if we and the extra exclusivity hypothesis that agents can have only one primary insurance. We consider exclusivity in example B3.

probably depends on whether (the reliable agents) have set up an inexpensive system for monitoring insurance policies.

Version B3

Rothschild and Stiglitz (1976) made the important observation that adverse selection in insurance markets might lead to the same kind of inefficient signalling that Spence had earlier postulated would arise in labor markets. In labor markets, Spence (1973) argued that agents with high ability would purchase expensive and unproductive education simply to signal that they were indeed of high ability. In insurance markets, Rothschild and Stiglitz argued, agents would commit themselves exclusively to contracts with low insurance in order to signal that they were reliable. Rothschild and Stiglitz went on to suggest that with signalling there might not be any equilibrium in insurance markets. In this they were mistaken, we believe, because they had an inadequate definition of equilibrium. But the important point, that signalling can be inefficient, remains intact.

Rothschild and Stiglitz proposed a severe signalling equilibrium in which agents can sell some contracts which commit them not to sell any other assets. We can capture this idea by adding to Version B2 additional assets $j = 101, 102, ..., 200$ which make the same promises as before, $A^j = (1, 1, 1)$, and with the same quantity constraints as before, $Q^h_{j+100} = Q^h_j = \frac{j}{30}$, for $j = 1, ..., 100$. The difference now is that if $\varphi^h_j > 0$ for some $j > 100$, then $\varphi^h_i = 0$ for all $i \neq j$.

We can incorporate exclusive quantity constraints into our diagram as follows. Imagine that there is an asset (call it 0) trading in the market which delivers a constant K in each state, and sells for $\pi_0 = K$. Suppose now that some asset j sells for π_j and has a quantity constraint Q_j. An agent who sells one unit of it gets π_j, which he can use to purchase asset 0, obtaining $(\pi_j/K)K = \pi_j$ in each state. The agent delivers 1 in his good state, and nothing in his bad state. On net he receives π_j in his bad state and gives up $1 - \pi_j$ in his good state. Implicitly that defines a price $q_j = (1 - \pi_j)/\pi_j$, indicating how much consumption must be given up in the good state to get one unit of consumption in the bad state. If Q_j is sold of the asset, then the agent gives up $Q_j - Q_j\pi_j$ in his good state, and gets $Q_j\pi_j$ in his bad state. As the price π_j rises, the agent gains equally in his good state and bad state. If the agent sells only asset j, and sells the maximum allowed, then his final consumption must lie on the Q_j-quantity straight line beginning at the point $(1-Q_j, 0)$ and rising at a 45° slope, passing through the point $(1 - Q_j/2,\ Q_j/2)$, and passing through the point $(1 - Q_j/3,\ 2Q_j/3)$ and terminating at $(1, Q_j)$. In fact, consumption must lie on the intersection of this line (which depends on Q_j, but is independent of π_j) and the π_j-price line starting at $(1, 0)$ and rising with slope $-1/q = -\pi_j/(1 - \pi_j)$. The price line depends on π_j but is independent of Q_j. Thus in Figure 6 reliable agents sold only one asset $j = 20$, with quantity constraint $Q_j = 2/3$, at a price $\pi_{20} = 1/2$. Their final consumption thus could be found at the point $(2/3, 1/3)$, which lies at the intersection of the 2/3-quantity line connecting $(1-2/3, 0)$ with $(1, 2/3)$ and the 1/2-price line connecting $(1, 0)$ and $(0, 1)$. Three quantity lines and three price lines are given in Figure 7 below.

The exclusivity assumption eliminates the primary/secondary equilibrium of Version B2 and also eliminates the "pooling" equilibrium of Version B1 from our model, even though in that equilibrium each agent indeed sells only one contract. Recall that in the pooling equilibrium, $K_{s0} = \frac{4}{9}$, $\forall s \in S$, and $\varphi_0^h = \frac{3}{5}$, for $h = 1, 2, 3$ and $\varphi_0^h = \frac{6}{5}$ for $h = 4, 5, 6$.. We shall now show that it is impossible to assign prices π_{100+j} for $j = 1, ..., 100$ in such a way that nobody will want to trade these latter assets. Consider contract (i.e., asset) 117 defined by its promise $A^{117} = (1, 1, 1)$ and its quantity constraint $Q_{117}^h = \frac{17}{30} < \frac{3}{5} < \frac{6}{5}$. The final consumption of any agent who sells asset $j = 117$ must lie on the quantity line in Figure 7 denoted by $Q_{117} = 17/30$. In order that nobody actually wants to sell it, the price must be low enough that the consumption point lies below both indifference curves.

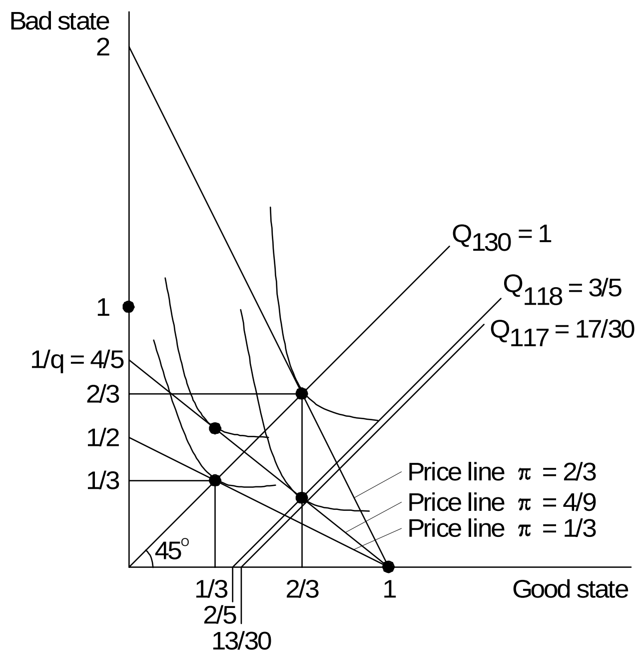

Figure 7

If π_{117} is too low, say $\pi_{117} \leq \pi_0 = 4/9$, then all agents strictly prefer not to sell A^{117} (because by the exclusivity condition and the low level of Q_{117}^h, they would be constraining themselves with no compensating advantage). By equilibrium condition (5), its K_{s117} would then all be 1, in which case $K_{s117}/\pi_{117} = 1/\pi_{117} \geq 1/\pi_0 > K_{s0}/\pi_0 = (1/3)/\pi_0$, demand would be huge, since all buyers would switch from asset 0 to asset 117. Hence if in equilibrium asset 117 is not traded, we must have at least one agent type indifferent to switching to asset 117. But along the quantity line $Q = 17/30$, the reliable indifference curve is always below the unreliable indifference curve. Hence π_{117} must make the reliable agents indifferent and π_{117} must be slightly more than $\pi_0 = 4/9$. (At a much higher price, say $\pi_{117} = \frac{2}{3} = \frac{3}{2}\pi_0$, reliable agents would sell positive amounts of it, since they can get more money selling $\frac{17}{30}$ units of asset 117 at a price of $\frac{3}{2}\pi_0$ then they can by selling $\frac{3}{5}$ units of asset 0 at a price of

π_0.) Small perturbations would induce only the reliable agents to sell, and they will all deliver in two out of the three states. Hence according to equilibrium condition (5), we must have $K_{s117} \geq \frac{2}{3} = \frac{3}{2}(\frac{4}{9})$. But then for demand to be zero, π_{117} must reflect this, hence $\pi_{117} \geq \frac{3}{2}\pi_0$, which is a contradiction. Our equilibrium refinement thus captures the same spirit as the Rothschild–Stiglitz definition of no entry equilibrium in eliminating the pooling equilibrium. A similar argument eliminates the primary/secondary equilibrium.

The only candidate equilibrium Rothschild and Stiglitz found is the "separating" candidate equilibrium in which each type of agent sells a different asset. Rothschild and Stiglitz observed that in such an equilibrium the unreliable types should feel unconstrained by the quantity restriction while the reliable types should feel quantity constrained. Moreover, the unreliable types should be indifferent to either of the two contracts, while the reliable types should strictly prefer their quantity constrained contract, because the price is more favorable, reflecting the fact that only they trade it, and because they are less bothered by the quantity constraint since they know themselves to be reliable. Indeed, once we impose the Rothschild–Stiglitz exclusivity restriction, we get exactly this sort of equilibrium, though not quite exactly because our menu of quantity constraints is not fine enough.

We claim that with the exclusivity condition ($\varphi_j^h > 0$ for at most one contract $j > 100$), there is essentially a unique equilibrium, which is defined as follows. Asset 0 trades at a price $\pi_0 = 1/3$. Only unreliable types trade asset 0: $\theta_0^h = \varphi_0^h = 1$ for $h = 4, 5, 6$ and $\theta_0^h = \varphi_0^h = 0$ for $h = 1, 2, 3$. It follows that $K_{s0} = \frac{1}{3}$ for all s, and that $x^h = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ for $h \in \{4, 5, 6\}$. For example, $x^4 = (1, 0, 0) + 1\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) - 1(1, 0, 0)$. The reader can check that unreliable agents are indifferent between $(x_G, x_B) = \left(\frac{1}{3}, \frac{1}{3}\right)$ and a point $(x'_G, x'_B) = \left(\frac{9}{10} - \varepsilon, \frac{2}{10} + 2\varepsilon\right)$ on the $\pi = 2/3$ price line connecting (1,0) and (0,2). The point (x'_G, x'_B) lies strictly between the quantity lines $Q_{109} = 9/30$ and $Q_{110} = 10/30$. On the other hand, reliable agents trade only asset 109, which is marketed at a price $\pi_{109} = 2/3$, setting $\theta_{109}^h = \varphi_{109}^h = \frac{3}{10} = \frac{9}{30} = Q_{109}^h$ for $h = 1, 2, 3$ while $\theta_{109}^h = \varphi_{109}^h = 0$ for $h = 4, 5, 6$. Thus $K_{s109} = 2/3$ for all s, and $x^1 = \left(\frac{2}{10}, \frac{9}{10}, \frac{9}{10}\right) = (0, 1, 1) + \frac{3}{10}\left(\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right) - \frac{3}{10}(0, 1, 1)$; $x^2 = \left(\frac{9}{10}, \frac{2}{10}, \frac{9}{10}\right)$; $x^3 = \left(\frac{9}{10}, \frac{9}{10}, \frac{2}{10}\right)$. If there were no quantity constraints, everyone would prefer to sell more of contract A^{109} at its higher price; but if an agent does, he is excluded from selling any other contract. The unreliable agents, who need lots of insurance, are lured away (though just barely) by contract 0, despite its low price, because it does not have a quantity constraint. The reliable agents would also prefer not to be constrained, but since for them insurance is less important, the constraint is less a burden, and they stick with selling contract 109 rather than contract 0, despite its quantity constraint, because of its high price. For assets $j = 1, ..., 100$, and assets $j = 131, ..., 200$, $K_{sj} = \frac{1}{3}$ and $\pi_j = 1/3$. No agent trades any other asset. To see how to price assets $100 + j$, for $1 \leq j \leq 30$ turn to Figure 8.

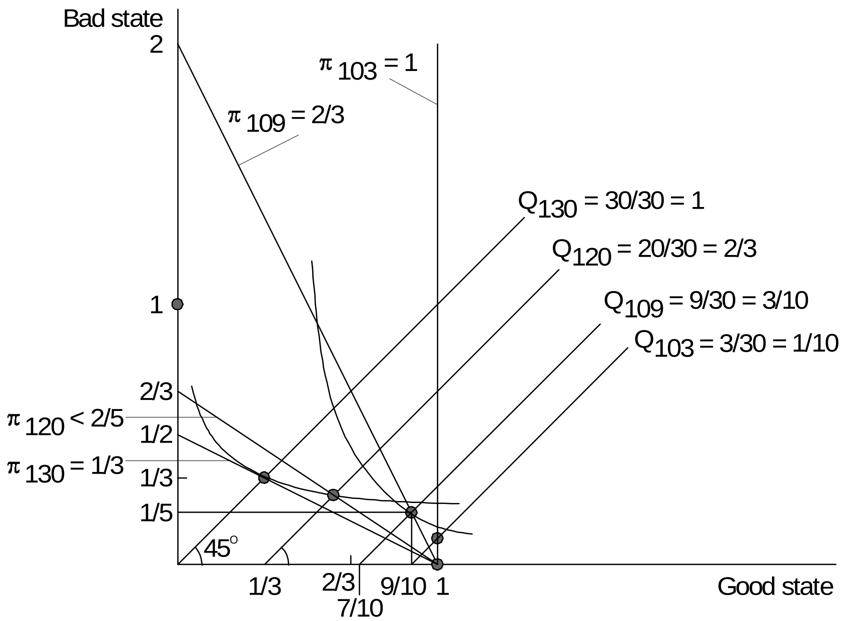

Figure 8

For assets $j = 10, ..., 30$, the quantity line Q_j intersects the unreliable indifference curve below the reliable indifference curve. Hence π_j is set so that the unreliable agents are exactly indifferent between selling only asset j and selling only asset 0. At this price the reliable agents will strictly prefer to stick to asset 109. Hence equilibrium condition (5) requires the $K_{sj} = \frac{1}{3}$. Thus for example for asset $100 + j = 120$, the price is chosen so that the π_{120} price line intersects the Q_{120} line exactly at the point where the Q_{120} line intersects the unreliable indifference curve. At this price the reliable agents do not wish to sell, while the unreliable agents are just indifferent to not selling. Since the price $\pi_{120} > 1/3$, no agent wants to buy, since a buyer could get the same delivery rate $K = 1/3$ at the lower price of 1/3 by buying asset $j = 0$ (or $j = 130$).

Asset $100 + j = 109$ is the first one whose quantity line $Q_j = 9/30 = 3/10$ passes through the reliable indifference curve before the unreliable indifference curve. At the corresponding price $\pi_{109} = 2/3$, consumption is 9/10 in the good state and 1/5 in the bad state, so the unreliable agents just barely prefer to stick to selling asset 0. Evidently if we took a finer grid of assets with quantity constraints evenly spaced 1/1000 units apart, rather than 1/30, we would have found that the unreliable agents were almost exactly indifferent between the two traded assets.

For assets $101, ..., 108$, the Q_j-line intersects the reliable indifference curve below the unreliable indifference curve. Hence we look for the point (x_G, x_B) where the Q-line intersects the reliable indifference curve. If that occurs with $x_G \leq 1$, then π_{100+j} is set so that the reliable agents are exactly indifferent between selling only asset $100+j$ and selling only asset 109. At this price the unreliable agents will strictly prefer to stick to asset 0. Hence equilibrium condition (5) requires the $K_{s,100+j} = 2/3$. If $x_G > 1$, then set $\pi_{100+j} = 1$. At that price and with such a low quantity constraint, both the reliable and unreliable agents would prefer not to sell the asset. Thus by equilibrium condition (5), we must set $K_{s,100+j} = 1$. But buyers will be content not to buy at this price, since they can already get 1 in each state by spending \$1 and buying 3 units of asset 0. An example here is asset $j = 103$.

Observe that this equilibrium construction works because for $Q \leq 9/30$, the Q-line intersects the reliable indifference curve below the unreliable indifference curve, whereas for $Q > 9/30$, the Q-line intersects the unreliable indifference curve before the reliable indifference curve. The reason for this is that the reliable indifference curve is steeper than the unreliable indifference curve. (This has been called the single crossing property). Had the grid of assets been continuous, we could have found an asset with quantity constraint Q^* such that the Q^*-line intersects the unreliable indifference curve exactly at the point it crosses the $\pi = 2/3$ line. In our example, with quantity constraints available only in multiples of $1/30$, we had to choose $Q^* = 9/30$ so that the Q^*-line intersects the $\pi = 2/3$ line just below the unreliable indifference curve. Thus if there were an asset with Q just slightly above Q^*, the separating equilibrium would be disrupted, since the Q-line intersects the reliable indifference curve before the unreliable indifference curve. A promise with such a constraint Q could not be priced in such a way that it would not be traded. If the price π were low enough so that the reliable agents did not want to sell it, then $\pi < 2/3$ (if $\pi \geq 2/3$ they would sell it since $Q > Q^* = 9/30$ and $\pi_{9/30} = 2/3$). But then a small perturbation would induce either nobody, or just the reliable agents, to sell it. Hence by equilibrium condition (5), the expected delivery rates K_s would be at least $2/3$. Hence agents would rush to buy it. However, the reader can check that for $Q = 10/30$, which is the smallest available constraint above $Q^* = 9/30$, the Q-line intersects the reliable indifference curve first, and equilibrium is as described.[21]

It is worthwhile to point out that constructing equilibrium can be difficult, and must exploit all the details of the economy (like the single crossing property). But the existence of equilibrium does not rely on such properties, as Theorem 7 shows.

Rothschild and Stiglitz correctly noted that the separating equilibrium allocation and price system is well-defined and feasible independent of the proportion of reliable agents. By contrast observe that the pooling equilibrium and the primary/secondary equilibrium both improve in utility terms as the proportion of reliable agents converges to 1, eventually Pareto dominating the separating equilibrium. Rothschild and Stiglitz went on to claim that if nearly all the agents are reliable, then the separating equilibrium could not be an equilibrium, because some contract such as $A^{130} = (1, 1, 1)$ with its more generous constraint $Q^h_{130} = 1$ would break the equilibrium. Their paper is not precise about how expectations are formed when assets are not traded, but the idea is that if it was expected that the sellers of contract A^{130} were in the same proportion as the population as a whole, then in a population consisting almost entirely of reliable agents, the corresponding K_{s130} would be nearly $2/3$ and the price it would fetch would be $\pi_{130} = 2/3$. This price (and its generous quantity

[21] Had the $Q = 10/30$ line intersected the reliable indifference curve first, then there would still be an equilibrium (as Theorem 7 demonstrates), but it would not be strictly separating. Some of the uneliable agents would sell the high price asset $Q = 110$, and its price would be less than $2/3$.

constraint) definitely would lure away sellers of both types, and so Rothschild and Stiglitz argue, justify the expectations $K_{s130} = 2/3$, and thus upset the separating "equilibrium."

However, such an expectation is hasty, since agents do not all have the same incentive to switch to the new contract. We have set the price in equilibrium of asset 130 at $\pi_{130} = 1/3$. At this price the unreliable agents are just indifferent to switching from asset 0 into asset 130, whereas the reliable agents are not close to wanting to sell asset 130. Even if some agent offered to buy asset 130 at a price of 1/2, only unreliable agents would rush to sell it. Not until the price reaches .53 would the reliable agents become interested in selling asset 130. Thus we feel justified in setting the expectations of delivery for asset 130 at $K_{s130} = 1/3$, and the price at $\pi_{130} = 1/3$. If an agent did for some remarkable reason offer to pay 2/3 for asset 130, he would be deluged with offers from sellers, so many in fact that he could never accommodate them all. A natural reaction would be to lower his buying price until the number of sellers fell. But as we just saw, to reduce demand sufficiently he would end up selling only to the unreliable types, as we have presumed.

We are in agreement with the concern of Rothschild and Stiglitz about the separating equilibrium when there is a high proportion of reliable agents. But the problem is not the nonexistence of equilibrium. (We constructed an equilibrium in our example, and in general, Theorem 7 guarantees that equilibrium exists.) The problem is its inefficiency.

13 Modigliani–Miller Theorem?

Virtually every example we have given so far illustrates the general failure of the Modigliani–Miller principle, when the default penalties are low enough to permit default. When the total debt issued by the firm is unobservable, as was the case for our agents in Sections 1–11 of this paper, then typically the firm will find a unique optimizing portfolio decision, as we confirm later.[22] When the firm debt is observable, but there is asymmetric information about the quality of the firm, as there was in the example B3 of the Rothschild–Stiglitz insurance markets, there will again typically be a signalling reason for the firm to want to choose a unique portfolio. We shall therefore concentrate in this section on the case where both the firm characteristics and its financing decisions are known to investors. To guarantee the existence of equilibrium, we need to appeal to Theorem 7 and the apparatus of nonconvex quantity constraints.

The Modigliani–Miller "theorem" asserts that the instruments a firm (or an industry) uses to finance its operations do not matter to its valuation, or to the real operations of the economy as a whole. The "theorem" operates at three levels. It asserts that a perfectly competitive, price-taking firm should perceive that its total value will be unaffected by its financing decision, just as a firm operating under constant-returns-to-scale in conventional general equilibrium feels indifferent to its scale. Second, it asserts that if a single firm, or even a nonnegligible (but strict)

[22] In this section we use the words firm and agent $h \in H$, or $t \in [0, H)$ interchangeably.

subset of firms in an industry does change its debt–equity ratio, then the rest of the (identical) firms in the industry can restore the old real equilibrium by altering their debt–equity ratios in the opposite direction. This is precisely analogous to the situation in conventional general equilibrium in which half the identical firms in a constant-returns-to-scale industry reduce their output by 10% while the other half raise their output by 10%, all without essentially disturbing the equilibrium. Finally, the Modigliani–Miller "theorem" asserts that even if all the firms in an industry reduce their debt–equity ratios, other agents (such as the investors in the industry) can adjust their portfolios to restore the old real equilibrium. Since we have a model with genuine, endogenous default, embedded in an economy-wide model of general equilibrium, it is natural to investigate under what circumstances the Modigliani–Miller "theorem" actually holds.

We show that all three levels of the Modigliani–Miller "theorem" hold when there is no possibility that any agent will default. Even with default, we shall confirm the (first level) partial equilibrium assertion that each individual firm must think it is indifferent between substituting debt for equity.

Moreover, even when a firm defaults with positive probability on its junior debt, *if it never defaults on its senior debt*, then the second level of Modigliani–Miller holds in our model. But the third level fails. Typically each industry debt–equity ratio is determinate, but not the equilibrium debt–equity ratio of any proper subset of firms in an industry (if the firms in the industry never default on their senior debt).

Finally, if firms also default with positive probability on their senior debt, then both the second and third levels of Modigliani–Miller fail. In the financial equilibrium we shall describe below, if half the identical firms in an industry change their debt–equity ratios by 10%, there will generally be no compensating change the rest of the industry, or the rest of the economy, can undertake which will restore the original equilibrium. In equilibrium, the debt–equity ratios chosen by the individual firms in an industry are determinate. From now on we refer more accurately to the Modigliani–Miller principle, instead of to a theorem.

In our model there is no equity, so we treat the analogous problem of senior debt versus junior debt, as we shall see. The question becomes: does it matter how a firm chooses to divide its promises among the different debt instruments?

When the default penalties are so high that there is no default the answer is no, at all three levels. There is a trivial multiplicity of $GE(A, \lambda, Q)$ equilibria in which agent h arbitrarily chooses his financing, with no real impact either on his fortunes or on the economy. In this case the Modigliani–Miller principle is a genuine theorem which we now state in somewhat more generality:

Theorem 8 *Let* $\mathcal{E}_A = ((u^h, e^h)_{h\in H},\ (A^j, ((\lambda^h_{sj})_{s\in S}, Q^h_{sj})_{h\in H})_{j\in J})$*; suppose* $\lambda^h_{sj} = \infty$ *and* $Q^h_j = \infty$, $\forall h,\ s,\ j$*. Let* $\mathcal{E}_B = \mathcal{E}_A$*, except that assets* B^j *replace* A^j*. Suppose* $(p,\ \pi,\ K,\ (x^h, \theta^h, \varphi^h, D^h)_{h\in H})$ *is a* $GE(A, \lambda, Q)$ *for* $\mathcal{E}_A$*, and that letting* $p\Box A^j \equiv (p_s \cdot A^j_s)_{s\in S}$ *and* $p\Box B^k \equiv (p_s \cdot B^k_s)_{s\in S}$*,* $\operatorname{span}[p\Box B] = \operatorname{span}[p\Box A]$*. Suppose for some industry (agent type)* $\bar{h}$*, there are vectors* $(\bar{\theta}^{\bar{h}}, \bar{\varphi}^{\bar{h}}) \in \mathbb{R}^J_+ \times \mathbb{R}^J_+$ *such that* $[p\Box A](\theta^{\bar{h}} - \varphi^{\bar{h}}) = [p\Box B](\bar{\theta}^{\bar{h}} - \bar{\varphi}^{\bar{h}})$*. Suppose* $\#H > 1$*.Then there is a* $GE(A, \lambda, Q)$*,*

$(p, \overline{\pi}, \overline{K}, (x^h, \bar{\theta}^h, \bar{\varphi}^h, \bar{D}^h)_{h \in H})$, *for* $\mathcal{E}_B$ *which gives the same consumption for each* $h \in H$*, and which respects the new financial decisions of agent* $\bar{h}$*. Moreover,* $\sum_{j \in J} \overline{\pi}_j(\bar{\theta}^{\bar{h}} - \bar{\varphi}^{\bar{h}}) = \sum_{j \in J} \pi_j(\bar{\theta}^{\bar{h}} - \bar{\varphi}^{\bar{h}})$*, for each* $h \in H$*. The theorem is also true if* $\#H = 1$*, provided that only a fraction* $\mu < 1$ *of agents of type* $\bar{h}$ *choose* $(\bar{\theta}^{\bar{h}}, \bar{\varphi}^{\bar{h}})$*.*

The theorem considers a situation in which there is no possibility of default, and a fraction (possibly all) of the agents of type $\bar{h}$ issue debt $\bar{\varphi}^{\bar{h}} \neq \varphi^{\bar{h}}$. If the net money promises under $\bar{\varphi}^{\bar{h}}$ are the same as under $\varphi^{\bar{h}}$, then those agents of type $\bar{h}$ are still optimizing, and moreover, other agents in the economy can adjust so as to maintain the same real equilibrium. The last part of the theorem asserts that if some proper fraction $\mu < 1$ of the agents of type $\bar{h}$ switch from $\bar{\varphi}^{\bar{h}}$ to $\varphi^{\bar{h}}$, then the compensation can take place entirely among the remaining fraction $1-\mu$ of agents of type $\bar{h}$ (leaving the rest of the agents acting as before).

The proof is left to the reader.

When there is no default, the Modigliani–Miller principle is elementary. In particular, when a small firm considers issuing more debt, there is no reason to contemplate that the price of its debt might change. After all, each successive debt is as good as the previous one, and altogether the firm makes an imperceptible change to the whole market. Without default, the first level of Modigliani–Miller is completely trivial, and Theorem 8 confirms that the last two levels hold as well.

When penalties are finite, and default occurs, even the first level of Modigliani–Miller becomes problematic. A firm that observably issues more debt can expect its bond prices to go down as investors factor in a higher chance of default. One might guess that the firm would not regard all such changes as neutral, and that indeed the situation is incompatible with perfect competition. Both guesses are mistaken. Once we introduce quantity constraints, in order to make the size of the debt issue observable, as in Section 12, we can capture the intuitive connection between the size of the bond issue and its interest rate (price), while at the same time preserving perfect competition. We shall also find that each firm will indeed regard its financing portfolio choices as a matter of indifference, though only one choice will be compatible with equilibrium.

Equity holders are residual claimants after the debt holders have been paid off; that role is played by junior debt in our model. The question we pose is whether a borrower (and the economy as a whole) should be indifferent to the division of loans between junior and senior debt. Suppose, for example, that we have two assets $A^i = A^j = (1, 1, ..., 1)$, but $\lambda^h_{si} > \lambda^h_{sj}$. Then debt j is junior to debt i for agent h in state s, because h will always pay back all of the i debt before he begins to pay back any of the j debt. So as not to worry about the difference in utility penalty from defaulting on junior or senior debt, we will take λ^h_{si} just infinitesimally bigger than λ^h_{sj}. It follows that agent h will choose (virtually) the same total default on i plus j no matter the proportion he sells them. We ask if equilibrium is unaffected by the proportion of senior debt and junior debt a firm issues?

We now proceed to embed this question in a full-blown general equilibrium model with default. Assume that every agent $j \in J$ makes the same promise $A^j = (1, ..., 1)$. Suppose that the assets $j \in J$ explicitly recognize the issuing industry, senior and junior debt, and quantity restrictions. To formalize this, take a finite set $Q \subset \mathbb{R}_+$ of quantity constraints. Define $J = H \times Q \times Q \times \{1, 2\}$. Each element $j \in J$ is then given by a tuple hq_1q_2d. The first element $h(j) \in H$ denotes which industry the issue belongs to: $Q_j^h = 0$ unless $h = h(j)$. The second element $q_1(j) \in Q$ denotes the maximum sale of senior debt issued by the firm, and the third element $q_2(j) \in Q$ denotes the maximum amount of junior debt issued by the firm. The last element $d(j) \in \{1, 2\}$ denotes whether j is senior debt, $d(j) = 1$, or junior debt, $d(j) = 2$. Thus $Q_j^h = q_1(j)$ if $h = h(j)$ and $d(j) = 1$, while $Q_j^h = q_2(j)$ if $h = h(j)$ and $d(j) = 2$. If $h = h(i) = h(j)$, $q_1(i) = q_1(j)$, $q_2(i) = q_2(j)$, and $d(i) = 1$ and $d(j) = 2$, then we call (i, j) a senior–junior debt combination. We impose another sales constraint that a firm can sell only one senior debt, and only the corresponding junior debt, that is, only one senior–junior debt combination. This can be written as $\varphi_i^h \varphi_j^h > 0 \Rightarrow q_1(j) = q_1(i)$ and $q_2(j) = q_2(i)$.

For simplicity we suppose that default penalties are the same for all assets, $\lambda_{sj}^h = \lambda_s^h$ for all $s \in S$, $h \in H$. We assume, however, that each h always pays back all his senior debt before making any payments on his junior debt. (This can be achieved formally by taking $\lambda_{si}^h > \lambda_{sj}^h$ for all $s \in S$ if (i, j) is a senior–junior debt combination, then taking limits as $\lambda_{si}^h \to \lambda_{sj}^h$. But it is simpler to just assume $\lambda_{si}^h = \lambda_{sj}^h$ and require the order of payment.) The reader can check that the resulting budget sets $B^h(p, \pi, K)$ are closed (though not convex). Hence by Theorem 7, equilibrium always exists in this model.

The restriction $Q_j^h = 0$ unless $h = h(j)$ guarantees that there is no asymmetric information about the characteristics of any firm issuing debt. The quantity constraints and the limitation to one senior–junior debt combination guarantee that the issuing firm's actions are observable, so that the price of debt can reflect the added risk of increased debt.

Consider an economy with assets and a budget set of the sort described above, and suppose that we have an equilibrium in which some agent h has actually sold some junior debt, $\varphi_j^h > 0$ for some $j \in J$. Let $i \in J$ be the senior debt corresponding to j, and suppose firm h has also issued $\varphi_i > 0$ of senior debt. Since we have already dealt with the no-default case, suppose there is at least one state $s \in S$ in which there is default. Then we must have $\pi_i \geq \pi_j$, since $K_{si} \geq K_{sj}$ for all $s \in S$. If $K_{si} = K_{sj} \in \{0, 1\}$ for all $s \in S$, then junior debt and senior debt are perfect substitutes, and trivially substituting one for the other leaves equilibrium intact. So let us confine attention to the case of non-trivial default; suppose for at least one state s, $K_{si} > K_{sj}$. Then $\pi_i > \pi_j$, and firm h will not bother to issue any junior debt unless $\varphi_i^h = Q_i^h$. It follows incidentally, that if $\varphi_i^h < Q_i^h$, so that the debt issuance of firm i is effectively unobservable, then there cannot be any junior debt issued with non-trivial default. As we stated at the outset of Section 13, if the firm debt is unobservable (i.e., without quantity constraints) the Modigliani–Miller theorem cannot possibly hold, in the presence of non-trivial default.

So suppose $\varphi_i^h = Q_i^h \equiv q_1(i)$, and consider another senior–junior debt combination

(a, b) such that $q_1(a) < q_1(b)$. (The case where $q_1(a) > q_1(i)$ is handled the same way.) It seems unlikely that agent h would voluntarily curtail the amount of senior debt he issued when he gets the better price for senior debt. The reason of course is that by doing so agent h gets to sell his junior and perhaps also his senior debt for a higher price, $\pi_b > \pi_j$ and $\pi_a \geq \pi_i$.

In fact, from equilibrium condition (5), we know that agent h must be indifferent to selling every senior–junior combination. If he strictly preferred not to sell (a, b), then by condition (5), deliveries would be expected to be 100%, which implies (from the fact that in the equilibrium nobody buys assets (a, b)) that the price $\pi_a = \pi_b \geq \pi_i > \pi_j$. But then agent h would strictly prefer to sell assets (a, b) to assets (i, j), a contradiction.

If $Q_a^h + Q_b^h = \varphi_i^h + \varphi_j^h$, then this indifference implies that $\pi_a Q_a^h + \pi_b Q_b^h = \pi_i \varphi_i^h + \pi_j \varphi_j^h$. Hence once we take into account quantity constraints on both senior and junior debt, the partial equilibrium Modigliani–Miller principle follows trivially from the definition of equilibrium.

But the economy as a whole is not indifferent between the two debt pairs, because a different pair of assets means a different way for *other* agents to spread risk.

If h defaults on his senior debt i in some states in a nontrivial default equilibrium, and if there are at least three states, then the span of the asset deliveries of h is almost surely different from what it would be if he sold assets a and b instead, with $Q_a^h < Q_b^h$. Suppose for example that there are three states, and that h is delivering (10,9,3) in total and defaulting on his senior debt only in state 3, and on his junior debt in states 2 and 3. Suppose $Q_i^h = 5 = Q_j^h$ and $Q_a^h = 4$, $Q_b^h = 6$. Then in the equilibrium h was delivering (5,5,3) on his senior debt, and (5,4,0) on his junior debts. Had he sold a and b instead, he would have delivered (4,4,3) and (6,5,0) on his senior and junior debt, respectively.[23] The linear spans of the two pairs of vectors are different.

The sum of the two vectors in each pair is the same, and hence there is no difference to agent h who contemplates selling one *pair* or the other. But a buyer may be purchasing exclusively senior debt i (or exclusively junior debt j). If a is issued instead, he gets a different asset, which might matter a great deal to him.

Our first reason for the failure of Modigliani–Miller is that if there is default on the senior debt, the span of the assets almost surely is changed by different proportions of senior and junior debt. The conventional argument in favor of Modigliani–Miller suggests that investors who had purchased the original issues of a firm could compensate in their own portfolios for any changes in the debt issuance of the firm. But that argument ignores the implications of default. There may be no available asset such an investor could purchase or sell that would make up for the change in the firm's debt deliveries. Thus when a firm defaults on its senior debt with positive probability, the second and third versions of the Modigliani–Miller principle fail.

So let us suppose finally that firms of type h do not default on their senior debt. Now the span of asset deliveries on i and j will be the same as the span of the asset deliveries on a and b. But again it might make a difference to the buyers of the assets

[23] Since the default penalties are all the same, $\lambda_{si}^h = \lambda_{sj}^h = \lambda_{sa}^h = \lambda_{sb}^h$, and total promises under (i, j) and (a, b) are the same, it follows that total deliveries must be the same.

that h has chosen to issue less senior debt. Consider the following diagram:

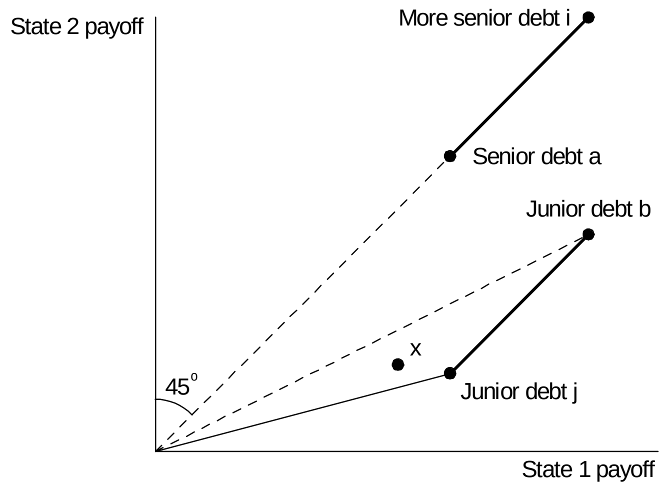

Figure 9

The dotted lines represent the new position of h (less senior debt and more junior debt). The point x represents the desired portfolio payoffs of some investor; x could be achieved by holding some nonnegative combination of senior and junior debt from the original mix. But if all the firms of type h chose to sell a and b instead of i and j, then to achieve x the investor would need to buy b and go short the senior debt. In other words, the investor would like to issue his own debt. But his default penalties may be very different from those of agent h. In particular, the market might presume he would default. Thus our second reason for the failure of the Modigliani–Miller principle is that investors cannot issue their own debt as a perfect substitute for corporate debt. It may be easier for some industry to leverage than for its investors. Thus the third variant of Modigliani–Miller fails even if the firms in industry h do not themselves ever default on their senior debt.

Finally, let us suppose that only a fraction (say half) of the firms in industry h switch from (i, j) to (a, b), and that none of the firms defaults on senior debt. For concreteness, suppose that a involves 10% less senior debt than i. We just saw that investors might not be able to compensate. But if the other half of the firms in industry h increased their senior debt by 10% over Q_i^h, then it is easy to see that equilibrium could be restored. Each investor would buy the same fraction of every firm's issuance as he was before. Thus we see that if the firms in an industry do not default on their senior debt, then the industry aggregate senior–junior debt ratio is determinate, but individual firms in the industry may have difference senior–junior debt ratios.

We summarize the conclusions of this section in Proposition 9 below. We do not call it a theorem because we have not formalized the word typical, which appears in the proposition.

Proposition 9 *If the debt issues of a firm are not publicly observable (i.e., without quantity commitments on the debt issuance), firms will typically think they have a unique optimal debt structure. A similar conclusion holds even when debt issuance is publicly observed, if there is asymmetric information about the characteristics of the firm. If debt issuance is public, and there is no asymmetric information about the quality of the issuing firm, then the firm will regard its senior–junior debt ratio as a matter of indifference. But if the firms occasionally default on their senior debt, then typically equilibrium will require a determinate senior–junior debt ratio for each firm. If the firms never default on their senior debt, but do occasionally default on their junior debt, then typically equilibrium will require a determinate aggregate senior–junior debt ratio for the industry, but identical firms in the industry may have different, indeterminate, senior–junior debt ratios.*

We conclude by noting that had we introduced collateral as another mechanism for enforcing promises (in addition to penalties), then we might have found that agents themselves, taking prices as given, would not be indifferent to their senior–junior debt ratios, and so even the first level interpretation of Modigliani–Miller would fail.

References

[1] Akerlof, G., 1970. "The Market for Lemons: Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84: 488–500.

[2] Allen, F. and D. Gale, 1988. "Optimal Security Design," *Review of Financial Studies*, 1: 229–263.

[3] Allen, F. and D. Gale, 1991. "Arbitrage, Short Sales, and Financial Innovation," *Econometrica*, 59: 1041–1068.

[4] Arrow, K. J., 1953, "Generalization des theories de l'equilibre economique general et du rendement social au cas du risque," *Econometrie*, Paris, CNRS, 81–120.

[5] Baron, D. P., 1976, "Default and the Modigliani–Miller Theorem: A Synthesis," *American Economic Review*, 66, 204–212.

[6] Cho, I. K. and D. Kreps, 1987. "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102: 179–221.

[7] Dubey, P., J. Geanakoplos and M. Shubik, 1990. "Default and Efficiency in a General Equilibrium Model with Incomplete Markets," Cowles Foundation Discussion Paper, No. 773.

[8] Dubey, P. and M. Shubik, 1979. "Bankruptcy and Optimality in a Closed Trading Mass Economy Modelled as a Noncooperative Game," *Journal of Mathematical Economics*, 6: 115–134.

[9] Gale, D., 1992. "A Walrasian Theory of Markets with Adverse Selection," *Review of Economic Studies*, 59: 229–255.

[10] Geanakoplos, J., 1996. "Promises, Promises," Cowles Foundation Discussion Paper No. 1123.

[11] Geanakoplos, J., H. Polemarchakis, 1986. "Existence, Regularity, and Constrained Suboptimality of Competitive Allocations When the Asset Market Is Incomplete," in W. Heller, R. Starr, and D. Starrett (eds.), *Essays in Honor of Kenneth Arrow*, Vol. 3, Cambridge University Press, pp. 65–99.

[12] Geanakoplos, J., and W. Zame, 1995. "Default, Collateral, and Derivatives," unpublished manuscript.

[13] Hellwig, M. F., 1981. "Bankruptcy, Limited Liability and the Modigliani–Miller Theorem," *American Economic Review*, 155–170.

[14] Hellwig, M. F., 1981. "A Model of Borrowing and Lending with Bankruptcy," *Econometrica*, 45, 1879–1905.

[15] Hellwig, M. F., 1987. "Some Recent Developments in the Theory of Competition in Markets with Adverse Selection," *European Economic Review Papers and Proceedings*, 31: 319–325.

[16] Helpman, Elhanan and Jean-Jacques Laffont. 1975. "On Moral Hazard in General Equilibrium Theory," *Journal of Economic Theory*, 10: 8–23.

[17] Modigliani, F. and M. H. Miller. 1958. "The Cost of Capital, Corporation Finance and the Theory of Investment," *American Economic Review*, 48: 261–297.

[18] Pesendorfer, W., 1995. "Financial Innovation in a General Equilibrium Model," *Journal of Economic Theory*, 65: 79–116.

[19] Prescott, E. and R. Townsend, 1984. "Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard," *Econometrica*, 52: 21–45.

[20] Radner, Roy, 1972. "Existence of Equilibrium of Plans, Prices, and Price Expectations in a Sequence of Markets," *Econometrica*, 40(2): 289–303.

[21] Riley, J., 1979. "Informational Equilibrium," *Econometrica*, 47: 331–359.

[22] Rothschild, M. and J. Stiglitz, 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90: 629–650.

[23] Shubik, Martin and Charles Wilson, 1977. "The Optimal Bankruptcy Rule in a Trading Economy Using Fiat Money," *Zeitschrift fur Nationalokomie*, 37(3–4): 337-354.

[24] Smith, V. L., 1972, "Default Risk, Scale and the Homemade Leverage Theorem," *American Economic Review*, 62, 66–76.

[25] Spence, M., 1973. "Job Market Signalling," *Quarterly Journal of Economics*, 87: 355–374.

[26] Stiglitz, J. E., 1974, "The Irrelevance of Corporate Financial Policy," *American Economic Review*, 64, 851–866.

[27] Stiglitz, J. and A. Weiss, 1981. "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, 72: 393–410.

[28] Wilson, C., 1977. "A Model of Insurance Markets with Incomplete Information," *Journal of Economic Theory*, 16: 167–207.

[29] Zame, W., 1993. "Efficiency and the Role of Default When Security Markets Are Incomplete," *American Economic Review*, 83: 1442–1164.